Agent-Based Models of Geographical Systems

Alison J. Heppenstall • Andrew T. Crooks
Linda M. See • Michael Batty
Editors

Agent-Based Models of Geographical Systems

Springer

Editors
Alison J. Heppenstall
School of Geography
University of Leeds
Woodhouse Lane
LS2 9JT Leeds
United Kingdom
a.j.heppenstall@leeds.ac.uk

Linda M. See
International Institute
of Applied Systems Analysis
Schlossplatz 1
2361 Laxenburg
Austria
see@iiasa.ac.at

Andrew T. Crooks
Department of Computational
Social Science Research l
George Mason University
University Drive 4400
22030 Fairfax Virginia
CSC Suite, Level 3
USA
acrooks2@gmu.edu

Michael Batty
Centre for Advanced Spatial Analysis
University College London
Torrington Place 1-19
WC1E 6BT London
United Kingdom
m.batty@ucl.ac.uk

ISBN 978-90-481-8926-7 e-ISBN 978-90-481-8927-4
DOI 10.1007/978-90-481-8927-4
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2011940794

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Part I
Computational Modelling: Techniques for Simulating Geographical Systems

Chapter 2
A Generic Framework for Computational Spatial Modelling

Michael Batty

Abstract We develop a generic framework for comparing spatial models whose dynamics range from comparative static equilibrium structures to fully dynamic models. In the last 40 years, a variety of spatial models have been suggested. Until the mid 1980s, most models were static in structure and tended to embrace detailed mechanisms involving spatial economics and social physics. Typical examples were Land Use Transportation Interaction (LUTI) models that embraced theories of spatial interaction and discrete choice modelling. During this earlier period, the problems of making these models dynamic and more disaggregate was broached but progress was slow largely because of problems in collecting requisite data and problems of increasing the complexity of such models to the point where they could be properly validated in traditional ways. 20 years or more ago, new modelling approaches from very different sources came onto the horizon: in particular, dynamic models based in Cellular Automata (CA) which were largely physical in nature and Agent-Based Models (ABM) providing explicit behavioural processes that often rested alongside these automata. Systems Dynamics Models (SDM), Spatial Econometric Models (SEM) and Microsimulation Models (MM) all informed the debate. It is tempting to see these models as all being of different genera but here we attempt to see them as part of an integrated whole, introducing a framework for their elaboration and comparison. After the framework is introduced, we review these six model types and choose three – CA, ABM and LUTI models – that we then work up in more detail to illustrate these comparisons. We conclude with the conundrums and paradoxes that beset this field.

M. Batty (✉)
Centre for Advanced Spatial Analysis (CASA), University College London (UCL), London, UK
e-mail: m.batty@ucl.ac.uk

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_2,

2.1 Antecedents: The Origins of Spatial Models

Digital computers appeared in the late 1940s largely as a result of developments in the logic of computing and the notion that large-scale numerical processing could be massively speeded up by reducing routine tasks to binary equivalents operating on equivalent electrical devices. Right from the beginning, scientific applications involved spatial as well as temporal problems and by the mid-1950s, rapid advances in digital computation led to computable problems in the human applications domain involving spatial systems such as cities and transportation. Mathematical theories of such spatial systems were slowly developing prior to the invention of the digital computer but there had been little focus on how such theories might be operationalised, tested through validation, and then used in forecasting. Digital computers were to provide the spark for such applications and in 1955, the first models of traffic flow were implemented in a digital environment in the Chicago Area Transportation Study (Plummer 2007).

These first models, unlike many if not most that have followed them, were specifically tailored to the problems in question and the way those problems were perceived. Transport flows were critical as the problems in question involved providing for new transport capacity, while land use location too was essential in a period of relatively rapid economic growth which involved the search for new locations for urban development. These early models were equilibrium-seeking rather than dynamic, aggregate at the level of populations involving spatial interactions, and built on conceptions of the city articulated using ideas from urban economics and social physics. They are usually now referred to as Land Use Transportation Interaction (LUTI) models. From these early attempts, as computers and their software developed, new generations of computable spatial models have become more generic in that the software developed for general classes of model has become ever more significant, thus elevating generic ideas about modelling through their software to a point where specific model types now tend to defer to generic modelling styles. In this chapter, indeed in this book, this notion of generic models and generic software is very much to the fore because agent-based models (ABM) and their close relatives cellular automata (CA) models represent classes and styles that are much wider in scope and applicability than the sorts of spatial systems to which they are applied.

Here we will outline as wide an array of spatial models as is possible in an integrated fashion, setting the scene for many of the more specific applications and developments in the chapters that follow. As it is rare in this field to see highly standardised applications which barely differ from case to case, each model application tends to be tailored in some specific way to the problem and its context such that model styles and structures become mixed. However what we will do is identify six distinct styles of spatial model that cover most of this array beginning with the original social physics and urban economic models that kick-started the field half a century ago. But before we introduce specific model types and show how these relate and evolve from one another, we will begin this review by examining model structures, identifying the key characteristics and themes that dominate model development. We will first focus on questions of abstraction and representation, noting the difference between the substantive components of any spatial model which we define as its population in contrast to the environment with which it interacts. In one sense, all models can be so defined and this serves as a basis on which to characterise the way populations which provide the objects or components of the spatial system under question, interact with one another and with their environment through a series of key processes. We will examine issues of representing spatial and temporal scale, aggregation, and constraints, and then we will look at processes of change, feedback, and dynamics. Many of these features and themes merge into one another and to an extent, any such categorisation of the key characteristics of spatial models is arbitrary. But these categories do enable us to sketch out the array of ideas that dominate the field which appear time and again in this book. Once we have introduced these ideas to set the context, we will examine six model types beginning with the simplest cellular automata, defining agent-based models, noting econometric, systems dynamics and microsimulation all of which involve generic approaches, concluding with notions about specific models that contain their own styles and features such as those that were the first to be developed in the land use transportation domain. To give focus to this review, we will then outline examples of CA, ABM and LUTI models in more detail, providing the reader with ideas about how such models are designed and used in practice.

2.2 Modelling as Computation: Abstraction and Representation

Half a century ago, the idea of a model was in its infancy. Scientific theory essentially was based on formal and systematic theories, often represented mathematically, whose testing was confined either to controlled experiments in the laboratory or to various categories of thought experiment. Computation changed all that. The idea that a scientific theory could then be translated into an intermediate form – called a 'model' – represented a way of enabling controlled experiments to be carried out not on the actual system of interest but on a computable abstraction of that system. The term model quickly entered the lexicon and it is now widely used to describe any kind of experimental context in which the computer is used as the medium for its exploration and testing. In fact, the term is now used even more generally to refer to any kind of abstraction that represents an obvious 'simplification of the real thing' and in this sense its meaning is no longer exclusively associated with computation (Lowry 1965; Batty 2007).

When computer models were first developed, the general assumption was that these were simply representations of the system on which testing would take place so that the theory on which the model was based could be tested against data. In general, it was assumed that the traditional canons of the scientific method in which theory was successively refined to withstand its falsification and to engender greater parsimony of explanation, would apply. Most spatial models from the 1950s onwards were predicated on the basis that their predictions would be tested against data taken from the system of interest and that the model would be tuned in such a way as to reproduce the system of interest within the computational environment in a way that was closest to the real thing. Goodness of fit was the main means of validation while calibration of the parameter values ensured that the model might be tuned most effectively to the system in question. The quest originally was thus to find some minimalist explanation for the system of interest in the belief that models should be as simple as possible while also generating predictions closest to our observations of the 'true' system. In fact, as in all science, this involves a trade-off.

Yet the complexity of human systems has meant that right from the first applications, there was continued pressure to develop greater and greater detail – to disaggregate the model's variables to the point where sufficient heterogeneity of the system might be represented in a manner useful to those who sought to use the model to make predictions. There were limits on what computation could offer and data concerning social systems has always been a problem but as computers got more powerful and as the world moved to a point where computation became all pervasive, our ability to model in detail changed by an order of magnitude. As the world moved online, new and richer data sources are becoming ever more available and this computational power combined with access to new and different data, meant that what we could model and represent began to change. Moreover, the key challenge in social systems is to know how much detail to represent and it would appear that the sorts of average behaviour that are characteristic of physical systems are rather different in the social world. Heterogeneity and hence greater detail is what seems to be required so that ever more plausible models can be constructed.

At the same time, as bigger and richer models have been built, their software has become more generic with general purpose simulation processes being articulated in software that can be adapted to many different types of problem. All this is fast leading to significant doubt that the scientific method taken from the classical traditions of physics has the same relevance to the social world as it does in the physical. Indeed even in science itself there is substantial questioning of the traditional canons of scientific inquiry as the quest for parsimony, simplicity, and homogeneity is increasingly being confronted by the need for plausibility, richness, and heterogeneity. The question turns on whether or not a simple, parsimonious model that can completely explain a limited set of system characteristics is as useful as one which contains many characteristics which are plausible in terms of the functioning of the system but cannot be proven as being of definitive explanatory value. In fact the problem is complicated by the predictability of many parsimonious models that are able to explain spatial behaviour as it can be observed but are unable to predict future behaviours which do not admit the same stability as those that are observed in the past. This is a deep problem that suggests that what we observe is considerably more ordered and structured at any point in time than that same set of observations at a future time. This is not just a problem in dynamics or equilibrium but one which is intrinsic to our ability to disentangle true explanation from the way we observe the world. Currently the received wisdom is that different models apply to different kinds of problem and problem context and that in the last analysis, models are useful to inform the debate through crystallising ideas.

In designing any model, the builder must decide what constitutes the structure of the system as distinct from the environment in which the system functions. In fact, this boundary problem is highly significant for it defines how the system relates to other systems and to the rest of the world in general. Very often the same model can be applied to different conceptions of the same system which is defined differently with respect to its environment. Here we will define the term environment rather more narrowly than its general use in systems theory where it refers to the rest of the world or the problem context. We will make a distinction between the wider environment within which the system sits relative to the rest of the world, and the local environment of the system which is the space-time nexus that pertains to the functions in question. In short, the system's environment here is the spatial tessellation of its cells or its locational referents which change through time. In contrast, we define the system in terms of its population, meaning its components and their functions that operate within this local environment. In essence, it is the population that constitutes the structure of the system and its functioning which operates in its space-time environment. The functioning takes place between the population and its environment and there are feedbacks in both directions, that is the population can influence the environment just as the environment can influence the system but these two aspects of the model are qualitatively quite different as we will see. In terms of how this *population-environment system* relates to the outside world often called the environment too, then the usual assumption is that although the environment of the outside world can influence the system, the system does not influence the outside world in terms of the operation of its model. This is the usual convention in systems theory.

In this review, we will attempt to represent all our models no matter how different using the same notational structure and to this end, we define an index of space as i or j and any interaction or relation between them as ij while we use k to define some attribute or feature of the population which pertains to different sectors. Time is indexed as t. Where we need to refer to more than two locations or two attributes or two time periods, we will define appropriate additional symbols as we proceed. We first define a spatial unit i at time t within the environment as A_{it}, and then an attribute or segment of the population at the same coordinates as N_{it}. The two matrices $\mathbf{A}$ and $\mathbf{N}$ contain the key elements of the system which interact with one another in ways that we make specific when we detail models of how populations function, interact and change and how these relate to the spatial system. We can write these feedback loops as $\mathbf{A} \Leftrightarrow \mathbf{N}$ to give some sense of the symmetry of these relations but at the same noting that $\mathbf{A}$ and $\mathbf{N}$ are generically different.

We can easily aggregate these discrete quantities into larger spatial units that we call Z_I where I is a spatial index to the number of cells i that are within Z, or into larger temporal units Q_T where T is the aggregate temporal index. Note that there are continuity and contiguity constraints that we need to be aware of when we aggregate over space and/or time. We thus define the appropriate units at larger scales as

$$\left.\begin{aligned} A_{IT} &= \sum_{i \in Z_I} \sum_{t \in Q_T} A_{it} \qquad and \\ N_{IT} &= \sum_{i \in Z_I} \sum_{t \in Q_T} N_{it} \end{aligned}\right\} \tag{2.1}$$

where there are likely to be conservation constraints in terms of size such as $A = \sum_i A_{it}, \forall t$ and $N = \sum_i \sum_t N_{it}$, the particular form of which are usually specified when the model is implemented. Functions defined on the population and the environment and the relations between them constitute the structure of the system and usually specify the dynamics of change through time. However to provide some sense of closure to this rather abstract form of representation, at any cross section in time, it is possible to define interactions between these components over space. For example, the populations might interact which we can specify in the following way without detailing the mechanisms. Then the interaction between spaces i and j can be written as $A_{ij} = A_{it} \otimes A_{jt}$ where the concatenation is specified according to some behavioural or physical principle embodied in the model.

It is worth noting that functions like this tend to be specified in systems theory independently of time so that the structure of the system is laid bare. There may be many such functions and before anything further can be said about a model structure, the mechanisms must be specified. What is important is that this framework is seen as being generic in that it can apply to a variety of different problems and problem contexts, to different systems be they physical or human, material or conceptual but with a slight bias towards the subject matter of this book which is agent-based models in the social sciences, particularly the geographical social sciences. Whether or not this is the best representation is not particularly relevant. Each model is developed in its own formal style and the purpose of this framework is to provide a template for assessing how different the array of models that we define here are from one another, not in terms of their substantive or behavioural similarities or differences. In this sense, the population and the environment can be very different. The only common point of reference is the fact that we make this distinction between these two sides of the model and specify space and time in the formal notation of cells and time instants, rather than in the continuous fashion that is often used to couch more theoretical statements of spatial models.

2.3 Feedback, Dynamics and Processes of Change

During the sweep of history over which spatial models have evolved, there has been a shift from simple, parsimonious models that simulate systems at a cross section in time and represent populations in aggregate form to more complex, richer models that deal directly with the time dimension and specify model functionality in terms of processes of change at a much more disaggregate level than their earlier counterparts. The switch has been occasioned by many forces. Already we have noted the growth in computation and the emergence of online data sources which have made much richer models possible but there has also been a sea change in the way we think about human systems. Complexity theory has raised the notion that systems are never in equilibrium, in fact their predominant condition is far-from-equilibrium and disequilibrium is their normal state. Moreover human systems have become ever more complex due to technological change, the demographic transition and increasing wealth, at least in the west and many newly developed countries. This has made spatial behaviours more complex, certainly in terms of movement and communication as well as locational preferences. All in all, dynamics has come firmly onto the agenda while the notion of explanation has shifted from aggregates to much more heterogeneous populations composed of individuals and groups that need to be understood at a much finer level of detail.

In generating behaviours, feedback is an important mechanism where we might specify this in functional terms as a dependence of population on itself or on the environment, that is $A_{it+1} = f(N_{it})$, $N_{it+1} = f(N_{jt})$, or $N_{it+1} = f(A_{it})$ and so on. Negative feedback tends to damp activity so that departures from some norm are restored, the classic example being a thermostat which controls the heat from a boiler to some environment. Positive feedback on the other hand accelerates the degree of change, sometimes with catastrophic consequences, but usually with beneficial impacts if some quantity such as income or even population is increasing. The best way to illustrate the effect of feedback is in terms of population growth and the basic equation which can be used to simulate positive feedback is

$$N_{it+1} = \alpha N_{it}, \tag{2.2}$$

where α is the rate of change defined as N_{it+1} / N_{it}. If the growth rate α is greater than 1, then this leads to exponential growth as we show in Fig. 2.1. If less than 1 then this leads to a decline to zero population but in both cases, the change is due to the compounding effect which can be easily seen if we generate a recursion on Eq. 2.2 up to time $t+T$ as

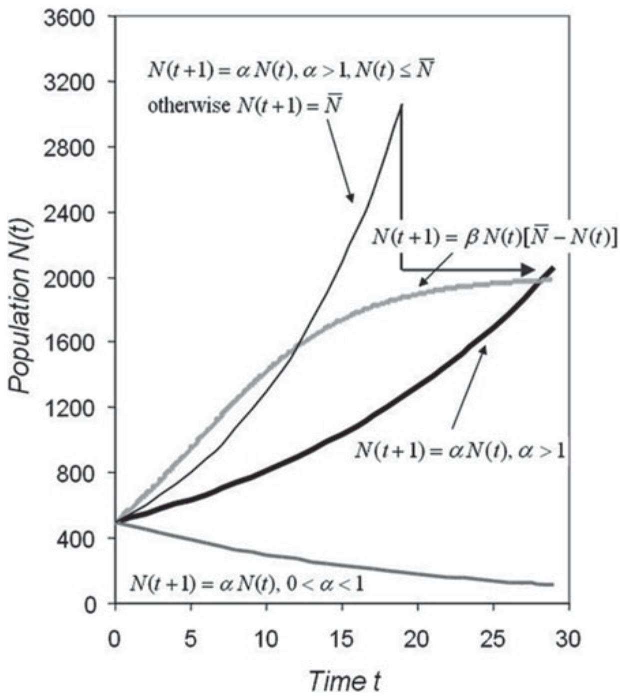
$$N_{it+T} = \alpha^T N_{it}. \tag{2.3}$$

Negative feedback can be shown when change is damped according to some threshold but it is more appropriate to show this as a moderation of exponential growth as encapsulated in the logistic equation. Then if we define a limit to population as say $\bar{N}_i$, then we write the logistic as

$$N_{it+1} = \beta N_{it}(\bar{N}_i - N_{it}), \tag{2.4}$$

where the rate β is moderated with respect to the scale of the growth. We also show this form in Fig. 2.1 where it is clear that the population grows exponentially at first and is then damped by the effect of the constraint $\bar{N}_i$. In fact if the damping effect is lagged leading to an oscillation around the limit value of $\bar{N}_i$, then the growth of population mirrors the sort of behaviour characteristic of systems dynamics models that were developed by Forrester (1969) in cases where resource limits dominate.

Fig. 2.1 Simple dynamics based on positive feedback



Another form of dynamics relates to variations across space in the manner we illustrated for spatial interaction in the previous section. If we add time to this kind of dynamics which involves spatial relations, associations, correlations or movements, then we can represent these as flows from i to j between times t and $t+T$. In fact the interaction N_{ij} which we associate with cross-sectional static models in the previous section does take place through time although the time is much shorter than the usual periods that are associated with spatial modelling. Only quite recently has our concern in understanding cities shifted to thinking of cities in real time for such a real time focus has previously been captured as a static snapshot of movements in the city as, for example, in transport and traffic modelling. However longer time periods are associated with flows such as migration where the variable $N_{i,j,t,t+T}$ is now associated directly with time. Mechanisms for such models are only specified when the precise form of model is defined and these are often based on activity patterns, distance, travel time and related cost structures that determine spatial associations. In fact, flows of this kind are also associated with networks which scale from topological relations down to physical infrastructures. Currently there is substantial activity in embedding such flow structures in their networks and this is beginning to be reflected in spatial models as is implicit in some of the contributions in this book. In the three examples we use to illustrate the computational model types below, flows and networks are significant. It is worth noting too at this point, that in spatial modelling, most focus has been on measurable physical and hence observable quantities that change through time but increasingly there are hidden flows and relations associated with the electronic world that are influencing how spatial systems change and develop. This too is a major challenge for spatial modelling.

Before we move the discussion on to classify different kinds of dynamics, it is worth noting that all the variables that we have introduced so far can be disaggregated down from their aggregate populations to more disaggregate components. Ultimately the disaggregation is down to individuals where we denote such atomic elements by the subscript k on populations, that is N_{it}^{k} is the $k'th$ individual or group in the appropriate space and at the relevant time in the system. In fact this notation can also be extended to any group which is a subset of the aggregate population such that the sum of the groups and/or individuals adds to the relevant aggregate variables, that is $\sum_k N_{it}^{k} = N_{it}$. In the models, particularly the agent-based models that follow, individuals will form the focus of the simulation where processes are specified for a class of individuals but are operated at the level of each individual in the simulation, usually specific to the space and time within which the individual is located as well as individuals 'proximal in some way' to the object in question.

As the concept of equilibrium has fallen into disrepute and as the spatial models have become more explicitly dynamic, different kinds of time scale and change have been identified which characterise spatial systems. In particular the notion of smooth change has given way to systems that clearly have discontinuities in their behaviour through time (as well as space) where such discontinuities represent thresholds that are crossed, for example, the step function also shown in Fig. 2.1. Here once the growing population reaches the limit, it precipitously declines to its initial value. This classification of dynamics extends all the way to behaviours that generate endogenous discontinuities as is characteristic of catastrophe and bifurcation theories. This portfolio of dynamic behaviours has also been enriched by smooth changes that lead to chaos, systems that behave in entirely unpredictable ways in terms of their initial conditions but are nonetheless deterministic and portray smooth and continuous change. Into this nexus has come the notion that change can generate surprising and novel behaviours. For example, edge cities that suddenly appear around well established metropolitan areas, segregation patterns that do not appear to be embedded in the logic of change but suddenly manifest themselves, and repercussions from changes in one element of the system that cascade and grow as they diffuse to other sectors are all examples of the sort of changes that many models of spatial systems now take as routine.

Dynamics in all these senses has added to the burden of modelling. Like disaggregation, dynamics enriches the model in that data demands become severe and often much of the change that needs to be simulated is hard to observe and match to data. In fact, the notion that dynamics leads to surprising changes is part and parcel of the insights that are coming from complexity theory where the routine operation of space-time processes from the bottom up leads to emergent patterns that only in hindsight can be explained. Such unanticipated behaviour is quite counter to the traditions of well-behaved dynamic systems that tend to converge to an equilibrium or steady state, that is where $N_{it} \rightarrow N_i$ in the limit of t.

The last feature of dynamics that we need to note before we begin to classify spatial models in these terms involves relationships to the external environment, either to the rest of the world or indeed exogenous changes to the population and/or environment that comprise the system in question, for it is by these means through which unusual dynamics can be stimulated. For example the population equation might be subject to external shocks, that is from Eq. 2.1 we might add a shock such as X_{it} leading to

$$N_{it+1} = \alpha N_{it} + X_{it} \tag{2.5}$$

which basically removes a degree of predictability from this particular model, dependent on the size and frequency of the external input. If the external input is once-for-all, its effects may die away but sometimes these kinds of shocks feed on one another and are enough to push the system into uncharted waters with quite unpredictable consequences. Moreover changes in the environment of the system, such as the addition of new capacity a_{it} in terms of land available, say, which we might mirror as

$$\left.\begin{aligned} A_{it+1} &= A_{it} + a_{it} \\ N_{it+1} &= f(A_{it+1}) \end{aligned}\right\} \tag{2.6}$$

can lead to equivalent unpredictability. Even in these simple cases, we can easily complicate the dynamics through additional functions that immediately show that any movement to a steady state is likely to be the exception rather than the rule.

2.4 Six Styles of Spatial Model

It is exceptionally hard to provide a completely comprehensive overview of spatial models in the human domain even with as narrow a focus as we adopt here which is mainly on cities. This is largely because model types shade into one another and many of the features that we have identified in the previous sections appear in more than one model. Different modelling styles merge into one another. Nevertheless various researchers have attempted to classify such models and it is worth noting some of these attempts before we outline our own focus on this field. In general as noted earlier, there has been a sea change from aggregate cross-sectional comparative static models of spatial systems to models that are disaggregate and dynamic. This has marked the transition from *land use transportation interaction* models (LUTI) to *cellular automata* (CA) and *agent-based models* (ABM). This has also represented a change in scale and focus and in the case of CA models, these shift the focus from social and economic processes to physical land development. ABM models are more generic still but in terms of urban modelling, most applications are at the fine spatial scale at the level of pedestrians, for example, and local movement, with only a handful of such models being developed for metropolitan areas. In fact as LUTI models have been disaggregated, then some of these such as ILUTE and UrbanSim have features that can be described as agent-based (Hunt et al. 2005).

The other four types of model that we will classify and define here are those based on less well entrenched applications and methodologies. *Spatial econometric models* (SEM) have been widely applied but often at a larger scale involving regions while *systems dynamics models* (SDM) have been proposed and implemented in some contexts but these have not found widespread application largely because they have not been generalised to spatial systems in any consistent manner. Last but not least there are *microsimulation models* (MM) of which there are several spatial variants and these also tend to merge into ABM at one level of specification. There are no general reviews of all six modelling styles but the author (Batty 2008) provides a discursive discussion of how LUTI models made the transition to CA and ABM during the last 30 years. The short review of LUTI, ABM and CA models also by Batty (2009) focuses on their structure, dynamics and aggregation properties. There are comprehensive reviews of ABM, CA, SDM, MM and some LUTI models by Haase and Schwartz (2009) and there are a series of reviews of operational land use models mainly in the US agencies such as the EPA (see Southworth 1995 for example). However apart from the review of CA models by Liu (2008), most of the reviews tend to be of LUTI models. In particular the chapters by Wegener (2005), Iacono et al. (2008) and Hunt et al. (2005) are good summaries of the state of the art to which the reader is referred. The essence of the models which are the subject of this book – mainly ABM, CA and MM – are contained in the relevant chapters in this section by Birkin and Wu (2011) (MM), Dearden and Wilson (2011) (LUTI-spatial interaction), Iltanen (2011) (CA) and Crooks and Heppenstall (2011) (ABM). In fact the focus is much more strongly on ABM than any other model type in this book although CA models, as we will see, provide an implicit form of ABM. This chapter and more generally this section, do however provide a useful overview of the field with the focus very much on situating ABM in the wider context of spatial modelling.

We will begin with generic models and only when we have reviewed most of these will we look at specific models with methodologies that are precisely configured to the systems and problems at hand. We will treat each model in terms of the eight characteristics which we identified in the previous two sections, namely, environment and population, scale and aggregation, conservation and constraint, disaggregation, feedback in space-time, dynamic type, emergence and convergence, and external inputs, and we will begin with CA models which are by far the simplest. In fact CA models are explicit and simple spatial dynamic models with little or no presumption about the form of the dynamics and rather simple notions about the effect of space. In their strictest form they simulate the spatial diffusion around a point where the diffusion is to immediate neighbours and time and space are treated as one. In this sense, the environment is treated as being synonymous with the population with each state of the system – i.e. the population – being directly associated with a spatial location at a point in time, in short $A_{it} = N_{it}$. Scale and level of temporal and spatial aggregation tend to be quite flexible in these models although for urban and land cover systems, both scales are large – often land parcels and census tracts and above, while temporal intervals are at least for one yearly periods. This however is not a major constraint. Such models do not strictly conserve quantities of population in that there is nothing intrinsic to such models that limits their growth or decline although often such models are subject to more macro-constraints provided by other models in their wider environment. The models can be fairly disaggregate but most applications divide the cell states into land use types limited to no more than a dozen. Feedback in space is extremely simplistic and often unrealistic in that the CA nearest neighbour influence principle which is essential for physical diffusion processes is often not a good analogue for spatial effects where there is action-at-a-distance. Such models do not tend to fall into any particular dynamic class, for if they produce unusual and discontinuous dynamic behaviours, this is likely to be due to external inputs rather than anything built into the model dynamics. Emergence is possible with such models, indeed essential to their original formulation although in urban applications this is generally not a specific focus. All in all, such models tend to simulate land development processes from the supply side or at best models of the balance between demand for and supply of land. They are not strongly socio-economic in that they do not embrace detailed demographics, and in this sense are essentially physicalist in tone.

ABM models have many of the characteristics of CA models except that the environment and population sides of the system are kept apart. The population sector is essentially that which contains these agents whose behaviour is specified in considerable detail. Agents tend to be mobile in a spatial sense and even if they do not physically move in space, they can be associated with different spaces and their change over time can reflect an implicit process of movement. In this sense, the environment is treated more passively than the population with the population driving any change in the environment, although in principle there is no priority for one or the other. A detailed specification of ABM in these terms is contained in Batty's (2005) book where the idea of an agent having a specific behavioural profile and acting on this purposively is central to their definition. In terms of aggregation and scale, ABMs tend to be at smaller scales than the region or the metropolis although some land cover models based on ABM are predicated at these larger scales. They tend not to be constrained in terms of conserving any key quantity although they may be structured to generate or conserve a certain level of population, especially if the focus is on movement in a fixed space as in pedestrian models. Their dynamics and relationships to the wider environment are similar to CA and they tend to be highly disaggregate down to the point where individuals constitute their basic units. Problems emerge when individuals are aggregated to groups or when the agents become agencies for then such models tend to be of more conceptual interest than of predictive practical use.

Like CA and ABM models, microsimulation models (MM) tend to be loosely structured in terms of their dynamics. Such models may even be cross-sectional rather than dynamic but the fact that the populations tend to be represented in terms of their basic units means that such models are usually temporally dynamic, i.e. individuals are represented in terms of their behaviour which is intrinsically dynamic. These kinds of models work on the premise that a population is described in terms of a distribution of characteristics – for example, an income distribution, and individuals are then selected from that distribution so such models are essentially random samples from a much larger universe or population. In this sense, the models can be at any scale but the distributions are usually composed of individuals in that any point sample from a distribution is associated with an individual. Point samples can of course be aggregated into large groups in space and time. There is not much more that can be said about such models for all their other characteristics will depend on the specific model characterisation once it has been worked up to the system in question. Quantities do tend to be conserved and sampling can be subject to some constraints while feedbacks depend on how different sectors in the model are configured in relation to one another. The model dynamics again tends to be straightforward and most models to date (see Birkin and Wu 2012) do not tend to reflect discontinuities of the kind associated with emergence of new structures. External inputs into such models are usually extensive as many of the drivers of such behaviour are reflected in the wider environment. Microsimulation models are essential tools for sampling large-scale populations where it is impossible to represent all the individuals explicitly and where some sense of the heterogeneity of the population needs to be represented in the model. The MoSeS model designed by Birkin and Wu (see this volume) is a good example of how MM is applied to human spatial systems where the focus is on demographics and its relationships to the provision of health and related social facilities at a fine spatial scale.

Spatial econometric models (SEMs) have been widely developed in the tradition of aggregate modelling (Anselin 1988). To an extent such models do not really distinguish between population and environment although the focus in such models is more on subsuming the environment into the population than the other way around in contrast to CA models. Such models are usually developed at a scale where statistical averages are stable and this means that the spatial and temporal units must be such that the data are appropriate for standard statistical inference. Quantities in such models tend to be conserved but within statistical limits although increasingly constraints are put on statistical models where it is essential to keep predictions within bounds. SEMs tend to be structured along rather formal lines where the standard model is linear, often simultaneous in that feedbacks between different model sectors are associated with different model equations, and the dynamics is often well-defined with the equilibrium properties of such models being well-known in terms of their stationarity. Emergent behaviours are not usually a feature of such models but the distinction between exogenous and endogenous variables as in much economic modelling is strong. In this book, these kinds of models are not reported although occasionally, econometric techniques are used in ABM, SDM, and MM.

Systems dynamics models (SDM) are very much in the tradition of the discrete population models that we illustrated earlier in Eqs. 2.1–2.4. In fact these models are based largely on coupled difference equations whose structure is such that they lead to exponential growth followed by damped oscillations around fixed resource limits. In this sense such models are heavily constrained. They can be quite disaggregate dealing with different sectors but the environment is entirely absorbed in the population as there is usually no spatial variation although some models have simply applied what happens in one space to many others. In terms of feedbacks, the entire behaviour of these models is structured around damped logistic growth reflecting repercussions through the model structure which leads to oscillations around the resource limits. In this sense, the dynamic behaviour of these models is well-defined. Links to the wider environment are structured in terms of control over resource limits. Progress with these models has been quite slow with only a limited number of applications largely due to the difficulty of articulating space within their structure. In fact as soon as space is introduced, these models begin to look rather different from traditional SDM and in this sense, they change in focus. Many of these model structures are more like model methodologies that can be merged together in the construction of more elaborate models, as for example, in models such as UrbanSim.

Our last class of models – land use transportation interaction or LUTI models – are quite different in structure. These models are essentially fashioned around ideas in spatial interaction and discrete choice theory, merged with notions about economic input-output analysis, multipliers and demographic modelling that all come together in what are largely aggregate cross-sectional model structures simulating the location of activities and their interactions at a single point in time. These models, like SEM, tend to merge environment into population and since their inception, they have become more disaggregate. Spatial constraints and the concatenation of activities are central to such structures. Various feedbacks between the sectors are incorporated but these usually reflect spatial not temporal effects. In terms of dynamics, such models struggle to embrace the wider portfolio of possibilities being, at best, incremental which essentially involve static models being applied to increments of time. That is, static model structures are used to model incremental change and such models do not attempt to explore longer term dynamics. In fact there are extensions of such models into dynamic frameworks such as those developed by Wilson (2008) but in general, the practicalities of limited temporal data have constrained such models in terms of dynamic simulation. This is an important issue as most of the other models we have described in this section simply assume that the lack of temporal data is not a constraint on their specification and application. In short, LUTI models build on social physics and urban economics which are essentially atemporal.

These model types and styles provide a wide range of possible structures from which to select appropriate forms for specific problems. Our summary shows at a glance the array of model types that we might draw upon in simulating spatial systems in the human domain. In the rest of this review, we will not detail all of these but we will focus on CA, ABM and LUTI models to give some flavour of how they might be developed and the way they are calibrated, validated, and verified in practice. This will set the scene for the rest of the review chapters in this section which take these models types further and develop specific issues with respect to their design and construction.

2.5 Cellular Automata: Physical Simulation Models of Urban Morphologies

CA models are by far the simplest of any urban model in that they merge entirely their populations with their environment. In essence, the components of the environment are identical to the objects comprising the population in the sense that the locational spaces that define the environment at any point in time, are equivalent to each elements of the population. In the simplest case, one cell in the environment is equivalent to one object in the population which in formal terms means that $A_{it} = N_{it}$. Now each cell in a CA model can take on more than one state which means that the population object can vary in its attributes. Again, the simplest form is that a cell can take on one of two states – it can be switched on or off which in urban terms might be compared to the cell being developed or not developed. This is often represented as

$$A_{it} = \begin{cases} 1\ \textit{if}\ i \text{ is developed} \\ 0, \text{otherwise} \end{cases} \tag{2.7}$$

In slightly more complicated CA models, there may be more than one population object in one cell but this probably is the interface between CA and ABM. If a cell has one population object only but that object can take on different attributes or changes in state, then this is still a CA model. In short, when a cell can take on more than two states, then this is usually used to reflect different changes in land cover such as land use types but it could also be associated with different changes in the population object such as its level of income, its age and so on. The formulation is entirely generic.

CA models in their strict sense have no action-at-a-distance except in the most restrictive sense. A cell is deemed to influence or be influenced by its nearest neighbours where near is defined as physically adjacent if the application is to some spatial system. This is the only way in which emergence can be charted in such models in that if the field of influence is wider than nearest neighbours in a regular sense, then it is impossible to trace any emergent effects on the ultimate spatial structure. Essentially CA in this manner is used to implement procedures that lead to fractal structures where patterns repeat themselves at different scales which only emerge when the system in question grows and evolves. We can illustrate strict CA in the following way. Assume that the set Z_I is the set of immediate neighbours on a regular square lattice. The usual neighbourhood is defined as the Moore neighbourhood – all cells at the eight compass points around the cell in question or the von Neumann neighbourhood which are the cells N, S, E and W of the central cell. Then we define a function F_{it} as the concatenation of effects in the Z_I neighbourhood, and if this function takes a certain value, this generates a change in state of the cell in question, cell i. Imagine that the rule – and there can be many, many different rules – is that if this function is greater than a certain threshold Ψ which is a count of the developed cells in the neighbourhood, then the cell changes state.

In the simplest case, it is developed if it is not already developed or its stays developed if already developed. Using the definition in Eq. 2.7, then

$$F_{it} = \sum_{j \in Z_I} A_{jt} \quad \text{and} \tag{2.8}$$

$$if \quad F_{it} > \Psi \quad then \quad A_{it+1} = 1. \tag{2.9}$$

It is very easy to show that this process leads to a regular diffusion starting from a single cell. If we assume that the threshold $\Psi = 1$, all the cells in original Moore neighbourhood around the seed cell get developed first, then all cells around those that have just been developed, and so on with the recursion simply leading to the growth of a square cellular region around the starting cell. In fact in this instance, space and time are collapsed into one which is the key criteria of regular physical diffusion. These ideas are developed in more detail in Batty (2005) to which the reader is referred for many illustrations of such basic strict CA models.

If the CA models are slightly more complicated in terms of their neighbourhood rules then various geometric fractals result while there can be key spatial orientations and biases introduced into the structures that are generated. However it is usual in CA modelling for the neighbourhoods, the rules and the process of generation to be entirely uniform. As soon as the notion of varying neighbourhoods over space and varying rules over time is introduced, the models are no longer CA. In fact many urban applications are not strictly CA models at all but cell-space models, motivated by physical land development problems and raster based GIS map algebras in that they do not generate emergent patterns in any recognisable form and they usually relax the constraints placed on both size of neighbourhood and uniformity of cell transition rules. In Fig. 2.2, we show three typical CA models generated using the Moore neighbourhood. The first is the simple diffusion from a source where any development in any adjacent cell spurs development of the cell in question, the second is simple diffusion from a source using a fractal generating rule where the pattern of cells developed determines the rule, and the third is based on a more complicated pattern of cells in the neighbourhood that steers the growth which in this instance is stochastic in a given direction. These are the kinds of structures that form the basis of such automata and all applications to real systems contain mechanisms of recursion built along the same lines as those used to generate the patterns in Fig. 2.2.

There are several ways in which the strict CA model has been relaxed in developing spatial applications. First it is easy to control the growth of developed cells by imposing some sort of growth rates with respect to different cells. If growth is one unit cell, then various external constraints can be used to control the growth but as in all cases where the homogeneity rules are relaxed, then the CA no longer can generate emergent patterns in quite the simple way in which those in Fig. 2.2 are generated. Moreover to introduce variety and heterogeneity into the simplest

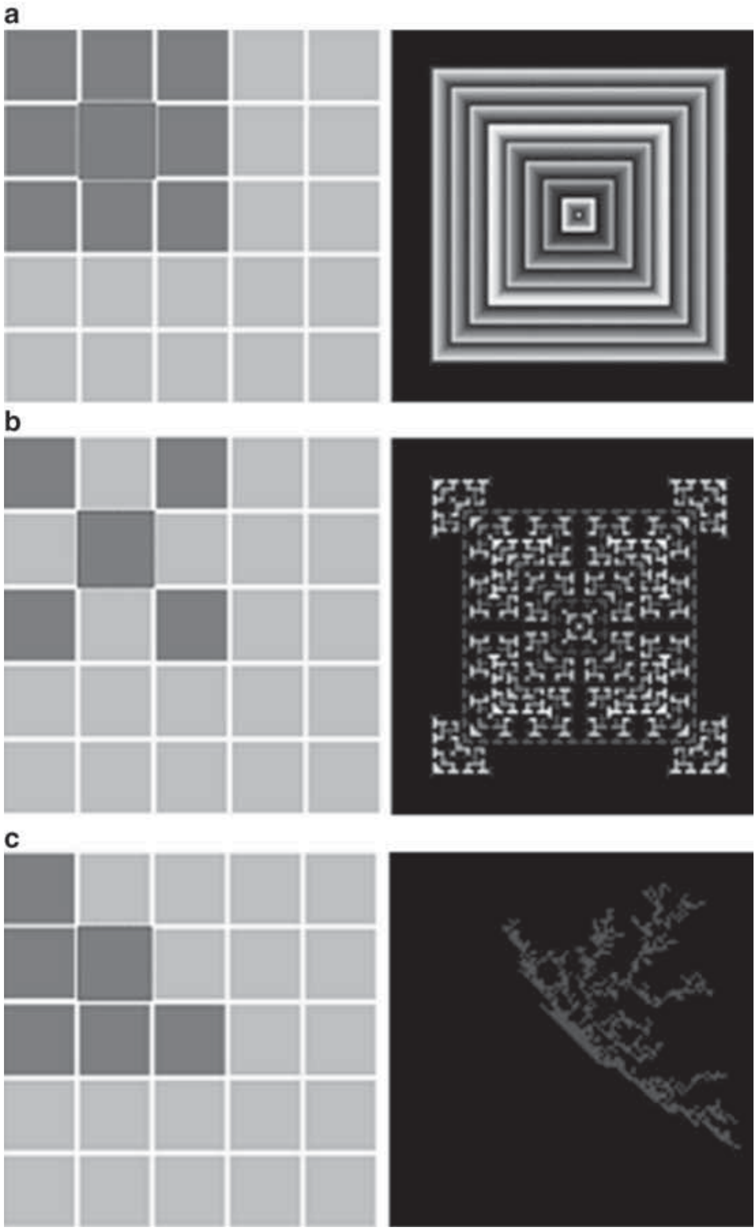


Fig. 2.2 Classic CA models (**a**) Nearest neighbour physical diffusion on a grid (**b**) Koch-like fractal diffusion (**c**) Oriented diffusion limited aggregation

models directly, sometimes the cellular count or concatenation of cells performed in the neighbourhoods is converted to a probability function which is then used to condition the development using a random number generator. For example the structure in Eqs. 2.8 and 2.9 now becomes

$$P_{it} = \sum_{j \in Z_I} A_{jt} / 8, \quad 0 \le P_{it} \le 1 \qquad \text{and} \tag{2.10}$$

$$if \quad rand\,(\Psi) < P_{it} \quad then \quad A_{it+1} = 1 \tag{2.11}$$

where P_{it} is a probability of development and $rand(\Psi)$ is a random number between 0 and 100, say which if less than the probability, implies the land should be developed. There are many adaptations that can be made in this manner but the most significant is related to relaxing the strict neighbourhood rule replacing this with some sort of action-at-a-distance. For example replacing F_{ij} in Eq. 2.8 with the gravitational expression for accessibility leads to

$$F_{it} = \sum_j A_{jt} \big/ d_{ij}^2 \tag{2.12}$$

and this provides a model which can predict development in proportion to accessibility, that is

$$A_{it+1} \propto F_{it} \tag{2.13}$$

This almost converts this cellular automata model to an accessibility potential models which lies at the core of spatial interaction theory and was first developed for these purposes at the very inception of land use transportation modelling (Hansen 1959). The question of course is how such a model might related to the extensive tradition of LUTI models that are in general far superior in their explanatory and predictive power than these kinds of CA model.

One of the major developments of these cellular models is to specify different cell states in terms of different land uses which we will disaggregate and notate as k, A_{it}^k being the appropriate land use k in cell i at time t. In several models, these land uses relate to one another as linkages which determine, to an extent, the locational potential for a site to be developed. Then we might write the change in state of the cell in question as a function of several land uses in adjacent cells where we use a functional notation to simply indicate that the change in question has to be specified in more detail once the model application is implemented. Then the new state of cell i at time t would be

$$A_{it+1}^k = f\left(A_{jt}^\ell, d_{ij}\right) \quad \forall \ell \tag{2.14}$$

where $j \in Z_I^k$ is a neighbourhood defined entirely generically and the field over which distance is defined is again specific to the zone in question. In fact this relaxes the strict CA quite dramatically and is characteristic of many applications (for reviews see Batty 2005, and Liu 2008). It is worth noting that the rules to define land use transitions generally vary the definition of the neighbourhood from the strict no action-at-a-distance principle to the gravitational one. This links different land use states and their densities and types to each land use in question, and also relates these links to different action-at-a-distances effects. These rules also pertain to constraints which are hard and fast on whether a cell can be developed or not. Above a given level, they define how land uses cannot relate to one another. Rules extend to the development of transport links in cells that ensure land use is connected, and structure the regeneration of cells according to various life cycle effects. All of these rule sets are featured in CA models and they are central for example to the SLEUTH, DUEM, METRONAMICA and related model packages that have been developed (Batty, and Xie 2005). They will feature in our brief reference to the DUEM model below.

A more generic CA like structure which is a lot closer to the differential model that dominates the dynamics of physical phenomena at much finer scales is based on a reaction-diffusion structure which might be written in the following way:

$$A_{it+1} = \alpha A_{it} + \beta \sum_{j \in Z_I} A_{it} + (1 - \alpha - \beta) X_{it} \tag{2.15}$$

where α and β are normalising parameters between 0 and 1 and X_{it} is an exogenous variable that reflects changes from the wider rest of the world environment that might be treated as error or noise in the system but more usually is treated as an exogenous shock or as an input that is not predictable by the model. To operationalise this structure, it may be necessary to impose various other constraints to ensure that variables remain within bounds but the essence of the structure is one where the first term on the right hand side is the reaction, the second the diffusion and the third the external input or noise. If we assume that $X_{it} = 0$, the evolution or growth is purely a function of the trade-off between how the system reacts and how activity within it diffuses. In fact, this is rather an artificial structure as change in absolute terms always needs to be controlled and in this sense, external inputs are always likely to be the case. Many CA models do not explicitly adopt this more general structure and a lot of applications have tended to simply scale the outputs of the developed cells to meet exogenous forecasts rather than introducing such exogeneity in more consistent and subtle ways as in the reaction diffusion model in Eq. 2.15.

There are many variants of CA models, examples of which are contained in the last section of this book but as we will see these do tend to merge into ABM. To conclude this section it is worth outlining a model that the author has worked with (see Batty et al. 1999, and Batty 2005). This is the Dynamic Urban Evolution Model (DUEM) which is a fine scale cellular model with several cells states reflecting land use as well as transport and a series of decision rules for changing states that relate one land use to another through its density and accessibility as well as their position in the life cycle of development. The model is largely a pedagogic tool rather than one which can be finely tuned to real situations although a number of applications have been made to the Ann Arbor region and the wider region of South East Michigan which is largely metro Detroit. The model is based on several land uses – residential, commercial, industrial, open space, vacant land and transport/road space – which are functions of the different density and accessibility rules as well as plot sizes which determine how land is developed. We have developed the model for

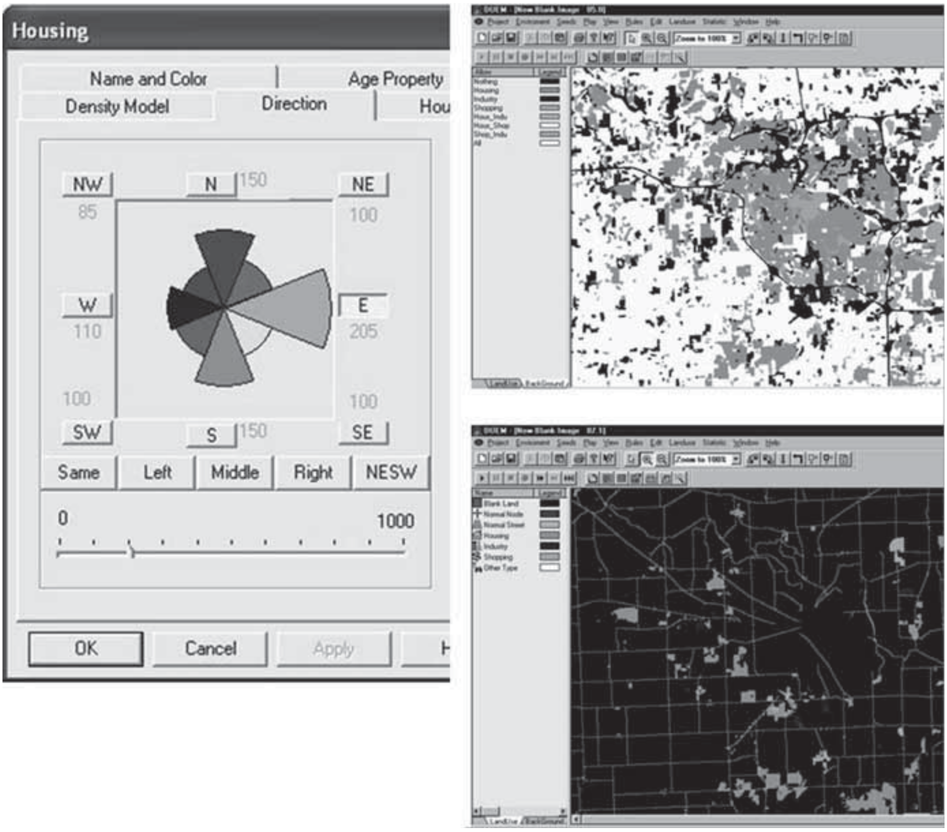


Fig. 2.3 Application of a typical CA model to simulating land use change 1985–1990 and 1990–1995 in Ann Arbor MI

the changes between 1985 and 1990, calibrating the model in a rather crude way. The rule set is large and thus we have not engaged in any kind of exhaustive calibration to find the best fit, although the fit to 1990–1995 from the calibration between 1985 and 1990 is reasonable. We show a segment of the typical interface to the models, showing developed land use in Ann Arbor in 1990, and changes predicted by the model from 1990 to 1995 in Fig. 2.3.

The real critique of CA models relates to their highly physicalist approach to urban structure and dynamics. Essentially these are supply side models, simulating the supply of land based on physical constraints. The notion of demand is foreign to these models as is the notion of interaction as reflected in transport. By abandoning the principles of uniformity, restricted neighbourhoods and homogeneity of states which it is often necessary to do once one applies these ideas, then the models often become poor equivalents to LUTI and other models. However in their favour is the fact that they are explicitly dynamic although dynamic processes other than physical land development do not feature very much in their formulations. Their dynamics is also rather straightforward and if surprising and novel forecasts do emerge, this is more by accident than design in the sense that these models tend to simulate a relatively smooth dynamics. There are not many points at which the kinds of disequilibrium crises and discontinuities which plague the modern world can enter into such models. They also owe a lot to GIS and remote sensing and it is no accident that they have been almost entirely developed by a very different set of researchers from those still working with traditional urban models in the LUTI tradition.

2.6 Agent-Based Models: Purposive Behaviour, Physical Movement and Temporal Change

As we have argued, at one level CA models can be seen as simplified varieties of ABM where the cells form the agents and the states their attributes. Unlike ABM, however, cells do not move and if they change their state, this change might be attributed to some movement but this movement remains implicit and is not formally simulated. ABM implies some form of movement or at least change between agents. Agents as objects in the population are defined individually as k but are made specific in terms of the locations where they exist i at time t. In fact agents may not physically move or indeed in non-spatial models, they may not even be defined in terms of location. If the model is simply one of examining relations between agents at a cross section in time, then such relations might solely be defined in terms of say N^k and N^l, the relation between them defining a link in a social network $N^{k\ell} = f(N^k, N^\ell)$. In fact throughout this book, the agents that are defined by various authors, exist in terms of location and time but very different kinds of relations exist across space. These imply movement or interaction from i to j, from time t to $t+1$ or a later time period $t+T$, from individual object k to ℓ as we have just defined in terms of social network links and any higher order combinations such as: links across space and time, space and different individuals, time and different individuals and across all three – space, time and individuals.

The key difference between CA and ABM is that the system is driven by the ABM where each individual object is endowed with purposive behaviour which conditions their specific and individual behaviour in contrast to aggregate models where this behaviour is part of an aggregate or collective. In this sense, the environment of the system is the space-time frame A_{it} which is relatively passive in comparison to the behaviour of the agents N_{it}^k. Nowhere in such models does $A_{it} = N_{it}$ or vice versa but as we have already implied earlier there are certainly feedback loops $A_{it} \Leftrightarrow N_{it}$ as well as the core loops between agents themselves which we define generically as $N_{it}^k \Leftrightarrow N_{jt+1}^\ell$. We assume in ABM models for spatial systems that the environment is not purposive, that is, no loops such as $A_{it} \Leftrightarrow A_{jt+1}$ exist. If such loops are required then the model would need to be reformulated and part of the environment may then enter the population. The movement of an agent is particularly important in spatial models because whereas in CA, these models tend to be bereft of spatial interaction, ABM models have found extensive application as models of fine scale movement at the pedestrian level for example (Batty 2003). We can formulate such a model in functional terms as

$$N^k_{jt+1} = f(N^k_{it}, N^\ell_{jt}, A^a_{it}, A^a_{jt+1}, z \in Z_I) \tag{2.16}$$

where the superscript a relates to some characteristic attribute of the cell. The functional Eq. 2.16 suggests that agents move through space across time but are influenced by other agents and other locations during such a move. The object N^ℓ_{jt} is in a different location from the moving object k and when the move takes place, a whole series of relations might exist between these two objects such as the visibility of one from another, avoidance of physical contact between one and the other, the clustering of the two or more objects through some social network, or the attributes of the other object being of importance to the locational move, and so on. In terms of the cells themselves, then an object moving from one cell to another would also take account of related cells in the system, usually in the neighbourhood of the move itself.

A good example might be shopping behaviour. An agent enters a shopping centre with a specific purpose to buy goods, encounters other agents along the way, avoids them, or follows them in terms of the crowd. The agent would be influenced by the provision of goods in different cells of the system and in this sense would move in relation to the existence of materials and products that were located in different cells of the system. This kind of characterisation can provide a baseline for movement with visibility, obstacle avoidance, the search for a location which matches the purpose for which the object or agent is moving, and so on. The agent may have a budget and when visiting different cells would exhaust this budget and end the trip once the movement had achieved its purpose. In terms of other moves, then if the agent were migrating over a longer time span in search of a job or house, then the characteristics of the job or house location would be encoded into the environment, in A^a_{it} but the job itself and maybe the actual house would also be part of the set of agents. In this sense, an agent need not be a human individual but an object in the built environment that in and of itself might be subject to change in type and location.

It is worth sketching a simple model of the development process to show how generic this kind of thinking can be. First we make a distinction between consumers k and producers ℓ with N^k_{it} the individual demanding to be housed and N^ℓ_{jt} the developer producing or supplying the housing. The characteristics of the site or cell under consideration for the production of housing is defined as A^a_{zt} where z is a different location but all the locations i, j, z define the cells in the system where consumers and producers carry out their activities. The sequence of actions in any one time period can be orchestrated as follows: first a producer examines all the sites in question which in terms of each site can be represented by $N^\ell_{jt} \leftrightarrow A^a_{jt}$. The decision to produce a house in cell j is then made with respect to the attributes of j but also the potential demand for site j which might be based on previous demand at that site N^k_{jt-1}. The decision is made and the house produced which alters the characteristics of the site A^a_{jt+1}. The production of the house at this site can be defined as a unit of development or level of development D_{jt+1} which a potential house buyer – consumer – will now react to. When the house has been developed, potential residents will examine its location and then decide to occupy it or not, that is $N^k_{jt+1} \rightarrow D_{jt+1}$ and if an evaluation threshold is crossed then the individual will occupy the house, that is the house will be occupied O_{jt+1}. Formally the consumer might evaluate a function which works out a new level of the attribute of the site A^a_{jt+1} which can be formalised as

$$A^a_{it+1} = \phi \sum_{j \in Z_I} O_{jt} + \theta \sum_{j \in Z_I} A^a_{jt} + \vartheta \sum_{j \in Z_I} D_{jt+1} + \varepsilon_{jt+1} \tag{2.17}$$

where the parameters ϕ, θ, ϑ determine the relative weighting and normalisation while the error term ε_{jt+1} is a way of introducing some noise or uncertainty into the locational choice. If the cell attribute value is now above a certain threshold Γ, then the house is occupied; if not it remains unoccupied and the systems move into the next time phase where the process begins once again. Then

$$O_{it+1} = \begin{cases} 1, & if \quad A_{it} = 0 \quad and \quad \sum_{j \in Z_I} A_{jt+1} \geq \Gamma \\ 0 & otherwise \end{cases} \tag{2.18}$$

In this way demand adjusts to supply and vice versa if the system is well specified. Of course this simple model could not be programmed from this formulation for there are other decisions that need to be made to make the process computable but this sketch suffices to show how demand and supply agents interact with their cell space environment to produce and then consume housing. Immediately it is clear that in such a model, although the rules are quite plausible, it is extremely difficult to collect data on such a decision-making process. Moreover at this level of disaggregation, there are many features of the development process that cry out for specification; for example, issues about housing finance and finance for land development, issues about distance from home to work and to other facilities, provision of budgets, life style issues, all crowd into such a model. In a sense, this is why ABMs are so hard to build and test because once this level of detail is broached, it is hard to control the aggregation in such a way as to produce testable propositions. It is worth noting that spatial interaction effects fall out of this model quite easily, thus connecting ABM directly to the LUTI models that we will deal with in the next and final section of this review. The gravitational model of trips can be specified in agent form as

$$T^{k\ell}_{ijt} = \frac{N^k_{it} N^\ell_{jt}}{d^2_{ij}} \tag{2.19}$$

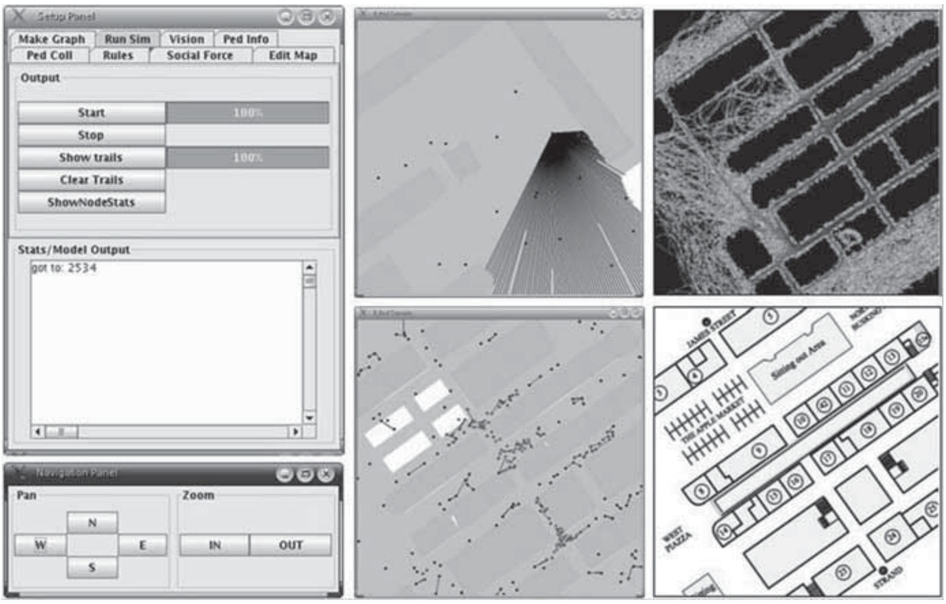


Fig. 2.4 Simulations from Ward's ABM pedestrian model of Covent Garden. *From top left to bottom right*: model control panel, visual lines of sight from a single pedestrian, flow intensity of all pedestrians, navigation panel, interactions of walkers, part of the map of the stalls in the shopping area

where we define k in terms of residence and o in terms of workplace. T_{ijt}^{ko} is the flow from i to j at the cross section and this can be lagged across time if so specified. We can also sum trips over i and j in terms of spatial interaction accounting and this serves to link these models to their aggregate equivalents. In fact, a sequence of locational decisions involving work and residence location in terms of an ABM might actually generate trips of these kinds through individual decisions rather than through this aggregate distance model. This does show that it is possible to begin to introduce social physics ideas into ABM with such connections to discrete choice modelling and microsimulation appearing extremely promising. Similar ideas of movement and spatial interaction are briefly introduced in Batty (2005) to which interested readers are referred.

The last thing we will do in this section is illustrate a typical example of ABM at the pedestrian movement level. In Fig. 2.4, we show a model built for the Covent Garden Market complex in central London by Ward (2007). This model is based on a simple social forces model in which agents have certain tasks to perform such as shopping and entertainment. They have two specific functions: to navigate in search of their goals which involves either purchasing entertainment or goods as efficiently as possible; and to move around the complex in more casual fashion. Most behaviour in this market is a combination of the casual and the formal but a key feature is the learning behaviour that must be built into navigating the space. This set of agents – walkers – is divided into different types dependent on purpose and how much exposure they have already had to the area. A substantial proportion of walkers are tourists. We do not have time to detail the model but it is clear that the nature of the problem imposes quite substantial differences between this application and others in terms of the composition of the agent population and the nature of the facilities in the complex. Again ABM is appropriate because of the rule-based behaviours in this kind of context and because navigation, obstacle avoidance and visibility calculations are important in simulating this type of mobility.

2.7 Land Use Transportation (LUTI) Models: Aggregate Behaviour in Spatial Equilibrium

Our last examples which are not part of the mainstream applications in this book except in the contribution below from Dearden and Wilson (2012), revert back to the origins of computational modelling in spatial systems which are in a rather different tradition from the new paradigms explored in the various contributions that follow. In fact, LUTI models have continued to be developed and strengthened and as we noted earlier, there has been a long quest to retain the advantages of simple aggregate models that can be calibrated against available data in contrast to the need for ever greater detail through disaggregation with the specification of temporal dynamics which move these models outside the equilibrium paradigm. In essence, when the time dimension is suppressed, the representation of environment and system is greatly simplified. The environment is simply indexed by space as A_i while the population is indexed as P_i^k where different activities k now refer to aggregates of populations covering employment, residential population, retail activity and so on. Just as CA models collapse population into the environment, LUTI models tend to collapse the environment into population: all the action in such models is, like ABM, focused on the aggregate with the environment in terms of cells, or zones as they are commonly called, being only relevant when various constraints on land availability and physical features of the space influence the simulation. In short, we can represent such models purely in terms of populations although distance and the attributes of space do occasionally enter the model framework from the environment.

We already have a simple form of LUTI model where spatial interactions are implicit in our development of CA in an earlier section. Equation (2.12) determines the function that converts a cell from one state into another, from undeveloped to developed for example, in terms of gravitational potential and we can write this more generally for any sector k as

$$N_i^k = \xi \sum_j A_j d_{ij}^{-\lambda^k} \tag{2.20}$$

where ξ is the relevant scaling constant, and λ^k is the friction of distance parameter for the gravitational potential. Equation (2.20) might apply to any sector although it is strongly physicalist in form being a function of only land (cell or zone) area A_j and geometric distance (or travel time/cost) d_{ij}. Without any obvious coupling, any LUTI model composed of several different population sectors such as types of residential housing, employment and so on would simply be a series of disconnected models. The most obvious way to connect sectors is to make each sector a function of all others in terms of composite accessibilities that might be written as

$$N_i^k = \xi \sum_{\ell} \sum_j N_j^{\ell} d_{ij}^{-\lambda^{\ell}} \tag{2.21}$$

where we note that the scaling constant is suitably adjusted and that the summation over sectors ℓ may or may not include the self-sector k, a decision that would depend on the precise model specification. In this sense then, the sectors are coupled through their relative spatial distributions.

In fact most LUTI models developed in the last 40 years have specified population as a function of explicit spatial interactions although the first models such as Lowry's (1964) were based on accessibility potentials as in Eqs. 2.20 and 2.21. Using an explicit spatial interaction model, then one of the simplest forms can be written as

$$N_i^k = \sum_j T_{ij}^k = \xi N_i^k \sum_{\ell} \sum_j N_j^{\ell} d_{ij}^{-\lambda^{\ell}} = \xi \sum_{\ell} \sum_j N_i^k N_j^{\ell} d_{ij}^{-\lambda^{\ell}} \tag{2.22}$$

We should note again that the summation is over sectors, that the scaling constant must be suitably adjusted and that there is immediate circularity in the model as the predicted variable appears on both sides of the equation. We do not have time here to dwell on this circularity but it can be resolved in many ways through model specification, balancing and iteration but in essence it reflects the reality of breaking into the spatial system at a cross section in time. In fact, in real applications, the use of appropriate balancing constraints resolves the issue (Batty 1976, 2008).

However the usual way of coupling such models is by assuming that the self-sector is not a function of the model or using another variable such as land area of the zone or cell A_i^k. Then substituting this for N_i^k in Eq. 2.22 and noting now that we will specify a two sector model where $k = 1$ is the first sector and $\ell = 2$, the second sector, then we can write equations for these two sectors as

$$\left. \begin{aligned} N_i^1 &= \sum_j T_{ij}^1 = \xi^1 A_i^1 \sum_j N_j^2 d_{ij}^{-\lambda^2} + X_i^1 \\ N_i^2 &= \sum_j T_{ij}^2 = \xi^2 A_i^2 \sum_j N_j^1 d_{ij}^{-\lambda^1} + X_i^2 \end{aligned} \right\} \tag{2.23}$$

Here we have extended the coupled model even further adding an exogenous input to each sector in the same manner that we did for the reaction-diffusion model earlier in Eq. 2.15 for the CA model. This structure is generic. It can be extended to many other sectors and it is at the basis of a whole class of LUTI models. For example the extended MEPLAN models developed by Echenique (2004) are based on this structure where there are explicit links to input-output models. The original extensions to the Lowry (1964) model were couched in these terms. The first equation in (2.23) was defined for total employment N_i^1 where X_i^1 was basic employment and the second equation was defined for total population N_i^2 where there was no exogenous population, that is $X_i^2 = 0, \forall i$. In short, this is the model structure suggested by Garin (1966) and Batty (1976).

This structure has been exploited in many ways. First it has been disaggregated to embrace many different classes of population with respect to residential population, housing and house types, industrial employment, retailing, commercial and related sectors such as education and health care. Second, relationships between the environment and population have been made in terms of land and density constraints, while third, the spatial interaction models have been extended in terms of utility maximising and route choice building on much more disaggregate individual-based models. In this sense, versions of LUTI models such as UrbanSim (Waddell 2002), ILUTE (Miller 2004) and DELTA (Simmonds 1999) begin to approach ABM illustrating that the line between modelling types and styles can become very blurred. Fourth, the models have been disaggregated to treat ever more zones and spatial units but of course, once these approach ABM, then locations are collapsed directly into individuals within the population and the notion of agents defined by zones has less relevance. Fifthly in many of these models, rule-based algorithms to sort out allocation as in CA models appear alongside more formal equation systems that determine locational distributions. Particularly where demand and supply are explicitly represented, then market clearing and the determination of prices that indicate how the model is balancing are often structured through rule-based mechanisms. As these models have extended their scope, then their formal parsimonious structures have been compromised. Their operation has become more ad hoc and pragmatic which appears to be a consequence of adding more and more detail and more and more sectors.

Dynamics has also been added to such models. At first, such static models were applied to forecast increments of change; that is the static model structure is used to assume that increments or decrements of change observed between two points in time such as $\Delta N_i = N_{it+1} - N_{it}$ become the focus of the prediction. In fact this is often simply a matter of scaling the equations to deal with net change. Many variants of this structure have been developed but there has not been much attention to breaking up the static structure into activities with different propensities to move. There are no models (to the authors knowledge, that is) where populations are divided into movers and stayers and these components dealt with in comparative static terms as different specifications of the equilibrium. Most extensions to dynamics have thus been ad hoc and in fact, there have been few developments of nonlinear dynamics of the kind described earlier involving catastrophes and bifurcations embedded directly into the structure of these models. There are examples where static models are embedded into dynamic frameworks but these are largely for pedagogic use and have never been fitted to real systems (see Dearden and Wilson 2012, this volume). The same might be said of Allen's (1997) work where embedding spatial interaction models into dynamics that lead to bifurcating behaviours in terms of locations are largely illustrative.

In terms of applications, the dominant model in urban and transport planning is still the LUTI model variant, largely because it deals explicitly with transport and housing in terms of their markets and the way they clear. Urban sprawl, for example, which CA models have attempted to simulate is highly dependent on transport and thus LUTI models are preferable as they deal directly with the drivers of sprawl. In North America, the dominant model was DRAM-EMPAL until quite recently when UrbanSim appears to have been more widely applied. Elsewhere MEPLAN and TRANUS have been developed, particularly in South America (Echenique 2004) while in Europe, there has been a mix of models. The focus is less on growth there and thus engagement with these kinds of formal model has been less intense although recently new waves of such models are being applied particularly in the London region. We will conclude our review with a brief summary of some of these models.

The MEPLAN structure developed as the LASER model has been used for 20 years for examining major transport proposals in the South East of England and this is now being supplemented with the LonLUTI model built on the back of the Delta model by Simmonds (1999). We have been developing residential location models as part of the integrated assessment of climate change, specifically flooding and pollution issues, in the Greater London region. This model is a standard structure of the kind presented here with a focus on heavy visualisation. A screen shot of typical output is shown at the top left of Fig. 2.5 where the focus on trip movements and their modal split is clear. It has now been extended using the structure in Eq. 2.23 where there are now three sectors being handled: population, retail and internal population-orientated employment with exogenous employment handled as a separate sector. This model is applied to the outer metropolitan area based on nearly 2,000 zones making the model quite large in spatial scale. The focus is still on fast and immediate visualisation and the current plan is for the model to be disaggregated and different modes to be added. The model is subject to capacity constraints in all sectors including trips and in this sense is quite comprehensive. We show a screen shot of the region in Fig. 2.5 at the top right and below, where it is clear that we are dealing with a complex polynucleated urban system based on a world city with some 14 million population. In contrast to the sort of pictures that we showed earlier for CA models in Ann Arbor (Fig. 2.3), it is clear that these models operate at a higher spatial scale although in the climate change applications, a CA-like model at 50 m grid square scale has been added to the integrated assessment to deal with populations at a much finer spatial scale than the LUTI configuration which is based on zones with an average of 10,000 persons. There is much more we could say about these models but interested readers are referred to this detail in Batty (2011), and Batty et al. (2011).

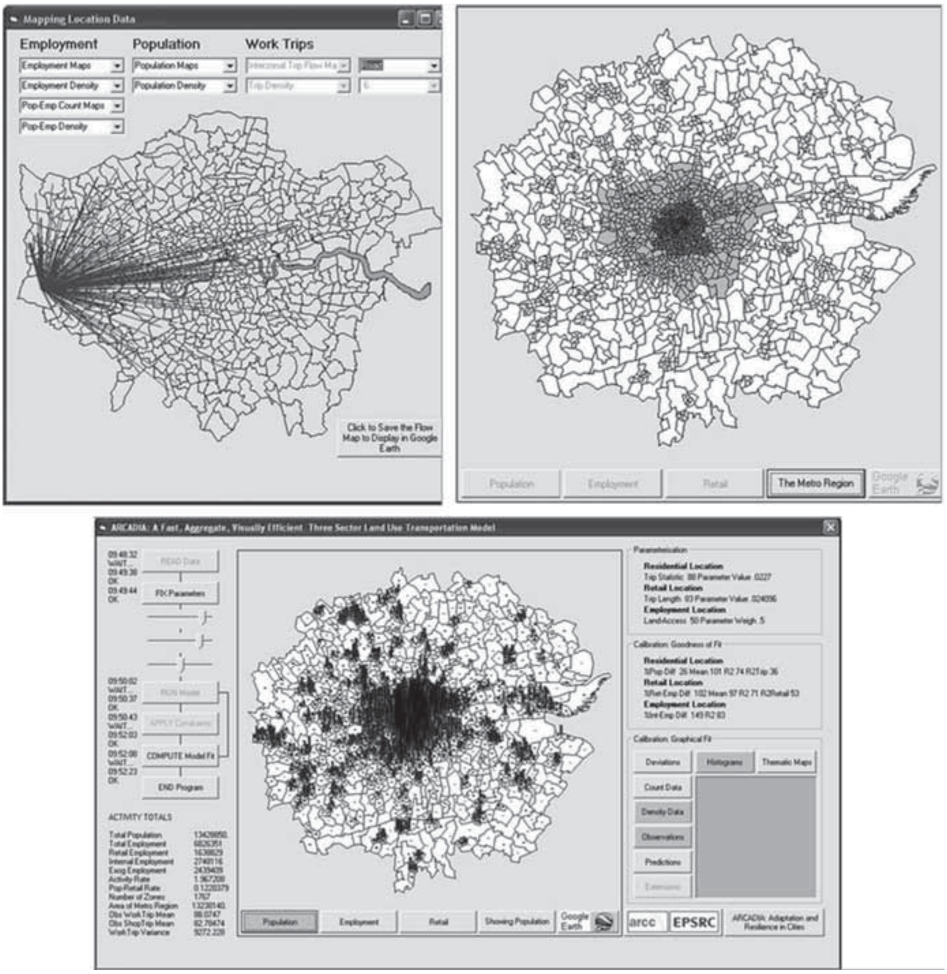


Fig. 2.5 LUTI models for the London region. *Top left*: Work trips from Heathrow in the Greater London residential location model: *Top right*: The nested model applications: *Bottom*: The interface to the 1767 zone London region model showing population histograms

2.8 Conclusions: Modelling Using Generic or Purpose-Built

The model framework developed in this chapter is designed so that readers might see the connections between a variety of model types at different levels of sectoral and temporal disaggregation. It is almost a non-sequitur that static cross-sectional models tend to be simpler to notate than dynamic models but what dynamic models add in terms of temporal richness, static models tend to compensate for in terms of sectoral feedback and strongly coupled activities. The framework we have introduced is certainly generic for the distinction between environment which is the space-time nexus and population which tends to be the driving force of all these models, is common to all spatial models of the kind developed in this book.

The level of aggregation although exceptionally important in terms of applications, is less important in terms of model structure. What we have not done here is dwell on methods of fitting different models within this framework to data and it is worth concluding with some remarks for this serves to polarise differences between the various models.

As the level of detail in terms of sectors, spatial-locational resolution, and temporal resolution increases, data demands generally increase and models become increasingly difficulty to validate in terms of being able to match all the model hypotheses – functions – to observed data. As temporal processes are added, this can become exceptionally difficult but even with cross-sectional static models, when we add mechanisms for coupling and for market clearing as is the case in many LUTI models, we face a severe problem of validation. Many processes in these models cannot be observed and in principle some of these may simply be unobservable. Thus the model-builder faces problems of convincing client and stakeholder groups, which may comprise other scientists, of the veracity of their simulations. This tends to force modelling back to the traditional canons of scientific inquiry where parsimonious and simple models are the main goal of scientific explanation. Occam's razor may still be the ultimate quest but in many social systems, evident complexity is so great that plausibility rather than validity may be the real quest. This tension is felt very heavily throughout this book although it is broached only gently by many of the authors who are clearly conscious of the weight of scientific credibility that these new approaches to social systems impose.

In fact cutting across this dilemma is the notion that as we improve our understanding of spatial systems, we might be able to generalise models to the point where generic software becomes dominant. In fact, quite the opposite is happening. As we learn more we consider each problem context to be more individualistic where the model has to be specifically tailored to the task in hand. Software engineers have in fact sought to develop ever more generic packages but these are often frameworks which guide the modeller rather than establish complete frameworks for the development of a specific model. Most general frameworks for ABM for example such as RePast and Netlogo, even MATLAB and Mathematica, do not extend to the point where very detailed spatial models can be built within their structures. LUTI models are a case in point. 30 years ago when spreadsheets were first developed it was perfectly possible to develop pedagogic versions of such models using that software but no real application would ever fit into such structures. To date, there is no standard software for such models. In fact herein is the dilemma. Most serious applications rather than proofs of concept or pedagogic demonstrations require specific software applications. Insofar as generic software can be used, this provides many of the basic routines but these still have to be assembled in situ by skilled programmers, notwithstanding the fact that downstream applications may emerge which are generic. But then such applications tend to be pedagogic, showing what has been done and any new application requires purpose-built software development. It is hard to see this situation changing in that the problems that we need to engage with always seem to outstrip previous applications and software already developed for these. The various contributions on this book clearly demonstrate this point.

References

Allen, P. M. (1997). *Cities and regions as self-organizing systems: Models of complexity*. London: Taylor and Francis.

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Dordrecht: Kluwer.

Batty, M. (1976). *Urban modelling: Algorithms, calibrations, predictions*. Cambridge: Cambridge University Press.

Batty, M. (2003). Agent-based pedestrian models. In P. A. Longley & M. Batty (Eds.), *Advanced spatial analysis: The CASA book of GIS* (pp. 81–106). Redlands: ESRI Press.

Batty, M. (2005). *Cities and complexity: Understanding cities with cellular automata, agent-based models, and fractals*. Cambridge, MA: The MIT Press.

Batty, M. (2007). Model cities. *Town Planning Review, 78*(2), 125–151.

Batty, M. (2008). Fifty years of urban modelling: Macro statics to micro dynamics. In S. Albevrio, D. Andrey, P. Giordano, & A. Vancheri (Eds.), *The dynamics of complex urban systems: An interdisciplinary approach* (pp. 1–20). Heidelberg: Physica.

Batty, M. (2009). Urban modeling. In R. Kitchin & N. Thrift (Eds.), *International encyclopaedia of human geography* (Vol. 12, pp. 51–58). Oxford: Elsevier.

Batty, M. (2011). Visually-driven urban simulation: Exploring fast and slow change in residential location. CASA working paper 164, University College London. Available at http://www.casa.ucl.ac.uk/working_papers/paper164.pdf

Batty, M., & Xie, Y. (2005). Urban growth using cellular automata models. In D. J. Maguire, M. Batty, & M. F. Goodchild (Eds.), *GIS, spatial analysis, and modeling* (pp. 151–172). Redlands: ESRI Press.

Batty, M., Xie, Y., & Sun, Z. (1999). Modeling urban dynamics through GIS-based cellular automata. *Computers, Environments and Urban Systems, 23*, 205–233.

Batty, M., Vargas-Ruiz, C., Smith, D., Serras, J., Reades, J., & Johansson, A. (2011). Visually-intelligible land use transportation models for the rapid assessment of urban futures. CASA working paper 163, University College London. Available at http://www.casa.ucl.ac.uk/working_papers/paper163.pdf

Birkin, M., & Wu, B. (2012), A review of microsimulation and hybrid agent-based approaches. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 51–68). Dordrecht: Springer.

Crooks, A. T., & Heppenstall, A. J. (2011). Introduction to agent-based modelling. In A. J. Heppenstall, A.T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 85–105). Dordrecht: Springer.

Dearden, J. & Wilson, A. (2012). The relationship of dynamic entropy maximising and agent-based approaches in urban modelling. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of Geographical Systems* (pp. 705–720). Dordrecht: Springer.

Echenique, M. H. (2004). Econometric models of land use and transportation. In D. A. Hensher, K. J. Button, K.E. Haynes, & P. R. Stopher (Eds.), *Handbook of transport geography and spatial systems* (pp. 185–202). Amsterdam: Pergamon.

Forrester, J. W. (1969). *Urban dynamics*. Cambridge: The MIT Press.

Garin, R. A. (1966). A matrix formulation of the Lowry model for intrametropolitan activity location. *Journal of the American Institute of Planners, 32*, 361–364.

Haase, D., & Schwarz, N. (2009). Simulation models on human–nature interactions in urban landscapes: A review including spatial economics, system dynamics, cellular automata and agent-based approaches. *Living Reviews in Landscape Research*, *3*, 5–35. Available at http://www.livingreviews.org/lrlr-2009-2

Hansen, W. G. (1959). How accessibility shapes land use. *Journal of the American Institute of Planners, 25*, 73–76.

Hunt, J. D., Kriger, D. S., & Miller, E. J. (2005). Current operational urban land-use-transport modelling frameworks: A review. *Transport Reviews, 25*(3), 329–376.

Iacono, M., Levinson, D., & El-Geneidy, A. (2008). Models of transportation and land use change: A guide to the territory. *Journal of Planning Literature, 22*, 323–340.

Iltanen, S. (2011). Cellular automata in urban spatial modelling. In A. J. Heppenstall, A. T Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of Geographical Systems* (pp. 69–84). Dordrecht: Springer.

Liu, Y. (2008). *Modelling urban development with geographical information systems and cellular automata*. Boca Raton: CRC Press.

Lowry, I.S. (1964). *A model of metropolis*. Memorandum RM-4035-RC. Santa Monica: The Rand Corporation.

Lowry, I.S. (1965). A short course in model design. *Journal of the American Institute of Planners, 31*, 158–165.

Miller, E. J. (2004). Integrated land use transport model requirements. In K. J. Button, K. E. Haynes, P. Stopher, & D. A. Hensher (Eds.), *Handbook of transport geography and spatial systems* (Handbooks in Transport, Vol. 5, pp. 147–162). New York: Elsevier Science.

Plummer, A.V. (2007). The Chicago area transportation study: Creating the first plan (1955–1962): A narrative. Available at http://www.surveyarchive.org/Chicago/cats_1954-62.pdf

Simmonds, D.C. (1999). The design of the DELTA land-use modelling package. *Environment and Planning B: Planning and Design, 26*, 665–684.

Southworth, F. (1995). *A technical review of urban land use-transportation models as tools for evaluating vehicle travel reduction strategies, ORNL-6881*. Washington DC: Center for Transportation Analysis Energy Division.

Waddell, P. (2002). UrbanSim: Modelling urban development for land use, transportation and environmental planning. *Journal of the American Planning Association, 68*, 297–314.

Ward, J. (2007). *Urban movement: Models of pedestrian activity*. Unpublished PhD thesis, CASA, University College London.

Wegener, M. (2005). Urban land use transportation models. In D. J. Maguire, M. F. Goodchild, & M. Batty (Eds.), *GIS, spatial analysis, and modeling* (pp. 203–220). Redlands: ESRI Press.

Wilson, A.G. (2008). Boltzmann, Lotka and Volterra and spatial structural evolution: An integrated methodology for some dynamical systems. *Journal of the Royal Society Interface, 5*, 865–871.

Chapter 3
A Review of Microsimulation and Hybrid Agent-Based Approaches

Mark Birkin and Belinda Wu

Abstract In this chapter we introduce an approach to individual based modelling of social and economic systems. Microsimulation models (MSM) appear similar to ABM through the representation of individual decision-making units, but there is a significant variation of emphasis between the two approaches. MSM are typically stochastic or rule-based, and with a strong applied policy focus. These characteristics are explored and elaborated through a number of examples. While MSM are often very rich in their representation of 'structures', ABM are usually better tuned to the analysis of 'behaviours'. We therefore argue that there is a strong logic to considering the MSM and ABM approaches as complementary and to begin a search for hybrids which might combine the best features of both approaches.

3.1 Introduction

Microsimulation models (MSMs) were introduced in the late 1950s by Guy Orcutt as a reaction to the failure of aggregate models to effectively represent the diversity of economic systems. Later developments have shown that the ambition of Orcutt's initial vision – the creation of a 'new type of economic system' (Orcutt 1957) – to be far from overstated. According to Gilbert and Troitzsch (2005), the distinguishing feature of MSMs is the desire to model interactions between the design and implementation of policies and individual decision making units (e.g. what is the effect of a changing tax regime on individual workers and their households). In contrast, cellular automata (CA) and agent-based models (ABMs) attempt to model the complexity of social systems with similar individual level representations, but

M. Birkin (✉) • B. Wu
School of Geography, University of Leeds, Leeds, UK
e-mail: m.h.birkin@leeds.ac.uk; b.wu@leeds.ac.uk

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_3,

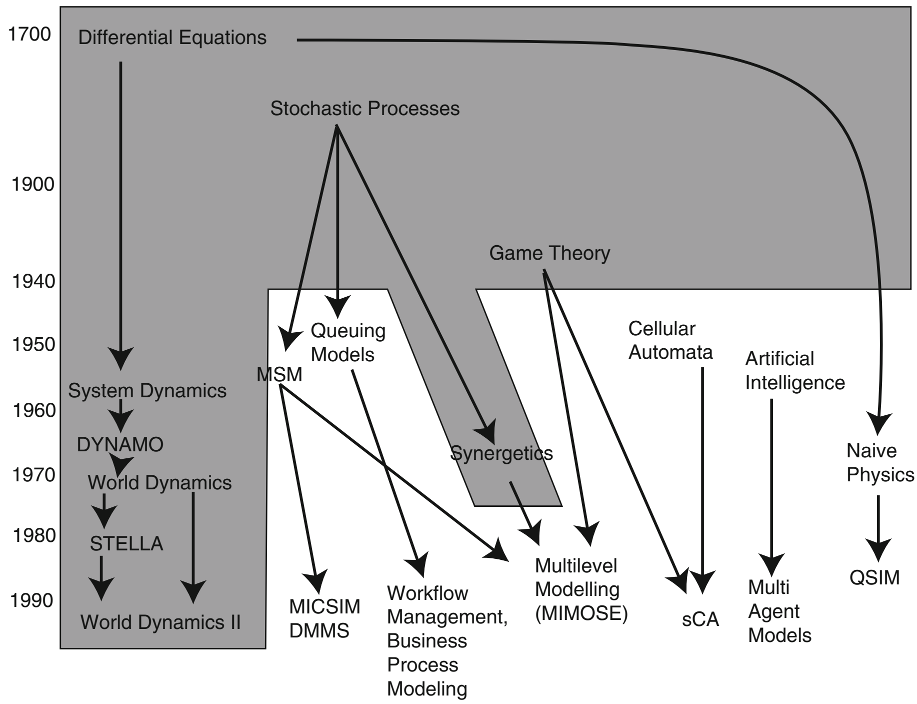


Fig. 3.1 Historical development of contemporary approaches (Source: Gilbert and Troitzsch 2005)

with a somewhat different emphasis. CA model social dynamics with a focus on the emergence of properties from local interactions while ABMs simulate more complex situations than the CA where the 'agents' control their own actions based on their perceptions of the environment. The relationship between these approaches is illustrated schematically in Fig. 3.1. An important feature here is the distinction between individual-based models and 'differential equation models', which focus on system level dynamics at the macro-scale. Figure 3.1 also seems to suggest a significant level of distinction between 'stochastic processes' (MSM), 'cellular automata' (CA), and 'artificial intelligence' (ABM).

The chapter is set out as follows. We first introduce microsimulation modelling as an approach to public policy analysis with a discussion of the most important characteristics and features of these models. The second half of the chapter deals with hybrid approaches through which we will explore the notion that some further fusion of these approaches could be desirable. Conclusions are drawn in the final section.

3.2 Microsimulation Models of Public Policy

Modern social science studies often require detailed information on the interactions between policy and the social-economic behaviours of people. MSMs capture such interactions through the simulation of distinctive behaviours and characteristics at the level of individual decision making units (Orcutt 1957). Advances in computing power and analytical techniques now allow great sophistication in the range of questions that MSMs can address.

A MSM works on the principle of creating small area microdata at a certain point in time and then generating future microdata from that basis (Ballas et al. 2005a). We start with a population of entities, set P, made up of individuals [P^1, P^2, …, P^n] where n is the number of individuals in the population sample. Each individual has a set of attributes, [a_1^t, a_2^t, … a_m^t], which describe the individual at time t. We therefore have an $n \times m$ array of person attributes. This array needs to be populated with reliable data or estimates (in the light of directly surveyed information, etc.). Then we update the population so that the baseline population [$P^1_{a_1^t a_2^t \ldots a_m^t}$, $P^2_{a_1^t a_2^t \ldots a_m^t}$, … $P^n_{a_1^t a_2^t \ldots a_m^t}$] changes to new sets with attributes/states at a point in time $t+1$, $t+2$, … and so on.

One of the most important advantages of MSM is that it enables us to examine the impact of policy changes on individual decision units, as it is based on unit records. This distinguishes MSM from the traditional mathematical models. Such models are often based on aggregated or averaged values and individual characteristics can often become blurred and even disappear in such models. MSM deals directly with social processes at the individual level, therefore it has been extensively used for various purposes in studies for which individual characteristics are important.

Although microsimulation modelling can be traced back to the pioneering work of Guy Orcutt (1957), the work of the Swedish geographer Torsten Hagerstrand was leading in a very similar direction at a similar time in the domain of migration and later innovation diffusion and location theory. Orcutt's research resulted in DYNASIM (Dynamic Simulation of Income Model) (Orcutt 1957; Orcutt et al. 1976), which has been used for a range of studies and inspired the development of many other MSMs. Among them is Steven Caldwell's (1998) CORSIM (Cornell Microsimulation Model), which models large scale government programs and is particularly strong in modelling the Social Security Programme. CORSIM constructs a database on the basis of a sample of 180,000 persons (70,000 families) from the 1960 US Census with demographic and economic attributes. CORSIM simulates changes of each individual unit (persons and families) on a yearly step. The resulting data are validated and aligned using the available external data before projections into the future.

CORSIM has a wide range of modules and therefore can be used extensively in different policy application domains. In 1995, CORSIM was selected by the Canadian government as a template for its own model development and the 'Canadianised' sister model was soon released as DYNACAN (Morrison 2003). In 1997, partly inspired by the Canadian strategy, the Swedish also selected CORSIM as the starting point of a new Swedish dynamic microsimulation model (SVERIGE), focused on exploring person-environment interactions (Rephann 1999).

3.3 Application Areas of Public Policy MSMs

In this section we aim to provide an introduction and overview to MSMs using examples from four domains in which the deployment of these models has been especially productive. The domains are tax and benefits, pensions, health and transport.

Of course, this range of domains is not entirely complete: for example, other examples can readily be found in anthropological systems, urban housing and local labour markets. It does however encompass the majority of work in microsimulation and can be used as a basis for understanding the nature of this research and its significance for individual-based modeling more generally. A synopsis of some important models is provided in Table 3.1.

Taxation and benefits is a core area building directly from Orcutt's original interest in economic systems. Tax-benefits models such as those shown in Table 3.1 aim to combine detailed representations of individual and household structures with well-defined rules about their financial entitlements. For example, if the annual (earned) income of an individual is £50 K, then the marginal rate of tax is 40%. Such models can then be used in a natural and conceptually straightforward way to examine the 'what if?' ramifications of changes in the rules (e.g. what happens if the marginal rate of taxation in the £50 K band is increased from 40% to 50%?). Because the representation of individual and household characteristics is so detailed in the MSM, this method is suitable for highly refined analysis which is often required here. Thus if housing benefits are assessed against, say, the income, occupation, and age of household heads, the composition of the family, and the tenure and physical size of the residence, then these characteristics and the associated benefit rates can all be represented relatively easily in the MSM.

Pension Microsimulations such as PRISM (Pension and Retirement Income Simulation Model) and PENSIM are typically used with a view to the future. In this way the National Insurance or other contributions of the existing workforce can be balanced against the specific entitlements of the retired population. Various policy responses to the impact of ageing populations in developed economies may be tested and evaluated. Similarly Health Microsimulations provide a powerful means to explore changes in the age and social composition of populations on the requirements for medical treatment and care. In addition to the national models outlined in Table 3.1, recent examples have begun to explore the implications of micro-demographic structure on morbidity and the spatial deployment of health care services such as diabetes (Smith et al. 2006, see also Smith, 2012), obesity (Edwards et al. 2011) and smoking-related illnesses (Tomintz et al. 2008). **Transport microsimulations** can be used for both transport policy assessment and simulation of a transport system or its components. The sheer breadth of these models is impressive, spanning all the way from microscopic simulation of individual vehicles to the representation of aggregate network conditions in a region or urban area.

The previous commentary, and the detail of Table 3.1, allows us to propose some useful conclusions about the technique of MSM. Firstly, flexible aggregation is a major strength of this approach. In the financial examples, we are primarily interested in the net effects of a rule e.g. how much benefit will the UK Exchequer derive from an increase in taxes, and what will be the distributional consequences of this change across social groups? The multiple application of rules across individual units is essentially a means to this end. Similarly in transport applications, the rules for individual vehicles may be quite detailed, but the ultimate objective is to say some-

Table 3.1 Comparison of static and dynamic MSM

Model name & domain	Origin	Description and example applications	Indicative reference(s)
(a) Tax-benefits			
TRIM (Transfer Income Model)	US	Simulates the major governmental tax, transfer, and health programs that affect the population; understand the potential outcomes of public policy changes such as welfare reform, tax reform, and national health care reform	Beebout and Bonina (1973)
POLIMOD	UK	Demonstrate how VAT, National Insurance Contributions and Local Taxes are calculated under different assumptions; entitlement to retirement pension and other non-means-tested social security benefits	Redmond et al. (1998)
STINMOD	Australia	Static micro-simulation model of the tax and transfer systems. The rules of government programs are applied to individuals and aggregated to calculate outcomes for income units, families, or households	Lambert et al. (1994)
ITEP	US	Calculates revenue yield and incidence of federal, state and local taxes by income group. It calculates revenue yield and proposed amendments to current law. To forecast future revenue and incidence the model relies on government or other widely respected economic projections	Ettlinger and O'Hare (1996)
EUROMOD	Europe	Tax-benefit model that covers 15 countries. It provides estimates of the distributional impact of changes to personal tax and transfer policy at either the national or the European Level	Sutherland (2001)
(b) Pensions			
PRISM	UK	Dynamic microsimulation of income from social security, earnings, assets, public and private occupational pensions and retirement savings plans	Kennell and Sheils (1990)
SfB3	Germany	Analyse pension reforms, the effect of shortening worker hours, distributional effects of education transfers, inter-personal redistribution in the state pension system	Galler and Wagner (1986)
PENSIM	UK	Simulate UK pensioners' incomes up to the year 2030 and to facilitate pension reform	Curry (1996)
DYNACAN	Canada	Projects the incidence, average levels and variation in private pensions into the future as a function of birth-year, age, and gender	Morrison (2003)

(continued)

Table 3.1 (continued)

Model name & domain	Origin	Description and example applications	Indicative reference(s)
(c) Health care			
PBS	Australia	Expenditure on pharmaceuticals by different types of households; resultant government outlays under the Pharmaceutical Benefits Scheme; and the remaining patient co-payment contributions	Walker et al. (1998)
LIFEMOD	UK	Model the lifetime impact of the welfare state through examination of health status over the life-course and implications for health care financing in the UK	Falkingham and Hills (1995), Propper (1995)
LIFEPATHS	Canada	A dynamic longitudinal microsimulation model of individuals and families which simulates the discrete events that together constitute an individual's life history	Wolfson and Rowe (1998)
(d) Transport			
DRACULA	UK	Simulate response of traffic to different network layouts and control strategies; measure network performance from outputs of the average travel time, speed, queue length, fuel consumption and pollutant emission	Liu et al. (1995)
PARAMICS	US	Microscopic simulation of a range of real world traffic and transportation problems handling scenarios ranging from a single intersection, to a congested freeway or the modelling of an entire city's traffic system	Laird et al. (1999)
VISSIM	Germany	Models traffic flow in urban areas as a discrete, stochastic, time step based microscopic model, with driver-vehicle-units as single entities. The model contains a psycho-physical car following model for longitudinal vehicle movement and a rule-based algorithm for lateral movements (lane changing)	PTV AG (2000)
AIMSUN	Spain	An evolutionary model of the transportation system which combines individual vehicle movements with network data such as traffic lights and detectors as well as road segments. Individual components are simulated in both continuous and discrete time-steps	Barceló et al. (1999)

(continued)

Table 3.1 (continued)

Model name & domain	Origin	Description and example applications	Indicative reference(s)
TRANSIMS	US	Predicts trips for individual households, residents and vehicles rather than for zonal aggregations of households. A regional microsimulation executes the generated trips on the transportation network, modelling the individual vehicle interactions and predicting the transportation system performance e.g. road speeds and motor vehicle emissions	TRANSIMS (1996)

thing about the underlying transport network and its configuration. This looks like a significant variation in emphasis from other approaches such as ABM where the interest in micro-level behaviours 'for their own sake' is much more fundamental.

It is clear that the range of applications for MSM is hugely varied in relation to both geography and substantive problem contexts. One outstanding feature which is common to all of the models in Table 3.1 is their policy-relevance. A recurring theme is the idea of 'what if?' simulations in which the impact of new policy rules on the whole system, or individual components and sub-groups can be assessed. Alternatively, the rules may stay relatively constant but the underlying conditions are changing fast, for example in the case of demographic changes which can have profound implications for taxation, pensions, health care and transport systems.

Finally, while many of the applications are described as 'dynamic' it is necessary to retain a healthy degree of skepticism as to the precision of this term. Often dynamics in these models will be little more than a cross-sectional or comparative static assessment of some globally assumed shift in the composition of the population. The incorporation of pure dynamics in which the individuals in the population actually evolve through time, whether through stochastic rules or more complex behavioural model processes is much more demanding and unusual. This feature is sufficiently important to demand further expansion in the next section.

3.4 Dynamic Microsimulation

Generally speaking, static MSMs do not have direct interaction of microanalytic units within the context of the model during the time period simulated. Static models normally are either deterministic or stochastic. In a static microsimulation, change of the demographic structure in the model is performed by static ageing techniques. Typically such techniques take a large representative sample with detailed information and apply modified laws to it to generate the synthetic demographic and economic characteristics expected in the future year. Simulations can estimate the impact of a change in the future year As the change of the demographic structure of the modelled population is performed by reweighting the age class according to external information, it is focused on what consequences of external information brings to the population and therefore it does not model the changes in population itself. A typical "What-if" Static MSM scenario would be: if there had been no poll tax in 1991, which communities would have benefited most and which would have had to have paid more tax in other forms? (Ballas et al. 2005b; Gilbert and Troitzsch 2005; Citro and Hanushek 1991)

Most tax-benefit MSMs are static. Examples of static microsimulations include models such as TRIM (Beebout and Bonina 1973), POLIMOD (Redmond et al. 1998), STINMOD (Lambert et al. 1994) and EUROMOD (Sutherland 2001). Descriptions of such models can be found in the previous section.

Dynamic MSM can be considered as a technique where entities change their characteristics as a result of endogenous factors within the model. Various degrees of direct interaction between micro population units can be found in dynamic MSMs. Such interaction typically includes processes such as birth and marriage etc. Dynamic microanalytic models rely on an accurate knowledge of the individuals and the dynamics of such interactions. In a dynamic MSM, the updating of the demographic structure is performed by ageing the modelled population individually (by asking "yes or no" questions on birth, death, marriage etc.) with transition probabilities according to life tables and/or exogenous time series. Thus the changes in the population itself are modelled and the simulation in 1 year may affect an individual unit's characteristics in the subsequent year. A typical future-oriented "what if" Dynamic MSM Scenario would be: if the current government had raised income taxes in 1997, what would the redistributive effects have been between different socio-economic groups and between central cities and their suburbs by 2011? (Birkin et al. 1996; Ballas et al. 2005b; Gilbert and Troitzsch 2005; O'Donoghue 2001).

A variety of models have been developed to explore the distributional consequences of demographic change, such as ageing, social mobility and labour market transitions. Thus the **DYNASIM** model ages individual and family characteristics by year, simulating demographic events as births, deaths, marriages and divorces and economic events as labour force participation, earnings, hours of work, disability onset, and retirement. It models a wide range of topics, including Social Security coverage and benefits, pension coverage and participation, benefit payments and pension assets, as well as home and financial assets, health status and living arrangements etc. (Favreault and Smith 2004).

In a similar way **DYNAMOD** (Harding 2002) uses discrete event simulation to age a 1% sample from the Australian census (about 160,000 individuals) on a monthly basis for up to about 60 years. Assets and superannuation have been adding to DYNAMOD to facilitate the research of the likely future retirement incomes of Australians.

SAGEMOD (Zaidi 2004) is a dynamic demographic/tax model which not only estimates incomes but also estimates a random-effects cross-sectional wage equation which included some individual wage history data with the error components. The impact of other labour market states (unemployed, inactive, student) in previous years has been investigated on the earnings of currently employed individuals.

Static and dynamic MSM each have their own strengths. Static models are regarded as more effective at times for specific short run projection purposes because of their greater simplicity and the often lower costs associated with building such models and obtaining computer generated model solutions. Another advantage of static models is that they have very detailed programme simulations. From the computational viewpoint, static MSMs demand less computing resource.

However, dynamic models feature more detailed and realistic population ageing. There is general acceptance that dynamic models provide a more realistic representation of micro population unit behaviour. Dynamic models are also viewed as better at producing realistic long-term estimates, which account for interim changes in economic and demographic trends (O'Donoghue 2001). Due to the interactions/interdependencies of the updating, one limitation is that dynamic MSMs are computationally demanding, even for high-speed modern machines (Ballas et al. 2005b; Gilbert and Troitzsch 2005; Citro and Hanushek 1991; McDonald, et al. 2006).

3.5 Spatial MSM

Spatial MSM is a special type of MSM that simulates virtual populations in given geographical areas (Ballas et al. 2005b). In a spatial MSM, local contexts can be taken into account when studying the characteristics of these populations. Such MSMs are concerned with the creation of large-scale datasets estimating the attributes of individuals within the study area and are used to analyse policy impacts on these microunits (Birkin and Clarke 1995; Clarke 1996). Spatial microsimulation models therefore have advantages over other microsimulation models in exploration of spatial relationships and analysis of the spatial implications of policy scenarios. A spatial MSM can be either static or dynamic.

Spatial MSM was first studied by Hägerstrand (1985) since the 1950s by first introducing the spatial and temporal dimensions into social studies. Wilson (1967), Clarke (1996) and Birkin and Clarke (1995) extended the theoretical framework over the years. Various spatial microsimulations have been developed, including both static and dynamic microsimulations. They allow data from various sources to be linked and patterns to be explored at different spatial scales with re-aggregation or disaggregation of the data. Furthermore they allow updating and projecting, which is of particular importance in forecasting future patterns (Clarke 1996; Ballas and Clarke 2001).

Examples of such models include: **SVERIGE** in Sweden (Rephann 1999). This dynamic population model is designed to study human eco-dynamics (the impact of human cultural and economic systems on the environment). Its main distinguishing characteristic is that it simulates spatial location and mobility of every individual in the data. The model took the CORSIM model framework as a starting point, adapting behavioural modules to be Swedish specific. The migration module attempts to model locational transitions to an accuracy of 100 m.

SimBritain (Ballas et al. 2005c) is a dynamic simulation attempting to model the British population at different geographical scales up to the year 2021. Datasets used in this model are the 1991 UK Census Small Area Statistics (SAS) data and the British Household Panel Survey (BHPS). Microdata for all wards in Britain have been generated through re-weighting the original BHPS data. Previous census data from 1971, 1981 and 1991 (SAS) have been used for projections of a set of small area statistics. Using these three time points, a trend curve was produced allowing tables to be predicted up to 2021.

SMILE in Ireland (Ballas et al. 2005a) is a dynamic spatial microsimulation model designed to analyse the impact of policy change and economic development on rural areas in Ireland. The core model of SMILE is a demographic model. It simulates the basic components of population change, fertility, mortality and internal migration and projects population change at the sub-county level.

HYDRA in the UK (Birkin et al. 2005) is a GRID enabled decision making support system for health service provision. Microsimulation can be run using different parameter sets by the user to find out the optimised location of services for specific queries (further details can be found in Wu and Birkin, 2012).

3.6 Towards a Hybrid Modelling Approach

Over the years, MSMs have been proved to be successful in modelling social systems, especially in facilitating public policy making and development. Large scale MSMs enable us to explore the interaction between policy changes and narrowly defined ranges of individuals or demographic groups, yet retain the heterogeneity in the population as revealed in the large household surveys. The capability of MSMs to replicate complex policy structures also allows us to forecast the outcomes of policy changes and 'what if' scenarios. However, there are also criticisms levelled at MSMs which include:

- MSMs require large datasets with high quality;
- Microsimulation model developments are normally computing intensive;
- Large scale microsimulations can take a long time and considerable effort to accomplish;
- Microsimulation only models one-direction interactions: the impact of the policy on the individuals, but not the impact of individuals on the policy;
- Microsimulation models are less strong in behavioural modelling; and
- It is difficult to validate MSMs (Krupp 1986; Williamson 1999; Citro and Hanushek 1991; O'Donoghue 2001; Gilbert and Troitzsch 2005).

Some of these limitations are better handed by individual-based models such as cellular automata (CA) and agent-based models (ABMs). More details about both modelling types can be found in Iltanen (2012) and Crooks and Heppenstall (2012). Although MSMs, CA and ABMs each have a different focus, they all model the studied system at individual levels, and there is some common ground among the three approaches. Firstly, all three approaches are simulations based on the global consequences of local interactions of members of a population. Unlike the aggregated models that often overlook the details at a more refined level, they provide a more effective and natural way to handle individual behaviours. Secondly these three approaches all track the characteristics of each individual through time, in contrast to traditional modelling techniques where the characteristics of the population are averaged together. Finally the emergence of global phenomena through local interactions in all IBMs (individual-based models) offers more than changes that are simulated on the basis of average data for the whole population as in traditional models.

With the advance in computing, the first three limitations have been improved greatly and new technologies such as ABM can provide the capability for behaviour modelling and allow us to study the interaction at both macro and micro levels, as well as interactions in both directions. However, despite the usefulness of ABM as described in the previous discussion, being a relatively new technology, it sometimes lacks more refined and well-established theories and concepts (Gilbert and Troitzsch 2005; Conte et al. 1998). ABM is also known as hard to validate. Many applications of agent systems to public or social policy domains involve the development of alternative scenarios to facilitate decision-making. However, there is no formal theory of scenarios and scenario analysis that tell us how to construct scenarios, how many scenarios to construct and how to reason between and across their outcomes. Developing formal theories of scenarios and rigorous methods of performance assessment for ABM will require collaboration between computer scientists, philosophers and decision theorists, as well as the domain experts to which these systems are applied.

Despite the work that remains to be done, agent-based social simulation can provide insight into the structure and effects of policies and norms and can assist in understanding and modifying interaction patterns where appropriate and possible (Luck et al. 2003).

To address the limitations of ABMs and MSMs as individual approaches, we suggest further development of a hybrid modelling approach that integrates the strengths of both approaches together. The main reasons for the proposal of such a hybrid approach include:

- MSM and ABM complement each other;
- Geography provides a bridge to link MSMs and ABMs;
- Previous attempts of hybrid approaches have resulted in fruitful outcomes;
- A hybrid approach may provide a new angle to view classical problems.

The following sections review various relevant hybrid modelling approaches.

3.6.1 ABM and MSM

It is generally agreed that MSMs provide important and effective tools for modelling in social science. Recent advances have helped to mitigate some of the major weaknesses of MSM as outlined above at the start of Sect. 3.6 (Holm et al. 1996).

In particular, high quality data is now much more widely available, and large scale process intensive simulations are better supported by the computational abilities of contemporary hardware. However the robustness of the behavioural basis to MSMs can still be questioned. According to Davidsson (2001) even dynamic MSMs fail to match up to ABM to the extent that such models do not justify the behaviours of each individual in terms of individual preferences, decisions, plans, etc. Furthermore, the interactions between individuals are not modelled in the simulation. Thus ABM is "well suited for the simulation of situations where there are a large number of heterogeneous individuals who may behave somewhat differently and is therefore an ideal simulation method for the social sciences" (Davidsson, 2001, p. 145). This view is endorsed by Jennings (2000) who highlights the advantage of ABM in modelling the intelligent behaviour of individuals by itself or in society. Interestingly, in relation to debates about processing capacity, Jennings also notes the potential of ABM for improving efficiency by distributing the control of the computation to multiple simpler units evolving through their interactions.

In addition to their capabilities for representing social behavior, the capacity of ABM to bring together diverse perspectives has been highlighted by Axelrod (2005). Social science is multi-disciplinary and social models often need to involve different disciplines. For instance a sustainability model would involve environmental, social, economic, and other disciplinary considerations. Such multidisciplinary tools are particularly valuable when the underlying mathematics are intractable. Taking the evolution of genes as an example, Axelrod pointed out that agent-based modeling could easily simulate the evolutionary effects of genes where application of mathematical equations is difficult. In this way, ABM can begin to reveal elements of the harmony between disciplines. For instance, Axelrod found that an agent-based model about military alignments could successfully predict strategic alignments of computer companies.

From an interdisciplinary perspective, David et al. (2004) also point out that ABM based social simulation originates in the intersection of the social and the computer sciences and this interdisciplinary character has encouraged collaborations from scientific fields. They also suggest that the wide interpretative scope of the theory of agents and the advances in computer capability have enlarged the communicative and interpretative room for ABM to interchange between different scientific fields and model interdisciplinary complex systems.

Nevertheless it is more constructive to view the relationship of ABM to MSM as one of complementarity rather than supremacy. One reason for this is the relative recency of the ABM paradigm, which can therefore profit from the more refined and well-established theories, concepts and models of social organizations and institutions developed within the social sciences (Conte et al. 1998). This rationale stresses that computational modeling is not just an applied tool, but a means for the production, testing and refinement of social theories. Such an eclectic view also allows for the development of more refined theories about social agents. For example, moving away from static and unsophisticated views of individual actors which overemphasise either rationality or simple social learning as a basis of behaviour. Hence we suggest that the fusion of microsimulation and agent perspectives is potentially an ideal combination in the study of both social structures and social behaviours.

3.6.2 GI Science and ABM

Torrens and Benenson (2005) proposed a new paradigm for integrating GIS and agent based simulation called Geographic Automata Systems (GAS). This system takes advantage of the formalism of automata theory and GIS to unite cellular automata and multi-agent systems techniques and provides a spatial approach to bottom-up modelling of complex geographic systems that are comprised of infrastructure and human objects. In this framework, geographic phenomena as a whole are considered as the outcomes of the collective dynamics of multiple animate and inanimate geographic automata. Geography serves as the binding force in merging CA and ABM (which are popularly confused in the geographic literature). Therefore automata become uniquely geographical, fusing CA and ABM but extending the concept to incorporate notions from GIS and Spatial Analysis.

Murphy (1995) believes that the evolution of GIS as a decision support system relies on improvements in technology, the creation of new analysis tools, and increased understanding of the interaction between decision support tools and the decision maker. He also points out that particularly fruitful areas may come from the use of artificial intelligence approaches for alternate representation of decision domains and knowledge. He thinks cooperation between the disciplines will be particularly beneficial in areas such as data quality, uncertainty representation and issues related to the management and sharing of large time-reliant and source-dependent data. Thus, a rewarding exchange may be possible between GI Science and decision support system research streams relating to the management, representation, and interpretation of complex multi-dimensional knowledge.

Gonçalves et al. (2004) suggest that GIS and ABM address space in different perspectives: GIS models geographic space and ABM models the behaviour of intelligent agents within geographic space. Gonçalves et al. propose a conceptual framework for integrating these different perspectives in the context of modelling and simulation of complex dynamic systems. They suggest that GIS enables the definition of a geographic region to be related with the phenomena in that region, but GIS do not seem to be appropriate to study dynamic phenomena in an area. Most ABM tools that use geographic information are not coupled with GIS. However, the simulation of the human behaviour with mobility in geographic space and intelligent behaviour has increased in the recent decades, which has led to a special interest in the integration of agent based models (mainly ABM) and GIS.

The authors proposed that in the hybrid model, ABM can be used to model the intelligent behaviour of entities, e.g. behaviour of people, animals, enterprises, etc., while GIS can be used to model geographic space. Intelligent agents move and reason within this environment. The authors also point out that GIS are already extensively used by people from the natural sciences, civil engineering, territory management authorities, urban planning, etc. Therefore there is no point not to give them what they already know plus the agents.

3.6.3 Unification of MSM, ABM and GI Science

As discussed in the previous section, attempts to bring MSM and ABM together (Caldwell et al. 1998; Rephann 1999) or ABM and GI Science together (Torrens and Benenson 2005) or MSM and GI Science together (Ballas et al. 2005b; Holm et al. 1996) have succeeded.

Given the characteristics of the agent based technology and geographical importance in social policies, researchers including Boman and Holm (2004) have promoted the study of social systems using a combination of different paradigms of MSM, ABM and 'time geography'. Boman and Holm argue that time geography provides a perspective to help unify the two paradigms of ABM as developed within computer science and MSM as developed within the social sciences. Time and space have important impacts on human activities in any social system. The authors suggest that time geography provides an alternative perspective on agents and collectives since it emphasises the importance of concurrent micro-level representation of agents and their relations to other agents. Time geography can also introduce a conceptual framework for analysing social micro-level interaction in time-space in MSM and ABM.

Boman and Holm (2004) attempt to unite the two paradigms through defining them and reasoning about the central concepts of each of them. They found that all three methodologies emphasise individual representation and computational solution. However many MSMs only apply a fairly aggregated and disconnected representation of individual behaviour, while ABM can provide the capacity to model individual adaptive behaviours and emergence of such behaviours. Their argument for a MSM-time geographic approach is that aggregation prior to analysis and modelling of trajectories over the state space of individuals with several attributes distorts not only individual but also aggregate results. Individual trajectory interactions and constraints need to be modelled individually to reflect the whole picture. On the other hand, MSM are developed with high estimation and validation ambitions, close to observables that facilitate empirical tests.

Therefore, developments based on a synthesis of the three paradigms can offer great potential in the advance of systems analysis methodology. Boman and Holm (2004) believe it gives a new angle to classical problems where we need to:

1. achieve consistency with the world outside a defined core system boundary;
2. simultaneously represent processes on different spatial and temporal scales;
3. enable agents to concurrently obey internal and external rules, and
4. integrate observable and postulated behaviour while preserving achievability of endogenous emergence (Boman and Holm, 2004: p. 108).

The potential benefits to the integration of MSM and ABM can be seen in relation to each of the application domains which were considered earlier. Whereas financial MSMs look at the stochastic consequences of changing rules, an agent perspective will perhaps provide some insight about new behaviours in response to a change in the background conditions. For example, changing the rules on housing benefits to unmarried partners might not just result in a change in payments, but could lead to fundamental shifts in patterns of marriage, cohabitation and family formation. A similar point could be made in relation to pension microsimulations. So if the retirement age is raised from 65 to 70, then the adjustment is probably much more complex than everyone simply agreeing to work 5 years longer. While demographic ageing is a major driver of changing health needs, the ability to provide care will be equally important. As the pressure on formal care increases then the value of informal care will also rise disproportionately, but this balance will presumably change as dependency rates become higher. Some study of the behavior of agents within social networks through which (informal) care is provided could be a fundamental component of a more effective model. Lastly, in relation to transport some of the boundaries here are already quite blurred to the extent that many systems already bridge all the way from the driving patterns of individual vehicles to strategic decisions about road networks and infrastructure provision (see the examples in Table 3.1). To the extent that individual behaviour is richly accommodated within these models then they start to look like ABM anyway, especially if individual interactions are accommodated. On the other hand, if individuals are characterized as 'agents' but their activity patterns are very predictable and well-defined then maybe they are not so different to MSM after all.

3.7 Conclusions

This review has suggested that MSM provides a powerful approach in modelling social systems and has a particular importance in public policy modelling studies. It has been widely used in a range of application domains, and major developments of MSMs have been experienced all over the world in the past few decades.

Although MSM has limitations such as requirements for both data and computational capacity, recent advances have rendered these issues as less significant. More importantly, new technologies such as ABM are naturally complementary for traditional MSMs. One advantage of using ABM is that it allows us to model these systems not only using traditional maths and statistics, but also using behavioural information, for which MSM has been criticised. The flexibility of ABM can also help us to achieve consistency outside a defined core system boundary. The usage of ABM enables us to generate the emergence of global complexity from relatively simple local actions and hence may also further reduce the computing requirements imposed by long-range interactions in a social system.

Geography has an important impact on human activities and therefore it is important to model the social system with its local context. The geography also provides a bridge to link MSM and ABM together, and the hybrid approach may provide an alternative way to study social problems.

As previously discussed, the success of hybrid approaches in modelling and simulating social systems provides the basis for the unification of MSM and ABM. A hybrid approach may offer a great potential for substantial advances in modelling social systems.

References

Axelrod, R. (2005). Agent-based modeling as a bridge between disciplines. In K. L. Judd & L. Tesfatsion (Eds.), *Handbook of computational economics, vol. 2: Agent-based computational economics* (Handbooks in Economics Series). Amsterdam: North-Holland.

Ballas, D., & Clarke, G. (2001). Modelling the local impacts of national social policies: A spatial microsimulation approach. *Environment and Planning C, 19*, 587–606.

Ballas, D., Clarke, G., & Wiemers, E. (2005a). Building a dynamic spatial microsimulation model for Ireland. *Population, Space and Place, 11*, 157–172.

Ballas, D., Rossiter, D., Thomas, B., Clarke, G., & Dorling, D. (2005b). *Geography matters: Simulating the local impacts of national social policies*. York: The Joseph Rowntree Foundation.

Ballas, D., Clarke, G., Dorling, D., Eyre, H., Thomas, B., & Rossiter, D. (2005c). SimBritian: A spatial microsimulation approach to population dynamics. *Population, Space and Place, 11*, 13–34.

Barceló, J., Casas, J., Ferrer, J., & García, D. (1999). Modeling advanced transport telematic applications with microscopic simulators: The case of AIMSUN2. In W. Brilon, F. Huber, M. Schreckenberg, & H. Wallentowitz (Eds.), *Traffic and mobility, simulation, economics, environment*. Berlin/Heidelberg: Springer.

Beebout, H., & Bonina, P. (1973). *TRIM: A microsimulation model for evaluating transfer income policy, working paper no 971–4*. Washington, DC: The Urban Institute.

Birkin, M., & Clarke, G. (1995). Using microsimulation methods to synthesize census data. In S. Openshaw (Ed.), *Census users' handbook* (pp. 363–87). London: GeoInformation International.

Birkin, M., Clarke, G., & Clarke, M. (1996). Urban and regional modelling at the microscale. In G. P. Clarke (Ed.), *Microsimulation for urban and regional policy analysis* (pp. 10–27). London: Pion.

Birkin, M., Dew, P., Macfarland, O., & Hodrien, J. (2005). HYDRA: A prototype grid-enabled spatial decision support system. *Proceedings of the First International Conference on e-Social Science*, National Centre for e-Social Science, Manchester.

Boman, M., & Holm, E. (2004). Multi-agent systems, time geography, and microsimulations. In M.-O. Olsson & G. Sjöstedt (Eds.), *Systems approaches and their application* (pp. 95–118). Dordrecht: Kluwer.

Caldwell, S., Clarke, G., & Keister, L. (1998). Modelling regional changes in US household income and wealth: A research agenda. *Environment and Planning C, 16*, 707–722.

Citro, C., & Hanushek, E. (1991). *Improving information for social policy decisions: The uses of microsimulation modeling, vol. 1, Review and recommendations*. Washington, DC: National Academy Press.

Clarke, G. P. (1996). Microsimulation: An introduction. In G. P. Clarke (Ed.), *Microsimulation for urban and regional policy analysis* (pp. 1–9). London: Pion.

Conte, R., Gilbert, N., & Sichman, J. (1998). MAS and social simulation: A suitable commitment. In J. Sichman, R. Conte, N. Gilbert (Eds.) Multi-agent systems and agent-based simulation (pp. 1–9). *Lecture Notes in Artificial Intelligence*, 1534.

Crooks, A. T., & Heppenstall, A. J. (2012). Introduction to agent-based modelling. In A. J. Heppenstall, A.T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 85–105). Dordrecht: Springer.

Curry, C. (1996). *PENSIM: A dynamic simulation model of pensioners' income, government economic service, Working Paper No. 129*. London: Analytical Services Division, Department of Social Security.

David, N., Marietto, M. B., Sichman, J.S., & Coelho, H. (2004). The structure and logic of interdisciplinary research in agent-based social simulation. *Journal of Artificial Societies and Social Simulation, 7*(3).

Davidsson, P. (2001). Multi agent based simulation: Beyond social simulation. In S. Moss & P. Davidsson (Eds.), *Multi agent based simulation* (Vol. 1979). Berlin/Heidelberg/New York: Springer. LNCS series.

Edwards, K., Clarke, G., Thomas, J., & Forman, D. (2011). Internal and external validation of spatial microsimulation models: Small area estimates of adult obesity. *Applied Spatial Analysis and Policy*. doi:10.1007/s12061-010-9056-2.

Ettlinger, M. P. & O'Hare, J. F. (1996). Revenue and incidence analysis of state and local tax Systems: Tools, technique and tradition. *The National Tax Association's 89th Annual Conference on Taxation*, USA.

Falkingham, J., & Hills, J. (1995). The effects of the welfare state over the life cycle. In J. Falkingham & J. Hills (Eds.), *The dynamic of welfare: the welfare state and the life cycle* (pp. 83–107). New York: Prentice Hall/Harvester Wheatsheaf.

Favreault, M. & Smith, K. (2004). A primer on the dynamic simulation of income model (DYNASIM3), An Urban Institute Issue Focus.

Galler, H. P., & Wagner, G. (1986). The microsimulations model of the Sfb3 for the analysis of economic and social policies. In G. H. Orcutt, J. Merz, & H. Quinke (Eds.), *Microanalytic simulation models to support social and financial policy*. Amsterdam: North-Holland.

Gilbert, N., & Troitzsch, K. G. (2005). *Simulation for the social scientist*. Berkshire: Open University Press.

Gonçalves, A. S., Rodrigues, A., & Correia, L. (2004). Multi-agent simulation within geographic information systems. *Proceedings of the Fifth Workshop on Agent-Based Simulation*, Lisbon.

Hägerstrand, T. (1985). *Time-geography: Focus on the corporeality of man, society, and environment. The science and praxis of complexity*. Tokyo: The United Nations University.

Harding, A. (2002). *Using microsimulation models in the policy process: DYNAMOD*. Paper presented at Modelling Policy in an Ageing Europe Conference, London, Jan 11, 2002.

Holm, E., Lingren, U., Makila, K., & Malmberg, G. (1996). Simulating an entire nation. In G. Clarke (Ed.), *Microsimulation for urban and regional policy analysis*. London: Pion.

Iltanen, S. (2012). Cellular automata in urban spatial modelling. In A. J. Heppenstall, A. T Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of Geographical Systems* (pp. 69–84). Dordrecht: Springer.

Jennings, N. R. (2000). On agent-based software engineering. *Artificial Intelligence, 117*, 277–296.

Kennell, D.L., & Sheils, J. F. (1990). PRISM: Dynamic simulation of pension and retirement income. In G. H. Lewis & R. C. Michel (Eds.), *Microsimulation techniques for tax and transfer analysis*. Washington, DC: The Urban Institute Press.

Krupp, H. (1986). Potential and limitations of microsimulation models. In G. H. Orcutt, J. Merz, & H. Quinke (Eds.), *Microanalytic simulation models to support social and financial policy*. Amsterdam: North-Holland.

Laird, J., Druitt, S., & Fraser, D. (1999). Edinburgh City Centre a micro-simulation case study. *Traffic Engineering and Control, 40*(2), 72–76.

Lambert, S., Percival, R., Schofield, D., & Paul, S. (1994). *An introduction to STINMOD: A static microsimulation Model*, NATSEM Technical Paper No 1. Canberra: University of Canberra.

Liu, R., Van Vliet, D., & Watling, D. P. (1995). *DRACULA: Dynamic route assignment combining user learning and microsimulation* (pp. 143–152). Paper presented at PTRC, vol E.

Luck, M., McBurney, P., & Preist, C. (2003). *Agent technology: Enabling next generation computing (a roadmap for agent based computing)*. Southampton: AgentLink II.

McDonald, P., Kippen, R., & Temple, J. (2006). Net transition probabilities: An approach to subnational level projections of households and housing demand based on census data. *Population, Space and Place, 12*, 479–495.

Morrison, R. J. (2003). Making pensions out of nothing at all. *The International Microsimulation Conference on Population, Ageing and Health: Modelling our Future*, Canberra.

Murphy, L. (1995). Geographic information systems: Are they decision support systems? HICSS (pp. 131). *28th Hawaii International Conference on System Sciences*, Honolulu.

O'Donoghue, C. (2001). Dynamic microsimulation: A methodological survey. *Brazilian Electronic Journal of Economics, 4*(2).
Orcutt, G. (1957). A new type of socio-economic system. *Review of Economics and Statistics, 58*, 773–797.
Orcutt, G., Merz, J., & Quinke, H. (Eds.). (1976). *Microanalytic simulation models to support social and financial policy*. Amsterdam: North Holland.
Propper, C. (1995). For richer, for poorer, in sickness and in health: The lifetime distribution of NHS health care. In J. Falkingham & J. Hills (Eds.), *The dynamics of welfare*. London: Prentice Hall Wheatsheaf.
PTV AG (2000). VISSIM user manual – Version 3.00, Karlsruhe, Germany.
Redmond, G., Sutherland, H., & Wilson, M. (1998). *The arithmetic of tax and social security reform: A user's guide to microsimulation: Methods and analysis*. Cambridge: Cambridge University Press.
Rephann, T. J. (1999). The education module for SVERIGE: Documentation V 1.0. Available at: http://www.equotient.net/papers/educate.pdf
Smith, D. M. (2012). Simulating spatial health inequalities. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 499–510). Dordrecht: Springer.
Smith, D. M., Clarke, G. P., Ransley, J., & Cade, J. (2006). Food access & health: A microsimulation framework for analysis. *Studies in Regional Science, 35*(4), 909–927.
Sutherland, H. (2001). EUROMOD: An Integrated European Benefit-tax Model. Final report, EUROMOD working paper No EM9/01. Cambridge: DAE University of Cambridge.
Tomintz, M., Clarke, G., & Rigby, J. (2008). The geography of smoking in Leeds: Estimating individual smoking rates and the implications for the location of stop smoking services. *Area, 40*, 341–353.
Torrens, P., & Benenson, I. (2005). Geographic automata systems. *International Journal of Geographical Information Science, 19*(4), 385–412.
TRANSIMS TRAVELOGUE (1996). TRANSIMS TRAVELOGUE (LAUR-96-1430). Available at: http://transims.tsasa.lanl.gov/PDF_Files/Travel4.pdf
Walker, A., Percival, R., & Fischer, S. (1998). Microsimulation model of Australia's pharmaceutical benefits scheme –Technical paper no. 15. Canberra: National Centre for Economic Modelling, University of Canberra.
Williamson, P. (1999). Microsimulation: An idea whose time has come? *The 39th European Congress of the European Regional Science Association*, Dublin, Ireland, Aug 23–27, 1999.
Wilson, A. (1967). A statistical theory of spatial trip distribution models. *Transportation Research, 1*, 253–269.
Wolfson, M. & Rowe, G. (1998). *LifePaths – Toward an integrated microanalytic framework for socio-economic statistics*. Paper presented to the 26th general conference of the International Association for Research in Income and Wealth, Cambridge.
Wu, B. M. & Birkin, M. H. (2012). Agent-based extensions to a spatial microsimulation model of demographic change. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 347–360). Dordrecht: Springer.
Zaidi, A. (2004). Estimation of earnings in the SAGE dynamic microsimulation Model. Technical note 10, ESRC SAGE Research Group. London: London School of Economics.

Chapter 4
Cellular Automata in Urban Spatial Modelling

Sanna Iltanen

Abstract Cities and urban dynamics are today understood as self-organized complex systems. While the understanding of cities has changed, also the paradigm in modeling their dynamics has changed from a top-down to a bottom-up approach. Cellular automata models provide an excellent framework for urban spatial modeling of complex dynamics and the accumulation of local actions. The first part of this chapter describes the basic concepts of cellular automata. The second part discusses the definition of complexity and the complex features of cellular automata. The history and principles of urban cellular automata models are introduced in the third part.

4.1 Preliminaries

The contemporary city, consisting of numerous strongly interconnected structures, multiple centers and continuous flows, although spatially scattered, has developed into a complex structure that cannot be understood with traditional methods. The interpretations about this new urbanity of the third modernisation (metapolisation) emphasizes continuous mutual competition between cities. Fast communication technologies, on one hand, connect cities and their districts together stronger than ever, and on the other hand, it enables scattering of their physical structure so that global centers do not by definition determine their development. Thus local dynamics has increasing meaning for competitiveness of cities (Ascher 2004). Cities of third modernity are considered as entities pursuing dynamic change in a state of continuous disequilibrium (Batty 2005) rather than entities pursuing some equilibrium state. For example, economic activity driven by comparative advantage

S. Iltanen (✉)
School of Architecture, Tampere University of Technology, Tampere, Finland
e-mail: sanna.iltanen@tut.fi

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_4, © Springer Science+Business Media B.V. 2012

searches continuously for new locations and modes, and thus produces a polycentric recentralized structure that disperses the traditional monocentric city.

Accompanying these new concepts and city structure, the paradigm has also changed in urban modeling: from aggregates to individuals and from equilibrium to far-from equilibrium. Complex models and concepts of Cellular Automata (CA) offer tools for understanding these dynamics.

The history of computing and CA are intertwined with each other and this affinity makes the foundations of CA-based modeling particularly firm. Attractiveness of CA is largely based on the simplicity of its basic concepts that are accessible to a wider audience but are still also intellectually fascinating. Due to the digital revolution through the 1990s when competent computing capacity and graphics became accessible for almost anyone, there was a rise in computational modeling of urban development. Numerous CA-based simulation methodologies for urban dynamics have been created during the past few decades. The process begun in the geographical sciences in the 1960s with so called raster models and continued as development of truly cellular models that were based on the idea of complexity. Understanding of urban entities as self-organizing systems, and the demand for tools to discern, control and predict these emergent phenomena, ensures interest towards computational modeling of urban development.

4.2 Basic Concepts of CA

4.2.1 Origins of CA

The history of CA leads back to John von Neumann's (1966) theory of self-reproducing automata and his co-operation with Stanislaw Ulam at the time when they were working with concepts of artificial life and idealizations of biological systems. The theory of self-replicating automata describes conceptual principles of a machine that was able to self-replicate. Alan Turing was also already working with automata in the 1930s and defined in his paper "On computable numbers, with an application to the Entscheidungsproblem", a simple abstract computer later known as the *Turing machine* (Turing 1936) where the idea of the automaton comes close to what we today consider as CA.

A cellular automaton is a dynamic discrete system and can be defined as a lattice (or array) of discrete variables or "cells" that can exist in different *states*. Usually the lattice is considered as infinite and the number of different states is finite. Cells change their states in discrete time steps according to local rules which define the cell's state on the basis of states of the cell itself and the neighboring cells in previous time steps. These *transition rules* are deterministic. Graphically, simpler forms of the cellular automaton lattices are represented as grid format but also other tessellations have been used. Due to the conditions described above, three fundamental features of CA have been defined: uniformity, synchronicity and locality. *Uniformity* means that all cell states are transformed by the same set of rules. *Synchronicity*

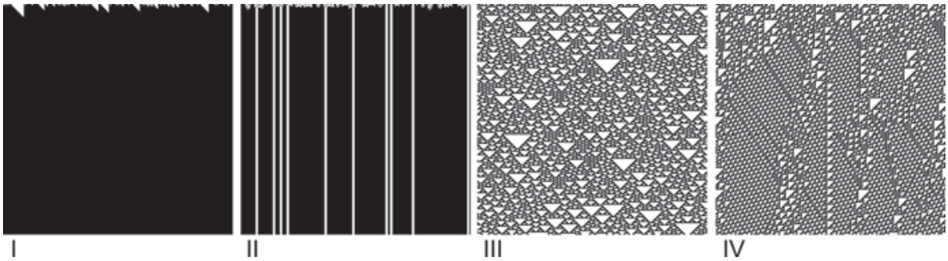


Fig. 4.1 Wolfram classes of 1-D CA dynamics

means that all the cell states are transformed simultaneously. *Locality* means that all the transformation rules are local in nature (Schiff 2008). In the next section, the characteristics of CA are discussed using one-dimensional CA as an example.

4.2.2 One-Dimensional CA

The simplest form of CA, i.e. elementary CA as Stephen Wolfram defines it, is usually considered as a one-dimensional CA consisting of an array of cells that can exist in two states 1/0 (or black/white or alive/dead) in which local rules are applied in the neighborhood of the cell itself and its immediate adjacent cells (r = 1). Thus the neighborhood of one cell consists of three cells, and since they are varying in two values, there are $2^3 = 8$ different neighborhood states. For each neighborhood state, a transition rule is defined. These rules can also be presented as eight-digit binary numbers, and thus $2^8 = 256$ possible transformation rules exist in a one-dimensional two state 'r = 1'-neighborhood cellular automaton.

Wolfram was one of the first who really systematically generated and examined the behavior of one-dimensional CA. In this work, which started in the early 1980s, he classified CA in four universality classes mostly according to the qualitative complexity in their behavior (Wolfram 1984). An analysis of the qualitative features of CA rules was mainly based on visually observable properties of CA evolution patterns (Wolfram 2002). The four Wolfram classes (Fig. 4.1) are as follows:

The class I – fixed – CA evolve to the homogenous state after a finite number of time steps independently from the initial state. Hence this class of automata is irreversible, which means that after a certain convergence point where all the cells have the same value, it loses all the information from the initial state. However, some exceptional configurations can be found that do not converge to a homogenous state, but the number of these exceptions approaches zero as the size of the automaton approaches infinity. Class 1 CA are comparable with dynamical systems that tend to a fixed-point attractor.

The class II – periodic – CA evolve to periodic structures that repeat after a fixed number of time steps. The size of the possible periods increases while the number of possible states increases. This class is naturally analogous with periodic behavior in dynamical systems.

The class III – chaotic – CA evolve to aperiodic patterns almost regardless of the initial states. In these chaotic automata, the number of initial cells that affect the value of a particular cell increases as new generations evolve. The class III CA are analogous with chaotic dynamical systems that are converging to strange attractors (Wolfram 1984).

The class IV – complex – CA evolve to complex localized structures. This class, with a mixture of chaos and randomness, is the most interesting one of the Wolfram classes. However, the definition for this class is not as rigorous as for the other classes. Localized structures that arise as the automaton progresses can move and interact, but the exact prediction of this behavior is impossible. For this class, no equivalent can be found in dynamical systems. The class IV CA behavior can also be defined as emergent, which is typical for complex systems in general (Wolfram 1984).

In his book "A New Kind of Science", Wolfram (2002) discusses the possibility that all CA can be divided into these four classes, which have been discovered by exploring one-dimensional CA. He also states that results obtained from idealized mathematical models can tell us some more general results about complex systems in nature.

4.2.3 Two-Dimensional CA

After discussing one-dimensional CA, one can ask: what happens if more dimensions are added? Wolfram discusses this question in his papers and his book "A New Kind of Science" and concludes that there is no remarkable difference in occurrence of complex phenomena as dimensions are added (Wolfram 2002). At least from a spatial modeler's point of view, two dimensions naturally look more interesting because of its similarity with maps. If the complexity of two dimensional CA are perceived by taking one dimensional slices, then the behavior of the automaton resembles pretty much pure one-dimensional CA. But what is maybe more interesting and a new feature after increasing the number of dimensions, is the *overall shape* of the pattern that emerges. There are many two-dimensional CA whose overall shape approximates a circle, but also rules that lead to more complicated overall shapes and it seems that usually these differences in overall shape are very sensitive to the initial configurations. Even more fascinating is when these shapes start to move in two dimensional space as in the most famous CA, John Conway's "Game of Life", which is discussed later.

Another thing that changes with the dimension of the automaton is the space of possible rule sets, and also the form of the neighborhood can vary in more than one dimension. The most typical form of two-dimensional CA is an orthogonal square lattice of cells. In this space, the locality is typically defined as two alternative neighborhoods: *von Neumann* and *Moore neighborhoods* (Fig. 4.2).

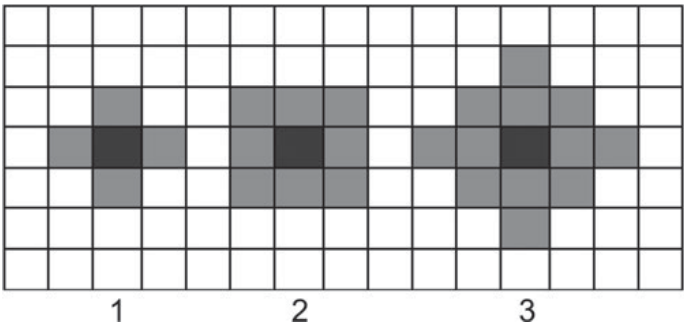


Fig. 4.2 Typical neighborhoods in 2D CA: (*1*) von Neumann 1-neigborhood; (*2*) Moore 1-neighborhood; and (*3*) von Neumann 2-neighborhood

Also, a few typical rule categories have been defined: *general*, *symmetric* and *totalistic* rules. General rule type means all the possible combinations in a given neighborhood, e.g. in a five-cell von Neumann neighborhood with two possible cell states, there are $2^{32} \approx 4 \times 10^9$ possible transition rules. The number of possible rules can be reduced if different symmetries – like rotational, reflectional or complete – are adopted. Sometimes only sums of cell values in the neighborhood are considered as in the Game of Life. This group of rules is called totalistic rules. If the value of the cell itself is taken into account, then the rule set belongs to the category of outer totalistic rules.

4.2.4 Game of Life

Developed by British mathematician John Horton Conway, the popular CA application the Game of Life was first published in Martin Gardner's (1970) column in the October 1970 issue of Scientific American. Operating in a two-dimensional lattice, the rules of the game are defined by two cell states and the eight-cell Moore neighborhood. The Game of Life belongs to the Class IV category of CA, and its rule set is an example of outer totalistic rules. There are three rules in the Game of Life:

- Rule 1 – Survival: a live cell with exactly two or three neighbors stays alive
- Rule 2 – Birth: a dead cell with exactly three live neighbors becomes a live cell
- Rule 3 – Death: owing to overcrowding or loneliness, in all other cases a cell dies or remains dead.

The popularity of the Game of Life rests on the outstanding variation of the behavior and in the patterns it can produce with these simple rules. It is also easily accessible to the general public through the internet. Several applications of the Game of Life in other tessellations, e.g. triangular, hexagonal, have been developed but they have not surpassed the original one in richness of behavior.

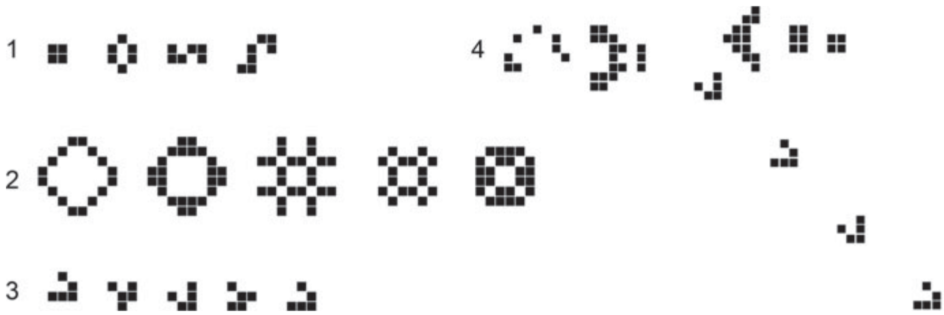


Fig. 4.3 Different life forms: (*1*) invariants, (*2*) oscillator (*5 steps*), (*3*) glider (*5 steps*) and (*4*) glider gun

4.2.5 Life Forms

Since the invention of the Game of Life, a significant amount of work and creativity has been devoted to the development of different "life forms" (Fig. 4.3). The very simple patterns, like one or two cell configurations, disappear after one generation in the Game of Life but there are a huge amount of patterns that continue their lives. Life forms that reach an unchangeable state are called *invariant* forms. Patterns showing periodic behavior between a fixed number of configurations are defined as *oscillators*. Oscillators with two periods are the most common but oscillators with more periods have also been developed. Configurations that not only repeat themselves but also move in the lattice are called *gliders*. One step more complex in the structure and behavior in the Game of life is represented by *glider guns* that are configurations constantly producing new gliders. Some of the glider like behavior is called *puffer trains* which are objects moving vertically and leaving stable configurations behind them. *Methuselah configurations* are initial patterns that achieve stable states after a remarkably long evolution, say after several hundreds of generations (Schiff 2008). There are still more mathematically interesting features of the Game of Life, which are not discussed here.

The life forms have proven that the Game of life is capable of self-reproduction. The self-reproducing system exploits the information that has been stored in it, in the form of instructions and the data to be copied (Casti 1997). In urban spatial modeling, this means that the occurrence of certain initial states is copied to other locations as the system evolves.

What makes CA a special case within other automata and agent based models is the stationary structure of the agents (cells). The automata offer a framework for abstraction of "behaving systems" in which agents, behaviors, relationships and time can be represented formally (Benenson and Torrens 2004). A number of definitions and characteristics of CA have been represented in the previous sections. However, it is not necessary to fulfill all of these conditions to achieve certain system dynamics. In spatial modeling, many conditions have been relaxed to achieve a better correlation with the system. CA have been tested in varying spatial tessellations, like triangular and hexagonal, as well as a graph form CA. The usage of different spatial tessellations has not shown any remarkable difference in automata behavior except some cellular automata classes (or types). In these (exceptional) classes, the neighborhood relations (i.e. tessellation) can change (or vary) as the system evolves (i.e. CA proceeds) (Benenson and Torrens 2004). The L-systems (Lindenmayer 1968) are an example of this kind of CA.

4.3 CA as Complex Systems

CA have become a standard example of complex systems, although there are no rigorous definitions of complex systems. However, among different disciplines under the umbrella of complexity science, the principle of emergence as an indication of a complex phenomenon is widely agreed (Holland 1998; Casti 2002). The emergence arises when simple interaction rules of objects at lower level create unforeseeable phenomena that cannot be derived straight from the objects' qualities at a higher level. As we have seen earlier, the CA obey this kind of self-organizing behavior. Despite fluctuating initial states, the class IV CA systems organize themselves through dynamical evolution, spontaneously generating complicated structures (Wolfram 1988). Irreducibility is another distinguishing characteristic of complex systems. They must be studied as a whole, as there are no means to explore the system or predict the behavior of the system by looking separately at the parts.

Casti (2002) describes three kinds of complex systems. The first one has a complex structure but the behavior of the system is simple; as an example he gives a mechanical clock. The second system has a simple structure but complex behavior, where the toy rotator is provided as an example. In the third type, both the structure and the behavior are complex, as in a human brain. Obviously it is the second type that is interesting and CA belong to this category. Casti (2002) also presents four "fingerprints of complexity": instability; irreducibility; adaptability; and emergence.

Instability refers to the modes of behavior of the system. For the complex system it is typical to have different modes of behavior depending upon small changes in the initial conditions or the interactions of the system. The four classes of CA can be interpreted as well as modes of behavior, and thus CA fulfill this criterion of complexity.

Irreducibility means that the system is infrangible, i.e. if the system is dismantled, it loses some of its essential characteristics. This is against the classical view of science where typically properties of the higher level system can be explained by properties of the parts and laws governing the behavior of the parts. In CA systems, irreducibility is engaged with the capability of universal computing. If some algorithm is used to effectively predict the behavior of the system, it should perform more sophisticated computation than the system itself, which is impossible for universal computers. Thus, because the class IV CA – the complex class – is considered to be a universal computer, they are computationally irreducible.

Adaptability becomes apparent in systems that consist of several intelligent agents. Typically these agents change their interaction rules on the basis of information rules. For example, in traffic models, one agent such as a driver can change one's decision rules according to the information about the environment. With CA, it is also possible to create adaptive agents by considering a cell as an agent and by creating for them an internal mechanism that controls the behavior of the cell.

Emergence is often considered to be the most fundamental property of complex systems. The organized behavior or structure that is generated at a global level as the system evolves according to simple local rules is an emergent phenomenon. This self organization cannot be predicted or derived from the properties of the isolated parts of the system. In CA this is a feature of the class IV systems, and self-organization is intrinsic (Casti 2002; Wolfram 1988).

Efforts have also been made to measure complexity. Ilachinski (2001) discusses a list of different measures of complexity that fall into static and dynamic classes. The four static measures are graph complexity, hierarchical complexity, Shannon's information and simplicial complexity, while the four dynamic measures are algorithmic complexity, computational complexity, logical depth and thermodynamic depth. The static measures refer to structural properties of an assembly of the system and the dynamic measures refer to the dynamic or computational effort that is required to describe the information content of an object or a state of the system (Ilachinski 2001). However none of these measures alone, or even together, delineate complexity unambiguously. Defining and observing is largely based on the human ability of visual perception regardless of all the mathematical and technical analysis that has been developed. If our standard methods of perception and analysis cannot find a short description of the phenomenon, it is considered complex (Wolfram 2002). Wolfram also discusses human pattern and texture recognition and goes even further by comparing this process to simple computer programs. The strong visual nature of the representations of CA models is clearly a strength and also one of the reasons for the success of CA in spatial and urban modeling.

From a wider perspective, complexity has influenced the predominant scientific world view. Kauffmann (2007) challenges the reductionist way of doing science and offers emergence instead. He discusses the power of creativity in nature, in the "biosphere" and in the "econosphere". Moreover, ontological phenomena, which exist in the universe, cannot be deduced from physics. He also states that "our inability to state how novel functionalities come to exist in nature is an essential limitation to the way Newton taught us to do science" (Kauffmann 2007). This comes close to the world of urban planners and architects, who under the functionalist tendency have dismantled the intermeshed traditional city structure to monofunctional enclaves. What was lost was the rich spectrum of connections in neighborhoods with mixed use and diverse functions that, for example, Jane Jacobs has written about in her book "The Death and Life of Great American Cities" (Jacobs 1961).

4.4 Urban CA

CA include intrinsic spatiality and therefore offer an excellent instrument for simulations of urban spatial dynamics. The huge number and popularity of urban simulation models based on CA is evidence of this usefulness. With a relatively simple structure and model construction, CA also provide support for large parameter spaces (Torrens 2009). A self-evident advantage is also the natural affinity with raster data in GIS and alternatively different urban morphological or functional tessellations, e.g. plots of land can be quite easily represented as cells in simulation models.

In urban modeling, the concept of CA is mainly understood in quite a broad sense, and the majority of the applications do not follow all of the conditions of strict CA. Some of the rigorously defined components of CA can be relinquished according to the requirements of the phenomenon that is being examined. Benenson and Torrens (2004) have defined these extensions as follows:

- Neighborhoods can vary in size and shape.
- The cell states can be defined in different ways: nominal, ordinal, continuous, fuzzy or multi-parameterized.
- Transition rules can be deterministic, stochastic, fuzzy, given by equations or other predicatives.
- Factors above-neighborhood level urban hierarchy can be used to control development in the model.

4.4.1 History of Urban CA

The history of urban and geographical CA models dates back to the 1950s and 1960s. Already in 1952, Hägerstrand (1952) had developed a high-resolution model of spatial diffusion, in which the dynamics were already based on local interaction. But the crucial step towards CA was not yet realized while geographical modeling concentrated on regional models. However, during the 1960s, some cell space models and raster models were introduced (Lathrop and Hamburg 1965; Chapin and Weiss 1968). Most of the models applied cellular presentation of urban space, and their principles were close to the idea of CA models. In cellular space, there was a certain state defined for each cell, which was updated at every time step. However, the raster models did not follow the bottom-up approach, at least not in the sense of how we understand this today. The transition rules in those models were mainly based on higher level functions and only some of them were based on neighborhood relationships.

The first true CA model was introduced by Tobler in his article "Cellular Geography" (1979), where he classified five types of models using a geographical array. The first four models were representations of earlier models, but the fifth model – the geographical model – had a new feature: the transition of a cell state was based on the von Neumann neighborhood. He also mentions the complex properties of Conway's Game of Life as an example of CA dynamics. Nevertheless the boom of CA based modeling did not begin until the late 1980s when the formal background of CA was established within mathematics, computing and natural sciences. Also, the development of computer graphics was crucial for CA to become common in spatial and urban modeling. One of the central papers was written by Helen Couclelis (1985) where she stated that CA combined with progress in system theories can be utilized in studying urban systems. She realized the possibilities of emergent characteristics of the global structure that arises from the dynamism of local events and presented a framework for cellular modeling of land use. By the end of the 1980s, several other papers concerning CA as simulation methods in urban dynamics were published (Itami 1988; Phipps 1989 among others).

The next level in the development of CA models took place when White and Engelen (1993) published the first constrained CA model. The idea of the constrained CA model was to combine micro and macro scale mechanisms in cell state transition rules. The constrained model enabled merging of traditional top-down and emerging bottom-up methodology. After this development stage, the interest towards the paradigm exploded rapidly. Numerous models based on CA have now been developed. There is no rigorous classification of models although Santé et al. (2010) have made a recent attempt at classifying over 30 urban CA models. In this chapter, some areas for distinguishing different models are outlined. More theoretical models, which focus on the fundamentals of the modeling mechanisms, can be distinguished from the more realistic simulation models whose intention is to generate plausible scenarios for real environments. The modeling methodologies used and the examined phenomena define their own reference groups.

4.4.2 Theoretical Urban CA Models

The development of theoretical urban CA models concentrates on revealing the properties and effects of the modeling techniques, where the interest is in the theory of CA in an urban context. Michael Batty writes in his book Cities and Complexity (2005) about hypothetical models. He has developed an extensive variety of models in this category with his collaborators. These models are simple idealized city models in which the growth starts from reduced initial conditions, typically from a single seed. The idea of the simple models is to reveal special features of growth mechanisms in their purest form in laboratory-like conditions.

One of the interesting and salient features of these theoretical models is how the concept of geographical potential appears in them. Lots of dynamics in urban development is based on "action at a distance" and Batty discusses "action at a distance" as an emergent phenomenon that arises as the influence of cell transitions propagates in the lattice as the system evolves (Batty 2005). This is a key issue in the theory of urban dynamics and in differences between strict CA models and more general urban models. The demand for simple strict CA models arises from the "action at a distance" question that can be enlightened by examining single cells in those models.

4.4.3 Real City CA Models

Several urban CA models have been developed with the intent to create future scenarios for real urban environments. Therefore, many of their features have a pragmatic explanation. For instance, they can be configured according to the availability of data. The division between models into a theoretical or real category is not that rigorous. Rather there is a spectrum of models between these extremes. The differences between urban CA models can be differentiated by how they are configured to the five basic elements of CA: spatial tessellation, cell states, neighborhoods, transition rules and time (Liu 2009). The most significant differences between urban CA models seems to be based on differences in the transition rules, as they actually define the logic of how the modeled phenomena are handled. All these features have been stressed differently depending upon the purpose for which the simulation model was created. In the following sections we will discuss some examples of models developed for the simulation of urban growth, land use, sprawl, gentrification, etc.

4.4.3.1 Land Use Change in Constrained CA

The constrained CA model developed by White and Engelen (1993, 2000) has been used to simulate land use change. The operational principle of the model is based on the transition potential of the cell, which is derived from the properties of the cell and its neighboring cells. The potential is based on the intrinsic properties of the cell and the influence of the neighbors weighted by distance from the central cell. All cells are then ranked by their potential and the macro scale mechanisms are applied by determining the overall amount of cells to be transformed according to demand for certain land use at an aggregate level. The aggregate level transition operations, which utilize population data, were developed separately from the CA model.

The land use change is represented as a transition of 16 different cell states that are classified as active, passive or fixed state categories. The transition potential of the cells is defined as the vector sum of the components of attraction or repulsion of other land uses, accessibility to transportation networks, and the intrinsic suitability for the particular land use and zoning regulations. In the model, the size of the neighborhood is relaxed to a circular template of 113 cells. The cell size in the model is 500 m × 500 m (White and Engelen 2000).

The principle of combining the above neighborhood structures into transition functions has also been introduced by others (Xie 1996; Batty and Xie 1997; Phipps and Langlois 1997). How the constraints are formulated varies between the different models. The challenge in constrained modeling is how to implement the constraints so that the local dynamics are not destroyed.

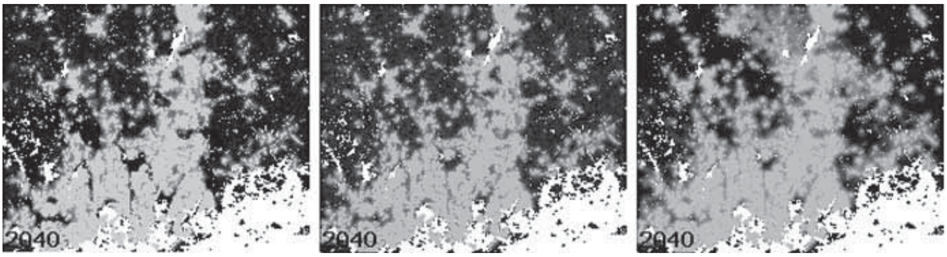


Fig. 4.4 Sleuth-model simulations of Helsinki city region. Three predictions for the hypothetical year 2040 with different input data. Taken from Iltanen (2008)

4.4.3.2 Diffusion-Based Urban Growth

The *Sleuth* model represents a diffusion-based view of urban development (Benenson and Torrens 2004). The model was developed to simulate urban growth for the San Francisco Bay area by Keith Clarke and colleagues during the 1990s (Clarke et al. 1997; Clarke and Gaydos 1998). The model is based on a self-modifying cellular automaton and can be calibrated according to predominant trends of urban development. The growth dynamics consist of four growth rules executed in the following order in every growth cycle: (1) spontaneous; (2) new spreading centre; (3) edge; and (4) road influenced growth. The spontaneous growth defines the random urbanization of a cell by giving a certain probability to every cell regarding urbanization. The new spreading centre growth determines with certain probability the newly urbanized cells to become a spreading center. The edge growth defines the growth on the edge of the existing urban structure by giving a certain probability to a cell to be urbanized if it has at least three neighbors. The road influenced growth is based on the urbanization in earlier steps, on the input data of the transportation infrastructure and a random walk component.

The model also includes an optional *Deltatron*-module, which simulates land use change. The core model can be used without this module. The number of newly urbanized cells, generated in the core model, is the driver for land use transitions. However, the *Deltatron*-module generates only nonurban land use transitions (Clarke 1997).

The calibration is carried out by using historical cross-sectional data as input to the model, and the Monte-Carlo method is used in iteration. The calibration phase produces five growth coefficients as a result. These growth coefficients control the growth rules that are typical for each simulation area and the input data used in the simulation. The input data needed for the model consist of five (or six if the *Deltatron*-module is implemented) layers: slope, land use (*Deltatron*), excluded, urban, transportation and auxiliary hill shade. The name *Sleuth* is comprised from the first letters of the layer names. After the calibration phase, the predictions (Fig. 4.4) can be executed using growth coefficients (Clarke et al. 1997).

The *Sleuth*-model combines a CA approach with different statistical methods to achieve higher realism in simulations. The features of the excluded layer enable the top-down control of growth to be combined with the bottom-up growth dynamics in a way that the level of top-down regulation can be defined by the user. The definition of urban and non-urban areas can be utilized in terms of density to catch sprawl like development (Iltanen 2008, 2011).

4.4.3.3 Urban Sprawl in CA

One interesting exploration concerning urban growth and polycentricity was introduced by Batty and Xie (1997). Their model was based on the idea of development potential, which is a driving force of urban growth. The positive feedback in land use transformations creates growing clusters that break the monocentric structure. This model was implemented in cellular space where the potential of the cell evolves on the basis of itself.

Different grades of urbanization and growth were modeled by Batty and Xie (1997). They also used an epidemic model and generalized it to a spatial context (Batty 2005).The model exploits aggregate models as a part of the simulation process, embedding them in the CA model *Duem* (Batty et al. 1999). The *Duem* model is a CA model that simulates urban growth and the land use change of five different categories. The five land uses of the model are housing, industry, commerce, services and vacant land. The transport network is also represented in cellular form. The model utilizes different decision methods and life-cycle processes of land use.

4.4.3.4 Fuzzy Urbanization

Fuzzy logic and fuzzy set theory have also been utilized in modeling urban growth. It has been argued that fuzzy methodologies are suitable for urban modeling since both physical factors and human decision making are characterized by uncertainty and fuzziness (Wu 1996; Liu 2009). Many urban conditions are continuous rather than discrete by their nature, which points to the appropriateness of using fuzzy logic in modeling urban dynamics. Fuzzy set theory has been developed to extend crisp set theory by defining membership of a set gradually instead of through a binary definition; 0 (=non member) or 1 (full member). Wu (1996) developed a methodology that utilized fuzzy logic in CA transition rules. He applied linguistic modeling with the idea to couple behavioral considerations of decision making to the simulation process. Liu (2009) developed an urban fuzzy constrained CA model in which fuzzy set theory has been used in the definition of cell states and their grade of urbanity. Liu (2009) found that more realistic simulation results were produced in terms of the human decision-making process. Moreover, fuzzy logic has been used in the representation of drivers and in the transition rules for an urban growth model in the city of Riyadh, Saudi Arabia (Al-Ahmadi et al. 2009a; 2009b; 2009c). One of the main advantages of using fuzzy logic was the ability to interpret the resulting model and the rulebase, and to understand which drivers are important and which rules fire most frequently during different periods of urban growth.

4.5 Conclusions

The increasing connectedness of urban structure, both locally and globally, makes it more and more difficult to understand and control the development of cities. CA models, as part of the modeling toolkit, can enlighten the complex interactions and relations in networked urban structure. The better we know the theoretical behavior of our models, the better we can adjust them to real world situations. Thus, there is still space for both theoretical and applied explorations of the models of urban dynamics. The knowledge concerning theoretical aspects of the model also enhances their transparency. This transparency is required for keeping the basis of the model simple enough to catch the complex features in the system.

The strength of CA models is fast processing of information and the illustrative nature of the results, which can be effectively interpreted by human visual perception. Many possibilities also lie in the exploitation of the urban morphological elements in CA modeling. New dimensions could be added to the modeling scheme by using suitable urban morphological elements to add more coherence between the model and reality. The quantitative analysis of urban morphological objects and configurations could be incorporated within the automata models and also the utilization of suitable morphological tessellations could be developed to achieve more sensitive representation of the environment.

Simulations do not necessarily represent the behavior of real urban systems, yet they reveal to us some essential mechanisms that are part of the overall dynamics. The models can be used as tools within urban planning to produce unforeseeable development paths and to help generate scenarios for the basis of decision making. By exploiting simulation models, suitable boundary conditions can be outlined to achieve eligible development, although the modeling always leaves the final state open. The challenge in the wider utilization of simulation models is a tradeoff between the ease of accessibility and understanding the inner logic of these models.

References

Al-Ahmadi, K., Heppenstall, A. J., Hogg, J., & See, L. (2009a). A fuzzy cellular automata urban growth model (FCAUGM) for the city of Riyadh, Saudi Arabia. Part 1: Model structure and validation. *Applied Spatial Analysis, 2*(1), 65–83.

Al-Ahmadi, K., Heppenstall, A. J., Hogg, J., & See, L. (2009b). A fuzzy cellular automata urban growth model (FCAUGM) for the city of Riyadh, Saudi Arabia. Part 2: Scenario analysis. *Applied Spatial Analysis, 2*(2), 85–105.

Al-Ahmadi, K., See, L., Heppenstall, A., & Hogg, J. (2009). Calibration of a fuzzy cellular automata model of urban dynamics in Saudi Arabia. *Ecological Complexity, 6*(2), 80–101.

Ascher, F. (2004). Métapolis: A third modern urban revolution. Changes in urban scale and shape in France. In L. Bölling & T. Sieverts (Eds.), *Mitten am rand. Auf dem weg von der vorstadt über die zwischenstadt zur regionalen stadtlandschaft. Zwischenstadt Band 1* (pp. 24–37). Wuppertal: Werlag Müller+Busmann Kg.

Batty, M. (2005). *Cities and complexity. Understanding cities with cellular automata, agent-based models and fractals*. London: Academic.

Batty, M., & Xie, Y. C. (1997). Possible urban automata. *Environment and Planning –B Planning & Design, 21*, S31–S38.

Batty, M., Xie, Y., & Sun, Z. (1999). Modeling urban dynamics through GIS-based cellular automata. *Computers, Environments and Urban Systems, 233*, 205–233.

Benenson, I., & Torrens, P. (2004). *Geosimulation – Automata-based modeling of urban phenomena*. London: Wiley.

Casti, J. L. (1997). *Would-be worlds: How simulation is changing the frontiers of science*. New York: Wiley.

Casti, J. L. (2002). "Biologizing" control theory: How to make a control system come alive. *Complexity, 4*(4), 10–12.

Chapin, S. F., & Weiss, S.F. (1968). A probabilistic model for residential growth. *Transportation Research, 2*, 375–390.

Clarke, K. C. (1997). Land transition modeling with deltatrons. Accessed Nov 2009, from http://www.geog.ucsb.edu/~kclarke/Papers/deltatron.html

Clarke, K. C., & Gaydos, L. J. (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., Hoppen, S., & Gaydos, L. J. (1997). A self-modifying cellular automata model of historical urbanization in the San Francisco Bay area. *Environment and Planning B: Planning and Design, 24*(2), 247–261.

Couclelis, H. (1985). Cellular worlds: A framework for modeling micro-macro dynamics. *Environment and Planning A, 17*, 585–596.

Gardner, M. (1970). The fantastic combinations of John Conway's new solitaire game "Life". *Scientific American, 223*(10), 120–123.

Hägerstrand, T. (1952). The propagation of innovation waves. *Lund Studies in Geography B, Human Geography, 4*, 3–19.

Holland, J. H. (1998). *Emergence – From chaos to order*. Oxford: Oxford University Press.

Ilachinski, A. (2001). *Cellular automata – A discrete universe*. New Jersey/London/Singapore/Hong Kong: World Scientific Publishing Co. Pte. Ltd.

Iltanen, S. (2008). *Urban generator – Kaupunkirakenteen kasvun mallinnusmenetelmä*. Tampere: Tampere University of Technology, Urban Planning and Design, Juvenes Print.

Iltanen, S. (2011). *Combining urban morphological network analysis with urban growth model SLEUTH*. Paper presented at Computers in Urban Planning and Urban management, Lake Louise 5 July to 8 July 2011.

Itami, R. (1988). Cellular worlds: Models for dynamic conception of landscapes. *Landscape Architecture, 78*(5), 52–57.

Jacobs, J. (1961). *The death and life of great American cities*. New York: Vintage Books, Random House.

Kauffmann, S. (2007). Beyond reductionism: Reinventing the sacred. *Zygon, 42*(4), 903–914.

Lathrop, G. T., & Hamburg, J. R. (1965). An opportunity-accessibility model for allocating regional growth. *Journal of the American Institute of Planners, 31*, 95–103.

Lindenmayer, A. (1968). Mathematical models for cellular interaction in development I. Filaments with one-sided inputs. *Journal of Theoretical Biology, 18*, 280–289.

Liu, Y. (2009). *Modelling urban development with geographical information systems and cellular automata*. New York: CRC Press.

Phipps, M. (1989). Dynamic behavior of cellular automata under the constraint of neighborhood coherence. *Geographical Analysis, 21*, 197–215.

Phipps, M., & Langlois, A. (1997). Spatial dynamics, cellular automata, and parallel processing computers. *Environment and Planning B: Planning and Design, 24*(2), 193–204.

Santé, I., Garcia, A. M., Miranda, D., & Crecente, R. (2010). Cellular automata models for the simulation of real-world urban processes: A review and analysis. *Landscape and Urban Planning, 96*(2), 108–122.

Schiff, J. L. (2008). *Cellular automata – A discrete view of the world*. Hoboken: Wiley.

Tobler, W. (1979). Cellular geography. In S. Gale & G. Ollson (Eds.), *Philosophy in geography* (pp. 379–386). Dordrecht: Kluwer.

Torrens, P. (2009). Cellular automata. In K. Rob & T. Nigel (Eds.), *International encyclopedia of human geography* (pp. 1–4). London: Elsevier.

Turing, A. (1936). On computable numbers with an application to the entscheidungsproblem. *Proceedings of the London Mathematical Society, 42*(2), 230–265.

von Neumann, J. (1966). Theory of self-reproducing automata. In A. W. Burks (Ed.), Urbana and London: University of Illinois Press, ISBN 0598377980.

White, R., & Engelen, G. (1993). Cellular-automata and fractal urban form – A cellular modeling approach to the evolution of urban land-use patterns. *Environment and Planning A, 25*(8), 1175–1199.

White, R., & Engelen, G. (2000). High-resolution integrated modeling of the spatial dynamics of urban and regional systems. *Computers, Environment and Urban Systems, 24*, 383–400.

Wolfram, S. (1984). Universality and complexity in Cellular Automata. *Physica D, 10*, 1–35.

Wolfram, S. (1988). Complex systems theory. Emerging syntheses in science. *Proceedings of the Founding Workshops of the Santa Fe Institute*, vol. 1 (pp. 183–189). Reading: Addison-Wesley.

Wolfram, S. (2002). *A new kind of science*. Champaign: Wolfram Media.

Wu, F. (1996). A Linguistic cellular automata simulation approach for sustainable land development in a fast growing region. *Computers, Environment and Urban Systems, 20*(6), 367–387.

Xie, Y. C. (1996). A generalized model for cellular urban dynamics. *Geographical Analysis, 28*(4), 350–373.

Chapter 5
Introduction to Agent-Based Modelling

Andrew T. Crooks and Alison J. Heppenstall

Abstract The application of agent-based modelling (ABM) to simulating dynamics within geographical systems has seen a considerable increase over the last decade. ABM allows the disaggregation of systems into individual components that can potentially have their own characteristics and rule sets. This is a powerful paradigm that can be exploited through simulation to further our knowledge of the workings of geographical systems. We present in this chapter an overview of ABM; the main features of an agent-based model are given, along with a discussion of what constitutes an agent-based model. The distinction between cellular automata (CA), microsimulation (MSM) and agent-based models are discussed along with the advantages and limitations of ABM for modelling geographical systems. We conclude with a brief discussion of important areas for further research.

5.1 Introduction

Many geographical systems are characterised by continual change and evolution through time and space. The impacts of interactions between individual agents (humans, cities or more abstract representations), or an individual agent and the environment (physical, social, information etc) can be felt at multiple scales as well as over differing timescales. Previous approaches to modelling the complexity of geographical systems have focused on representing these systems as static aggregations of populations, rational aggregate behaviour and flows of information. Examples

A.T. Crooks (✉)
George Mason University, MS 6B2, Fairfax, VA 22030, USA
e-mail: acrooks2@gmu.edu

A.J. Heppenstall
University of Leeds, Leeds LS2 9JT, UK
e-mail: a.j.heppenstall@leeds.ac.uk

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_5,

of these "traditional approaches" include multiple regression, location-allocation and spatial interaction models; Batty (2012) provides a detailed discussion of the different approaches applied within geographical modelling.

While the utility of these approaches are exemplified within the academic literature, one of the central criticisms that can be levelled at them is treatment of all geographical components as largely homogeneous entities, for example, populations modelled with the same characteristics. Over the course of the twentieth century geography has incorporated ideas and theories from other disciplines including economics, mathematics and computer science. These ideas have strengthened the significance of both modelling and understanding the impact of individual agents and the heterogeneity of geographical systems at different spatial and temporal scales. Simulating these processes and their impacts 'realistically' presents a significant challenge for the twenty-first century geographer.

To understand geographical problems such as sprawl, congestion and segregation, researchers have begun to focus on bottom-up approaches to simulating human systems, specifically researching the reasoning on which individual decisions are made. One such approach is agent-based modelling (ABM) which allows one to simulate the individual actions of diverse agents, and to measure the resulting system behaviour and outcomes over time. The distinction between these new approaches and the more aggregate, static conceptions and representations that they seek to complement, if not replace, is that they facilitate the exploration of system processes at the level of their constituent elements.

Essential to the progression of ABM has been the development of automata approaches more generally. An automaton is a processing mechanism with characteristics that change over time based on its internal characteristics, rules and external input. Automata process information input to them from their surroundings and their characteristics are altered according to rules that govern their reaction to these inputs. Two classes of automata tools – cellular automata (CA) and agent-based models – have been particularly popular, and their use has dominated the research literature.

The purpose of this chapter is to provide an overview to ABM. The key features of an agent-based model will be presented along with a discussion of what constitutes an agent-based model and brief overviews of the main areas of consideration when undertaking modelling. The distinction between CA, microsimulation (MSM) and ABM approaches are briefly outlined. The advantages and disadvantages of ABM for simulating geographical systems are then discussed before an overview of geographical applications are given. We conclude the chapter with a summary and discussion of areas that require further consideration.

5.2 What Is an Agent?

There is no universal agreement amongst researchers on the precise definition of the term 'agent' with researchers continually debating whether definition should be by an agent's application or environment; however definitions do tend to agree on more points than they disagree (Macal and North 2005). Diversity in their application makes agent characteristics difficult to extract from the literature in a consistent and concise manner as often an agent-based model is described from the perspective of its constituent parts (Bonabeau 2002).

From a pragmatic modelling standpoint, there are several features that are common to most agents (Wooldridge and Jennings 1995 – extended and explained further by Franklin and Graesser 1996; Epstein 1999; Macal and North 2005). These are briefly presented below:

- **Autonomy**: agents are autonomous units (i.e. governed without the influence of centralised control), capable of processing information and exchanging this information with other agents in order to make independent decisions. They are free to interact with other agents, at least over a limited range of situations, and this does not (necessarily) affect their autonomy.
- **Heterogeneity**: agents permit the development of autonomous individuals e.g. an agent representing a human could have attributes such as age, sex, job etc. Groups of agents can exist, but they are spawned from the bottom-up, and are thus amalgamations of similar autonomous individuals.
- **Active**: agents are active because they exert independent influence in a simulation. The following active features can be identified:
  - **Pro-active/goal-directed**: agents are often deemed goal-directed, having goals to achieve (not necessarily objectives to maximise) with respect to their behaviours.
  - **Reactive/Perceptive**: agents can be designed to have an awareness or sense of their surroundings. Agents can also be supplied with prior knowledge, in effect a 'mental map' of their environment, thus providing them with an awareness of other entities, obstacles, or required destinations within their environment.
  - **Bounded Rationality**: throughout the social sciences, the dominant form of modelling is based upon the rational-choice paradigm (Axelrod 2007). Rational-choice models generally assume that agents are perfectly rational optimisers with unfettered access to information, foresight, and infinite analytical ability (Parker et al. 2003). However, agents can be configured with 'bounded' rationality (through their heterogeneity). This allows agents to make inductive, discrete, and adaptive choices that move them towards achieving goals.
  - **Interactive/Communicative**: agents have the ability to communicate extensively. For example, agents can query other agents and/or the environment within a neighbourhood.
  - **Mobility**: agents can 'roam' the space within a model. Juxtaposed with an agent's ability to interact and their intelligence, this permits a vast range of potential uses. However, agents can also be fixed.
  - **Adaptation/Learning**: agents can also be designed to be adaptive, producing Complex Adaptive Systems (CAS; Holland 1995). Agents can be designed to alter their state depending on previous states, permitting agents to adapt with a form of memory or learning. Agents can adapt at the individual level (e.g. learning alters the probability distribution of rules that compete for attention), or the population level (e.g. learning alters the frequency distribution of agents competing for reproduction).

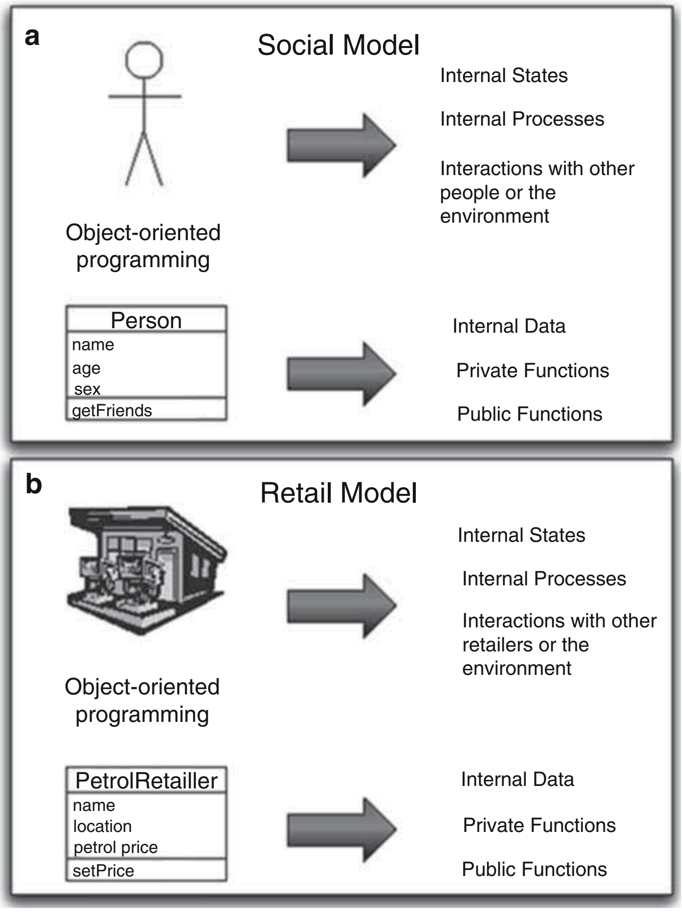


Fig. 5.1 Representation of (**a**) human agent and (**b**) petrol retailer agent alongside that of representation within an object-orientated environment

This list is not exhaustive or exclusive; within an application agents can possess other characteristics and for some applications, some features will be more important than others (Wooldridge and Jennings 1995). Often, there are many different types of agents within one simulation.

5.2.1 What Does an Agent Look Like?

Agents can be representations of any type of autonomous entity. These could be, for example, people, buildings, cars, land parcels, water droplets or insects. Figure 5.1 shows the similarities between the concept of (i) a "social" human and (ii) a grocery retailer and of a representation of an agent within an object-orientated program (see Abdou et al. 2012; Crooks and Castle 2012; Grimm and Railsback 2012 who explore

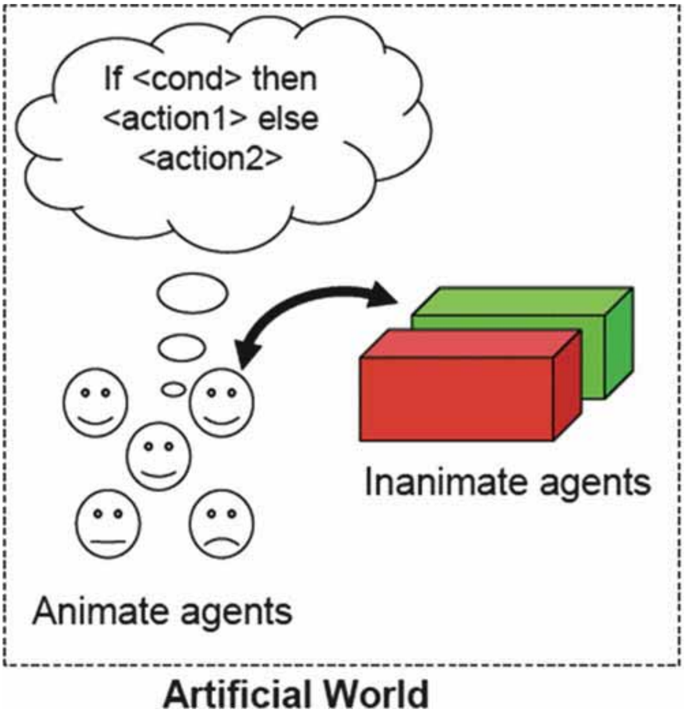


Fig. 5.2 ABM: the creation of artificial worlds populated by agents (Adapted from Cederman 2004)

constructions of agent-based models in more detail). It should be noted that ABM is not the same as object-oriented simulation, although the object-oriented paradigm provides a suitable medium for the development of agent-based models. For this reason, ABM systems are invariably object-oriented (Gilbert and Terna 2000).

A collection of multiple, interacting agents, situated within a model or simulation environment such as represented by the artificial world as shown in Fig. 5.2 is termed an agent-based model. Here, agents can be representations of animate entities such as humans that can roam freely around an environment or be inanimate, such as a petrol retailer, that has a fixed location but can change state.

5.2.2 Rules, Behaviour and Relationships

Each of the inanimate and animate agents outlined above can possess rules that will affect their behaviour and relationships with other agents and/or their surrounding environment. Rules are typically derived from published literature, expert knowledge, data analysis or numerical work and are the foundation of an agent's behaviour. One rule-set can be applied to all agents or each agent (or categories of agents) can have its own unique rule set. For example, the retail petrol agents in Heppenstall et al. (2006) all operated on the same basic rule set based on a desire to maximise profits. Further work saw different types of retailer agents, for example supermarkets, international, national and independent stations, given their own "realistic" rule-sets based on published behaviour, data analysis and numerical analysis (Heppenstall et al. 2006).

Rules are typically based around 'if-else' statements with agents carrying out an action once a specified condition has been satisfied. However, rules can also be put into action in ignorance of the actions of other agents. Agents can also be imbedded with a notion of learning and thus 'intelligence' through evolutionary computation (see Heppenstall et al. 2007 for further details). More recently there has been a move towards incorporating behavioural frameworks within agent-based models to better represent human behaviour. For example, Malleson et al. (2010) used the PECS (Physical conditions, Emotional states, Cognitive capabilities and Social status) framework to represent the motivations and desires of criminals. This type of work marks a move towards a more sophisticated handling of agent behaviour. Kennedy (2012) provides an overview of different frameworks for handling human behaviour in agent-based models.

Agents can interact with each other and amongst themselves and with the environment. Relationships may be specified in a variety of ways, from simply reactive (i.e. agents only perform actions when triggered to do so by some external stimulus e.g. actions of another agent) to goal-directed (i.e. seeking a particular goal). The behaviour of agents can be scheduled to take place synchronously (i.e. every agent performs actions at each discrete time step, all change occurs simultaneously), or asynchronously (i.e. agent actions are scheduled by the actions of other agents, and/or with reference to a clock).

5.2.3 Agent Environments

Environments define the space in which agents operate, serving to support their interaction with the environment and other agents. For example, depending on the space defined for agent interactions, proximity may be defined by spatial distance for continuous space, adjacency for grid cells, or by connectivity in social networks. Agents within an environment may be spatially explicit, meaning agents have a location in geometrical space, although the agent itself may be static. For example, within a route navigation model, agents would be required to have a specific location for them to assess their route strategy. Conversely, agents within an environment may be spatially implicit; this means that their location within the environment is irrelevant.

In a modelling context, agent-based models can be used as experimental media for running and observing agent-based simulations. To this extent, they can be thought of as a miniature laboratory where the attributes and behaviour of agents, and the environment in which they are housed, can be altered and the repercussions observed over the course of multiple simulation runs, thus providing a tool to 'think with.' The ability to simulate individual actions of many diverse agents and measure the resulting system behaviour and outcomes over time (e.g. changes in patterns of traffic flow), means agent-based models can be useful tools for studying the effects on processes that operate at multiple scales and organisational levels (Brown 2006). In particular, the roots of ABM are within the simulation of human social behaviour and individual decision-making (Bonabeau 2002). In this sense ABM has transformed social science research by allowing researchers to replicate or generate the emergence of empirically complex social phenomena from a set of relatively simple agent-based rules at the micro-level (Balan et al. 2003).

5.3 Individual-Based Models

Agent-based models fall into the broad category of individual based models. Within this category are also the closely related techniques of CA and MSM. This section clarifies the scope of these other techniques and emphasises the distinction from agent-based models. CAs and MSM are more fully explained in Iltanen (2012) and Birkin and Wu (2012).

5.3.1 Cellular Automata

The basic features of CA are well-known from the research literature. A CA is a discrete dynamic system, the behaviour of which is specified in terms of local relations. The space in a CA system is divided into a lattice or grid of regularly-space cells of the same size and shape, usually square. Each cell has a value either 0 or 1 or on a scale from 0 to 1. The state of a cell and its behaviour is determined by the state of other cells in close proximity at a previous time step, by a set of local rules and by the cell itself (Benenson and Torrens 2004; Torrens 2003; Wolfram 2002).

An important feature of a CA is that the automata's location does not move; they can only change their state. The position of the cells and their neighbourhood relations remain fixed over time. In contrast, agents can be either fixed in location or free to 'roam' around their environment. Unlike agents, CAs cannot have more than one attribute; for example, a cell could be occupied or unoccupied, but the cell could not contain multiple attributes such as building type, date built etc.

Both CA and agent-based models, model the complexity of social systems with similar individual level representations. However, they differ in their emphasis; CA model social dynamics with a focus on the emergence of properties from local interactions while agent-based models simulate more complex situations where agents control their own actions based on their knowledge of the environment (Birkin and Wu 2012).

In practice, CA and ABM have often been applied separately to explore a wide variety of geographical phenomena. This is particularly evident in urban modelling. For example, CA models are commonly applied to represent possible land-use changes (for example, White et al. 1997; Landis and Zhang 1998) while ABM are often applied to crowd dynamics and traffic simulation (see Benenson and Torrens 2004 for further reviews). However, models are increasingly being developed using a combination of CA and ABM techniques to produce flexible and powerful models, and the distinction between them is increasingly becoming blurred.

5.3.2 Microsimulation

MSM is a well established methodology that works on the principle of creating small area microdata at a point in time, and then generating future microdata from that basis (Ballas et al. 2005). MSM has been extensively applied to modelling the effects of different policy scenarios on individual 'units', for example understanding the implications of a factory closure on individual households.

As with CA and ABM, MSM operates at the level of the individual, is able to simulate the global consequences of local interactions whilst allowing the characteristics of each individual to be tracked over time. However, crucially in contrast to ABM, MSM only models one-direction interactions: the impact of the policy on the individuals, but not the impact of individuals on the policy and interactions between individuals are not simulated. Furthermore MSM models do not have the behavioural modelling capability of ABM.

Birkin and Wu (2012) see the relationship between ABM and MSM as complimentary; linking the two approaches can help overcomes inherent limitations in both approaches, for example problematic validation in ABM and the absence of real behavioural modelling in MSM. Examples of the hybridisation of these approaches can be found in the work of Boman and Holm (2004) and more recently Wu et al. (2008).

5.4 Constructing an Agent-Based Model

Creation of an agent-based model can be facilitated through the use of an object-orientated language, modelling toolkits and platforms. Here we briefly discuss these approaches describing their advantages and disadvantages. For a more detailed discussion, the reader is directed to Crooks and Castle (2012).

Frequently used programming languages are Java and C++. While programming from the ground up allows complete control over every aspect of the agent-based model, this can be a time-consuming option unless the researcher is an experienced programmer. Model implementation can be cumbersome and considerable time can be spent on non content-specific aspects such as graphical user interfaces (GUI's), visualisation and data importing.

Toolkits do not require substantial coding experience and provide conceptual frameworks and templates that allow the user to design a customised model. Prominently used toolkits include the highly popular Repast, SWARM and MASON, although Crooks and Castle (2012) note that there are over 100 toolkits currently available. These toolkits are often supported by libraries of pre-defined methods and functions that can be easily incorporated into an agent-based model and linked into other software libraries, for example geographical information systems (GIS) such as OpenMap or GeoTools. Using a toolkit can greatly reduce the model construction time allowing more time to be dedicated to research. However, drawbacks include a substantial time investment on behalf of the researcher to learn the how to design and implement a model in the toolkit and the programming language the software uses. After this investment of time, it is possible that the desired functionality is not available.

In addition to toolkits, there is a steady increase of available software for constructing agent-based models. Notable examples include NetLogo and AgentSheets. Utilisation of such software is particularly useful for rapid development of basic or prototype models. The major drawback using software is that researchers are restricted to the design framework supported by the software and maybe unable to extend or integrate additional tools.

5.5 Working with Agent-Based Models

Once a model has been designed at the level of abstraction deemed necessary for the purpose of the model and an appropriate toolkit or software chosen for implementation (see Grimm and Railsback 2012; Crooks and Castle 2012 for more of a discussion), several other important issues need to be considered. These revolve around gaining an understanding and communicating the inner workings of the model but also considerations with respect to verification, calibration and validation of the model itself. It is to these issues that we now turn.

5.5.1 Verification, Calibration and Validation

One of the greatest challenges of utilising ABM concerns the issues of verification, calibration and validation (Crooks et al. 2008). “Verification is the process of making sure that an implemented model matches its design. Validation is the process of making sure that an implemented model matches the real-world.” (North and Macal 2007, pages 30–31). Verification is thus as a much a matter of testing the logic of the model through its computer programme as testing its formal logic. It involves checking that the model behaves as expected which is something that is often taken for granted. Validation relates to the extent that the model adequately represents the system being modelled (Casti 1997) and in this sense, it involves the goodness-of-fit of the model to data. However, the validity of a model should not be thought of as binary event (i.e. a model cannot simply be classified as valid or invalid); a model can have a certain degree of validity (Law and Kelton 1991), which of course is encapsulated by various measures of fit.

In contrast, calibration involves fine-tuning the model to a particular context and this means establishing a unique set of parameters that dimension the model to its data. This is not validation *per se* but calibration can often involve validation because the parameters are often chosen so that performance of the model related to data is optimal in some way, in terms of some criterion of goodness-of-fit, for example. This is a large subject area and suffice it to say, many if not most agent-based models suffer from a lack of uniqueness in parameter estimation due to the fact that their assumptions and processes tend to outweigh the data available for a complete assessment of their goodness-of-fit.

Concerns have been raised pertaining to verification and validation by numerous researchers (e.g. Batty and Torrens 2005; Crooks et al. 2008) and can be considered limitations to ABM (see Sect. 5.7). Ngo and See (2012) present a more detailed discussion of how verification, calibration and validation issues can be addressed while Evans (2012) raises awareness of error and uncertainty with respect to input data, parameterisation, and model form and offers guidance to minimising and understanding such errors. These issues are only mentioned here to stress to the reader that these are important and need to be considered when working with agent-based models.

5.5.2 Communication and Visualisation

Agent-based models tend to be overtly visual and this is extremely helpful as visualisation is one of the most effective ways to communicate key model information with regard to ABM (North and Macal 2007). Some argue that by making models more visual they become more transparent (Batty 2007) but also by visualising key model processes, helps to convey the model clearly and quickly (Kornhauser et al. 2009) and thus aiding with the verification and validation of model outputs. For example, via the GUI of the model we are able to track the simulation history as advocated by Axelrod (2007). Through this we can observe and explain how aggregate outcomes emerge from the local interactions of many individuals. Moreover, there are also qualitative evaluations of model validity that might be made from visualising outcomes of such models. For example, Mandelbrot (1983) argues that models which generate spatial or physical predictions that can be mapped or visualised must 'look right'.

Patel and Smith (2012) provide a review of tools, techniques and methods for such visualizations in the second and third dimensions. Such tools as game engines and virtual worlds (see Crooks et al. 2009) provide a highly visual and immersive medium for ABM and has the potential to greatly aid in the communication and understanding of agent-based models. The dynamic and real-time visualisation and communication options (especially those in virtual worlds) provided by agent-based models allows us to address the challenge modellers face on how we might communicate and share agent-based models with all those we seek to influence. In the past, model results were mainly presented through the discussion of the model outcomes via static charts or screen shoots. However, as one of the key aspects of ABM lies in their dynamics, real-time visualisation of models and their outcomes can capture this and in a sense, show how micro interaction of individuals leads to more aggregate outcomes.

However, visualisation alone does not address all the issues relating to the communication of agent-based models. We also need methods to convey the model structure and key model parameters that allow for replication of such models. Replication of models allows others to gain confidence about the model and its underlying assumptions (see Crooks et al. 2008). Replication can be aided through the releasing of the source code of the model, along with the data and parameters presented in a paper or by providing a detailed description of the model such as the "ODD" (Overview, Design concepts, Details) protocol (Grimm and Railsback 2012).

5.6 Advantages of Agent-Based Modelling

The way we currently conceptualise and model human geographical systems, in particular the evolution of cities, has changed, from the aggregate to disaggregate, and from the static to the dynamic as highlighted in Table 5.1. ABM provides us with tools to explore this change in approach. There are three main claimed advantages of the agent-based approach over traditional modelling techniques, such as top-down techniques of non-linear dynamical systems in which related state variables are aggregated (e.g. through differential equations). The agent-based approach: (i) captures emergent phenomena; (ii) provides a natural environment for the study of certain systems; and (iii) is flexible, particularly in relation to the development of geospatial models.

Traditional urban models focused on modelling the system of interest top-down in contrast to model developers who divided the city into a few 100 units, while assuming average behaviour of individuals. Through their ability to describe behaviour and interactions of a system's constituent parts from the bottom-up,[1] ABM provides

Table 5.1 The changes in modelling techniques using ABM over traditional modelling of the 1960s and 1970s (Source: Bernard 1999)

Traditional modelling	Agent-based modelling
Deterministic (one future)	Stochastic (multiple futures)
Allocative (top-down)	Aggregative (bottom-up)
Equation based formulas	Adaptive agents
Do not give explanations	Explanatory power
Few parameters	Many parameters
Spatially coarse	Spatially explicit
Environment given	Environment created
You react to them	You learn from them

[1]While ABM deals with individual and not aggregate behaviour, often it is neither feasible nor desirable to model complete agent heterogeneity. Instead agents are often given a representative behaviour; thus we move from average aggregate behaviour to average individual behaviour. However, greater heterogeneity can be introduced by adding 'noise' to such agents.

an alternative approach. Bonabeau (2002) has identified a non-exhaustive list of conditions where ABMs can be useful for capturing emergent behaviour:

1. Interaction between agents is complicated, non-linear, discontinuous, or discrete (i.e. the behaviour of an agent can be altered dramatically, even discontinuously, by other agents). This can be particularly useful if describing discontinuity of individual behaviour, for example, using differential equations;
2. The ability to design a heterogeneous population of agents with an agent-based model is significant. Agents can represent any type of unit. Unlike agent-based models, aggregate differential equations tend to smooth out fluctuations. This is important because under certain conditions, fluctuations can be amplified: a system can be linearly stable but susceptible to large perturbations. Heterogeneity also allows for the specification of agents with varying degrees of rationality. This offers advantages over approaches that assume perfectly rational individuals, if they consider individuals at all;
3. The topology of agent interactions is heterogeneous and complex. Aggregate flow equations usually assume global homogeneous mixing, but the topology of an interaction network can lead to significant deviations from predicted aggregate behavior and,
4. Agents exhibit complex behaviour, including learning and adaptation.

Furthermore, the ability of agent-based models to describe the behaviour and interactions of a system allows for system dynamics to be directly incorporated into the model. This represents a movement away from the static nature of earlier styles of urban and regional modelling (see Batty 1976). However, while time in ABMs is still discrete, i.e. it still moves in 'snapshots', the time steps may be small enough to approximate real time dynamics. Additionally different processes occur over different time periods, for example, long term economic cycles, daily commuting and hour by hour social interaction. Agent-based models can incorporate these different scale time processes into a single simulation by using a variety of automata clocks designed to mimic the temporal attributes of the specific urban process under study (Torrens 2003), thus allowing the modeller to realistically simulate urban development or a particular geographical phenomenon (O'Sullivan 2001).

In many cases, ABM is a natural method for describing and simulating a system composed of real-world entities especially when using object-orientated principles (Gilbert and Terna 2000). The agent-based approach is more akin to 'reality' than other modelling approaches. Agent-based simulations provide an opportunity to represent and test social theory which cannot easily be described using mathematical formulae (Axelrod 1997). The models often map more naturally to the structure of the problem than equation-based models (Parunak et al. 1998) by specifying simple behavioural and transition rules attached to well defined entities, therefore providing a medium for the infusion of any geographic theory or methodology into the model. In particular, the agent-based approach can be useful when it is more natural to describe the constituent units of a system under some of the following conditions (Bonabeau 2002):

1. The behaviour of individuals cannot clearly be defined through aggregate transition rates (e.g. the decision to move);
2. Individual behaviour is complex. Although hypothetically any process can be explained by an equation, the complexity of differential equations increases exponentially as the complexity of behaviour increases. Describing complex individual behaviour with equations can therefore become intractable;
3. Activities are arguably a more natural way of describing a system than processes; and,
4. Agent behaviour is stochastic. Points of randomness can be applied strategically within agent-based models, rather than arbitrarily within aggregate equations.

Finally, the agent-based approach to modelling is flexible, particularly in relation to geospatial modelling. Notably, spatial simulations benefit from the mobility that agent-based models offer. An agent-based model can be defined within any given system environment (e.g. a building, a city, a road network, a computer network, etc). Therefore agent-based models are essentially without scale. It is the phenomena of interest which drives the scale to be used, for example, from the micro movement of pedestrians within a building during an evacuation (e.g. Gwynne et al. 2001), to the movement of cars on a street network (e.g. Nagel 2003) to the study of urban growth (e.g. Brown et al. 2005). Additionally as ABM allows for the representation of individual objects, it is therefore possible to combine these objects to represent phenomena at different scales within the same model. Furthermore, agents have the ability to physically move within their environment, in different directions and at different velocities. Agent mobility makes ABM very flexible in terms of potential variables and parameters that can be specified. Neighbourhoods can also be specified using a variety of mechanisms such as well understood geographical relations such as market catchments areas, travel to work zones, walking distance buffers etc.

The implementation of agent interactions can easily be governed by space, networks, or a combination of structures (as highlighted in Alam et al. 2012) This would be far more complex to explain by mathematics, for example (Axtell 2000). Significantly, agent-based models can regulate behaviours based on interactions at a specific distance and direction (thus allowing for action-at-a-distance). In addition, agent-based models also provide a robust and flexible framework for tuning the complexity of agents (i.e. their behaviour, degree of rationality, ability to learn and evolve, and rules of interaction). Another dimension of flexibility is the ability to adjust levels of description and aggregation. It is easy to experiment with aggregate agents, sub groups of agents, and single agents, with different levels of description coexisting within a model. Thus, the agent-based approach can be used when the appropriate level of description or complexity is unknown, and finding a suitable level requires exploration.

5.7 Limitations of Agent-Based Modelling

The enthusiasm of adopting the ABM approach for modelling geographical systems is curtailed by some limitations. Although common to all modelling techniques, one issue relates to the purpose of the model; a model is only as useful as the purpose for which it is constructed. A model has to be built at the right level of abstraction for every phenomenon, judiciously using the right amount of detail for the model to serve its purpose (Couclelis 2002). If the level of abstraction is too simple, one may miss the key variables. Too much detail, and the model will have too many constraints and become overly complicated. Abdou et al. (2012) provide useful advice for designing and building an agent-based model. This remains an art more than a science (Axelrod 2007). Axtell and Epstein (1994) provide practical guidelines for the evaluation of model performance depending on the level of model abstraction.

The nature of the system being modelled is another consideration. For example, a system based on human beings will involve agents with potentially irrational behaviour, subjective choices, and complex psychology (see Kennedy 2012, for an overview of how behavioural frameworks can be implemented in agent-based models). These factors are difficult to quantify, calibrate, and sometimes justify, which complicates the implementation and development of a model, as well as the interpretation of its simulation outputs. However, the fundamental motivation for modelling arises from a lack of full access to data relating to a phenomenon of interest. Often, the target itself is neither well-understood nor easy to access. The development of agent-based models offers a means to increase the utility of simulation models, by closely tailoring the model and subsequent analysis to the needs of end users (Parker et al. 2003). In particular, the visual communication often provided by spatially explicit models, especially those coupled with GIS, can be effective at depicting formal model results to a wide range of users (Axtell 2000). Nevertheless, a model's output must be interpreted appropriately. Varying degrees of accuracy and completeness in the model inputs determine whether the output should be used purely for qualitative insight, or accurate quantitative forecasting. Crooks and Castle (2012) review the purpose of different ABM approaches in more detail especially relating to explanatory and predictive (descriptive) modelling approaches.

By their very definition, agent-based models consider systems at a disaggregated level. This level of detail involves the description of potentially many agent attributes and behaviours, and their interaction with an environment. The only way to treat this type of problem in agent computing is through multiple runs, systematically varying initial conditions or parameters in order to assess the robustness of results (Axtell 2000). There is a practical upper limit to the size of the parameter space that can be checked for robustness, and this process can be computationally intensive, thus time consuming. Although computing power is increasing rapidly, the high computational requirement of ABM remains a limitation when modelling large systems (see Parry and Bithnell 2012).

In this sense, agent-based models have the potential to suffer from similar limitations of the first generation of urban models such as Lee's (1973) criticism of data hungriness. However, this can be overcome by considering what level of abstraction is needed to examine the phenomena of interest (for example, is 'all the detail needed?'). Or a series of smaller models could be created examining specific aspects of the system. There is also a lack of personal data both for the present and the past. For example in the UK, the smallest measure of individual data from the census is the Output Area which contains around 125 households (notwithstanding access to personal data, see Benenson et al. 2002) which can be obtained through MSM techniques as demonstrated in Birkin and Wu (2012).

Critics of complexity theory point out that the wide variety of surprising behaviour exhibited by mathematical and computational models are rarely found in the real-world. In particular, ABMs are very sensitive to initial conditions and to small variations in interaction rules (Couclelis 2002). This path dependence means that using ABM for prediction can be challenging (see Batty and Torrens 2005). Consequently, modellers of complex systems are never likely to enjoy the intellectual comfort of 'laws' as seen in the physical or chemical worlds (Wilson 2000). Despite this, and the other limitations that have been highlighted, ABM is a useful tool for exploring systems that exhibit complex behaviour. They highlight uncertainty behind modelling geographical systems and provide a technique to explore such uncertainty through their ability to generate possible futures rather than generating definitive models with strong predictive assumptions (Epstein 1999). Complexity theory has brought awareness of the subtle, diverse, and interconnected facets common to many phenomena, and continues to contribute many powerful concepts, modelling approaches and techniques (see Manson et al. 2012 for further information). In this vein, Section 5.8 explores general ABM applications before focussing on agent-based models of geographical systems.

5.8 Applications of Agent-Based Models

It is impractical to comprehensively review the full range of ABM applications within this chapter, and even examination of a representative sample presents a challenging exercise. ABMs have been developed for a diverse range of subject areas, such as: archaeological reconstruction of ancient civilisations (Axtell et al. 2002; Kohler et al. 2000); understanding theories of political identity and stability (Lustick 2002); understanding processes involving national identity and state formation (Cederman 2001); biological models of infectious diseases (Yang and Atkinson 2005); growth of bacterial colonies (Kreft et al. 1998); single- (Emonet et al. 2005) and multi-cellular level interaction and behaviour (Athale and Deisboeck 2006); alliance formation of nations during the Second World War (Axelrod and Bennett 1993); modelling economic processes as dynamic systems of interacting agents (Tesfatsion 2006); company size and growth rate distributions (Axtell 1999); geographical retail markets (Heppenstall et al. 2006), size-frequency distributions for traffic jams

(Nagel and Rasmussen 1994); price variations within stock-market trading (Bak et al. 1999); voting behaviours in elections (Kollman et al. 1992); identifying and exploring behaviour in battlefields (Ilachinski 1997); spatial patterns of unemployment (Topa 2001); trade networks (Epstein and Axtell 1996); business coalitions over industry standards (Axelrod 2006); social networks of terrorist groups (North et al. 2004), to name but a few. These examples can be constructed as lying on a continuum, from minimalist academic models based upon ideal assumptions, to large scale commercial decision support systems based upon real-world data. In relation to the focus of this chapter, the remainder of this section concentrates on the origin of ABM applied to urban phenomena, particularly in a geographical context.

Despite the advantages of ABM as a tool for simulation, ABM has only recently been adopted for geospatial research. Thomas Schelling is credited with developing the first social agent-based model in which agents represent people, and agent interactions represent a socially relevant process. Schelling's (1971) model demonstrated that stark segregation patterns can emerge from migratory movements among two culturally distinct, but relatively tolerant, types of household. Yet ABM did not begin to feature prominently in the geographical literature until the mid-1990s when Epstein and Axtell (1996) extended the notion of modelling people to growing entire artificial societies. Epstein and Axtell's Sugarscape model demonstrated that agents could emerge with a variety of characteristics and behaviours suggestive of a rudimentary society (e.g. in terms of patterns of death, disease, trade, health, culture, conflict, war, etc).

The above two models have inspired a number of modelling efforts with respect to urban simulation and it is to this that we now turn briefly. Further information and summaries of ABM applications are presented in Parts 3 and 4 of this book. As stated previously, ABMs can be applied to any scale, from the atomic to the global. How one defines an agent depends on what phenomena one is interested in.

Numerous efforts have been made to apply ABM to environmental modelling especially land-use and land-cover change models (see Parker 2005; Parker et al. 2012; Magliocco 2012; for further details). ABM specifically pertaining to urban phenomena including dynamics in Latin American cities (Barros 2012) urban housing dynamics (Benenson et al. 2002), urban growth and residential location (Torrens 2006; Brown et al. 2005; Liu and Feng 2012), and gentrification (Jackson et al. 2008). The impact of policy on geographical areas has also been investigated through ABM, for example education planning (Harland and Heppenstall 2012) and crime simulation (Malleson 2012). Due to the ability of agents within ABMs to move, they are also commonly used to simulate traffic movement (e.g. Beuck et al. 2008). Additionally, numerous applications have been developed specifically to study micro-scale phenomena such as pedestrian models (Johnasson and Kretz 2012), which explore how agents move around their environment. Other useful examples of spatially explicit agent-based models include: the simulation of pedestrians in the urban centres (Haklay et al. 2001), the examination of crowd congestion at London's Notting Hill carnival (Batty et al. 2003), and emergency evacuation of buildings (e.g. Gwynne et al. 2001).

5.9 Conclusion

Successfully replicating the processes and dynamics that occur within geographical systems is highly challenging. There are a potentially infinite number of individual components linked together by often unknown interconnected processes that play out at different spatial and temporal scales. The notion of bottom-up modelling advocated by ABM allows the results of local phenomena to be understood and measured at a global level. While established methods, such as SI (Spatial Interaction) modelling, treat populations as aggregate homogeneous components, ABMs potentially allow every individual to be assigned their own characteristics. This is a powerful paradigm that holds great promise for facilitating greater understanding of geographical systems.

This chapter has provided a general introduction to ABM. Along with a presentation of the main characteristics of ABM, the distinction between ABM, CA and MSM have been discussed. Important considerations when working with ABM, for example validation, verification and visualisation, were presented along with the advantages and limitations of this approach for geographical systems. The chapter concluded by exploring a diverse range of geographical applications of ABM.

References

Abdou, M., Hamill, L., & Gilbert, N. (2012). Designing and building an agent-based model. In A. J. Heppenstall, A.T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 141–166). Dordrecht: Springer.

Alam, S. J., Geller, A., & Tsvetovat, M. (2012). Networks in agent-based social simulation. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 199–216). Dordrecht: Springer.

Athale, C. A., & Deisboeck, T.S. (2006). The effects of EGF-receptor density on multi-scale tumour growth patterns. *Journal of Theoretical Biology, 238*(4), 771–779.

Axelrod, R. (1997). *The complexity of cooperation: Agent-based models of competition and collaboration*. Princeton: Princeton University Press.

Axelrod, R. (2006). Agent-based modelling as a bridge between disciplines. In L. Tesfatsion & K. L. Judd (Eds.), *Handbook of computational economics: Agent-based computational economics* (Vol. 2, pp. 1598–1583). Amsterdam: North-Holland.

Axelrod, R. (2007). Simulation in the social sciences. In J. P. Rennard (Ed.), *Handbook of research on nature inspired computing for economy and management* (pp. 90–100). Hershey: Idea Group.

Axelrod, R., & Bennett, S. D. (1993). A landscape theory of aggregation. *British Journal of Political Science, 23*(2), 211–233.

Axtell, R. (1999). *The emergence of firms in a population of agents: Local increasing returns, unstable nash equilibria, and power law size distributions,* Working paper no. 3. Washington DC: Center on Social and Economic Dynamics (The Brookings Institute).

Axtell, R. (2000). *Why agents? On the varied motivations for agent computing in the social sciences*, Working paper 17. Washington DC: Center on Social and Economic Dynamics (The Brookings Institute).

Axtell, R., & Epstein, J. M. (1994), Agent-based modelling: Understanding our creations (pp. 28–32). *The Bulletin of the Santa Fe Institute*, Winter.

Axtell, R., Epstein, J. M., Dean, J. S., Gumerman, G. J., Swedlund, A. C., Harburger, J., et al. (2002). Population growth and collapse in a multiagent model of the Kayenta Anasazi in long house valley. *Proceedings of the National Academy of Sciences of the United States of America (PNAS), 99*(3), 7275–7279.

Bak, P., Paczuski, M., & Shubik, M. (1999). *Price variations in a stock market with many agents,* Discussion paper 1132. New Haven: Cowles Foundation, (Yale University). Available at http://cowles.econ.yale.edu/P/cd/d11a/d1132.pdf

Balan, G. C., Cioffi-Revella, C., Luke, S., Pamait, L., & Paus, S. (2003). MASON: A Java multi agent simulation library. In C. M. Macal,, M. J. North & D. Sallach (Eds.) *Proceedings of Agent 2003 Conference on Challenges in Social Simulation* (pp. 49–64). Chicago: University of Chicago.

Ballas, D., Clarke, G., & Wiemers, E. (2005). Building a dynamic spatial microsimulation model for Ireland. *Population, Space and Place, 11*(3), 157–172.

Barros, J. (2012). Exploring urban dynamics in Latin American cities using an agent-based modelling approach. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 571–589). Dordrecht: Springer.

Batty, M. (1976). *Urban modelling: Algorithms, Calibrations, Predictions*. Cambridge: Cambridge University Press.

Batty, M. (2007). Model cities. *Town Planning Review, 78*(2), 125–178.

Batty, M. (2012). A generic framework for computational spatial modelling. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.) *Agent-based models of geographical systems* (pp. 19–50). Dordrecht: Springer.

Batty, M., & Torrens, P. M. (2005). Modelling and prediction in a complex world. *Futures, 37*(7), 745–766.

Batty, M., Desyllas, J., & Duxbury, E. (2003). Safety in numbers? Modelling crowds and designing control for the Notting Hill Carnival. *Urban Studies, 40*(8), 1573–1590.

Benenson, I., & Torrens, P. M. (2004). *Geosimulation: Automata-based modelling of urban phenomena*. London: Wiley.

Benenson, I., Omer, I., & Hatna, E. (2002). Entity-based modelling of urban residential dynamics: The case of Yaffo, Tel Aviv. *Environment and Planning B, 29*(4), 491–512.

Bernard, R. N. (1999). Using adaptive agent-based simulation models to assist planners in policy development: The case of rent control, Working paper 99-07-052. Santa Fe, NM: Santa Fe Institute.

Beuck, U., Rieser, M., Strippgen, D., Balmer, M., & Nagel, K. (2008). Preliminary results of a multi-agent traffic simulation for Berlin. In S. Albeverio, D. Andrey, P. Giordano, & A. Vancheri (Eds.), *The dynamics of complex urban systems: An interdisciplinary approach* (pp. 75–94). New York: Springer Physica.

Birkin, M., & Wu, B. (2012). A review of microsimulation and hybrid agent-based approaches. In A. J. Heppenstall, A.T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 51–68). Dordrecht: Springer.

Boman, M., & Holm, E. (2004). Multi-agent systems, time geography, and microsimulations. In M.-O. Olsson & G. Sjöstedt (Eds.), *Systems approaches and their application* (pp. 95–118). Norwell: Kluwer.

Bonabeau, E. (2002). Agent-based modelling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America (PNAS), 99*(3), 7280–7287.

Brown, D. G. (2006). Agent-Based Models. In H. Geist (Ed.), *The earth's changing land: An encyclopaedia of land-use and land-cover change* (pp. 7–13). Westport: Greenwood Publishing Group.

Brown, D. G., Page, S. E., Riolo, R., Zellner, M., & Rand, W. (2005). Path dependence and the validation of agent-based spatial models of land use. *International Journal of Geographical Information Science, 19*(2), 153–174.

Casti, J. L. (1997). *Would-be-worlds: How simulation is changing the frontiers of science*. New York: Wiley.

Cederman, L. E. (2001). Agent-based modelling in political science. *The Political Methodologist, 10*(1), 16–22.

Cederman, L. E. (2004). Computational models of social systems. Available at http://www.cederman.ethz.ch/teaching/archive/compmodels/ss2004/slides/compmodels-20040330.pdf
Couclelis, H. (2002). Modelling frameworks, paradigms, and approaches. In K. C. Clarke, B. E. Parks, & M. P. Crane (Eds.), *Geographic information systems and environmental modelling* (pp. 36–50). London: Prentice Hall.
Crooks, A. T., & Castle, C. (2012). The integration of agent-based modelling and geographical information for geospatial simulation. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 219–252). Dordrecht: Springer.
Crooks, A. T., Castle, C. J. E., & Batty, M. (2008). Key challenges in agent-based modelling for geo-spatial simulation. *Computers, Environment and Urban Systems, 32*(6), 417–430.
Crooks, A. T., Hudson-Smith, A., & Dearden, J. (2009). Agent street: An environment for exploring agent-based models in second life. *Journal of Artificial Societies and Social Simulation, 12*(4). Available at http://jasss.soc.surrey.ac.uk/12/4/10.html
Emonet, T., Macal, C. M., North, M. J., Wickersham, C. E., & Cluzel, P. (2005). AgentCell: A digital single-cell assay for bacterial chemotaxis. *Bioinfomatics, 21*(11), 2714–2721.
Epstein, J. M. (1999). Agent-based computational models and generative social science. *Complexity, 4*(5), 41–60.
Epstein, J.M., & Axtell, R. (1996). *Growing artificial societies: Social science from the bottom up*. Cambridge: MIT Press.
Evans A. J. (2012). Uncertainty and error. In A. J. Heppenstall, A.T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 309–346). Dordrecht: Springer.
Franklin, S., & Graesser, A. (1996). Is it an agent, or just a program?: A taxonomy for autonomous agent. In *Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages* (pp. 21–35). Springer.
Gilbert, N., & Terna, P. (2000). How to build and use agent-based models in social science. *Mind and Society, 1*(1), 57–72.
Grimm, V., & Railsback, S.F. (2012). Designing, formulating and communicating agent-based models. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 361–377). Dordrecht: Springer.
Gwynne, S., Galea, E. R., Lawrence, P. J., & Filippidis, L. (2001). Modelling occupant interaction with fire conditions using the building EXODUS evacuation model. *Fire Safety Journal, 36*(4), 327–357.
Haklay, M., O'Sullivan, D., Thurstain-Goodwin, M., & Schelhorn, T. (2001). "So go downtown": Simulating pedestrian movement in town centres. *Environment and Planning B, 28*(3), 343–359.
Harland, K., & Heppenstall, A.J. (2012). Using agent-based models for education planning: Is the UK education system agent-based. In A. J. Heppenstall, A.T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 481–497). Dordrecht: Springer.
Heppenstall, A. J., Evans, A. J., & Birkin, M. H. (2006) Application of multi-agent systems to modelling a dynamic, locally interacting retail market. *Journal of Artificial Societies and Social Simulation*, *9*(3). Available at http://jasss.soc.surrey.ac.uk/9/3/2.html
Heppenstall, A. J., Evans, A. J., & Birkin, M. H. (2007). Genetic algorithm optimisation of a multi-agent system for simulating a retail market. *Environment and Planning B, 34*(6), 1051–1070.
Holland, J. H. (1995). *Hidden order: How adaptation builds complexity*. Reading: Addison-Wesley.
Ilachinski, A. (1997). *Irreducible Semi-Autonomous Adaptive Combat (ISAAC): An artificial-life approach to land combat*. Alexandria: Center for Naval Analyses.
Iltanen, S. (2012). Cellular automata in urban spatial modelling. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 69–84). Dordrecht: Springer.
Jackson, J., Forest, B., & Sengupta, R. (2008). Agent-based simulation of urban residential dynamics and land rent change in a gentrifying area of Boston. *Transactions in GIS, 12*(4), 475–491.
Johnasson, A., & Kretz, T. (2012). Applied pedestrian modelling. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 451–462). Dordrecht: Springer.

Kennedy, B. (2012). Accounting for human behaviour in agent-based models. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 167–179). Dordrecht: Springer.

Kohler, T.A., Kresl, J., Van Wes, Q., Carr, E., & Wilshusen, R.H. (2000). Be there then: A modeling approach to settlement determinants and spatial efficiency among late ancestral Pueblo populations of the Mesa Verde Region, U.S. Southwest. In T.A. Kohler & G. J. Gumerman (Eds.), *Dynamics in human and primate societies: Agent-based modeling of social and spatial processes* (pp. 145–178). Oxford: Oxford University Press.

Kollman, K., Miller, J. H., & Page, S. E. (1992). Adaptive parties in spatial elections. *American Political Science Review, 86*(4), 929–937.

Kornhauser, D., Wilensky, U., & Rand, D. (2009). Design guidelines for agent based model visualization. *Journal of Artificial Societies and Social Simulation*, *12*(2). Available at http://jasss.soc.surrey.ac.uk/12/2/1.html

Kreft, J. U., Booth, G., & Wimpenny, W. T. (1998). BacSim, a simulator for individual based modelling of bacterial colony growth. *Microbiology, 144*(12), 3275–3287.

Landis, J., & Zhang, M. (1998). 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, 25*(6), 795–824.

Law, A. M., & Kelton, D. (1991). *Simulation modelling and analysis* (2nd ed.). New York: McGraw-Hill.

Lee, D. B. (1973). Requiem for large-scale models. *Journal of the American Institute of Planners, 39*, 163–178.

Liu, Y., & Feng, Y. (2012). A logistic based cellular automate model for continuous urban growth simulation: A case study of the Gold Coast City, Australia. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 643–662). Dordrecht: Springer.

Lustick, I. (2002). PS-I: A user-friendly agent-based modelling platform for testing theories of political identity and political stability. *Journal of Artificial Societies and Social Simulation, 5*(3). Available at http://jasss.soc.surrey.ac.uk/5/3/7.html

Macal, C. M., & North, M. J. (2005). Tutorial on agent-based modelling and simulation. In M. E. Euhl, N. M. Steiger, F. B. Armstrong & J. A. Joines (Eds.), *Proceedings of the 2005 Winter Simulation Conference* (pp. 2–15), Orlando.

Magliocco, N. R. (2012). Exploring coupled housing and land market interactions through an economic agent-based model (CHALMS). In A. J. Heppenstall, A.T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 543–568). Dordrecht: Springer.

Malleson, N. S. (2012). Using agent-based models to simulate crime. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 411–434). Dordrecht: Springer.

Malleson, N. S., Heppenstall, A. J., & See, L. M. (2010). Simulating burglary with an agent-based model. *Computers, Environment and Urban Systems, 34*(3), 236–250.

Mandelbrot, B. (1983) *The fractal geometry of nature*. Freeman: San Francisco.

Manson, S. M., Sun, S., & Bonsal, D. (2012). Agent-based modeling and complexity. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 125–139). Dordrecht: Springer.

Nagel, K. (2003). Traffic networks. In S. Bornholdt & H. Schuster (Eds.), *Handbook of graphs and networks: From the genome to the internet* (pp. 248–272). New York: Wiley.

Nagel, K., & Rasmussen, S. (1994). Traffic at the edge of Chaos. In R. Brooks (Ed.), *Artificial life IV* (pp. 222–236). Cambridge, MA: MIT Press.

Ngo, T. A., & See, L. M. (2012). An agent-based modelling application of shifting cultivation. In A.J. Heppenstall, A. T. Crooks, L.M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 611–627). Dordrecht: Springer.

North, M. J., & Macal, C. M. (2007). *Managing business complexity: Discovering strategic solutions with agent-based modelling and simulation*. New York: Oxford University Press.

North, M. J., Macal, C. M., & Vos, J. R. (2004). Terrorist organization modelling. *North American Association for Computational Social and Organizational Science Conference 2004*, Pittsburgh.

O'Sullivan, D. (2001). Exploring spatial process dynamics using irregular cellular automaton models. *Geographical Analysis, 33*(1), 1–18.

Parker, D. C. (2005). Integration of geographic information systems and agent-based models of land use: Challenges and prospects. In D. J. Maguire, M. Batty, & M. F. Goodchild (Eds.), *GIS, spatial analysis and modelling* (pp. 403–422). Redlands: ESRI Press.

Parker, D. C., Manson, S. M., Janssen, M. A., Hoffmann, M. J., & Deadman, P. (2003). Multi-agent systems for the simulation of land-use and land-cover change: A review. *Annals of the Association of American Geographers, 93*(2), 314–337.

Parker, D. C., Brown, D. G., Filatova, T., Riolo, R., Robinson, D. T., & Sun, S. (2012). Do land markets matter? A modeling ontology and experimental design to test the effects of land markets for an agent-based model of ex-urban residential land-use change. In A. J. Heppenstall, A.T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical s ystems* (pp. 525–542). Dordrecht: Springer.

Parry, H. R., & Bithnell, M. (2012). Large scale agent-based modelling: A review and guidelines for model scaling. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.) *Agent-based models of geographical systems* (pp. 525–542). Dordrecht: Springer.

Parunak, H. V. D., Savit, R., & Riolo, R. L. (1998). Agent-based modelling vs. equation-based modelling: A case study and users' guide. *Proceedings of Multi-Agent Systems and Agent-Based Simulation (MABS'98)* (pp. 10–25), Paris.

Patel, A., & Hudson-Smith, A. (2012). Tools, techniques and methods for macro and microscopic simulation. In A. J. Heppenstall, A. T. Crooks, L.M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 379–407). Dordrecht: Springer.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology, 1*(1), 143–186.

Tesfatsion, L. (2006). Agent-based computational economics: A constructive approach to economic theory. In L. Tesfatsion & K. L. Judd (Eds.), *Handbook of computational economics: Agent-based computational economics* (Vol. 2, pp. 831–880). Amsterdam: North-Holland.

Topa, G. (2001). Social interactions, local spillovers and unemployment. *The Review of Economic Studies, 68*(2), 261–295.

Torrens, P. M. (2003). Automata-based models of urban systems. In P. A. Longley & M. Batty (Eds.), *Advanced spatial analysis: The CASA book of GIS* (pp. 61–81). Redlands: ESRI Press.

Torrens, P. M. (2006). Simulating sprawl. *Annals of the Association of American Geographers, 96*(2), 248–275.

White, R., Engelen, G., & Uljee, I. (1997). The use of constrained cellular automata for high-resolution modelling of urban land use dynamics. *Environment and Planning B, 24*(3), 323–343.

Wilson, A. G. (2000). *Complex spatial systems: The modelling foundations of urban and regional analysis*. Harlow: Pearson Education.

Wolfram, S. (2002). *A new kind of science*. Champaign: Wolfram Media. Champaign, IL.

Wooldridge, M., & Jennings, N. R. (1995). Intelligent agents: Theory and practice. *Knowledge Engineering Review, 10*(2), 115–152.

Wu, B., Birkin, M., & Rees, P. (2008). A spatial microsimulation model with student agents. *Computers Environment and Urban Systems, 32*(6), 440–453.

Yang, Y., & Atkinson, P. M. (2005). An integrated ABM and GIS model of infectious disease transmission. In S. Batty (Ed.), *Computers in Urban Planning and Urban Management (CUPUM)*, London.

Chapter 1
Perspectives on Agent-Based Models and Geographical Systems

Michael Batty, Andrew T. Crooks, Linda M. See, and Alison J. Heppenstall

Abstract This chapter guides the reader to the material in this book. It begins by outlining the meaning and rationale for agent-based models/modelling (ABM), focusing on their history, how they evolved and how they sit within the broader context of modelling and simulation for geographical systems. Three themes which we see essential to ABM are then outlined, namely the question of detail versus model and data parsimony of which ABM represents the former, questions of model validation that flow from this, and lastly issues about the extent to which ABM is a generic or specific style in terms of applications. We examine the essence of such models in terms of the way behaviour is modelled using various rules, and then we discuss technical issues such as computation, visualization, error, and schemes for model design. All this sets the context for the various chapters that follow. We conclude by explaining briefly what is contained in each chapter and by guiding the reader in how best to use this book.

M. Batty (✉)
Centre for Advanced Spatial Analysis (CASA) University College London, London, UK
e-mail: m.batty@ucl.ac.uk

A.T. Crooks
Krasnow Institute for Advanced Study, George Mason University, Fairfax, VA, USA
e-mail: acrooks2@gmu.edu

L.M. See
International Institute of Applied Systems Analysis (IIASA), Laxenburg, Austria
e-mail: see@iiasa.ac.at

A.J. Heppenstall
School of Geography, University of Leeds, Leeds, UK
e-mail: A.J.Heppenstall@leeds.ac.uk

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_1, © Springer Science+Business Media B.V. 2012

1.1 A Little Bit of History

It has been over 50 years since the first attempts were made to explain geographical systems using formal tools from mathematics. In the 1950s spurred on by rapid developments in digital computation that were immediately grasped as a new media in which to conduct virtual 'experiments' on human and physical systems, geographical theory underwent a radical shift in emphasis. Systems were articulated using ideas from physics and biology which found expression in a wide array of mathematical formalisms widely exploiting the role of analogy and metaphors involving representations and processes in the physical and natural sciences. Yet from the beginning, there was an assumption, usually implicit, that for geographical theory to be meaningful, it must deal with aggregates, ironing out the noise and much of the variation that was associated with spatial systems. Particularly in human geography, there was a tacit assumption that populations needed to be represented as aggregates whose attributes were averaged over different variational characteristics in such a way that homogeneity might be ascribed to their behaviour in space and time. This was not pursued in the mistaken belief that populations were actually homogeneous but in the interests of simplicity and parsimony, the search for such regularity appeared to be the correct way forward.

The experience 50 years on has been salutary. It has been exceptionally hard to find theories and models that are robust enough to withstand the sort of testing and validation that is associated with the harder sciences, particularly with classical physics. In the effort to simplify and distil the essence of geographical systems and their processes into the same kinds of simple but powerful laws that characterize the physical world, formal theory has tended to reduce geography to the point where much of the richness and variety that we associate with the world is defined away, often leaving simplistic notions that are both obvious and banal. Those developing such models are well aware of these limits through the tortuous process that has beset the field during these years. Yet in all of this, slowly but surely the idea that we need to represent geographical systems at a much more elemental level has taken hold. There has always been resistance to the idea that we should search for some atomic element or unit of representation that characterizes the geography of a place, and the social sciences in particular have been reluctant to consider the notion that models of their systems should be postulated and tested at the individual level. But as progress with aggregate models of homogenous populations has faltered, there has been a perceptible shift from aggregate to disaggregate, from populations to individuals, from macro to micro. In this, the notion of an 'agent' has become the focus of this new quest.

If you define a social system as a collection of agents, then immediately you tend to consider agents as individuals in a wider population, individuals that act purposively, that learn and innovate, thus introducing ideas that are hard to consider using more aggregate styles of representation. Agents generate actions that occur in time as well as space, that influence their wider environments and that cooperate as well as conflict with one another over the use of space. Defining many agents in a population immediately gives some sense of their diversity, and in this way, any heterogeneity in the system is directly picked up. It is easy to see why the idea of agent-based modelling (ABM)[1] has become so popular in the last two decades for it begets a style of modelling that has the capability of reflecting the richness of the world in a way that appears essential to any good explanation of how spatial structures such as cities, regions, the global system itself as well as all its physical components evolve and change. The power of the agent paradigm is consistent too with the fact that as our world has become more complex, largely due to increasing wealth and innovation in technology; data about ourselves is becoming more available, particularly through online sources. Ways of handling such data using ever powerful methods of computation are going hand in hand with these developments, all leading to the notion that simulating worlds composed of agents rather than their aggregates might now represent a feasible and productive way forward.

We consider that a book which synthesizes our collective wisdom on agent-based models in geographical systems is both opportune and timely; opportune because there is much to say about how we are beginning to build agent-based models and how geography imposes its own requirements on such developments, timely because so far there are few, if any, reviews of the state-of-the-art in this area, and those wishing to enter and contribute to the field require as much source material as we can muster. Here we have collected together a series of contributions that cover a very wide range of issues and approaches to agent-based modelling, beginning with a review of modelling styles and types that inform the field, moving then to more conceptual approaches, and then to methods and techniques that are involved in designing and constructing such models. These form the first three parts of the book and thence prepare the reader for a multitude of applications which we organize in terms of the scale of the agent – micro or macro. These form the last two parts which to an extent also correspond to spatial scale. They constitute about half the contributions contained herein, thus balancing theory, method and technique evenly with applications.

1.2 Essential Themes

There are many themes that we will identify in this opening chapter to which we will alert readers. In no sense do we consider these to be exhaustive but there are three we consider essential to an appreciation of all that follows: these involve the dramatic differences between the style of modelling which has dominated geographical theory and applications in the past from those which we consider now form the cutting edge. It is quite clear that ABMs represent geographical systems at a level of richness and variety that is an order of magnitude greater than their aggregate precursors. ABMs usually have many more components – where we think of these

---

[1] ABM is also taken to mean Agent-Based Model (s) as well as Modelling.

as agents themselves – than their aggregates, and this means that their attributes are specified at a level of detail that is associated with each individual agent. Interactions between agents are usually involved and thus the level of representation grows exponentially as the number of agents increases. Even if the number of agents is quite limited, often in cases where ABMs are used for pedagogic experiments, then the level of detail for each agent and their interaction is still substantial in comparison with their aggregates. In short, ABMs break the basic rule of science that theory must be parsimonious – as simple as possible – and that a theory or model is better than any other if it performs equally well but is simpler; this is Occam's razor. In fact, the argument for ABMs is quite the opposite. For many systems, we have plausible but non-testable hypotheses about how we think the system works, and if we exclude these simply because we cannot test them against data, then we are guilty of distorting our theory simply due to the expediency of not being able to test it using classical means: against independent data. This issue is of enormous significance for it throws into doubt the whole process of developing and testing models of geographical systems, indeed of testing and validating or falsifying any theory.

The conventional process of theory development in science begins with observation, proceeds to an induction of some theory from that data, and then proposes some hypothesis that is testable against some other independent set of observations, usually in a different time and a different place. This is the classic process of experimentation where the experiment is repeated and validated (or not), the theory then being refined (or rejected) in entirely different situations by independent scientists. In this book, most authors who are applying ABM to real situations do assume that their models must be validated against real data. Most however are also uncomfortable with this process for usually their models are only testable to a degree and much of what is specified in the model associated with the behaviours of agents is simply non testable in that data on processes, decisions and actions is not available and/or observable. Outcomes of agent behaviours may be testable but the processes involving such behaviour are not.

Accordingly ABM has seen the process of model testing being elaborated in much more detail than traditionally associated with aggregate modelling. In particular, tests for plausibility, experiments with running models under many different sets of initial conditions, sensitivity testing of model parameters as well as traditional algorithms used to maximize the goodness of fit have come to dominate the process. Added to this, the idea that models which are richer by an order of magnitude than their counterparts should be verified as well as validated against data has become significant. This means that models should be run to test whether they are behaving as their originators intended and this has little or nothing to do with how well they might reproduce observable data. Surrounding this discussion is the notion too that models are no longer built for prediction per se but as much to inform general scientific inquiry as well as any debate between stakeholders over what the future might hold (Epstein 2008). In short, these kinds of model are as much to structure debate and dialogue as to provide measures of how the future might turn out. This is a controversial issue that is increasingly important to social science as well as science itself as the classical canons of scientific inquiry melt away into the vestiges of history.

There is a third theme that relates to ABM and marks a major difference from the past. Just as the term model has come to embrace theory, the term computation has come to embrace model. Since digital computers became the environment in which this type of modelling is possible, methods of computation have come to influence the construction of models as much as theory has done. In this sense, modelling has become more generic rather than specific with generalized approaches to modelling for many different types of system being developed during the last 30 years. Initially models of geographical systems were tied very closely to theory and each individual model contained a sufficient amount of its originator's personal knowledge of the problem to be quite distinct in terms of its computation. Of course as soon as computer programs of any scope and size became available for specific classes of model, there was a demand to generalize the program to any application. In fact, the very act of model development presupposes that simulations would emerge which would be generalizable to different situations. Indeed a true test of any model has always been predicated on the basis of taking the model elsewhere and evaluating its performance on independent data (Lowry 1965). In this sense, computation itself needs to be generic.

The experience however has been somewhat different from this notion that good models are entirely generalizable for it would appear that only the simplest of models meet this criterion, and when they do, they tend to be of pedagogic value only. Most spatial models tend to be developed for very specific situations whose data and context is sufficiently different from any other for the model to be only usable in any immediate sense for the problem at hand. Moreover in the past, models have tended to be closer to theory than to generic computation but as more experience has been gained with modelling, generic approaches have been fashioned. In geographical modelling, the spatial dimension has been so strong as to inhibit the development of generic modelling until quite recently but there are now sufficiently different frameworks of a generic nature available for model-builders to consider adopting a framework first and then adapting this to the particular theory and problem that define the simulation that is required, rather than the other way around.

Agent-based modelling is one of the most important generic modelling frameworks to have been developed to date. It has emerged largely due to the convergence of object-oriented programming ideas in computer science with the need to represent the heterogeneity involved in many kinds of physical and human system at much greater levels of detail, issues that we have already noted in some detail above. Although geographical models were best represented by specific land use transportation interaction (LUTI) models tailored very specifically to urban theories based on urban economics and social physics, as soon as formal modelling began, generic approaches appeared, as for example in systems dynamics which was based on general ideas about formulating models as partial difference equations subject to capacity constraints. These, as Batty (2012) shows in the next chapter, did not find much favour in geographical analysis largely because they were hard to adapt to spatial systems but other approaches based on econometrics for example, have formed the basis of some spatial models, although this style of modelling is specific to economic analysis, notwithstanding its generalization to mainstream statistical modelling. As we recount in the chapters that follow, cellular automata (CA) modelling developed before ABM but the software used to implement these styles of models is quite elementary. Although some generic modelling packages such as SLEUTH and METRONAMICA have been developed (see Iltanen 2012), generic CA packages for geographical systems are not widely available. Microsimulation models are even more specific, notwithstanding their almost tool-like focus, and generic software has not appeared, again perhaps due to their focus in our field on space which is hard to embrace.

It may even be worth making a distinction between generic or specific models with respect to the way they are formulated and the software and tools which tend to be used in different model types. The problem is that in some senses tools such as those that exist in econometrics and statistics can be elevated to entire model systems while model approaches like microsimulation often feature as tools in generating data. In short, microsimulation can be used in spatial interaction models as can agent-based approaches. In such cases, the model in its traditional format is augmented by the addition of agents or a decomposition using synthetic data analysis techniques which are core to microsimulation. For example, in some of the social physics models that are examined towards the end of the book such as those involving rank size (Gulden and Hammond 2012), spatial interaction (Dearden and Wilson 2012) and population change (Pumain 2012), agent-based approaches are used in their implementation but their structure is one dictated by the original model framework not by ABM itself. Even more confusing is the fact that model systems merge into one another and this is very clear in the case of CA and ABM, but as we will see, microsimulation models can transition into ABM as shown in Wu and Birkin (2012). In fact Torrens (2012) augments CA and ABM with GIS and calls these geographical automata systems (GAS).

Only ABM has developed very general packages which can be applied to a wide array of systems and problems. For example, the packages that are popular range from sophisticated programming systems such as SWARM, plug-in Java-based environments like Repast and MASON, and simpler scripting languages like NetLogo (and its originator StarLogo). A good review of these tools is given by Crooks and Castle (2012) where they show that to an extent these packages encapsulate CA models. In several of the contributions that follow, CA represent the environment in which agents behave in spatial terms. The other feature that is important when generic modelling packages are used is that their generalizability is always limited in some way. This can also force the modelling effort to embrace tools and techniques that are not suited to the system in hand and if certain functions are absent, it can lead to models that lack certain key components that more specialized software will enable. In fact, it is now so easy to customize many of these packages and to add other software as plug-ins using standard methods of linkage that most generic software is capable of being easily extended to deal with system specifics. However the downside of all this effort is that models which are the most effective tend to be those that involve considerable programming effort. We have not yet reached and may never do so the point where model users can specify a model for a problem type and simply assume that it is computable from generic software.

1.3 Structural Rules, Behaviour, and Dynamics in ABM

Agents almost by definition are purposive. They are endowed with behaviours that are usually proscribed in a series of rules that are activated under different conditions. This is in the manner of stimulus and response (or push and pull, or some such reactive logic), and in this sense, agents always engender change. Dynamics which may not be explicit but is almost invariably implicitly temporal, thus comes onto the agenda and in this sense, ABM deals with dynamic modelling. This is in stark contrast to LUTI models for example which are comparatively static for the most part or microsimulation models which as Birkin and Wu (2012) note, can be either static or dynamic. A particularly simple kind of ABM is in fact a CA where the transition from one state of a cell to another state– in each geographic area – is based on a set of rules that might be seen as representing how the state of the cell behaves as all the cells around it change. This somewhat anthropomorphic interpretation of CA might be appropriate if the cell contains an individual who is fixed in location but whose attributes define their state which is continually changing. A good example of this is the simplest model of segregation due to Schelling (1978) where the cell state is an individual with one view or another, who may then change their view dependent on the number of surrounding cells with individuals holding similar or different views. Here the cell is the agent; the agent does not move in space but does move in terms of their opinion. Indeed CA models are excellent examples of structures where many rules of a relatively simple nature in and of themselves combine to generate extremely complex behaviours when operated on a large lattice of cells (Batty 2005).

Agent behaviours may be reactive (sometimes called passive) or proactive (anticipatory). Invariably such behaviours are engendered by the agents in question scanning their environment in which other agents exist. More complicated forms of ABM involve different classes of agent, with agents being a mixture of types along the spectrum from reactive to proactive. Agents may be any distinct object in the system that is involved in changes of state, ranging from actual individuals in human systems to elements of the built environment. Moreover unlike agent types can interact with one another. In fact, in object-orientated programming, any element in the computation can be an object which is endowed with properties. In particular in visual programming, all the various elements of the graphical user interface are agents or objects. This ability to define different types of objects gives ABM its power but it also defines its limits in that it is hard to see a completely general system where any kind of agent might be defined in terms of generic properties and attributes of any other.

Yet despite these constraints, it is possible to see very wide ranges of problem being simulated using ABM. The more specific involve literal interpretations of agents as individuals in the human population such as those used in pedestrian and crowd modelling, the best examples here being those discussed by Patel and Hudson-Smith (2012) and Johansson and Kretz (2012). At the other extreme, ABM can be used to simulate interactions between groups of humans or even groups of policies that do not have a direct association with specific individuals as, for example, in a whole range of land cover models such as those used in developing countries where land and aid are key to development. In all these cases, at the heart of ABM lie processes of change which in our context have an impact on the geography of the system in question. These rules embody the key elements of the processes involved reflecting the way agents operate which lie at the core of the model. Agents interact with one another and with their environment, changing each other and their environment and in this sense, ABM is able to deal with open systems in a way that more specific modelling approaches cannot. These processes cannot be prescribed outside the modelling context except to say that they reflect a wide range of techniques. Simple rules of logic as in CA models are rather standard but many criteria are also built on algebraic functions that in geographical systems often relate time and space, action at a distance and across time. In fact the many contributions in this book show this variety in the way model processes are articulated, ranging from the standard algebraic formulations of micro economic theory (see Magliocca 2012) all the way to the rule-based logics used by Liu and Feng (2012) in their extension of CA modelling for urban development.

There are three elements related to dynamics and behaviour that are worth flagging as these appear many times in the various contributions that follow. First there is the question of cognition that relates to how agents perceive change in their wider environment and how they learn. Learning is often simulated through simple exposure to events over time and by watching what the majority do. In ABM, navigation and way finding in geographical space tend to be the most obvious elements in which the cognitive apparatus of the agent is utilized. There is little formal theory about how agents might best learn as the rule-based structure of many ABMs mean that such behaviours are defined in ad hoc empirical ways that are often tested using trial and error experiments. Second there is the question of scale. Behaviours occur across many scales but in their most elemental, these lie at the finest scales where the individual is located. Various ABMs and certainly CA models assume some principles of self similarity which operate across spatial scales and lead to the emergence of patterns at higher levels consistent with fractal structure. This is central to complexity theory. As ABMs are applied to coarser spatial scales, models change in focus and often even in type as the agent paradigm weakens although it is more likely that the way the model operates and the processes that are defined change rather than the framework itself. ABMs become less predictive and more speculative as scale changes from finer to coarser, from small scale to large.

The last point worth noting is that ABMs deal almost by definition with interactions, with their environment but also with inter-agent links. This introduces directly the concept of networks which appear implicitly in many of the contributions presented in this book. In fact, we do not emphasize networks very strongly in this book and there are no specific contributions apart from those dealing with movement of pedestrians and more aggregate populations. In a sense, this mirrors the fact that only quite recently have researchers in the geographical sciences begun to grapple with networks (although these have been implicit in spatial interaction and LUTI models for many years). One of the main developments in network science is their linkage with epidemiological models where propagations of rumour, innovation, disease, indeed any process that spreads through space and time can be cast in an ABM framework, as illustrated in Simoes (2012). We will pick such issues up in our conclusions when we anticipate the future of this field where we see agents moving across networks as being central to new applications.

1.4 Computation, Calibration, Error and Uncertainty

Before we launch into a brief guide to the contributions that follow, we will address a series of more technical questions that pervade any and every approach to modelling. In principle, ABMs can generate enormous data requirements in that the assumption is that every agent in a population that in the past was treated in aggregate (or not at all), must be represented explicitly in some computable form. This can give rise to massively parallel computation where agents are farmed out for individual processing on multiple processors but it also leads to simplifications which involve aggregation into super-individuals in the manner suggested by Parry and Bithell (2012). Moreover computation is massively increased because each agent has to be tracked and in situations where there are thousands of such agents, it is usually necessary to visualize their behaviours so some sense of the order and pattern generated in their simulation can be evaluated. We have not yet mentioned visualization but in these new generations of model, both CA and ABM, visualization has become essential based on links to GIS, CAD and other multimedia systems as noted by Patel and Hudson-Smith (2012).

Data requirements notwithstanding, most ABM so far, with the exception of large transport models such as TRANSIMS (Casti 1997) and MATSIMS (Rieser et al. 2007), do not appear to use intensive computational facilities or generate massive demands for parallel or related high performance computing. This is partly because many of the processes that characterise ABM cannot be matched with real world data and thus are never testable, despite the fact that most ABM have multiple parameter sets that make a complete enumeration of their possible solutions impossible. There are proposals to build extensive global models of entire populations such as that suggested by Epstein (PACER 2011) where some 6.5 billion individuals are being simulated with respect to their abilities and exposures to generate global pandemics. Visualization is essential for such models and this can set up severe computational demands. However most ABMs run in desktop environments and tend to be more pedagogic in focus due to the fact that once the number of assumptions which are non-testable yet plausible begins to dominate model structure, the models themselves become more like devices on which to develop thought experiments, to inform debate rather than to predict actual futures.

We have already noted the problem of calibration which has been extended dramatically during the last two decades to embrace not only validation and fine tuning through calibration but extensive sensitivity testing, checks for plausibility, verifiability of the model's implementation, and various aggregation checks against different layers of data. Error and uncertainty are key to models that have many processes and multiple assumptions for a good model might minimize error and reduce uncertainty as much as it might optimize its goodness of fit against actual data. In this sense, ABMs cover a wider range of issues in terms of their validation than other more parsimonious models for there are many issues that need to be judged qualitatively and have no equivalent in quantitative evaluation. Evans (2012) outlines the key issues involved in exploring ABM in terms of error and uncertainty defining a cornucopia of possible sources of error, noting contrasts between accuracy and precision and defining issues involving risk and uncertainty as these come to characterize actual models and their outputs. In one sense, all the models introduced here address these issues but few do so explicitly, as much because the line needs to be drawn between what is possible, what is worthwhile and what is feasible in terms of the level of resources related to the modelling effort.

One last issue involves the actual process of model design. Many chapters that follow deal with different approaches to model construction but it is Grimm and Railsback (2012) who address the issue directly in outlining a procedural approach to evaluating models and this is immediately applicable to the design of a good ABM. They review ABM using a structure which provides Overview (O), Design (D), and Details (D) which they term ODD. From this structure, they are able to derive design patterns that enable model-builders to produce a scheme for Pattern Oriented Modeling (POM). This guides the designer in developing good ABMs based on a considered view of how entities, states, and processes need to be incorporated into the best model possible. This scheme is gaining ground in this field and others writing in this book are beginning to use it.

1.5 The Structure and Rationale for What Follows

We have divided the book into two main sections which in turn are divided in parts. In the first half of the book which is organized in three parts, we review ABM in Part 1 with respect to other related but different approaches, then in Part 2 in terms of their conceptual structure, and lastly in Part 3 in terms of the tools and techniques used to operationalize such models. In the second half of the book, we deal with model applications and divide these into two. Part 4 deals with micro models which are the true preserve of ABM while Part 5 deals with macro models, largely how macro patterns of spatial development and interaction often structured around other model frameworks, are implemented using ABM. As we noted above, the division into micro and macro applications tends to be one of sectoral or topical aggregation rather than spatial scale although there is some correlation between them.

In Part 1, Batty (2012) begins with an overview of models in general attempting to compare ABM and CA with other approaches such as LUTI, microsimulation, and systems dynamics models. This is followed by Birkin and Wu's (2012) more detailed review of microsimulation models which are close in spirit if not in structure to ABM, while Iltanen (2012) attempts the same review for CA models. In this sense, we establish that the wider class of ABM dealt with in this book includes microsimulation and CA which in one sense are extreme variants of this general domain. This part is concluded with a survey of ABM itself by Crooks and Heppenstall (2012) who examine the history, scope and focus of the field so far, noting the correspondence between all three model types: CA, microsimulation, and ABM.

Part 2 deals with more conceptual issues. O'Sullivan et al. (2012) provide a somewhat oblique perspective on ABM, explaining a little about how ABMs actually work but also cautioning the reader to identify conditions under which this style of modelling is most appropriate. Manson et al. (2012) take this further when they relate ABM to the wider domain of the complexity sciences, arguing that this is one of the main tools to simulate systems which operate from the bottom up and generate emergent patterns at coarser spatial and more aggregate scales. Abdou et al. (2012) then provide a blow by blow account of how to design and build an ABM. They set this context by exploring two well known models – Sugarscape which is the spatial ABM developed by Epstein and Axtell (1996) and Schelling's (1978) model of residential segregation both of which illustrate how emergence occurs in such systems. But they reserve their key example to the construction of car-following models that generate traffic jams of the classic kind that are pictured using what traffic engineers have for many years referred to as the 'fundamental diagram' – the relationships between speed and flow, which in turn shows how as flow increases so does speed only to level off after a flow threshold has been reached and then decline when the traffic jams: another example of an emergent phenomenon. Kennedy (2012) provides a useful exploration of cognition in ABM introducing some key issues involving the simulation of behaviour and this is followed by Ngo and See's (2012) discussion of methods of calibrating and validating an ABM which are far more detailed and inquisitive than methods used for traditionally more macro, aggregative and parsimonious models. This part is concluded by Alam et al. (2012) who broach the question of networks in ABM, reviewing issues of interaction, which involve specifying neighbourhood sizes, segregation rules and the way ideas and diseases propagate.

In Part 3, Crooks and Castle (2012) begin with a detailed review of ABM in terms of its software and the generic packages that have been developed to implement a range of model types. They conclude that space is not that well represented by such models, although GIS can now be linked to most of these packages. Stanilov (2012) then presents a more reflective essay on how space is incorporated in CA and ABM and this is followed by Parry and Bithell's (2012) chapter on computational issues that they discuss through the medium of model scaling which is akin to aggregation which preserves the role of the agent. Evans (2012) then deals with error and uncertainty and Wu and Birkin (2012) show how microsimulation can be augmented by ABM, showing exactly how these two frameworks are consistent and of course complementary to one another. The last two chapters which conclude this part and the first half of the book are those by Grimm and Railsback (2012) who introduce their ODD framework noted above and by Patel and Hudson-Smith (2012) who deal with models of crowding which use both macro and microscopic simulation but which illustrate quite clearly the need for good visualization in this field.

The second half of the book deals with applications which demonstrate the concepts, principles, and techniques that are dealt with in Parts 1–3. Part 4 deals with micro ABMs which cover crime, pedestrian movement, educational demand and supply, health, housing choice and land. These are all sectors that can be described in fine spatial detail and where populations are disaggregated to the level where individuals are explicitly represented in terms of their spatial behaviours. Malleson (2012) begins with his model of burglary that involves modelling how burglars select residential homes to rob and learn from the experience. Mobile crime is a key feature of these models. Torrens (2012) then shows how GIS can be added to ABM in his models of pedestrian movement while Johansson and Kretz (2012) provide a detailed review of the various models involved. Rand (2012) explores how micro and macro ABMs fuse into one another while Harland and Heppenstall (2012) and Smith (2012) outline how the education and health sectors can be simulated using the notion of agents being matched and allocated to school and health facilities. Jordan et al. (2012) examine diversity in housing markets using ABMs while Parker et al. (2012) explore how land markets can be modelled in the context of residential land use changes, specifically urban sprawl. Magliocca (2012) concludes this section with a foray into how a housing market model can be developed using ABM, an example of where urban economic theory provides the overarching structure which can be implemented by defining individuals engaged in demand and supply as agents.

In Part 5, the focus shifts to macro models, which are both spatially and sectorally orientated to aggregates but with these aggregates being applicable to space and sectors not the individual agents that populate them. Barros (2012) develops various ABMs of the peripherization growth process in Latin American cities using CA representations where the focus is on developing analogues of real growth patterns which manifest the sort of inequalities that characterize such cities. Simoes (2012) develops a robust model of the spread of mumps in Portugal that is implemented using standard epidemiological models in a spatial and network context. Ngo et al. (2012) show how land use and farming interests and policies in a Vietnamese village can be simulated using ABM and then Banos and Genre-Grandpierre (2012) explore a CA-ABM like model of idealized network systems with traffic flow which, like Abdou, Hamill and Gilbert's paper earlier in the book, mirrors how jams build up in spatial networks. Liu and Feng (2012) then develop an extended CA model of urban growth which is illustrative of how error and uncertainty can be incorporated into such models while Cabrera et al. (2012) examine how ABM lies at the basis of land cover models of agriculture in a developing countries context. The book is concluded with three papers that deal with traditional social physics models which can be implemented using ABM. First Gulden and Hammond (2012) show how a variant of a network model of cities linking to one another can be used to generate city size distributions that mirror familiar power laws. Dearden and Wilson (2012) implement their Boltzmann-Lotka-Volterra models that link spatial interaction to constrained logistic growth by running the model through agents rather than aggregates. Finally Pumain (2012) explores her SimPop model framework, which she and her colleagues have been developing for over a decade, showing how agent interactions and actions generate the distributions and sizes of cities that have existed in Europe from the thirtieth century.

1.6 A Guide for the Reader

Many readers will be familiar with agent-based modelling to some degree and will wish to dip into the contributions that follow in an order that they will be able to determine from the titles and abstracts of the various chapters. But for those who are new to this field, we have organized the contributions beginning with more general overviews of the field, and then filling in more technical detail as we proceed. The first three parts provide a reasonable primer on ABM for those who have not explored the field before and the last two parts provide examples of applications to geographical systems. For those who wish to learn quickly about the field, then the contributions in Part 1 provide overviews, in Chap. 2 of six related modelling styles and types of which ABM is one, in Chap. 3 of microsimulation and in Chap. 4 of CA that are those styles of model that are closest to ABM, and lastly in Chap. 5 of ABM itself. If readers then wish to concentrate on filling in more detail about ABM, we advise them to look at Chaps. 6, 7, 8, 12 and 17 which focus exclusively on ABM and how such models can be defined, constructed and implemented. The rest of the contributions in the first three parts expand this overview to include related models and more technical details while Parts 4 and 5 deal with applications which are self explanatory. In no sense, do we as editors consider this set of contributions to be any kind of finished product. ABM is a work in progress and this represents as good a snapshot that we can currently assemble (in 2012) of this world as it is developing.

References

Abdou, M., Hamill, L., & Gilbert, N. (2012). Designing and building an agent-based model. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 141–166). Dordrecht: Springer.

Alam, S. J., Geller, A., & Tsvetovat, M. (2012). Networks in agent-based social simulation. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 199–216). Dordrecht: Springer.

Banos, A. & Genre-Grandpierre, C. (2012). Towards new metrics for urban road networks: Some preliminary evidence from agent-based simulations. In A. J. Heppenstall, A.T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 627–642). Dordrecht: Springer.

Barros, J. (2012). Exploring urban dynamics in Latin American cities using an agent-based simulation approach. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 571–589). Dordrecht: Springer.

Batty, M. (2005). *Cities and complexity: Understanding cities with cellular automata, agent-based models and fractals*. Cambridge, MA: The MIT Press.

Batty, M. (2012). A generic framework for computational spatial modelling. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 19–50). Dordrecht: Springer.

Birkin, M. & Wu, B. (2012). A review of microsimulation and hybrid agent-based approaches. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 51–68). Dordrecht: Springer.

Cabrera, R., Deadman, P., Vanwey, L.K., Brondízio, E.S., & Moran, E. (2011). Exploring demographic and lot effects in an ABM/LUCC of agriculture in the Brazilian Amazon. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 663–676). Dordrecht: Springer.

Casti, J. L. (1997). *Would-be worlds: How simulation is changing the frontiers of science*. New York: Wiley.

Crooks, A. T. & Castle, C. (2012). The integration of agent-based modelling and geographical information for geospatial simulation. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 219–252). Dordrecht: Springer.

Crooks, A. T. & Heppenstall, A. J. (2012). Introduction to agent-based modelling. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 85–105). Dordrecht: Springer.

Dearden, J. & Wilson, A. (2012). The relationship of dynamic entropy maximising and agent-based approaches in urban modelling. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 705–720). Dordrecht: Springer.

Epstein, J.M. (2008). Why model? *Journal of Artificial Societies and Social Simulation*, *11*(4), 12. Available at http://jasss.soc.surrey.ac.uk/11/4/12.html. Accessed 16 June 2011.

Epstein, J. M., & Axtell, R. (1996). *Growing artificial societies: social science from the bottom up*. Cambridge, MA: The MIT Press.

Evans, A. (2012). Uncertainty and error. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 309–346). Dordrecht: Springer.

Grimm, V. & Railsback, S.F. (2012). Designing, formulating and communicating agent-based models. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 361–377). Dordrecht: Springer.

Gulden, T. R. & Hammond, R. A. (2012). Beyond Zipf: An agent-based understanding of city size distributions. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 677–704). Dordrecht: Springer.

Harland, K. & Heppenstall, A. J. (2012). Using agent-based models for education planning: Is the UK education system agent based? In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 481–497). Dordrecht: Springer.

Iltanen, S. (2012). Cellular automata in urban spatial modelling. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 69–84). Dordrecht: Springer.

Johansson, A. & Kretz, T. (2012). Applied pedestrian modeling. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 451–462). Dordrecht: Springer.

Jordan, R., Birkin, M., Evans, A. (2012). Agent-based modelling of residential mobility, housing choice and regeneration. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 511–524). Dordrecht: Springer.

Kennedy, W. G. (2012). Modelling human behaviour in agent-based models. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 167–179). Dordrecht: Springer.

Liu, Y. & Feng, Y. (2012). A logistic based cellular automata model for continuous urban growth simulation: A case study of the Gold Coast City, Australia. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 643–662). Dordrecht: Springer.

Lowry, I.S. (1965). A short course in model design. *Journal of the American Institute of Planners, 31*, 158–165.

Magliocca, N. R. (2012). Exploring coupled housing and land market interactions through an economic agent-based model (CHALMS). In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 543–568). Dordrecht: Springer.

Malleson, N. (2012). Using agent-based models to simulate crime. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 411–434). Dordrecht: Springer.

Manson, S. M., Sun, S., & Bonsal, D. (2012). Agent-based modeling and complexity. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 125–139). Dordrecht: Springer.

Ngo, T. N, & See, L. M. (2012). Calibration and validation of agent-based models of land cover change, In A. J. Heppenstall, A.T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 181–196). Dordrecht: Springer.

Ngo, T. A., Drake, F., & See, L. M. (2012). An agent-based modelling application of shifting cultivation. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 611–627). Dordrecht: Springer.

O'Sullivan, D., Millington, J., Perry, G. & Wainwright, J. (2012). Agent-based models – because they're worth it? In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 109–123). Dordrecht: Springer.

PACER (2011). National Center for the Study of Preparedness and Catastrophic Event Response (PACER). Available at http://www.orau.gov/dhssummit/factsheet/PACER2011.pdf. Accessed 16 June 2011.

Parker, D.C., Brown, D.G., Filatova, T., Riolo, R., Robinson, D.T., & Sun, S. (2012). Do land markets matter? A modeling ontology and experimental design to test the effects of land markets for an agent-based model of ex-urban residential land-use change. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 525–542). Dordrecht: Springer.

Parry, H. R. & Bithell, M. (2012). Large scale agent-based modelling: A review and guidelines for model scaling. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 271–308). Dordrecht: Springer.

Patel, A. & Hudson-Smith, A. (2012). Agent tools, techniques and methods for macro and microscopic simulation. In A. J. Heppenstall, A.T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 379–407). Dordrecht: Springer.

Pumain, D. (2012). Multi-agent system modelling for urban systems: The series of SIMPOP models. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 721–738). Dordrecht: Springer.

Rand, W. (2012). Business applications and research questions using spatial agent-based models. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 463–480). Dordrecht: Springer.

Rieser, M., Nagel, K., Beuck, U., Balmer, M., & Rümenapp, J. (2007). Truly agent-oriented coupling of an activity-based demand generation with a multi-agent traffic simulation. *Transportation Research Record, 2021*, 10–17.

Schelling, T.C. (1978). *Micromotives and macrobehavior*. New York: W.W. Norton.

Simoes, J. A. (2012). An agent-based/network approach to spatial epidemics. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 591–610). Dordrecht: Springer.

Smith, D. M. (2012). Simulating spatial health inequalities. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 499–510). Dordrecht: Springer.

Stanilov, K. (2012). Space in agent-based models. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 253–269). Dordrecht: Springer.

Torrens, P. (2012). Urban geosimulation. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 435–450). Dordrecht: Springer.

Wu, B. M. & Birkin, M. H. (2012). Agent-based extensions to a spatial microsimulation model of demographic change. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 347–360). Dordrecht: Springer.

Part II
Principles and Concepts of Agent-Based Modelling

Chapter 6
Agent-Based Models – Because They're Worth It?

David O'Sullivan, James Millington, George Perry, and John Wainwright

Abstract We address the question of when the relative complicatedness of spatial agent-based models (ABMs) compared to alternative modelling approaches can be justified. The spectrum of ABM types from simple, abstract models to complicated models aspiring to realism makes a single answer impossible. Therefore we focus on identifying circumstances where the advantages of ABMs outweigh the additional effort involved. We first recall the reasons for building *any* model: to simplify the phenomena at hand to improve understanding. Thus, the representational detail of ABMs may not always be desirable. We suggest that critical aspects of the phenomena of interest that help us to assess the likely usefulness of ABMs are the nature of the decisions which actors make, and how their decisions relate to the spatio-temporal grain and extent of the system. More specifically, the heterogeneity of the decision-making context of actors, the importance of interaction effects, and the overall size and organization of the system must be considered. We conclude by suggesting that there are good grounds based on our discussion for ABMs to become a widely used approach in understanding many spatial systems.

D. O'Sullivan (✉) • G. Perry
School of Environment, University of Auckland, Auckland, New Zealand
e-mail: d.osullivan@auckland.ac.nz; george.perry@auckland.ac.nz

J. Millington
Department of Geography, King's College London, Strand, London, UK
e-mail: james.millington@kcl.ac.uk

J. Wainwright
Department of Geography, University of Sheffield, Sheffield, UK
e-mail: j.wainwright@sheffield.ac.uk

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_6,

6.1 Introduction

In this chapter we critically examine the usefulness of agent-based models (ABMs) in geography. Such an examination is important because although ABMs offer some advantages when considered purely as faithful representations of their subject matter, agent-based approaches place much greater demands on computational resources, and on the model-builder in their requirements for explicit and well-grounded theories of the drivers of social, economic and cultural activity. Rather than assume that these features ensure that ABMs are self-evidently a good thing – an obviously superior representation in all cases – we take the contrary view, and attempt to identify the circumstances in which the additional effort that taking an agent-based approach requires can be justified. This justification is important as such models are also typically demanding of detailed data both for input parameters and evaluation and so raise other questions about their position within a broader research agenda.

One inspiration for our approach is found in a brief but challenging commentary by Helen Couclelis (2002). Noting that ABMs add to "the well-known problems of modeling a highly complex, dynamic spatial environment" (pp. 4–5), the additional difficulties of "modeling highly complex, dynamic decision-making units interacting with that environment and among themselves in highly complex, dynamic ways". She continues:

> The question is whether the benefits of that approach to spatial modeling exceed the considerable costs of the added dimensions of complexity introduced into the modeling effort. (Couclelis 2002, pp. 4–5)

Couclelis offers her own answer, when she goes on to say: "The answer is far from clear and, in my mind, it is in the negative" (p. 5). However, Couclelis does leave the door open to further discussion. Others such as Gross and Strand (2000) have argued that capturing micro-scale complexity requires models with the complex micro-structures that the agent-based approach incorporates; in short, a complex world requires structurally complex models. These contrasting perspectives make it clear that an open question remains: under what circumstances is the extra effort of these data- and theory-intensive models rewarded, and why? The aim of this chapter is therefore to establish under which circumstances ABMs really are worth it!

6.2 Horses for Courses: Different Agent Models for Different Purposes

There are many possible ways of classifying ABMs (see Crooks and Heppenstall 2012 for a brief overview). In geographical applications, at the most abstract level, an ABM consists of agents interacting with and in an environment. Various typologies can be constructed on the basis of the nature of the agents and of the environmental representation. Couclelis (2002, p. 4) offers one such classification based on whether the agents and the environment are 'designed' or 'analyzed'. This terminology is somewhat confusing (it derives from an engineering perspective), but may be clearer if we replace 'designed' with *theoretically derived* and 'analyzed' with *empirically derived*. Couclelis goes on to consider the purpose of these different possible combinations of agent and environment type.

An alternative approach to classifying ABMs is to consider three broad styles of model (see O'Sullivan 2008). Arguably, the bulk of academically orientated work to date using ABMs continues to be in the realm of simple abstract models where the focus is on exploring the collective implications of individual-level decision making. Schelling's book title *Micromotives and Macrobehaviour* (Schelling 1978) captures the intention of this approach well (and is discussed by Birkin and Wu 2012). The 'Schelling model' of residential segregation is the most familiar example of this style (Schelling 1969), and has spawned a cottage industry of variants and explorations of how various minor changes to the assumptions underlying the model affect the outcomes (see Fossett 2006, for a detailed exploration of some aspects of the model). In the same vein are Epstein and Axtell's (1996) *Sugarscape* models, Axelrod's work on iterated game theoretic models (Axelrod 1997) and many ABMs of economic behaviour (see Tesfatsion and Judd 2006). Examples of this abstract approach in an urban context include Batty's work on how simple movement and resource exploitation actions on heterogeneous landscapes produce characteristic settlement size distributions (Batty 2005, Chap. 8), and a preliminary model of sprawl presented by Brown and Robinson (2006). The abstract approach is also common in other fields such as biology (see, for example, Ehrlich and Levin (2005)). It is this style of work which is largely responsible for excitement in some quarters around the potential of 'complexity science' to answer general questions about the nature of systems in a wide range of specialist fields (e.g. Bar-Yam 1997).

A second type of ABM is more detailed and locates virtual model agents in a representation of the real world setting of interest. Typically, such models operate at a regional or landscape scale, although this is dependent on the issue(s) that a particular model is addressing. A common application for this flavour of ABM is land-use and cover change (LUCC), often in the context of climate-change scenarios. A recent special issue of *Landscape Ecology* (Milne et al. 2009) gives a sense of the diversity of models in this context, and also of the importance of integrating ABMs with those other approaches. Examples of the type we have in mind are the work of Millington et al. (2008) and Matthews (2006). Here, the goal of developing a model is to understand how expected or possible changes in the behaviour of individual entities arising from the changing policy environment affect landscape-level variables that feedback to both agent behaviour and resulting system-level outcomes (such as, for example, climate change). A different context for models of this kind is the attempt to understand how an urban streetscape or a complicated building design affects the behaviour and paths followed by pedestrians interacting in that environment (Haklay et al. 2001; Helbing et al. 2001; Kerridge et al. 2001). The common thread linking these settings is that the interactions among agents may have more or less dramatic effects on the overall outcomes of the model. In both cases, agent actions change the decision-making environment of other agents, albeit at different spatio-temporal scales, and in different ways. In a LUCC model, more or less permanent changes in the environment are made by agent actors, and these collectively affect future decision-making for all agents at the scale of the whole model. In a pedestrian model, the urban or built environment is fixed, and the agents themselves are a salient and rapidly changing feature of the environment, which affects agents, often only locally.

Thirdly, some of the most ambitious models aim at detailed (i.e. "realistic", although see Dietrich et al. 2003, pp. 7–8, for a more extended consideration of realism in models) representations of both the geographical setting and the processes unfolding in that setting. Such models tend to be driven by the concerns of policy- and decision-makers and revolve around urban, economic, and demographic management applications. The most obvious example of this style of model is the TRANSIMS ABM of urban traffic where every individual vehicle in a large urban system is represented second-by-second (Cetin et al. 2002). Closely related to TRANSIMS is EpiSims, which takes the same approach to epidemic processes in detailed representations of social networks (Toroczkai and Guclu 2007). When models become this large, it becomes difficult to get to grips with their overall structure, or even to consider them as truly single models. The 'model' becomes a framework in which subsystem models are integrated. An example of this approach which has evolved over many years is the SIMPOP family of urban growth models (Sanders et al. 1997; Bretagnolle et al. 2009). The modular and extensible structure of such models is an attempt to cope with the difficulties inherent in extending the scope of individual-based models as they grow to encompass large scale continental or global systems, a problem which is also encountered in using and interpreting general circulation models of global climate.

This last category makes it clear that any typology of ABMs is necessarily highly schematic. The three types of ABM we have discussed are more like points along a continuum of increasing size and complexity than discrete categories. The value of developing such a typology at all is to realize that ABMs are built for a wide variety of reasons across a wide range of disciplines. ABMs, like all models, may be used to explore theories and their possible implications, to understand how particular theories may play out in particular contexts, and to assist in risk-assessment, or policy- and decision-making. This complicates answering the question of whether or not ABMs are useful in any particular application, although it suggests that the answer is "it depends!" (on context, on purpose, on application, and so on). Even so, it is possible to be more specific about the situations where agent approaches are likely to justify the additional effort and cost that their development, analysis and use entail.

6.3 Are Modellers Agent-Based Because They Should Be or Because They Can Be?

While there has been a lot of excitement in recent years about the potential of agent-based methods, it is important to remember that none of the cases cited above is one where agent models are the only possible approach. In most cases, ABMs are a relatively late arrival in a field where there is considerable previous experience with styles of model that adopt a more aggregated approach, and these aggregated models continue to be widely used. Thus, for example, land-use transport models, which are calibrated and run based on transport analysis zones, are much more widely deployed by city governments worldwide than ABMs at the individual vehicle level simulating morning and evening rush hours; see Wegener (2004). What, if anything do ABMs add, and by extension, when should we prefer ABMs over more traditional methods?

At times, it appears that the main motivation for adopting an agent-based approach is simply because it can (now) be done. While the tools available for ABMs (Railsback et al. 2006) are not yet as accessible or as well developed as those for more established approaches such as systems dynamics (Muetzelfeldt and Massheder 2003; Deaton and Winebrake 2000; Eberlein and Peterson 1994), ABMs have surprisingly quickly become a viable approach for the spatial model builder. The increasing ease with which ABMs can be developed, coupled with their intuitively satisfying representational approach, in which each software agent represents an 'actor' whether an individual person (or animal or plant) or an institution (often the barely more aggregated household) has led to widespread enthusiasm for the approach. The appeal is undeniable: it appears obvious that individual-level decision making is the fundamental driver of social systems, or more broadly that the individual-level behaviours of plants and animals drive environmental change. Setting to one side the thorny question of whether or not social phenomena are distinctive in kind from the merely aggregate actions of individuals (see O'Sullivan and Haklay 2000), and hence also the question of whether it is the case that social and environmental systems really are driven entirely by individual-level decision-making, if we *can* represent systems at the 'atomic' level on which they operate, then surely we *should*?

In our view this stance ignores the motives for developing models in the first place. Put simply, the need for a model arises when understanding the world itself is too hard! The danger of wholesale adoption of ABMs is that we simply replace one difficult to understand phenomenon – the world itself – with an equally hard to understand model. This is the difficulty that Couclelis identifies in her commentary. A model that advances our understanding is one that represents what are considered in a particular context the key features of a system and thus enables us to improve our understanding of how that system works. Any gain in understanding of the system resulting from the modelling process derives from our ability to analyze the model and experiment with it. If the model is too complicated to analyze, all we have done is to replace one poorly understood object of study with another, which we know to be incomplete! There are good reasons to believe that using disposable 'fast-and-frugal' models will result in more rapid learning than highly detailed ones (Carpenter 2003), and in most, if not all cases, ABMs are not a 'fast-and-frugal' option.

Considering such issues is at the heart of all model building. However, ABMs are one aspect of a recent trend towards more complicated and detailed models. This trend flies in the face of longstanding conventions in modelling and simulation, which hold that simpler, more parsimonious models are preferable to complicated ones, all other things being equal. The search for parsimony in models is often presented as a logical consequence of Ockham's razor (see Perry 2009). That is not a position we wish to defend. First, it is clear that Ockham's admonishment to avoid the 'unnecessary multiplication of entities' was never intended to guide the development of simulation models! Second, there is no *a priori* reason for assuming that the world is a simple place, when it is patently not!

Careless application of the principle of Ockham's razor might lead us to conclude that a less complicated model is more convincing, just because it is less complicated, although this is not a logically defensible point of view. Ockham's razor is an argument about the capacity of different descriptions of reality to explain observed phenomena, not grounds for always preferring simpler explanations to more complicated ones. Even so, there are good pragmatic reasons for preferring parsimonious models. Such models are much easier to learn from than models with many parameters and sub-models. They are easier and more cost-effective to parameterize, and they are also much less vulnerable to the propagation of errors due to the uncertainties in estimating multiple interrelated parameters (again, see Carpenter 2003).

Based on this observation, the important question is to determine what features agents bring to a model *which make a difference that matters*. This concern is similar to the argument made by Andrew Sayer in his consideration of "the difference that space makes" in explaining social systems (Sayer 1985). Although he is discussing the role of space in social theory, Sayer's arguments seem to us to apply with equal force to the evaluation of models. The basis of the argument is the distinction to be made between *necessary* and *contingent* features of a theory. Some aspects of any phenomena we wish to explain are absolutely central – that is, necessary – to the nature of that phenomena, while others are peculiar to occurrences of those phenomena in particular contexts – that is, contingent on those particular occurrences. A less philosophical way to express the same idea is simply to ask, which features of the phenomena we are interested in are essential? Asking this question is really what building a model is all about. Answering this question in the context of ABMs should focus our thinking on the issue of what the agents in a model are, what they do, and following from this, when they are necessary to any representation of the phenomena of interest. In the remainder of this chapter, we sketch out the circumstances in which agents are more likely to be necessary to an adequate model. In our conclusions, we briefly revisit the idea of contingency and its relevance to this issue.

6.4 What Are Agents? And What Do They Do?

These considerations bring us to the basic question of what adopting an agent-based representation in a model achieves in terms of a simulation. There is general agreement (amidst much debate about finer points!) on the basic characteristics of agents in spatial models. More detailed consideration of the meaning of the defining characteristics of such agents can be found in Crooks and Heppenstall (2012). We consider the most fundamental characteristics of agents in spatial models to be goal-direction and autonomy (Jennings et al. 1998). However, more specific definitions of the concept may add any of flexibility, 'intelligence', communication, learning, adaptation or a host of other features to these two. In practice, whatever way we describe their characteristics, agent actions in models revolve around exercising *choice* among available options in order to achieve defined *goals*.

The outcome of an agent making a particular choice is some difference in either the location of the agent (i.e. the agent moves) or in the environment. In the latter case, the agent alters the attributes of its current location in some way. Depending on the model context, this may involve the agent exploiting resources at its current location (and hence depleting the supply of those resources at that location); altering the state of the location (e.g. changing the land use); acquiring the land at its present location; or, perhaps simply updating its current 'map' of the environment. In each case, there may be an accompanying change in the state of the agent itself, such as when resource exploitation increases the agent's wealth or energy resources.

This account of spatial ABMs (and it is important to note that there are many examples in the literature of aspatial ABMs) has several implications:

- Agents may be mobile, but this is not a necessary feature (models of trees in forests are among the most common types of ABM). However, it is important that each agent has a different relationship with the spatial environment, most simply in terms of a location in the environment. If all agents have the same spatial relationship with the environment (if, for example, every agent has an equal capability to alter every location in the model regardless of the agent's specific location or every agent sees and responds to an aggregate 'average' of the environment), then it makes little sense to formulate the model as an agent model;
- Agents may change their spatial relationship with the environment over time, which may be by moving, or it may be by alteration, acquisition or disposal of locations; and
- Agents are able to evaluate spatial configurations. This ability may be as simple as determining that the availability of some resource at the current location is sufficient for some purpose, or is greater than at neighbouring locations. Alternatively, it may involve a complicated evaluation of the spatial distribution of resources (including other agents) with respect to the current location, relative to a number of alternative locations.

This framework for thinking about agents in a spatial ABM may be illuminated by considering some examples (see also Fig. 6.1):

1. *Pedestrian or other mobile agents* in a model of an urban streetscape or complex building. The primary choice made by such agents is to determine, with respect to their intended destinations, which among the possible next locations they should move to. In most models of this kind, the location of other agents is an important element in the choice, but the decision will also be affected by the agents' local physical environments (e.g. building geometries).

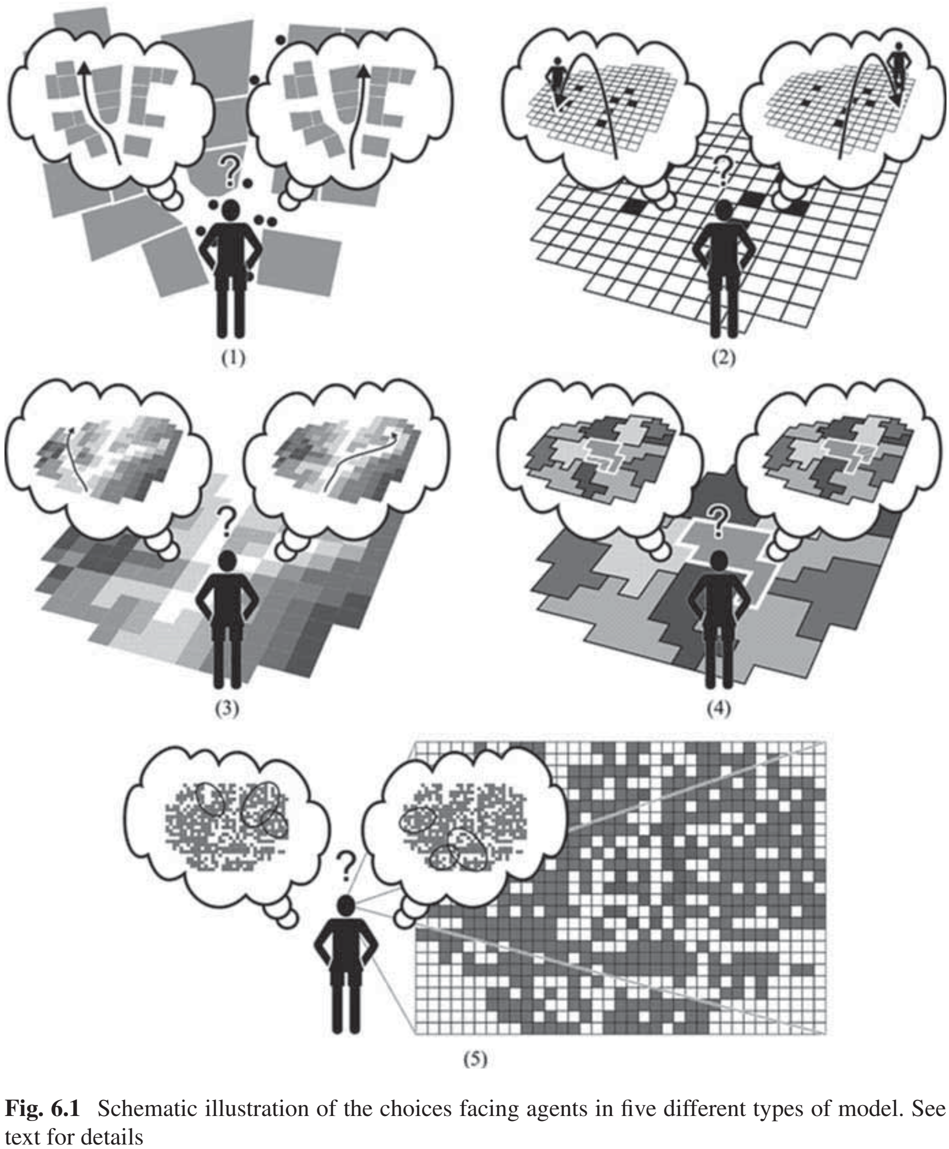


Fig. 6.1 Schematic illustration of the choices facing agents in five different types of model. See text for details

2. *Residential agents* in a 'Schelling-style' model are also primarily concerned with movement, although it is movement of a rather different kind. They evaluate their current and potential new locations, and if one of the new locations considered is preferable in some way to their current location, then they may move there. Again, the locations of other agents in this model influence the choices made by each agent, but the nature of the environment itself does not.
3. *Hunter-gatherer agents* in a model of resource exploitation in which establishment of permanent settlements is an outcome will probably combine aspects of the two previous types of agent, in that they evaluate competing locations, and

will choose to stay or go depending on the resources at those locations. Rather differently to the previous two cases, however, the actions of these agents will alter the environment directly, not just in terms of the location of the agents themselves.

4. *Farmer/land-use change agents*, like the previous type, alter the environment itself, but unlike them are unlikely to move in the process. They may alter their relationship with the spatial environment by acquiring or disposing of land as one aspect of the management of their resources.
5. *Property developer agents* in an urban growth or development model are unlikely to be explicitly spatially located in the way that agents in the previous examples are. Like farmer agents, they will have some attachment to a 'territory', which they are able to grow, change, or reduce by acquisitions, development actions, or sales into a property market. Such agents are likely to have a relatively sophisticated ability to evaluate spatial configurations of currently owned locations relative to the various land uses and land values in the model.

Aspects not explicitly considered in these examples, but highly relevant in practice, are the spatial and temporal *grain* of the model representation, and the relationship between the two. By grain we mean the extent of the smallest units of space and time which are explicitly represented in a model. A fine-grained model might represent second-by-second developments at spatial resolutions of a metre or less; traditionally, such models have been seen as unable to consider extended spatio-temporal domains. A coarse-grained model might operate on large units of space (say several square kilometres) over time periods of a year or more. Grains much coarser than this seem unlikely in practice because ABMs are about the choice-making behaviour of individual living actors. While contemporary societies occasionally aspire to decision-making that takes into account time horizons longer than a year or so (and simulation models are seen as central to this decision-making; see Clark et al. 2001), it is rare for choices to be 'locked in' over much longer time frames than this. Similarly, it is difficult to imagine an ABM model that would be recognizable as such where spatial agents act on 'local' spatial knowledge more wide-ranging than a few kilometres.

Note that we adopt the concept of grain here in preference to spatio-temporal scale because the latter often also implies the overall extent or scope of a model. While the grain of the representation in a model and its overall scope are not independent, it is increasingly common to see unexpected combinations particularly of fine grains with wide extents (for example, Epstein 2009, refers to an epidemic model that explicitly represents the whole population of the Earth as individual agents).

Although as geographers we might wish to grant representation of the spatial aspects priority over temporal aspects, temporal considerations are of at least equal importance, not least because the two are interlinked (both conceptually and computationally). Decisions are usually made by agents over some timeframe of interest, which may in turn imply a relevant spatial grain.

In a pedestrian model this timeframe might be second-by-second, as pedestrians adjust their course to avoid obstacles (including other agents). More generally, mobile agents (whether human or some other animal) will be making decisions at time scales dictated by their mobility on the one hand and their perception of the nature of the spatial distribution of resources on the other. The decision-making timeframe combined with the speed of movement of the agents then effectively dictates a sensible spatial grain for a model of this type – with plausible models capturing spatial detail down to or below the metre range for human agents. In resource exploitation models the timeframe of interest is dictated by context. In a model of hunter-gatherer behaviour with only limited storage of resources, daily or weekly activity patterns and decisions will predominate, and this, combined with rates of movement, will govern how we represent spatial aspects of both the agents *and* their environment. In cases where the mobility of agents is less dominant, as in the farmer or property developer examples above, the linkage between the temporal and spatial grain is less direct, but nevertheless remains important. The key issue in these cases is how rapidly agents change the environment, and how quickly those changes affect the later decisions of other agents. A monthly, seasonal or annual timeframe is likely to be the most appropriate in these cases, since the outcomes of planting or development decisions that take appreciable times to unfold will affect further decision making. In these cases the spatial grain is a product of the amount of change which can be effected by individual agents over the chosen time frames. This in turn will be dependent on organizational features of the agents themselves in particular if they are institutional actors. For example, where property developers are small businesses, we may be interested in development at the level of individual land parcels. Where we are interested in larger corporate actors, the spatial extent of agent actions may be much larger.

In the one highly abstract case we consider above, that of 'Schelling-style' residential relocation models, these considerations are a lot less clear-cut. In such cases, questioning the spatial and temporal grain can contribute to conclusions that may be considered very unflattering to the model under examination; see, for example, Goering (2006). The essentially theoretical, abstract nature of the model comes to the fore and the spatio-temporal grain of the representation is of less relevance than its structure and the overall system tendencies it points to.

6.5 So When Do Agents Make a Difference?

The emphasis we have placed on decision making by agents and the related choice of the spatial and temporal grain in a model helps to address our original question about when it is appropriate to adopt an agent-based representation in a model. If the decisions at the heart of a model are made in local contexts, which depend in turn on the spatio-temporal grain of the model in such a way that every agent decision reduces to the same decision, then an aggregated statistical or mathematical representation may be sufficient. The classic examples from game theory, such as Prisoner's Dilemma and the Tragedy of the Commons fit this template well, and continue to shed light on the overall structure of many social systems and coordination problems.

Where agents' preferences and (spatial) situations differ widely, and where agents' decisions substantially alter the decision-making contexts for other agents, there is likely to be a good case for exploring the usefulness of an agent-based approach. This argument focuses attention on three model features: *heterogeneity* of the decision-making context of agents, the importance of *interaction effects*, and the overall *size* and *organization* of the system.

If agents are the same throughout the system, then, other things being equal, an aggregate approach is likely to capture the same significant features of the system as an agent-based approach. However, it is important to extend our concern with heterogeneity to encompass not just agents but to agents in their (spatial) decision-making contexts. A population of identical agents in diverse contexts can produce somewhat unexpected outcomes as a result of different choices being made in those different contexts, which then alter the options available to all agents at subsequent times. 'Schelling-style' models exemplify this. The opposite case, where every agent makes its choices in the same context but heterogeneity in the agents may produce dramatically different results depending on the degree of heterogeneity, is less familiar. An example is provided by Rand et al. (2002), whose abstract model of urban growth shows that the existence of even small numbers of households with a preference for aesthetic over urban amenity can dramatically accelerate exurban sprawl.

In both of these cases, agent actions result in changes to the decision making context for other agents, an indirect and weak form of agent-to-agent interaction. Some form of agent interaction is necessary at a minimum if an agent-based approach is to be justified. If each agent's decisions make no difference to the subsequent decision-making contexts of other agents, then the generalized pay-off matrices of classical game theory are again likely to provide a sufficient representation of social systems. The stronger any interaction effects are, then the more important it will be to consider agent-based or other disaggregated approaches. In a pedestrian model, interaction is direct. Each pedestrian agent is a significant element in the local environment of many other agents, and decisions made by one agent immediately alter the local decision-making environment of nearby agents. Where the contexts for decision-making are more general, based on aggregate system measures, so that each individual's decisions make only minor differences to the choices of others, then the case for an agent-based approach is less clear.

By the system size, we mean the total number of agents in the system. This aspect relates to the previous point. In large systems, other things being equal, unless interaction effects are strong and direct, it may not be necessary to adopt ABM approaches. In such cases, mean-field approaches provide appropriate representations of system dynamics (Berec 2002). This consideration is closely related to one of the earliest characterizations of the idea of system complexity by Warren Weaver (1948), who distinguishes middle-sized systems of "organized complexity" from small systems of only a few elements on the one hand, and large systems of disorganized complexity explicable in statistical terms (gases are the obvious example) on the other. Systems of organized complexity are those where interaction among elements – more than that, iterative or hierarchical *organization* of the elements – renders statistical explanation inadequate. He wryly notes that the size range of such systems is very broad: "large compared to two, but small compared to the number of atoms in a pinch of salt" (Weaver 1948, p. 539). Taking only system size into account, this aspect may appear redundant in determining the viability of agent-based approaches since all social systems (that we know of!) fall into this broad 'middle' range.

To resolve this issue, we must delve more deeply into the idea of system organization. Where systems are sufficiently 'organized', it may be that intermediate levels of organization are durable enough to form the atomic units on which we should focus in a model, rather than individuals. This fact is already implicit in cases where households rather than individuals are the agents in a model. Similarly, in economic models, firms are often recognized as the appropriate units for representation. In models of large collections of individual actors, perhaps the most important question for the would-be agent-based modeller to ask is not "is an ABM appropriate?" (where the presumption often is that agents should represent individual actors). A more important question may be, "what should the agents in an ABM of this system represent?" If the interactions among individual actors in the real world are substantially channelled via institutions or other social or spatial structures, perhaps it is those institutions or social or spatial structures that should be represented as agents in an ABM rather than the individuals of which they are formed. One way to think about this is to see that in choosing to represent not individual actors as agents but instead some other intermediate level aggregate entity, we are effectively reducing the system size to a point where actions of individual agents make a difference, thus justifying the approach.

All three of our system criteria favouring the adoption of ABM – heterogeneity, interaction, and the combined effects of system size and organization ('middle-numbered-ness') – call for considerable prior knowledge and insight about system characteristics on the part of those developing models. Thus, it would be wrong to draw any universal conclusions from our account to a statement about the usefulness of agent-based approaches in general. Instead, we strongly recommend careful consideration of the system features we have discussed before simply assuming that an agent-based representation is inherently superior. Where consideration of these aspects suggests that an agent-based representation is indeed necessary, then it is worth noting that the resulting model is often one where a full explanation of the model behaviour calls for a historical account of the events in the model. If agents are necessary in the model because they are differentiated from one another, because they interact meaningfully with one another, and because they are able to make a difference to system level outcomes, then in describing and understanding the model, it is likely that Sayer's (1985) contingent effects will be significant. Thus particular agent-agent interactions will matter, and a detailed account of the model 'history' may be necessary for a complete understanding of any particular model run. The difference from the real world target system we seek to understand, is that a model allows repeated runs and enables a probabilistic or general account of the system behaviours and tendencies to be developed.

Our discussion relies on *a priori* understanding or analysis of the system structure, or *post hoc* assessment of whether the resulting model demonstrates the historical-contingent features that would suggest it was the right choice of approach. Neither is a particularly satisfactory or systematic way to decide whether or not to embark on the demanding and potentially costly development of an ABM approach in a particular case. Given the complex nature of the systems and problems involved, it is difficult to see how at least piloting ABM and alternative approaches can be completely avoided (another reason for preferring simple models to complicated ones?), but recent approaches do suggest ways in which the usefulness of ABMs can be assessed, such as pattern-oriented modeling (Grimm et al. 2005; Grimm and Railsback 2012) and the comparison of mean-field and individual-based models (Iwasa 2000).

While we cannot make sweeping general claims from our discussion, it seems clear that human settlement systems are often strong candidates for agent-based representations. This claim is based on the criteria for the usefulness of ABMs that we have identified: heterogeneity, interaction, and system size and organization. Similar arguments can be made about human-environment systems more generally, even in prehistoric settings where the degree of organization of the social systems may be rather more limited. While other approaches remain useful, arguments against building ABMs based on the extra effort involved can be countered because the potential for insight and understanding from building and using such models makes those efforts worth it.

References

Axelrod, R. M. (1997). *The complexity of cooperation: Agent-based models of competition and collaboration*. Princeton: Princeton University Press.

Bar-Yam, Y. (1997). *Dynamics of complex systems*. Reading: Perseus Books.

Batty, M. (2005). *Cities and complexity: Understanding cities with cellular automata, agent-based models and fractals*. Cambridge: MIT Press.

Berec, L. (2002). Techniques of spatially explicit individual-based models: Construction, simulation, and mean-field analysis. *Ecological Modelling, 150*, 55–81.

Birkin, M., & Wu, B. (2012). A review of microsimulation and hybrid agent-based approaches. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 51–68). Dordrecht: Springer.

Bretagnolle, A., Pumain, D., & Vacchiani-Marcuzzo, C. (2009). The organization of urban systems. In D. Lane, D. Pumain, S. E. van der Leeuw, & G. West (Eds.), *Complexity perspectives in innovation and social change* (pp. 197–220). New York: Springer.

Brown, D. G., & Robinson, D. T. (2006). Effects of heterogeneity in residential preferences on an agent-based model of urban sprawl. *Ecology and Society, 11*, 46–65.

Carpenter, S. R. (2003). The need for fast-and-frugal models. In C. D. Canham, J. J. Cole, & W. K. Lauenroth (Eds.), *Models in ecosystem science* (pp. 455–460). Princeton: Princeton University Press.

Cetin, N., Nagel, K., Raney, B., & Voellmy, A. (2002). Large-scale multi-agent transportation simulations. *Computer Physics Communications, 147*, 559–564.

Clark, J. S., Carpenter, S. R., Barber, M., Collins, S., Dobson, A., Foley, J. A., et al. (2001). Ecological forecasts: An emerging imperative. *Science, 293*, 657–660.

Couclelis, H. (2002). Why I no longer work with agents: A challenge for ABMs of human-environment interactions. In D. C. Parker, T. Berger, & S. M. Manson (Eds.), *Agent-based* *models of land-use and land-cover change: Report and review of an international workshop,* Irvine, Oct 4–7, 2001 (pp. 3–5). Louvain-la-neuve, Belgium: LUCC International Project Office.

Crooks, A. T. & Heppenstall, A. J. (2012). Introduction to agent-based modelling. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 85–105). Dordrecht: Springer.

Deaton, M. L., & Winebrake, J. J. (2000). *Dynamic modeling of environmental systems*. New York: Springer.

Dietrich, W. E., Belugi, D. G., Sklar, L. S., Stock, J. D., Heimsath, A. M., & Roering, J. J. (2003). Geomorphic transport laws for predicting landscape form and dynamics. In P. R. Wilcock & R. M. Iverson (Eds.), *Prediction in geomorphology* (Geophysical monograph, Vol. 135, pp. 1–30). Washington DC: American Geophysical Union.

Eberlein, R. L., & Peterson, D. W. (1994). Understanding models with VenSim. In J. D. W. Morecroft & J. D. Sterman (Eds.), *Modeling for learning organizations* (pp. 339–348). New York, NY: Productivity Press.

Ehrlich, P. R., & Levin, S. A. (2005). The evolution of norms. *PLoS Biology, 3*, e194. doi:10.1371/journal.pbio.0030194.

Epstein, J. M. (2009). Modelling to contain pandemics. *Nature, 460*, 687.

Epstein, J. M., & Axtell, R. (1996). *Growing artificial societies: social science from the bottom up*. Washington, DC: The Brookings Institute.

Fossett, M. A. (2006). Ethnic preferences, social distance dynamics, and residential segregation: Theoretical explorations using simulation analysis. *Journal of Mathematical Sociology, 30*, 185–274.

Goering, J. (2006). Shelling [sic] redux: How sociology fails to make progress in building and empirically testing complex causal models regarding race and residence. *Journal of Mathematical Sociology, 30*, 299–317.

Grimm, V. & Railsback, S. F. (2012). Designing, formulating and communicating agent-based models. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 361–377). Dordrecht: Springer.

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., et al. (2005). Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science, 310*, 987–991.

Gross, D., & Strand, R. (2000). Can agent-based models assist decisions on large-scale practical problems? A philosophical analysis. *Complexity, 5*, 26–33.

Haklay, M., Schelhorn, T., O'Sullivan, D., & Thurstain-Goodwin, M. (2001). "So go down town": Simulating pedestrian movement in town centres. *Environment and Planning B: Planning & Design, 28*, 343–359.

Helbing, D., Molnár, P., Farkas, I. J., & Bolay, K. (2001). Self-organizing pedestrian movement. *Environment and Planning B: Planning and Design, 28*, 361–383.

Iwasa, Y. (2000). Lattice models and pair approximation in ecology. In U. Dieckmann, R. Law, & J. A. J. Metz (Eds.), *The geometry of ecological interactions: Simplifying spatial complexity* (pp. 227–252). Cambridge: Cambridge University Press.

Jennings, N. R., Sycara, K., & Wooldridge, M. (1998). A roadmap of agent research and development. *Autonomous Agents and Multi-Agent Systems, 1*, 7–38.

Kerridge, J., Hine, J., & Wigan, M. (2001). Agent-based modelling of pedestrian movements: The questions that need to be asked and answered. *Environment and Planning B: Planning and Design, 28*, 327–341.

Matthews, R. (2006). The People and Landscape Model (PALM): Towards full integration of human decision-making and biophysical simulation models. *Ecological Modelling, 194*, 329–343.

Millington, J., Romero-Calcerrada, R., Wainwright, J., Perry, G. W. L. (2008). An agent-based model of Mediterranean agricultural land-use/cover change for examining wildfire risk. *Journal of Artificial Societies and Social Simulation*, *11*(4), 4. Available at: http://jasss.soc.surrey.ac.uk/11/4/4.html

Milne, E., Aspinall, R. J., & Veldkamp, T. A. (2009). Integrated modeling of natural and social systems in land change science. *Landscape Ecology, 24*, 1145–1147.

Muetzelfeldt, R. I., & Massheder, J. (2003). The Simile visual modelling environment. *European Journal of Agronomy, 18*, 345–358.

O'Sullivan, D. (2008). Geographical information science: Agent-based models. *Progress in Human Geography, 32*, 541–550.

O'Sullivan, D., & Haklay, M. (2000). Agent-based models and individualism: Is the world agent-based? *Environment and Planning A, 32*, 1409–1425.

Perry, G. L. W. (2009). Modelling and simulation. In N. Castree, D. Demeritt, D. Liverman, & B. Rhoads (Eds.), *A companion to environmental geography* (pp. 336–357). Hoboken: Wiley.

Railsback, S. F., Lytinen, S. L., & Jackson, S. K. (2006). Agent-based simulation platforms: Review and development recommendations. *Simulation, 82*, 609–623.

Rand, W., Zellner, M., Page, S. E., Riolo, R., Brown, D. G., & Fernandez, L. E. (2002). The complex interaction of agents and environments: An example in urban sprawl. In C. Macal & D. Sallach, (Eds.), *Proceedings of Agent 2002: Social Agents: Ecology, Exchange, and Evolution* (pp. 149–161). Chicago: University of Chicago and Argonne National Lab. Available at http://www.dis.anl.gov/pubs/2002%20Proceedings.pdf

Sanders, L., Pumain, D., Mathian, H., Guerin-Pace, F., & Bura, S. (1997). SIMPOP: A multi-agents system for the study of urbanism. *Environment and Planning B: Planning and Design, 24*, 287–305.

Sayer, A. (1985). The difference that space makes. In D. Gregory & J. Urry (Eds.), *Social relations and spatial structures* (pp. 49–66). London: Macmillan.

Schelling, T. C. (1969). Models of segregation. *American Economic Association Papers and Proceedings, 59*, 488–493.

Schelling, T. C. (1978). *Micromotives and macrobehavior*. New York: Norton.

Tesfatsion, L., & Judd, K. L. (2006). *Handbook of computational economics, volume II: Agent-based computational economics*. Amsterdam: Elsevier/North-Holland.

Toroczkai, Z., & Guclu, H. (2007). Proximity networks and epidemics. *Physica A: Statistical Mechanics and Its Applications, 378*, 68–75.

Weaver, W. (1948). Science and complexity. *American Scientist, 36*, 536–544.

Wegener, M. (2004). Overview of land-use transport models. In D. A. Hensher & K. Button (Eds.), *Transport geography and spatial system* (pp. 127–146). Kidlington: Pergamon/Elsevier Science.

Chapter 7
Agent-Based Modeling and Complexity

Steven M. Manson, Shipeng Sun, and Dudley Bonsal

Abstract Complexity theory provides a common language and rubric for applying agent-based processes to a range of complex systems. Agent-based modeling in turn advances complexity science by actuating many complex system characteristics, such as self-organization, nonlinearity, sensitivity, and resilience. There are many points of contact between complexity and agent-based modeling, and we examine several of particular importance: the range of complexity approaches; tensions between theoretical and empirical research; calibration, verification, and validation; scale; equilibrium and change; and decision making. These issues, together and separately, comprise some of the key issues found at the interface of complexity research and agent-based modeling.

7.1 Introduction

Complexity theory and the accompanying trappings of complex systems provide the theoretical basis for agent-based models (ABMs). While modelers are usually interested in addressing specific theoretical questions and working in particular substantive areas, they almost invariably draw on complexity concepts when using an agent-based approach. The relationship between ABM and complexity is mutually beneficial. While complexity has much to offer ABM in terms of underlying concepts, modeling advances complexity by making real many of the often fuzzy concepts on which complexity science relies. Advances in ABM are allowing modelers to move beyond studying complex systems in just metaphorical or rhetorical terms by giving them the tools to represent complex phenomena. Many disciplines are using ABM to enhance understanding of the interplay of complexity concepts, ranging from policy fields

S.M. Manson (✉) • S. Sun • D. Bonsal
Department of Geography, University of Minnesota, Minneapolis MN, USA
e-mail: manson@umn.edu; sunx0170@umn.edu; bons0015@umn.edu

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_7,

(Carrillo-Hermosilla 2006; Gatti et al. 2005; McKelvey 1999) to the natural sciences (Brose et al. 2004; Phillips 2006; Rind 1999) through to the social sciences (Batten 2000; Manson and O'Sullivan 2006; Sampson et al. 2002) and into the humanities and arts (Nowotny 2005; Portugali 2006).

When the theoretical questions regarding complexity are combined with the broadly applicable research allowed using ABM, a number of issues stand out, including:

- Reconciling a range of complexity approaches
- Navigating the tension between theoretical and empirical research
- Implementing calibration, verification, and validation of models
- Dealing with scale
- Balancing the corollaries of equilibrium and change
- Representing features of decision making.

These issues, together and separately, comprise some of the key points of contact and contention among the various components of complexity research and ABM. Ongoing examination of these issues is spurring further ABM research that illuminates phenomena studied in the physical environment, social systems, and their combination via human-environment research.

7.2 Complexity Approaches

Complexity theory is less a singularly defined, discrete conceptual entity than an interdisciplinary focus for which individual fields and researchers use a common set of queries, concepts, and approaches. Given this lack of a single, identifiable core, the terms 'complexity theory' and 'complexity sciences' can therefore fittingly refer to an array of research methods. In order to provide an organizational schema to this diverse field, we identify three streams of complexity research: algorithmic complexity, deterministic complexity, and aggregate complexity (cf. Byrne 1998; Cilliers 1998; Lissack 2001; Manson 2001). At its simplest, algorithmic complexity conceives of any system in terms of the computational or heuristic processes needed to replicate system behavior. Deterministic complexity envisions a system through the lens of nonlinear dynamics and chaos theory, in order to try to capture system dynamics via a small set of mathematical equations. Aggregate complexity focuses on how complex systems arise from interactions among individual entities. It is this final kind of complexity, aggregate complexity, that most ABM researchers tend to invoke when modeling, although algorithmic and deterministic complexity make their own contributions to ABM.

Complexity in any of its above-mentioned forms typically applies to a system, a set of entities connected to each other and the external environment in a way that gives it an overall identity and behavior. An ABM in its most basic form represents a system of such discrete entities. Systems can be of almost any scale, from atoms bound together in a molecule to households in an economy to planets in the solar system. The key to modeling any of these systems, and therefore the key to complexity research and ABM, is the capture of core characteristics among system entities and, critically, their interrelationships. An ecosystem, for example, is self-contained in terms of much of its structure and function but also has many connections to the larger climatic, geophysical, and biotic environment. The model must also have system boundaries that set it apart from its larger context. An urban area, for example, can be defined in a number of ways, but most models focus on elements of the built environment such as buildings and populations (e.g., workers, homeowners) that have relationships via migration, capital flows, and environmental relationships with the larger world.

Algorithmic complexity focuses on representing systems in computational and mathematical terms. The component fields of computational complexity theory and information theory examine the difficulties of computing the solution to problems and representing a system or reproducing its behavior (Chaitin 1974; Gell-Mann 1994). At its most useful, algorithmic complexity provides a number of different measures of how a system is composed and represented. One helpful side effect is that some measures will identify problems that cannot be solved mathematically or computationally with our current state of knowledge, but that may yield to simulation or heuristic approximations. Beyond these instances, the use of algorithmic complexity in complexity research and ABM has been limited given the greater interest in deterministic and aggregate complexity (O'Sullivan 2004).

Deterministic complexity is defined by approaches that use sets of mathematical equations to describe the state and trajectory of system dynamics. Deterministic complexity is so called because it finds for complex systems a few key variables and equations to describe system state and evolution; in this sense, system behavior is 'determined' by these equations and variables. Positive and negative feedback are important components of deterministic complexity, spurring changes that self-reinforce or diminish over time, respectively. Given the potential for such feedback, deterministically complex systems exhibit both sensitivity and nonlinearity. The former refers to how systemic changes can result from small perturbations while the latter refers to how these small changes can give rise to disproportionately large changes in system structure or behavior (Phillips 2003). The combination of sensitivity and nonlinearity is exemplified by the 'butterfly effect,' where slight variations in initial model parameters, due to the displacement of air by butterfly wings, can lead to large meteorological changes in a modeled weather system (Lorenz 1973). The elements of sensitivity and nonlinearity are further adopted and extended by aggregate complexity for the modeling of agent-based systems.

Aggregate complexity focuses on how complex systems arise from the local interactions of system entities. With this perspective, the structure and dynamics of a system such as a city must be understood as driven by individual components and their relationships. In a city, these entities are people, households, firms, and organizations whose relationships are defined by exchanges of matter, energy, and information. These entities have relationships with other entities and play multiple roles within the city. Some of the stronger relationships give rise to larger aggregations (e.g., families, neighborhoods) that may act as entities in and of themselves. This potential for larger entities and behaviors to arise out of local interactions is seen as a form of self-organization, whereby entities and their relationships are sufficiently strong yet flexible enough to allow the overall system to adapt to a changing environment (Easterling and Kok 2002). In some settings, self-organization leads to self-organized criticality, where the system rapidly reconfigures entities and internal relationships, in response to internal perturbation or external shocks (Bak 1996). Self-organization is related to the concept of emergence, whereby system characteristics or behavior result less from additive effects of system entities and their behavior and more from synergistic relationships among entities (Funtowitcz and Ravetz 1994; Holland 1998). One important kind of emergence is supervenience, where changes in system structure or behavior at one level of aggregation are driven by changes at a lower one (Sawyer 2002). In sum, aggregate complexity demonstrates how system entities and their relationships define the behavior of sub-systems and the system as a whole through self-organization and its offshoots, self-organized criticality, emergence, and supervenience.

While it is useful to denote various types of complexity – algorithmic, deterministic, aggregate – it is also important to note that complexity draws on many conceptual antecedents. Since much of current complexity research, particularly aggregate complexity, relies on notions of synergy and holism, it reflects philosophies tracing back to Aristotle's definitions of unity being more than the sum of parts and Whitehead's philosophy of organism, which contends that understanding nature requires more than recourse to fixed laws, and instead identifies it as a system that is continually evolving (Whitehead 1925). More recent antecedents include cybernetics and feedback (Wiener 1961), neural networks and other biological analogs (McCulloch and Pitts 1943), work in computing including cellular automata (von Neumann 1966), and importantly, general systems theory, which holds that many systems have underlying similarities (von Bertalanffy 1968). Complexity departs from earlier related work by focusing on how systems emerge from the simple and local interactions among system entities. While complexity shares with much previous work the assumption that systems can exist in equilibrium, it also actively explores the possibility of perpetual or repeated disequilibrium or near-chaotic behavior. In many respects, then, complexity draws on key features of holism and synergy while also focusing on evolution and the balance between equilibrium and disequilibrium.

7.3 Issues of Complexity and ABM

7.3.1 Tensions Between Theoretical and Empirical Modeling

ABMs are valuable for both theoretical exploration and empirical investigation of complex systems. For theoretical inquiry, modeling serves as a means to better understand how elements of interest and the relationships among them contribute to overall system behavior over time. For empirical investigation, modeling is a vehicle for presenting all known and necessary initial conditions – defined in large part by system entities and their relationships – in order to determine how they have brought about an observed state and how they could bear on future possible states. ABMs also offer many opportunities to combine theoretical and empirical approaches, although not without raising issues regarding the model's simplicity and complexity.

Theoretical inquiry with ABMs usually entails running "controlled experiments" that may spur the discovery of laws about complex processes (O'Sullivan 2004: 288). Purely theoretical ABMs are based on hypotheses that specify certain rules for the behavior of actor agents and their interaction with the environment. When using ABMs to model urban transportation, for example, actor behavior may be defined by utility maximization, as measured by housing quality or work proximity, and transportation cost minimization determined by distance to housing and work and modal choice. While many theoretical models are built for illustrative purposes, such as confirming what their underlying theories predict, some models generate convincing, and sometimes surprising, theoretical implications. Work on racial segregation simulation based on Schelling models, for example, continues to spur debate (Fossett 2006). ABMs contribute to the longstanding use of computer simulation to allow examination of many possible futures or pasts for a given system (Manson and O'Sullivan 2006).

Empirical models focus more than theoretical ones on using actual data to simulate real-world phenomena, although the two foci can be complementary. The increasing number of theoretical models, the growing volume of empirical data, and the use of lab experiments to create rules of agent behavior have all contributed to recent expansion in the development of empirical models (Janssen and Ostrom 2006). These models usually extend aspects of theoretical models using empirical data and have the ability to make predictions and prescriptions under different demographic, economic, and policy scenarios. Since one of the aims of creating empirical ABM is to accurately describe real-world processes, a tension exists between the descriptive power granted by specificity and the desire to generalize to other settings. A model must therefore maintain a balance between fitting the empirical data and highlighting the processes of interest (Manson 2007).

The relationship between theoretical and empirical foci in ABMs highlights how the modeling of empirically complex phenomena with relatively simple or foundational rules is a difficult task. For example, because it is impossible to completely simulate all aspects of natural or human organization without reduction and simplification, all urban complexity models will have a theoretical component (Irwin et al. 2009). Similarly, although complexity theory seeks to capture underlying dynamics, we still face a world where it is difficult to divine many characteristics of the economic state of a city beyond a few years. Any model that attempts to capture the necessary specificity of the myriad system entities may be regarded less for its complexity than for its complication (Torrens and O'Sullivan 2001). When adding a large number of features to a model, the modeler strays from the notion that a small number of rules describing the behavior of agents will lead to complex systems. This challenge arises when modeling urbanization and land change, for example, as

ABMs become more common and sophisticated. Modelers can represent many entities and relationships at the risk of moving away from the ethos of generating complex outcomes based on simple conditions and rules (Parker et al. 2003).

7.3.2 Calibration, Verification, and Validation

Agent-based complexity models require careful and thorough evaluation, which is comprised of calibration, verification, and validation (Manson 2003). Calibration is the adjustment of model parameters and specifications to fit certain theories or actual data. Verification determines whether the model runs in accordance with design and intention, as ABMs rely on computer code susceptible to programming errors. Model verification is usually carried out by running the model with simulated data and with sensitivity testing to determine if output data are in line with expectations. Validation involves comparing model outputs with real-world situations or the results of other models, often via statistical and geovisualization analysis. Model evaluation has more recently included the challenge of handling enormous data sets, both for the incorporation of empirical data and the production of simulation data. Modelers must also deal with questions concerning the relationship between pattern and process at all stages of calibration, verification, and validation. Ngo and See (2012) discuss these stages in ABM development in more detail.

Empirical ABM modelers struggle to obtain the data necessary for proper calibration. From a practical standpoint, simulating a complex system such as an urban housing market requires initializing a range of key components including agents, organizations, and the environment. Modelers rarely have the necessary individual-level data, however, to populate agents such as households, and may similarly be missing information on organizational dynamics or features of the environment. They typically have either a limited set of random samples (e.g., household surveys, phone interviews) or more often, spatially aggregated data at various scales that are collected for other purposes by different government agencies (e.g., census data, regional economic information). Exogenous parameters (e.g. for urbanization, drivers such as population growth rates, interest rates, and federal taxes) can often be derived from actual data, but sometimes are the results of educated guesses, simple linear interpolation, or extrapolation (Brown and Robinson 2006; Torrens 2007).

Proper calibration and validation also entails the integration and reconciliation of data across multiple scales and formats. In ABMs involving both human and environmental elements, for instance, integrating vector and raster data that describe human and natural phenomena respectively at different scales can create problems like ecological fallacies (drawing incorrect inference on individuals from aggregated data) or inappropriate classification when assigning attributes and aggregating features. There are also broader conceptual issues that arise when reconciling data from different scales (e.g., household data vs. census information vs. regional socioeconomic statistics) and linking these observed data to the agents of interest (e.g., households versus parcels versus neighbourhoods). In response to these issues, modelers may need to generate individual data from random samples or from aggregate data, such as census data. Promising approaches include iterative proportional fitting procedures, where tabular data are modified to new levels of aggregation, or Monte Carlo simulation, where multiple probabilistic draws are taken on a sample data set (Wheaton et al. 2009).

While complexity modelers often lack sufficient data for ABM calibration and validation, they also face challenges when generating simulation data. Understanding the dynamics of the attributes of different kinds of agents of even a moderately sophisticated simulation demands great effort to visualize, analyze, and replicate the modeled phenomenon or process (Janssen 2009). The nature of intermediate attribute and behavioral data of actor agents, for example, is rarely discussed in the literature, although such data are potentially useful for the validation of agent behavior and the social processes that produce such behavior. Here, complexity theory can allow the researcher to triangulate among different approaches and viewpoints, because it focuses on identifying generic features of complex systems without getting the inquiry mired in a need to address ontological or epistemological questions (O'Sullivan 2004).

In terms of broader validation challenges, distinctions between theoretical and empirical approaches lead to questions concerning pattern and process. Patterns that are often generated in complexity models, including fractals and information-theory measures, may not reveal much about the processes that generate them, much less whether the processes are complex in the sense meant by deterministic or aggregate complexity. The potential disconnect between pattern and process may influence how the modeler chooses between empirically-driven explanation and description (which usually tilts toward pattern) versus theoretically driven discovery and hypothesis generation (which is often biased towards process). A number of authors, for example, incorporate variables and rules into a model that bring about a community pattern for the Anasazi civilization in the southwest United States previously determined by archeologists and historians (Axtell et al. 2002; Dean et al. 2000). The ABM identifies how discrete entities and their relationships give rise to higher-level systemic processes, but this focus on scale raises the specter of equifinality, where different variables and processes may lead to the same outcome, or similarly, where only a few key variables determine model outcomes (Janssen 2009). For theoretical models, the modeler has more leeway to set initial conditions and formulate iterative rules that can illuminate a theoretical question, although validation becomes difficult in the absence of empirical data. Axelrod's (1997) culture dissemination model, for example, demonstrates how regions adopt or reject the cultural practices of neighboring regions. The model results, while not reflecting the real world in detail, elicit interesting questions about interactions between actors across space and over time.

7.3.3 Scale

ABM researchers pay close attention to the spatial, temporal, and organizational scale of the simulation process. As noted above, one of the hallmarks of scalar properties in ABMs is emergence, the phenomenon of processes occurring at one level that are not evident based on a summing up of lower-level processes.

Emergent properties are implicitly scalar, as seen in how humans function based on the workings and interactions of the component organs, the flight patterns of a flock of birds arising from the actions of individual birds, and the traffic gridlock that occurs based on the decisions of individual drivers (Mainzer 1996). Importantly, emergence is often unintentional. Drivers and their vehicles do not generally seek to create gridlock, for example, but their actions and subsequent interactions readily create traffic jams. ABM modelers can draw on several bodies of work to help define and understand scale and emergence as well as adding context to notions of non-linearity and sensitivity. In addition, scale offers an entry point to the modeling of networks using ABM.

One approach to defining scale levels and emergence is provided by hierarchy theory, wherein actors and systems, through their functions and interactions, form larger systems. A regional housing market, for example, can have several subregions as housing market areas; each housing market area also has housing submarkets; a housing sub-market might then include several cities or several school districts with similar socioeconomic characteristics; within such a sub-market exist smaller neighborhoods defined by residents' activity and interaction patterns. Under this formulation, scale levels should be considered as defined by interactions and relationships among entities, but importantly, it is up to the analyst to define these levels instead of taking them as pre-defined (O'Neill 1988). Similar frameworks exist for the emergence of scale from interactions among entities, such as when institutions arise from the interrelationships of individuals (Ostrom 2005) or, more broadly, when human-environment systems such as agriculture or forestry exist at multiple scales of analysis (Easterling and Polsky 2004).

When drawing upon hierarchy theory, the modeler can identify the system's constitutive hierarchies, wherein the components of a subsystem have emergent properties only when they are brought together to form a higher-level system (Gibson et al. 2000). When considering emergent properties in collective behavior, an implicit assumption is made by the modeler that the lower-level processes are individually not as complex as the collective outcome, yet simultaneously each individual entity may be constitutive of emergent properties based on processes one level further down. The modeler can therefore create a series of models that nest these processes within one another, thereby modeling a hierarchically ordered system.

Notions of scale levels defined by constitutive hierarchies provide a useful counterweight to non-linearity and sensitivity as conceived by deterministic complexity and aggregate complexity. When determining both the spatial and temporal scales of inquiry, one may discern linear associations or limits that coincide with scale levels. Identifying the flapping of a butterfly's wings as a cause of super-regional weather phenomena like hurricanes is a powerful idea, but may not account for a large set of temporal conditions that, in concert with the wings, led to the hurricane. Hence, a claim that the butterfly was necessary does not mean that it was sufficient. In regard to social processes that may seem non-linear, such as the ways that a massively distributed photograph or website video of an individual event may influence national or international policy, one must still consider the communication infrastructure and the social networks that represent a series of steps from one hierarchical scale level to the next, with each step imposing filters and meaning. In contrast to the conception of emergence being merely a bottom-up process, co-evolutionary processes play out when entities understand they are constitutive of the system and can modify it. In addition to understanding how social norms emerge from personal interrelationships, for example, it is necessary to determine how emergent norms feed back onto individuals (Ostrom 2005).

One rapidly emerging form of scale in ABM research, mirroring trends in scale and complexity research more broadly, is the notion of networks defining scales (Manson 2008). Networks have interesting scalar properties that are increasingly important to ABM as researchers combine modeling with the rubrics of graphing and topology. The study of small world networks reveals that when just a few additional links between distant nodes are added to a network where most links are otherwise based on proximity, the connectedness of the entire network greatly increases (Watts 2003). Barabasi and Albert (1999) find that many real networks are self-organizing and scale-free, as they follow a power-law distribution due to inherent processes of growth and preferential attraction of new nodes to well-connected ones. They cite examples of scale-free networks that include the World Wide Web, the electrical power grid of the western United States, and citations that link scientific journal papers. Advances in our understanding of networks arise from a confluence of pertinent data and ABMs, as seen with the joining of a variety of social science databases and decision-making agents in the context of economics and politics (Skvoretz 2002).

7.3.4 Equilibrium and Change

Researchers of all stripes have long modeled many systems under the assumption of equilibrium. Agent-based modeling, by focusing on complex dynamics, provides an opportunity to understand the degree of explanatory power that the assumption of equilibrium has for a given system. Deterministic complexity often does not regard equilibrium as a necessary feature, even if a model of system dynamics can capture whether equilibrium is attainable given the initial conditions and process interactions. For example, ABMs are increasingly used to investigate processes such as the spread of smallpox or cultural memes, where the spatiotemporal dynamics rather than system equilibrium are the phenomenon of interest (Epstein 2006). Issues of equilibrium and change lend further context to concepts of sensitivity and nonlinearity in complex systems by offering commentary on system resilience and the potential for dynamic movement among basins of attraction.

Dramatic changes wrought in a system because of its inherent sensitivity and nonlinearity of interactions are countered by the system's resilience, the ability to adjust to disturbance and reorganize without significantly changing its functions or structure, and its transformability, the ability to create a new system configuration when adjustment is not possible (Walker et al. 2004). A system can be highly resilient despite a high degree of instability when it is self-organizing (Holling 1996).

Also, resilience is a scale-dependent characteristic, both temporally and spatially. A system resilient in one span of time may be compromised in a longer span, while a resilient community may endure at great cost to its larger, encompassing region (Levin and Lubchenco 2008).

Deterministic and aggregate complexity research addresses the dynamics of non-equilibrium states found in complex systems. The lower-level, bottom-up forces create processes that are constantly adapting to environmental changes and undergoing organizational transformation. The interactions that give rise to these changes are non-linear and subject to novelty (Holland 1995), resulting in a system sensitive to the introduction of new components and fluctuations of component states. Despite the ever-changing nature of system behavior and structure, it may gravitate toward one of multiple basins of attraction (Holling 1973). Coupled human-environment systems have multiple attractors, as seen when a coupled population-phosphorus system has one attractor situated at a high population state with a balance of economic and ecological drivers, and a low population state representing a restored ecological system (Chen et al. 2009).

Just as scale levels can attenuate non-linearity and sensitivity, complex systems embody a tension between sensitivity to initial conditions and a dynamic movement between basins of attraction. Certain states may experience positive feedback, gravitating to an attraction basin that will not accommodate robust sub-systems and diverse inputs. Decreases in biological diversity and threats to the viability of ecosystem services, for example, represent a state where resilience is low and more vulnerable to disturbance (Folke 2006). Human institution research recognizes the sensitivity of changes to rules in organizational structure, wherein small changes via policy can bring about "a nontrivial probability of error" (Ostrom 2005: 243). Complex systems are susceptible to 'imaginable surprise' where seemingly unexpected system configurations are in fact understandable when we allow for complex features such as nonlinearity and sensitivity (Schneider et al. 1998). Sensitivity, as with resilience, is either scale-dependent, such that the system may be regarded as sensitive as it moves from one attraction basin to another, or independent as these attractors, over longer time periods, characterize the typical system states regardless of initial conditions. The ability of ABM to represent these complex systems offers great potential for exploring emergence and surprise in human systems, such as the recent financial crisis in the global economy (Farmer and Foley 2009).

7.3.5 Decision Making

Decision making is the engine of many ABMs, particularly those involving human actors, and in turn it has many ties to complexity. It has long been a core concern of many fields, including geography, economics, management, and psychology. ABMs have helped draw out the similarities and differences among different decision-making theories by emphasizing the importance of developing basic rules for agents to follow, leading to research focused on how such rules embody their decision-making strategies. Agents in an ABM usually pursue certain goals set by the modeler with given resources and constraints. Commuters want to minimize their commuting time, for example, while homebuyers want to purchase the best house within their budget, and parents want to move into neighborhoods with quality public schools. Standard decision-making theory is a logical starting point for modeling these decisions, as seen in the wide use of multi-criteria evaluation and Cobb-Douglas utility functions to enable simulated agents to make decisions regarding parcel development and household migration (Brown et al. 2005; Parker and Filatova 2008). While recognizing the value of assumptions such as utility maximization in classical economics, ABMs have also opened the door to other forms of decision-making theory. Behavioral economics, for example, emphasizes the importance of concepts like incomplete information, bounded rationality, reinforcement over time, expected utility, and market anomalies (Arthur 1991; Simon 1997b).

ABMs illustrate how actor agents make decisions to achieve predefined goals in an environment shaped by all agents, and more importantly, how these individual decisions lead to macro patterns that are not predicted by perfect rationality. The concept of bounded rationality, introduced by Herbert Simon, depicts the actor whose decision making is bounded three ways (1997b). The first represents the "skill, habits, and reflexes" (Simon 1997a: 46) that exist beyond our conscious grasp, and presumably, beyond rational decision making. The second is the actor's set of purposes and values, which may differ from those of someone else in an otherwise similar decision making scenario. The third bound is limited information, wherein the actor lacks certain facts or skills that would contribute to a fully informed decision. Representing these three bounds is nascent in ABM but arguably it is this form of modeling that is well suited to advance our understanding of bounded rationality because agents can represent various features of boundedness such as limited computational capacity or rules of thumb (Chen 2005; Dawid 1999; Edmonds and Moss 1997; Manson 2006). In particular, ABMs allow various decision-making strategies, including from rules-of-thumb or heuristics for adapting to a changing environment (Gigerenzer and Selton 2001). Axelrod (1997), for example, sees actors in his cultural dissemination model as not making rational decisions as such, but simply adapting to their environment. More broadly, decision-makers use heuristics to make 'non-rational' decisions, based on the manner in which possible choices are framed (Tversky and Kahneman 1974).

The distinction between an individual decision and a collective one allows for a more sophisticated mechanism to model the choices of actors. Simon notes that decisions "are not made by 'organizations' but by human beings behaving as members of organizations" (Simon 1997a: 281). Social network conceptions of social contagion, for example, address the process of collective decision-making wherein actors receive ideas the way that they may be exposed to the carrier of a disease. Thresholds may be established in which the actor accepts the idea after being exposed to it a given number of times (Granovetter 1978; Watts 2003). When people make migration decisions, for example, they not only want to physically move closer to the friends and relatives in their network, but their criteria for quality housing, their perception of specific neighborhoods, and their knowledge of vacancies are all influenced by the available information in their network (Clark 2008). Social influences on decision making are also evident in the role of reflexivity, such that the past and future are incorporated into present thinking. ABMs address a core question: how does one account for actors that are aware of how their actions may feed into collective outcomes? A person may stay away from social events that are expected to be too crowded or too sparsely attended, for example, as a function of past experience (Arthur 1991). In short, actors often play an expectations game when they act in order to avoid being part of an undesired collective outcome or in order to prevent that outcome from happening (Gilbert 1995).

7.4 Conclusion: Complex Agents, Complex World

Complexity and ABMs offer much to each other. ABM research draws on a range of concepts and approaches from algorithmic, deterministic, and aggregate complexity. In turn, modeling brings to complexity a large number of actual complex systems and attendant theories to advance complexity science. ABMs offer a virtual laboratory that helps researchers navigate between theoretical and empirical research. And while ABM faces many challenges in calibration, verification, and validation, it offers new ways to think about relationships between data and theory, pattern and process. Complexity and ABMs, separately and jointly, are also advancing our conceptualization of scale in a range of complex systems, alongside issues of sensitivity, nonlinearity, resilience, equilibrium, and change. Finally, ABMs are a very promising technique, alongside other approaches, for modeling and understanding decision making. In sum, one may take heart from the many challenges facing researchers working at the intersection of agent-based modeling and complexity science because they arise from the vast potential and promise of these two worlds meeting.

Acknowledgments This work is supported by the National Aeronautics and Space Administration (NASA) New Investigator Program in Earth-Sun System Science (NNX06AE85G) and the National Science Foundation (0709613). Responsibility for the opinions expressed herein is solely that of the authors.

References

Arthur, W. B. (1991). Designing economic agents that act like human agents: A behavioral approach to bounded rationality. *American Economic Review, Papers and Proceedings, 81*(2), 353–359.

Axelrod, R. M. (1997). *The complexity of cooperation: Agent-based models of competition and collaboration*. Princeton: Princeton University Press.

Axtell, R. L., Epstein, J. M., Dean, J. S., Gumerman, G. J., Swedlund, A. C., Harburger, J., et al. (2002). Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House Valley. *Proceedings of the National Academy of Sciences, 99*(Suppl 3), 7275.

Bak, P. (1996). *How nature works: The science of self-organized criticality*. New York: Copernicus Books.

Barabasi, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*(5439), 509.

Batten, D. F. (2000). *Discovering artificial economics: how agents learn and economies evolve*. Boulder: Westview Press.

Brose, U., Ostling, A., Harrison, K., & Martinez, N. D. (2004). Unified spatial scaling of species and their trophic interactions. *Nature, 428*(6979), 167–171.

Brown, D. G. & Robinson, D. T. (2006). Effects of heterogeneity in residential preferences on an agent-based model of urban sprawl. *Ecology and Society, 11*(1). http://www.ecologyandsociety.org/vol11/iss11/art46/

Brown, D. G., Riolo, R., Robinson, D. T., North, M., & Rand, W. (2005). Spatial process and data models: Toward integration of agent-based models and GIS. *Journal of Geographical Systems, 7*, 1–23.

Byrne, D. (1998). *Complexity theory and the social sciences*. London: Routledge.

Carrillo-Hermosilla, J. (2006). A policy approach to the environmental impacts of technological lock-in. *Ecological Economics, 58*(4), 717.

Chaitin, G. J. (1974). Information-theoretic computational complexity. *IEEE Transactions on Information Theory, IT20*, 10–15.

Chen, S.-H. (2005). Computational intelligence in economics and finance: Carrying on the legacy of Herbert Simon. *Information Sciences, 170*(1), 121–131.

Chen, Y., Irwin, E. G., & Jayaprakash, C. (2009). Dynamic modeling of environmental amenity-driven migration with ecological feedbacks. *Ecological Economics, 68*(10), 2498–2510.

Cilliers, P. (1998). *Complexity and postmodernism: Understanding complex systems*. New York: Routledge.

Clark, W. A. V. (2008). Geography, space, and science: Perspectives from studies of migration and geographical sorting. *Geographical Analysis, 40*(3), 258–275.

Dawid, H. (1999). On the convergence of genetic learning in a double auction market. *Journal of Economic Dynamics and Control, 23*(9–10), 1545–1567.

Dean, J. S., Gumerman, G. J., Epstein, J. M., Axtell, R. L., Swedlund, A. C., Parker, M. T., et al. (2000). Understanding Anasazi culture change through agent-based modeling. In T. A. Kohler & G. J. Gumerman (Eds.), *Dynamics of human and primate societies: Agent-based modeling of social and spatial processes* (pp. 179–205). New York: Oxford University Press.

Easterling, W. E., & Kok, K. (2002). Emergent properties of scale in global environmental modeling – Are there any? *Integrated Assessment, 3*(2–3), 233–246.

Easterling, W. E., & Polsky, C. (2004). Crossing the divide: Linking global and local scales in human-environment systems. In E. Sheppard & R. McMaster (Eds.), *Scale and geographic inquiry: Nature, society and method* (pp. 66–85). Oxford: Blackwell.

Edmonds, B., & Moss, S. (1997). Modelling bounded rationality using evolutionary techniques. *Lecture Notes in Computer Science, 1305*, 31–42.

Epstein, J. M. (2006). *Generative social science: studies in agent-based computational modeling*. Princeton: Princeton University Press.

Farmer, J. D., & Foley, D. (2009). The economy needs agent-based modelling. *Nature, 460*(7256), 685–686.

Folke, C. (2006). Resilience: The emergence of a perspective for social-ecological systems analyses. *Global Environmental Change, 16*(3), 253–267.

Fossett, M. (2006). Ethnic preferences, social distance dynamics, and residential segregation: Theoretical explorations using simulation analysis. *Journal of Mathematical Sociology, 30*(3–4), 185.

Funtowitcz, S. O., & Ravetz, J. R. (1994). Emergent complex systems. *Futures, 26*(6), 568–582.

Gatti, D. D., Di Guilmi, C., Gaffeo, E., Giulioni, G., Gallegati, M., & Palestrini, A. (2005). A new approach to business fluctuations: Heterogeneous interacting agents, scaling laws and financial fragility. *Journal of Economic Behavior & Organization, 56*(4), 489.

Gell-Mann, M. (1994). What is complexity? *Complexity, 1*(1), 1–9.

Gibson, C., Ostrom, E., & Ahn, T.-K. (2000). The concept of scale and the human dimensions of global change: A survey. *Ecological Economics, 32*(2), 217–239.

Gigerenzer, G., & Selton, R. (2001). *Bounded rationality: The adaptive toolbox*. Cambridge, MA: MIT Press.

Gilbert, N. (1995). Emergence in social simulation. In N. Gilbert & R. Conte (Eds.), *Artificial societies: The computer simulation of social life* (pp. 144–156). London: UCL Press.

Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology, 83*(6), 1420.

Holland, J. H. (1995). *Hidden order: How adaptation builds complexity*. Reading: Addison-Wesley.

Holland, J. H. (1998). *Emergence: From chaos to order*. Philadelphia: Perseus Press.

Holling, C. S. (1973). Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics, 4*, 1–23.

Holling, C. S. (1996). Engineering resilience versus ecological resilience. In P. Schulze (Ed.), *Engineering with ecological restraints* (pp. 31–43). Washington, D.C.: National Academy Press.

Irwin, E., Jayaprakash, C., & Munroe, D. (2009). Towards a comprehensive framework for modeling urban spatial dynamics. *Landscape Ecology, 24*(9), 1223–1236.

Janssen, M. A. (2009). Understanding artificial Anasazi. *Journal of Artificial Societies and Social Simulation*, *12*(4), 13. Available at: http://jasss.soc.surrey.ac.uk/12/4/13.html

Janssen, M. A., & Ostrom, E. (2006). Empirically based, agent-based models. *Ecology and Society*, *11*(1). http://www.ecologyandsociety.org/vol11/iss12/art37/

Levin, S. A., & Lubchenco, J. (2008). Resilience, robustness, and marine ecosystem-based management. *BioScience, 58*(1), 27–32.

Lissack, M. (2001). Special issue: What is complexity science? *Emergence, 3*(1).

Lorenz, E. N. (1973). *Predictability: Does the flap of a butterfly's wings in Brazil set off a tornado in Texas?* Paper presented at the AAAS Convention of the Global Atmospheric Research Program, Cambridge, MA, Dec 29, 1972.

Mainzer, K. (1996). *Thinking in complexity: The complex dynamics of matter, mind, and mankind*. New York: Springer.

Manson, S. M. (2001). Simplifying complexity: A review of complexity theory. *Geoforum, 32*(3), 405–414.

Manson, S. M. (2003). Validation and verification of multi-agent models for ecosystem management. In M. Janssen (Ed.), *Complexity and ecosystem management: The theory and practice of multi-agent approaches* (pp. 63–74). Northampton: Edward Elgar.

Manson, S. M. (2006). Bounded rationality in agent-based models: Experiments with evolutionary programs. *International Journal of Geographic Information Science, 20*(9), 991–1012.

Manson, S. M. (2007). Challenges in evaluating models of geographic complexity. *Environment and Planning B: Planning & Design, 34*(2), 245–260.

Manson, S. M. (2008). Does scale exist? An epistemological scale continuum for complex human-environment systems. *Geoforum, 39*(2), 776–788.

Manson, S. M., & O'Sullivan, D. (2006). Complexity theory in the study of space and place. *Environment and Planning A, 38*(4), 677.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas imminent in nervous activity. *Bulletin of Mathematical Biophysics, 5*, 115–137.

McKelvey, B. (1999). Complexity theory in organization science: Seizing the promise or becoming a fad? *Emergence, 1*(1), 5–32.

Ngo, T. A., & See, L. M. (2012). Calibration and validation of agent-based models of land cover change. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 181–196). Dordrecht: Springer.

Nowotny, H. (2005). The increase of complexity and its reduction: Emergent interfaces between the natural sciences, humanities and social sciences. *Theory Culture Society, 22*(5), 15.

O'Neill, R. V. (1988). Hierarchy theory and global change. In T. Rosswall, R. G. Woodmansee, & P. G. Risser (Eds.), *Scales and global change* (pp. 29–45). New York: Wiley.

Ostrom, E. (2005). *Understanding institutional diversity*. Princeton: Princeton University Press.

O'Sullivan, D. (2004). Complexity science and human geography. *Transactions of the Institute of British Geographers, 29*(3), 282–295.

Parker, D. C., & Filatova, T. (2008). A conceptual design for a bilateral agent-based land market with heterogeneous economic agents. *Computers, Environment and Urban Systems, 32*(6), 454–463.

Parker, D. C., Manson, S. M., Janssen, M., Hoffmann, M. J., & Deadman, P. J. (2003). Multi-agent systems for the simulation of land use and land cover change: A review. *Annals of the Association of American Geographers, 93*(2), 316–340.

Phillips, J. D. (2003). Sources of nonlinearity and complexity in geomorphic systems. *Progress in Physical Geography, 27*(1), 1–23.

Phillips, J. D. (2006). Deterministic chaos and historical geomorphology: A review and look forward. *Geomorphology, 76*(1–2), 109.

Portugali, J. (2006). Complexity theory as a link between space and place. *Environment and Planning A, 38*(4), 647–664.

Rind, D. (1999). Complexity and climate. *Science, 284*(5411), 105–107.

Sampson, R. J., Morenoff, J. D., & Gannon-Rowley, T. (2002). Assessing "neighborhood effects": Social processes and new directions in research. *Annual Review of Sociology, 28*, 443–478.

Sawyer, R. K. (2002). Emergence in sociology: Contemporary philosophy of mind and some implications for sociological theory. *American Journal of Sociology, 107*, 551–585.

Schneider, S., Turner, B. L., II, & Garriga-Morehouse, H. (1998). Imaginable surprise in global change science. *Journal of Risk Research, 1*(2), 165–185.

Simon, H. A. (1997a). *Administrative behavior: A study of decision-making processes in administrative organizations*. New York: Free Press.

Simon, H. A. (1997b). *Models of bounded rationality*. Cambridge, MA: MIT Press.

Skvoretz, J. (2002). Complexity theory and models for social networks. *Complexity International, 8*(1), 47–55.

Torrens, P. (2007). Simulating sprawl. *Annals of the Association of American Geographers, 96*(2), 248–275.

Torrens, P., & O'Sullivan, D. (2001). Cellular automata and urban simulation: Where do we go from here? *Environment and Planning B, 28*(2), 163–168.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.

von Bertalanffy, L. (1968). *General systems theory: Foundation, development, applications*. London: Allen Lane.

von Neumann, J. (1966). *Theory of self-reproducing automata*. Champaign-Urbana: University of Illinois Press.

Walker, B., Holling, C. S., Carpenter, S. R., & Kinzig, A. (2004). Resilience, adaptability and transformability in social-ecological systems. *Ecology and Society*, *9*(2). http//www.ecologyandsociety.org/vol9/iss2/art5/

Watts, D. J. (2003). *Six degrees: The science of a connected age*. New York: W.W. Norton and Co.

Wheaton, W. D., Cajka, J. C., Chasteen, B. M., Wagener, D. K., Cooley, P. C., Ganapathi, L., et al. (2009). *Synthesized population databases: A US geospatial database for agent-based models*. Research Triangle Park: RTI International.

Whitehead, A. N. (1925). *Science and the modern world*. New York: Macmillan.

Wiener, N. (1961). *Cybernetics: Or, control and communication in the animal and the machine*. Cambridge, MA: MIT Press.

Chapter 8
Designing and Building an Agent-Based Model

Mohamed Abdou, Lynne Hamill, and Nigel Gilbert

Abstract This chapter discusses the process of designing and building an agent-based model, and suggests a set of steps to follow when using agent-based modelling as a research method. It starts with defining agent-based modelling and discusses its main concepts, and then it discusses how to design agents using different architectures. The chapter also suggests a standardized process consisting of a sequence of steps to develop agent-based models for social science research, and provides examples to illustrate this process.

8.1 What Is Agent-Based Modelling?

Agent-based modelling is a *computational* method that enables a researcher to create, analyze, and *experiment* with *models* composed of *agents* that interact within an *environment*. Let us shed some light on the core terms italicized in this definition.

A *model* is a simplified representation of a "target" system that expresses as clearly as possible the way in which (one believes) that system operates. This representation can take several forms. For example, in mathematical and statistical modelling, the model is a set of equations (e.g., a regression equation). A graphical network of nodes and edges can model a set of friendships. *Computational* methods, such as agent-based modelling, involve building models that are computer programs. The program (i.e., the model) represents

M. Abdou (✉)
Department of Social Science Computing, Faculty of Economics and Political Science, Cairo University, Cairo, Egypt
e-mail: moabdou71@yahoo.com

L. Hamill • N. Gilbert
CRESS, Department of Sociology, University of Surrey, Guildford, UK
e-mail: lynne.hamill@surrey.ac.uk; n.gilbert@surrey.ac.uk

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_8,

the processes that are thought to exist in the social world (Macy and Willer 2002). For example, we might build a model to study how friends ("agents") influence each other's purchasing choices. Such processes are not easy to represent using mathematical equations because what one agent buys will influence the purchasing of a friend, and what a friend buys will influence the first agent. This kind of mutual reinforcement is relatively easy to model using agent-based modelling.

Another advantage of agent-based modelling when doing social research is that it enables a researcher to use the model to do *experiments*. Unlike natural sciences, such as physics and chemistry, conducting experiments on the real system (people for example) is impossible or undesirable. Using a computer model, an experiment can be set up many times using a range of parameters. The idea of experimenting on models rather than on the real system is not novel. For example, it is a better idea to use a model of an aeroplane to test flying under various conditions than to use a real aircraft (where the cost of experimentation is very high).

Agent-based models (ABMs) consist of *agents* that interact within an *environment*. Agents themselves are distinct parts of a program that represent social actors (e.g., persons, organizations such as political parties, or even nation-states). They are programmed to react to the computational environment in which they are located, where this environment is a model of the real environment in which the social actors operate.

In the following, we present two simple examples of ABMs, Sugarscape and Schelling's model of residential segregation, to illustrate the main concepts of agent-based modelling used in the remaining sections of this chapter. A general introduction to agent-based modelling is presented in Crooks and Heppenstall (2012).

8.1.1 Sugarscape

Sugarscape (Epstein and Axtell 1996) is a simple example of an ABM that yields a range of interesting results about the distribution of wealth in a society. The model represents an artificial society in which agents move over a 50×50 cell grid. Each cell has a gradually renewable quantity of 'sugar', which the agent located at that cell can eat. However, the amount of sugar at each location varies. Agents have to consume sugar in order to survive. If they harvest more sugar than they need immediately, they can save it and eat it later (or, in more complex variants of the model, can trade it with other agents). Agents can look to the north, south, east and west of their current locations and can see a distance which varies randomly, so that some agents can see many cells away while others can only see adjacent cells.

Agents move in search of sugar according to the rule: look for the unoccupied cell that has the highest available sugar level within the limits of one's vision, and move there. Agents not only differ in the distance they can see, but also in their 'metabolic rate', the rate at which they use sugar. If their sugar level ever drops to zero, they die. New agents replace the dead ones with a random initial allocation of sugar. Thus there is an element of the 'survival of the fittest' in the model, since those agents that are relatively unsuited to the environment because they have high metabolic rates, poor vision, or are in places where there is little sugar for harvesting, die relatively quickly of starvation. However, even successful agents die after they have achieved their maximum lifespan, set according to a uniform random distribution.

Epstein and Axtell (1996) present a series of elaborations of this basic model in order to illustrate a variety of features of societies. The basic model shows that even if agents start with an approximately symmetrical distribution of wealth (the amount of sugar each agent has stored), a strongly skewed wealth distribution soon develops. This is because a few relatively well-endowed agents are able to accumulate more and more sugar, while the majority only barely survive or die.

8.1.2 Schelling's Model of Residential Segregation

Another simple example is Schelling's model of residential segregation (1971). Schelling was interested in the phenomenon of racial residential segregation in American cities, and he aimed to explain how segregation could happen, and how these segregationist residential structures, such as ghettos, may occur spontaneously, even if people are relatively tolerant towards other ethnic groups, and even when they are happy with being a minority in their neighbourhoods.

A city in Schelling's model is represented by a square grid of cells each representing a dwelling. A cell can be in any of three colours: white, black, or grey according to whether it is occupied by a white agent, a black agent, or is empty. The simulation starts by randomly distributing the agents over the grid. Schelling supposed that people have a 'threshold of tolerance' of other ethnic groups. That means that agents are 'content' to stay in their neighbourhood as long as the proportion of their neighbours (which are the eight cells to the north, north-east, east, south-east, south, south-west, west and north-west) of the same colour as themselves is not less than this threshold. For example, with 50% threshold of tolerance, agents would be happy to stay in place as long as at least four of their eight neighbours are of the same colour; otherwise, they try to move to another neighbourhood satisfying this proportion.

Figure 8.1 shows the result of the simulation with 2,000 agents. The upper-left panel shows the starting random allocation of black and white agents over the grid, and the other three panels show the final configurations after running the simulation with tolerance thresholds of 37.5% (at least three of an agent's eight neighbours must be of the same colour for the agent to be content), 50% (four of eight), and 75% (six of eight). Clustering emerges even when agents are happy to be a minority in their neighbourhood (with 37.5% threshold), and the sizes of these emergent clusters increase with increasing levels of tolerance threshold.

In the following, we discuss the core concepts of "agents" and their "environment" in more detail.

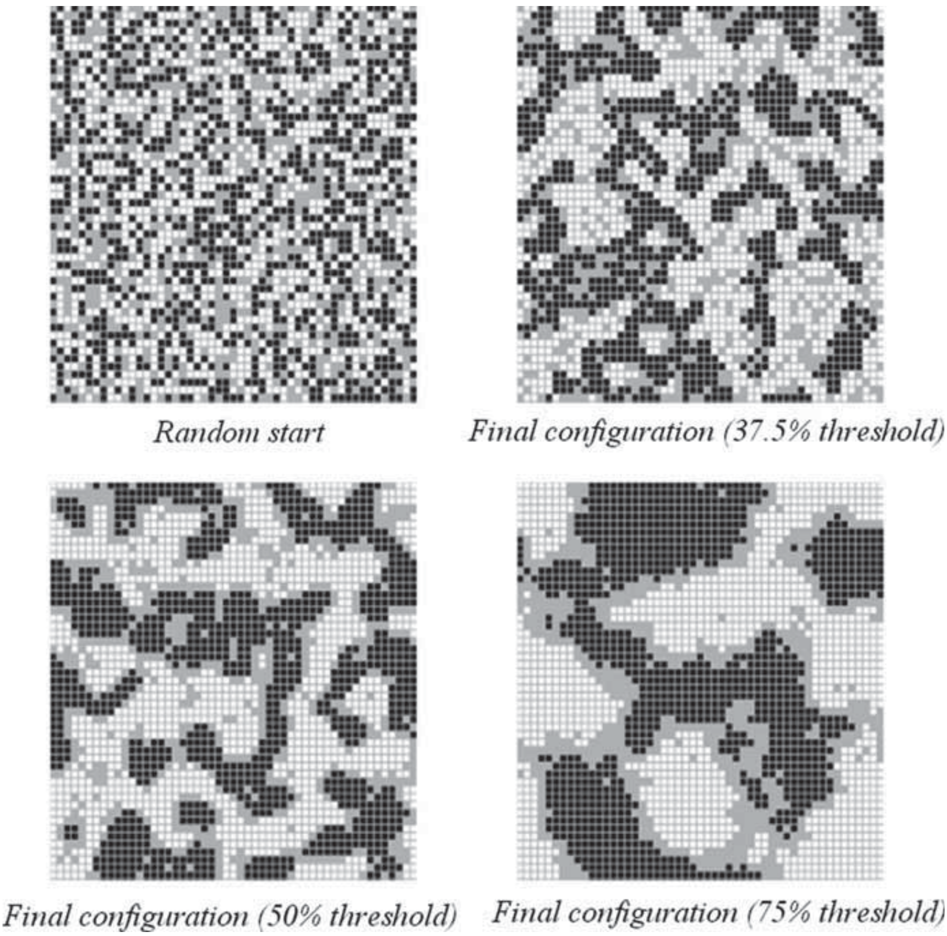


Fig. 8.1 The result of the simulation of the Schelling model

8.2 Agents

Applied to social science research, the concept of agency is usually used to indicate the purposive nature of human activity. It is thus related to concepts such as intentionality, free will, and the power to achieve one's goals. In agent-based modelling, agents are conventionally described as having four important characteristics:

- *Perception*. Agents can perceive their environment, including other agents in their vicinity. In the Sugarscape model, for example, agents can perceive the amount of sugar the current cell has.
- *Performance*: They have a set of behaviours that they are capable of performing such as moving, communicating with other agents, and interacting with the environment. In the Sugarscape model, they move and consume sugar.

- *Memory.* Agents have a memory in which they record their previous states and actions.
- *Policy*. They have a set of rules, heuristics or strategies that determine, given their present situation and their history, what they should do next, e.g. looking for cells with the highest level of sugar.

Agents with these features can be implemented in many different ways. Different architectures (i.e. designs) have merits depending on the purpose of the simulation. Nevertheless, every agent design has to include mechanisms for receiving input from the environment, for storing a history of previous inputs and actions, for devising what to do next, for carrying out actions and for distributing outputs. In the following, we describe three common approaches to agent architecture: using an object-oriented programming language directly, using a production rule system, and using learning approaches.

8.2.1 Object-Oriented Programming

The idea of object-oriented programming (OOP) is crucial to agent-based modelling, which is why almost all ABMs are built using an OOP language, such as Java, C++, or Visual Basic. A program developed in an OOP language typically consists of a collection of *objects*. An object is able to store data in its own *attributes,* and has *methods* that determine how it processes these data and interacts with other objects. As you might have noticed, there is an affinity between the idea of an agent and an object; it is natural to program each agent as an object.

The concept of 'class' is basic to OOP. A class is an abstract specification of an object. For example, a program might include a class called "Customer" to represent a customer of a firm in a model of business. A Customer might have a set of attributes such as name, address, and types of product (s)he likes. In the Sugarscape model, we can create a class named "Agent" with attributes such as *age*, *wealth* (the amount of sugar), *life-expectancy* (the maximum age that can be reached), *metabolism* (how much sugar an agent eats each time period), and *vision* (how many cells ahead an agent can see). A class also usually has some methods to describe its activities (e.g., move, eat sugar, save and die).

As the program runs, classes are *instantiated* to form objects. For example, the Customer class might be instantiated to yield two objects representing two customers: the first with name John Smith and the other with name Sara Jones (along with their other attributes). Although the two customers have the same methods and the same set of attributes, the values of their attributes (e.g., their names and addresses) differ.

When using OOP to design an ABM, one creates a class for each type of agent, provides attributes that retain the agents' past current state (*memory*), and adds suitable methods that observe the agents' environment (*perception*) and carry out agent actions (*performance*) according to some rules (*policy*). In addition, one needs to program a scheduler that instantiates the required number of agents at the beginning of the simulation and gives each of them a turn to act.

8.2.2 Production Systems

One of the simplest, yet effective, designs for an agent is to use a production system. A production system has three components:

1. *A Set of Rules of Behaviour*. These rules determine what an agent will do. Usually, a rule consists of two parts: a condition, which specifies when the rule is to be executed ('fire'), and an action part, which determines what is to be the consequence of the rule firing. Example of rules in Sugarscape include:
   - IF there is any sugar at the current cell, THEN eat it;
   - IF *sugar level* of the current cell exceeds *metabolism,* THEN add the extra sugar to *wealth;* and
   - IF age exceeds *life-expectancy,* THEN die.
2. *A Working Memory*. An agent's memory is represented by variables that store its current or previous states. For example, an agent's memory might store its current location and wealth (the amount of sugar). Rules can include actions that insert facts into the working memory (e.g. I am holding some sugar) or conditions that test the state of the working memory (e.g. IF I am holding sugar, THEN eat it).
3. *A Rule Interpreter*. The rule interpreter considers each rule in turn, fires those for which the condition is true, performs the indicated actions for the rules that have fired, and repeats this cycle indefinitely. Different rules may fire on each cycle either because the immediate environment has changed or because one rule has modified the working memory in such a way that a new rule begins to fire.

Using a production system, it is relatively easy to build reactive agents that respond to each stimulus from the environment with some action. A simple production system can be constructed from a toolkit such as JESS (the Java Expert System Shell, http://www.jessrules.com/) (Friedman-Hill 2003). There are also some much more elaborate systems that are based on psychologically plausible models of human cognition, such as Soar (Laird et al. 1987; Wray and Jones 2006; Ye and Carley 1995), CLARION (Sun 2006), and ACT-R (Taatgen et al. 2006).

8.2.3 Learning

Production-system-based agents have the potential to learn about their environment and about other agents through adding to the knowledge held in their working memories. The agents' rules themselves, however, always remain unchanged. For some models, it is desirable to create agents that are capable of more fundamental learning: where the internal structure and processing of the agents adapt to changing circumstances. There are two techniques commonly used for this: *artificial neural networks (ANNs)* and evolutionary algorithms such as the *genetic algorithm (GA)*.

ANNs are inspired by analogy to nerve connections in the brain. An ANN consists of three or more layers of neurons, with each neuron connected to all other neurons in the adjacent layers. The first layer accepts input from the environment, processes it and passes it on to the next layer. The signal is transmitted through the layers until it emerges at the output layer. Each neuron accepts inputs from the preceding layer, adjusts the inputs by positive or negative weights, sums them and transmits the signal onward. Using an algorithm called the *back propagation of error*, the network can be tuned so that each pattern of inputs gives rise to a different pattern of outputs. This is done by training the network against known examples and adjusting the weights until it generates the desired outputs (Garson 1998). Using ANNs, it is possible to design agents and train them to identify objects such as letters or words, or recognize voices and pictures.

In contrast to a production system, an ANN can modify its responses to stimuli in the light of its experience. A number of network topologies have been used to model agents so that they are able to learn from their actions and the responses of other agents (e.g. Hutchins and Hazlehurst 1995; Terna 1997).

Another way of enabling an agent to learn is to use an evolutionary algorithm. These are also based on a biological analogy, drawing on the theory of evolution by natural selection. The most common is the genetic algorithm (GA). This works with a population of individuals (agents), each of which has some measurable degree of 'fitness', using a metric defined by the model builder. The fittest individuals are 'reproduced' by breeding them with other fit individuals to produce new offspring that share some features taken from each parent. Breeding continues through many generations, with the result that the average fitness of the population increases as the population adapts to its environment.

Sometimes, it is desirable to use both techniques of learning, GAs and ANNs, in the same ABM. For example, one may need to create a large population of ANNs (each corresponding to one agent). The agents are initialized with a random set of connection weights and are set a task such as gathering "food" from a landscape. An agent's perception of whether there is food in front of it is fed into the ANN inputs, and the outputs are linked to the agent's action, such as move and eat. The agent is given an initial quantity of energy, some of which is used on every time step. If the energy declines to zero, the agent "dies" and it is removed from the simulation. An agent can boost its energy by eating food, which is scattered around the landscape.

Because of the random connection weights with which an agent's ANN is initialized, most agents will not succeed in finding and eating food and will quickly die, although some will succeed. Those more successful agents reproduce, giving their offspring similar connection weights as their own (but with slight mutation). Gradually, the population of agents will learn food harvesting behaviour (Acerbi and Parisi 2006; Gilbert et al. 2006).

8.2.4 The Environment

The environment is the virtual world in which agents operate. In many models, the environment includes passive objects, such as landscape barriers, "roads" down which agents may travel, resources to provide agents with energy or food (as in the Sugarscape model), and so on. These can be programmed in much the same way as agents, but more simply, because they do not need any capacity to react to their surroundings. For example, the environment in the Sugarscape model can be implemented by creating a class, called "Cell", which has two attributes: *location*, which is the *xy* position of a cell, and *sugar level*, which indicates the amount of sugar the cell has. Then 2,500 (50×50) objects of this class are instantiated at the start of the simulation with their proper locations and random values for their sugar levels.

Environments may represent geographical spaces, for example, in models concerned with residential segregation where the environment simulates some of the physical features of a city, and in models of international relations, where the environment maps states and nations. Models in which the environment represents a geographical space are called *spatially explicit*. In other models, the environment could represent other types of space. For example, scientists can be modelled in "knowledge space" (Gilbert et al. 2001). In spatial models, the agents have coordinates to indicate their location. Another option is to have no spatial representation at all but to link agents together into a network in which the only indication of an agent's relationship to other agents is the list of agents to which it is connected by network links (Scott 2000). It is also possible to combine both. Think, for example, of a railway network.

8.3 Developing ABMs in Social Science Research

Research in agent-based modelling has developed a more or less standardized research process, consisting of a sequence of steps. In practice, several of these steps occur in parallel and the whole process is often performed iteratively as ideas are refined and developed.

8.3.1 Identifying the Research Question

It is essential to define precisely the research question (or questions) that the model is going to address at an early stage. The typical research questions that ABMs are used to study are those that explain how regularities observed at the societal or macro level can emerge from the interactions of individuals (agents) at the micro level. For example, the Schelling model described earlier starts with the observation that neighbourhoods are ethnically segregated and seeks to explain this through modelling individual household decisions.

8.3.2 Review of Relevant Literature

The model should be embedded in existing theories and make use of whatever data are available. Reviewing existing theories relating to the model's research question is important to illuminate the factors that are likely to be significant in the model. It is also useful to review comparable phenomena. For example, when studying segregation, theories about prejudice and ethnic relations are likely to be relevant.

All ABMs are built based on assumptions (usually about the micro-level). These assumptions need to be clearly articulated, supported by the existing theories and justified by whatever information is available.

8.3.3 Model Design

After the research question, the theoretical approach and the assumptions have been clearly specified, the next step is to specify the agents that are to be involved in the model and the environment in which they will act.

For each type of agent in the model, the attributes and behavioural rules need to be specified. As explained in Sect. 8.2, an attribute is a characteristic or feature of the agent, and it is either something that helps to distinguish the agent from others in the model and does not change, or something that changes as the simulation runs. For example, in Sugarscape, an agent's *life-expectancy* (the maximum age that an agent can reach), *metabolism* (how much sugar an agent eats each time), and *vision* (how many cells ahead an agent can see) are examples of attributes that do not change, while *age* and *wealth* (the amount of sugar an agent has) are changeable attributes.

The agent's behaviour in different circumstances also needs to be specified, often as a set of condition-action rules (as explained in Sect. 8.2). This specification can be done in the form of two lists: one which shows all the different ways in which the environment (including other agents) can affect the agent, and one showing all the ways in which the agent can affect the environment (again, including other agents). Then the conditions under which the agent has to react to environmental changes can be written down, as can the conditions when the agent will need to act on the environment. These lists can then be refined to create agent rules that show how agents should act and react to environmental stimuli.

It will also be necessary to consider what form the environment should take (for instance, does it need to be spatial, with agents having a definite location, or should the agents be linked in a network) and what outputs of the model need to be displayed in order to show that it is reproducing the macro-level regularities as hoped (for example, the wealth distribution in the Sugarscape model, and the size of clusters of dwellings of the same colour in Schelling's model).

Once all this has been thought through, one can start to develop the program code that will form the simulation.

8.3.4 Model Implementation

After the model has been designed, and when the agents and environment are fully specified, the next step is to convert the design into a computer program. Most ABMs involve two main parts or *procedures*:

- ***Setup Procedure***. The Setup procedure initializes the simulation (and is therefore sometimes called the initialization procedure). It specifies the model's state at the start of the simulation, and it is executed once at the beginning. This part of the program might, for example, lay out the environment and specify the initial attributes of the agents (e.g., their position, wealth and life expectancy in the Sugarscape model).
- ***Dynamics Procedure***. This procedure is repeatedly executed in order to run the simulation. It asks agents in turn to interact with the environment and other agents according to their behavioural rules. This will make changes in the environment and invoke a series of action-reaction effects. For example, in Schelling's model of segregation, the dynamics procedure may ask all 'unhappy' agents to move from their neighbourhood. When an unhappy agent moves to a new place (where it feels happy), this may make some other agents (that were happy in the previous step) unhappy and want to move, and so on. The dynamics procedure may contain a condition to stop the program (e.g., if all agents are happy in Schelling's model).

An important decision is whether to write a special computer program (using a programming language such as Java, C++, C#, or Visual Basic) or use one of the packages or toolkits that have been created to help in the development of simulations. It is usually easier to use a package than to write a program from scratch. This is because many of the issues which take time when writing a program have already been dealt with in developing the package. For example, writing code to show plots and charts is a skilled and very time-consuming task, but most packages provide some kind of graphics facility for the display of output variables. On the other hand, packages are, inevitably, limited in what they can offer, and they are usually run more slowly than specially written code.

Many simulation models are constructed from similar building blocks. These commonly used elements have been assembled into *libraries* or *frameworks* that can be linked into an agent-based program. The first of these to be widely used was Swarm (http://www.swarm.org/), and although this is now generally superseded, its design has influenced more modern libraries, such as *RePast* (http://repast.sourceforge.net/) and *Mason* (http://cs.gmu.edu/~eclab/projects/mason/).

Both RePast and Mason provide a similar range of features, including:

- A variety of helpful example models
- A sophisticated scheduler for event-driven simulations
- A number of tools for visualizing on screen the models and the spaces in which the agents move
- Tools for collecting results in a file for later statistical analysis
- Ways to specify the parameters of the model and to change them while the model is running

- Support for network models (managing the links between agents)
- Links between the model and a Geographic Information System (GIS) so that the environment can be modeled on real landscapes (see Crooks and Castle 2012).
- A range of debugged algorithms for evolutionary computation (Sect. 8.2.3), the generation of random numbers and the implementation of ANNs.

Modelling environments provide complete systems in which models can be created, executed, and the results visualized without leaving the system. Such environments tend to be much easier to learn, and the time taken to produce a working model can be much shorter than using the library approach, and so they are more suited to beginners. However, the simplicity comes at the price of less flexibility and slower speed of execution. It is worth investing time to learn how to use a library based framework if you need the greater power and flexibility they provide, but often simulation environments are all that is needed.

NetLogo (Wilensky 1999) is currently the best of the agent-based simulation environments. (NetLogo will be briefly introduced in Sect. 8.4). This is available free of charge for educational and research use and can be downloaded from http://ccl.northwestern.edu/netlogo/. It will run on all common operating systems: Windows, Mac OS X and Linux. Other simulation environments include StarLogo (http://education.mit.edu/starlogo/) and AgentSheets (http://agentsheets.com), which are more suited to creating very simple models for teaching than for building simulations for research.

Table 8.1 provides a comparison between Swarm, RePast, Mason, and NetLogo on a number of criteria. The choice of the implementation tool depends on several factors, especially one's own expertise in programming and the complexity and the scale of the model. NetLogo is the quickest to learn and the easiest to use, but may not be the most suitable for large and complex models. Mason is faster than RePast, but has a significantly smaller user base, meaning that there is less of a community that can provide advice and support. A full discussion of the environments is presented in Crooks and Castle (2012).

8.3.5 Verification and Validation

Once we have a 'working' simulation model, it has to be verified and validated before using it to answer the research questions or to build theories about the real social world (model verification and validation are discussed in detail by Ngo and See (2012)). As Balci (1994) explains, *"model validation deals with building the right model … [while] model verification deals with building the model right"* (pp. 121–123).

It is very common to make errors when writing computer programs, especially complicated ones. The process of checking that a program does what it was planned to do is known as 'verification'. In the case of simulation, the difficulties of verification are compounded by the fact that many simulations include random number generators, which means that every run is different and that it is only the

Table 8.1 A Comparison of Swarm, RePast, Mason and NetLogo

	Swarm	RePast	Mason	NetLogo
License[a]	GPL	GPL	GPL	Free, but not open source
Documentation	Patchy	Limited	Improving, but limited	Good
User base	Diminishing	Large	Increasing	Large
Modelling language(s)	Objective-C, Java	Java, Python	Java	NetLogo
Speed of execution	Moderate	Fast	Fastest	Moderate
Support for graphical user interface development	Limited	Good	Good	Very easy to create using "point and click"
Built-in ability to create movies and animations	No	Yes	Yes	Yes
Support for systematic experimentations	Some	Yes	Yes	Yes
Ease of Learning and Programming	Poor	Moderate	Moderate	Good
Ease of Installation	Poor	Moderate	Moderate	Good
Link to geographical Information System	No	Yes	Yes	Yes

Source: Gilbert (2008)

[a]*GPL* General Public License, http://www.gnu.org/copyleft/gpl.html

distribution of results which can be anticipated by the theory. It is therefore essential to 'debug' the simulation carefully, preferably using a set of test cases, perhaps of extreme situations where the outcomes are easily predictable.

While verification concerns whether the program is working as the researcher expects, validation concerns whether the simulation is a good model of the real system, the 'target'. A model which can be relied on to reflect the behaviour of the target is 'valid'. A common way of validating a model is to compare the output of the simulation with real data collected about the target. However, there are several caveats which must be borne in mind when making this comparison. For example, exact correspondence between the real and simulated data should not be expected. So, the researcher has to decide what difference between the two kinds of data is acceptable for the model to be considered valid. This is usually done using some statistical measures to test the significance of the difference. While goodness-of-fit can always be improved by adding more explanatory factors, there is a trade-off between goodness-of-fit and simplicity. Too much fine-tuning can result in reduction of explanatory power because the model becomes difficult to interpret. At the extreme, if a model becomes as complicated as the real world, it will be just as difficult to interpret and offer no explanatory power. There is, therefore, a paradox here to which there is no obvious solution. Despite its apparently scientific nature, modelling is a matter of judgement.

8.3.6 Some Practicalities

Two important practical issues to consider are how big the model should be and how many runs should be done.

8.3.6.1 How Big?

How many agents should be used? Over how big a space? There is little guidance on this question, because it depends on the model. The model must be sufficiently large to permit enough heterogeneity and opportunities for interaction. But more agents mean longer run times.

It is often best to start programming with just a few agents in a small environment. Then, when the program is working satisfactorily, increase the scale until one feels there is a satisfactory balance between the size and the stability of the output. Some ABMs use millions of agents (see Parry and Bithnell 2012), but for most purposes, this is unnecessary and impractical. One should probably aim for at least 1,000 agents unless there is good reason to use fewer.

8.3.6.2 How Many Runs?

Because of the stochastic nature of agent-based modelling, each run produces a different output. It is therefore essential to undertake more than one run. The question is, how many runs? The more runs, the more confidence one can have in the results, but undertaking too many runs wastes time and there is more data to analyze. Basic statistical theory suggests 30 is sufficient and frequently, 30 or 50 runs are undertaken, e.g. Epstein (2006). Again, there is no clear guidance on this topic. However many runs are done, it is worth quoting the standard deviation to provide some indication of the variability.

8.4 Examples

This section presents two simple models based on models in NetLogo's library: Traffic Basic and Segregation (Wilensky 1997a, b). The models are taken from NetLogo version 4.0.2.

8.4.1 A Basic Traffic Model

This is a very simple model developed from Wilensky's basic NetLogo traffic model (1997a). It is not possible to give a full introduction to NetLogo here: there are tutorials on the NetLogo website and books such as Gilbert (2008). However, for those unfamiliar with NetLogo, an explanation of what the program is doing is provided alongside the code (see Box A).

Sect. 8.3 identified five stages to developing a model: identifying the research question, reviewing the literature, designing and implementing the model and finally verifying and validating it.

Stage 1: Identifying the research question
The research question to be addressed is the relationship between the level of congestion and the speed and smoothness of traffic flow.

Stage 2: Reviewing the literature
Because the main purpose of this model is to demonstrate agent-based modelling, it is sufficient to note here that it is a well recognised fact that traffic jams can arise without any obvious cause. In general, a good literature review is essential to support the model.

Stage 3: Model design
The environment is a road and the agents are drivers represented by cars. The drivers change their speed according to whether there are other cars in front so as to remain within set speed limits. The program records the speed of the vehicles and the number of vehicles queuing at any one time.

Stage 4: Model implementation
The set-up procedure involves setting the parameters and creating the agents and their environment. The environment – the road – is built and the cars are created, distributed randomly along the road and randomly allocated a speed, determined by three parameters, set by sliders on the interface:

- the number of cars (nOfCars): minimum 2, maximum, 30
- the minimum speed (minSpeedLimit): 0–0.5
- the maximum speed (maxSpeedLimit): 0.5–1.

The details are shown in Box A, and a sample of the result is illustrated in Fig. 8.2.

Box A: Setting up the model

Explanation	Code
Agents are cars	`to setup`
Agents' attributes	`breed [cars car ]`
	`cars-own [speed queuing]`

Set everything to zero
Make the road.
	`to setup`
	`  clear-all`
	`  ask patches`
	`    [ if ( pycor < 1 )`
	`    and ( pycor > -1)`
	`      [ set pcolor white ] ]`

(continued)

Box A: (continued)

Explanation	Code
Generate required number of cars	`create-cars nOfCars`
Set the cars' shape.	`ask cars`

Distribute them randomly along road. Set direction of movement.	`[ set shape "car"`
	`setxy random-xcor 0`
	`set heading 90`

Set speed randomly within the speed limits. Call up procedure.	`set speed`
	`minSpeedLimit +`
	`random-float`
	`( maxSpeedLimit )`
	`separate-cars`
	`]`
	`end`

Procedure to ensure only one car occupies the same patch of road.	`to separate-cars`
	`if any? other cars-here`
	`[ fd 1`
	`separate-cars ]`

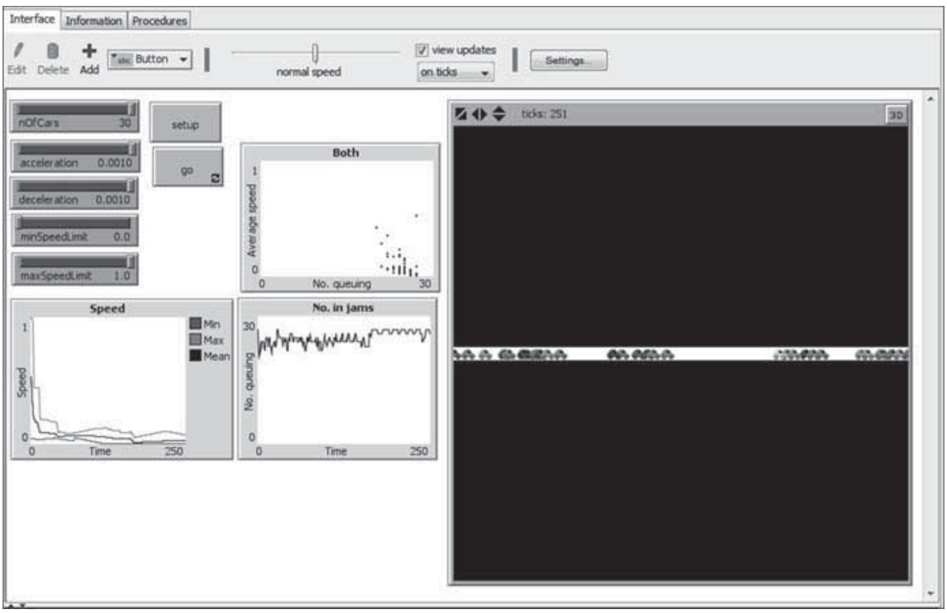


Fig. 8.2 Road with cars distributed randomly along it

Next the dynamic processes must be defined. All the cars move forward in the same direction. If the drivers see another car not far in front, they decelerate, at a rate set by the slider on the interface (decelerate), and if they catch up with the vehicle in front, slow to its speed, which may require rather abrupt deceleration! If they see no car within a specified distance, they accelerate again, set by a slider on the interface (accelerate). The rate of acceleration is small but sufficient to allow the cars to speed up to the maximum speed limit if the road is clear. Both deceleration and acceleration are allowed to vary between 0 and 0.001 in increments of 0.0001. The simulation is halted after 250 steps. The details are shown in Box B.

Box B: Running the model

Explanation	Code
Stop the program after 250 steps.	`to go` `if ticks > 250 [ stop ]`
Reset queuing indicator at start of each step	`ask cars [ set queuing “No” ]`

If a car catches up with the one in front it slows to match its speed.	`ask cars`
	`[ if any? cars-at 1 0`
	`[ set speed`
	`( [speed] of one-of cars-at 1 0 )`
	`set queuing “Yes” ]`
	`]`

If there is no car immediately in front but there is one a little further ahead, the car decelerates. Otherwise, it accelerates.	`ask cars with [queuing=“No” ]`
	`[ ifelse`
	`any? cars-at 5 0`
	`[ set speed speed - deceleration ]`
	`[ set speed speed + acceleration ]`
	`]`
	`]`

To keep the cars within speed limits.	`ask cars`
	`[ if speed < minSpeedLimit`
	`[ set speed minSpeedLimit ]`
	`if speed > maxSpeedLimit`
	`[ set speed maxSpeedLimit ]`

Cars move forward at the speed determined.	`fd speed` `]`
Time moves forward.	`tick`

Stage 5: Verifying and validating
To verify and validate the model requires outputs to be produced. Here three graphs are drawn:

- to show the minimum, average and maximum speeds
- to show the number of queuing cars, and
- to plot the number queuing against the average speed.

The details are in Box C.

Verification and validation are discussed in Sect. 8.3.5 above and in detail in Ngo and See (2012). In this example, one simple method of verification is setting the minimum and maximum speeds to the same value and checking that all the

Box C: Generating the output (continuing the 'go' procedure)

Explanation	Code

Name the plots	`plot-speed`
	`plot-jams`
	`plot-both`

End the "to go"	`end`

Plots minimum, mean and maximum speeds.	`to plot-speed`
	`set-current-plot "Speed"`
	`set-current-plot-pen "Min"`
	`plot min [speed] of cars`
	`set-current-plot-pen "Mean"`
	`plot mean [speed] of cars`
	`set-current-plot-pen "Max"`
	`plot max [speed] of cars`
	`end`

Plots the number queuing.	`to plot-jams`
	`set-current-plot "No. in jams"`
	`plot count cars with`
	`[ queuing = "Yes" ]`
	`end`

Plots the mean speed against the number queuing.	`to plot-both`
	`set-current-plot "Both"`
	`plotxy`
	`count cars with`
	`[ queuing = "Yes" ]`
	`mean [speed] of cars`
	`end`

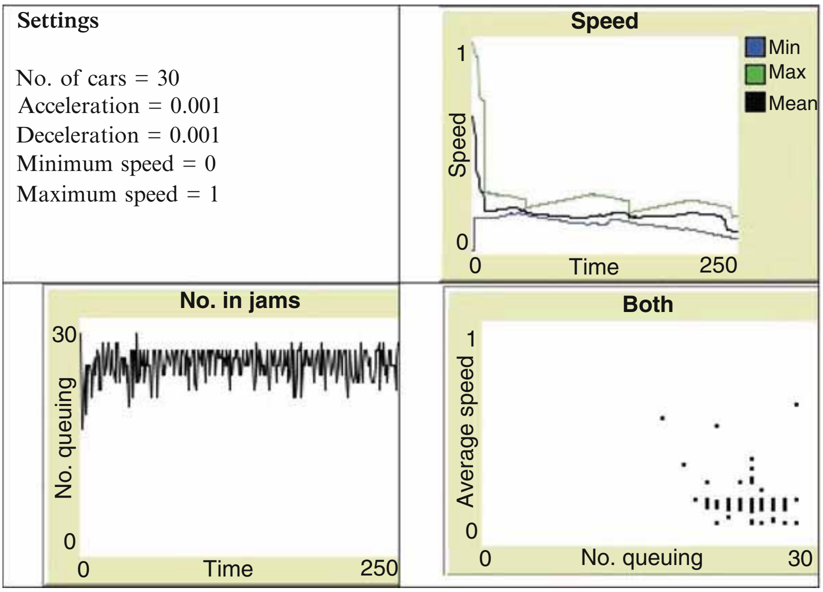


Fig. 8.3 Sample of results

drivers do adopt the same speed. By watching the movement of the cars on the screen, it can be seen that, for example, there is no overtaking, as there should not be. Also the queuing status of individual cars can be checked: if there is no car immediately in front, it should not be "queuing".

Even a simple model like this can produce a wide range of scenarios and reproduce observed characteristics of traffic flows. For example, Fig. 8.3 shows what can happen if the road is near full-capacity with 30 cars, speeds are allowed to vary from 0 to 1, and drivers accelerate and decelerate at the maximum rates. The top right plot shows that the maximum speed drops quickly, but maximum, average and minimum speeds fluctuate. As a result, the number queuing constantly changes, albeit within a small range, as shown in the bottom left hand panel. However, a reduction in the number queuing does not necessarily increase the average speed of the traffic: the bottom right hand panel shows that there is no clear relationship between the average speed and the number queuing.

8.4.2 Segregation Model

The segregation model can be found in the Social Science section of NetLogo's library (Wilensky 1997b).

Stage 1: Identifying the research question
As explained in Sect. 8.1.2, Schelling tried to explain the emergence of racial residential segregation in American cities. The main research question of Schelling's

models can be formulated as: *can segregation be eliminated (or reduced) if people become more tolerant towards others from different ethnic/racial groups?*

Stage 2: Reviewing the literature
Theories of intergroup relations (Sherif 1966) are relevant when discussing the emergence of residential segregation. Some of these theories are Social Identity and Social Categorization Theories (Tajfel 1981), Social Dominance Theory (Sidanius et al. 2004), and System Justification Theory SJT (Jost et al. 2004). The Contact Hypothesis (Allport 1954), which implies that inter-group relations decrease stereotyping, prejudice and discrimination, is also relevant. Reviewing literature on how to measure segregation is clearly essential (Massey and Denton 1988).

Stage 3: Model design
As explained in Sect. 1.2.2, the environment is a city that is modelled by a square grid of cells each representing a dwelling. A household (agent) would be 'happy' to stay at its place as long as the proportion of its neighbours of the same colour as itself is not less than its threshold of tolerance. Agents keep changing their places as long as they are not happy. Box D presents the complete code of the segregation model.[1]

Stage 4: Model implementation
Lines 1–30 of Box D initialize the model. The first line creates an agent type (breed in NetLogo's language) called 'household' to represent the main agent of the model. The attributes of agents (households) include the following (lines 2–7):

- *happy?:* indicates whether an agent is happy or not
- *similar-nearby:* how many neighbours with the same colour as the agent
- *other-nearby:* how many neighbours with a different colour
- *total-nearby*: total number of neighbours.

There are two global[2] variables (lines 8–12): the first is *percent-similar,* which is the average percent of an agent's neighbours of its own colour. This variable gives a measure of clustering or segregation. The second variable, *percent-unhappy*, reports the number of unhappy agents in the model. There are another two variables determined by sliders (so that the model user can change their values on each run as desired): the number of agents, *number*; and agent's threshold, *%-similar-wanted* (which is the same for all agents).

The *setup* procedure (lines 14–30) (which is triggered when the user presses the *setup* button, see Fig. 8.4) creates a number of agents (households), half black and half white, at random positions. The *setup* procedure also calls another two procedures: *update-variables* that updates the agents' variables, and *do-plots* that updates the model's graphs (both procedures will be explained later).

---

[1] There are minor differences between the code of the original model in NetLogo's library and the code presented here.

[2] *Global* variables are defined (or declared) outside any procedure, and they can be accessed or refer red to from any place in the program. In contrast, *local* variables are defined inside a procedure, and can be accessed only within this procedure. The variables *similar-neighbors* and *total-neighbors* (lines 75–76) are local variables.

Box D: Code of Segregation Model

```
Breed [households household]
households-own[
  happy?      ;indicates whether the house-
   hold is happy or not
  similar-nearby ;how many neighbours with
   the same colour as mine?
  other-nearby ;how many neighbours with
   different colour?
  total-nearby ;sum of previous two variables
]
globals[
  percent-similar ;average percent of a
   household's neighbours
     ;of the same colour as that household?
  percent-unhappy ;percent of the households
   are 'unhappy'
]

to setup
  clear-all ;clear any variables or plots from
   previous runs
  if number>count patches
    [ user-message (word "This pond only has
   room for " count patches " households.")
      stop ]
  ;; create households on random patches.
  ask patches [set pcolor 7] ;; patches are
   initialized in grey
set-default-shape households "square"
  ask n-of number patches
    [sprout-households 1
      [ set color black ] ]
  ask n-of (number / 2) households
    [ set color white ]
  update-variables
  do-plots
end

to go
  if all? households [happy?] [ stop ] ;keep
   running as long as
   there                       ;are unhappy
   agents
```

(continued)

Box D: (continued)

```
  move-unhappy-households
  update-variables
  tick
  do-plots
end

to move-unhappy-households
  ask households with [ not happy? ];; only moves unhappy agents
    [ find-new-spot ] ; find new patch to move to
end

to find-new-spot
  rt random-float 360
  fd random-float 10
  if any? other households-here
    [ find-new-spot ] ;; keep going until we
   find an unoccupied patch
  move-to patch-here ;; move to center of patch
end

to update-variables
  update-households
  update-globals
end

to update-households
  ask households [
    ;; in next two lines, we use "neighbors" to test the eight patches
    ;; surrounding the current patch
    set similar-nearby count (households-on neighbors)
      with [color=[color] of myself]
    set other-nearby count (households-on neighbors)
    with [color !=[color] of myself]
    set total-nearby similar-nearby+other-nearby
    set happy? similar-nearby>= ( %-similar-wanted * total-nearby / 100 )
  ]
end
```

(continued)

Box D: (continued)

```
72 to update-globals
73   let similar-neighbors sum [similar-nearby]
     of households
74   let total-neighbors sum [total-nearby] of
     households
75   set percent-similar (similar-neighbors /
     total-neighbors) * 100
76   set percent-unhappy (count households with
     [not happy?]) / (count households) * 100
77 end
78
79 to do-plots
80   set-current-plot "Percent Similar"
81   plot percent-similar
82   set-current-plot "Percent Unhappy"
83   plot percent-unhappy
84 end
85
86
87
```

The dynamic process (which starts when the user presses the *go* button, see Fig. 8.4) is implemented using a simple behavioural rule for an agent in this model: *IF I'm not happy THEN I move to another place.* As the *go* procedure (lines 32–38) shows, the simulation will continue to run until all agents became happy with their neighbourhood (or the user forces it to stop).

The model provides two plots to present the two global variables *percent-similar* and *percent-unhappy* visually. Figure 8.4 shows the user interface and plots of the segregation model.

Stage 5: Verifying and validating
Like the previous traffic example, a simple verification method is to use extreme values for the model's parameters. For example, when setting the agents' threshold, *%-similar-wanted,* to zero and running the model, no agents move as they are all happy regardless of the percentage of neighbours of the same colour. On the other hand, setting this parameter to 100 makes most of the agents unhappy and they keep moving from their places.

Regarding validation, the main objective of the basic Schelling model is to explain an existing phenomenon rather than to replicate an existing segregation pattern in a specific city, and the model was successful in this regard. It provides a plausible answer to a puzzling question: *why these segregation patterns are so*

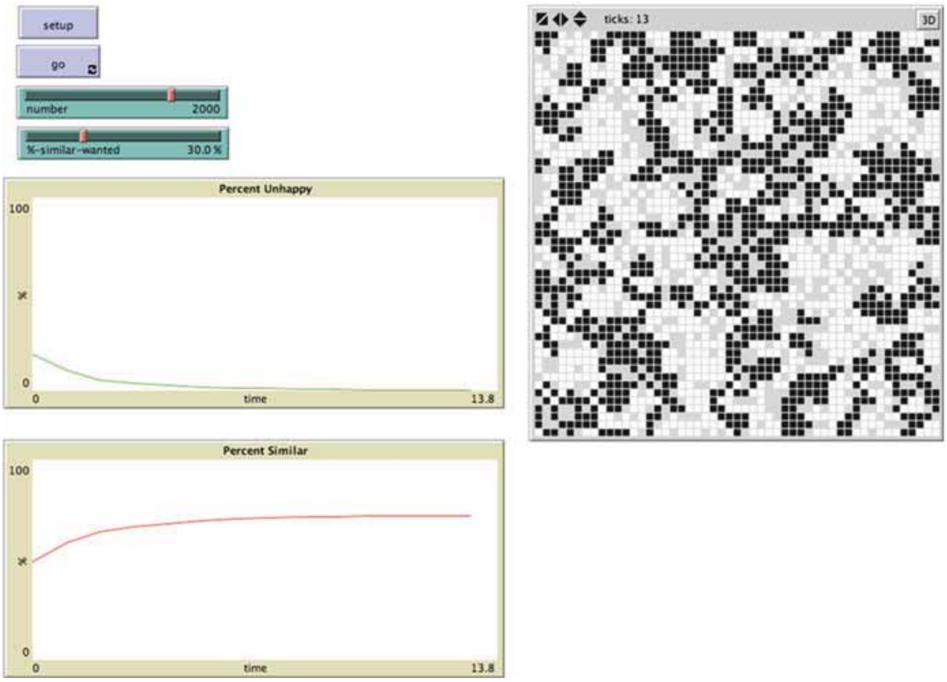


Fig. 8.4 User interface and plots of the segregation model

persistent regardless of the observed decline in ethnic prejudice. However, some attempts have been successful in extending the basic segregation model to replicate existing city segregation structures.

8.5 Conclusions

In this chapter, we discussed the process of designing and building an ABM. We recommended a set of standard steps to be used when building ABMs for social science research. The first, and the most important, step in the modelling process is to identify the purpose of the model and the question(s) to be addressed. The importance of using existing theories to justify a model's assumptions and to validate its results was stressed.

Recommended Reading

Edmonds, B., & Scott, M. (2005). From KISS to KIDS – An 'anti-simplistic' modelling approach. In P. Davidson, B. Logan, & K. Takadema (Eds.), *Multi-agent and multi-agent-based simulation. MABS, 2004*. New York/Berlin: Springer.

Epstein, J. M. (2006a). *Generative social science*. Princeton: Princeton University Press.

Epstein, J. M. (2008). Why model? *Journal of Artificial Societies and Social Simulation*, *11*(4), 12. Available at: http://jasss.soc.surrey.ac.uk/11/4/12.html

Gilbert, N., & Troitzsch, K. G. (2005). *Simulation for the social scientist*. Milton Keynes: Open University Press.

Gilbert, N. (2008a). *Agent-based models*. London: Sage.

Moss, S. (2008). Alternative approaches to the empirical validation of agent-based models. *Journal of Artificial Societies and Social Simulation*, *11*(1), 5. Available at: http://jasss.soc.surrey.ac.uk/11/1/5.html

References

Acerbi, A., & Parisi, D. (2006). Cultural transmission between and within generations. *Journal of Artificial Societies and Social Simulation 9*(1). Available at: http://jasss.soc.surrey.ac.uk/9/1/9.html

Allport, G. W. (1954). *The nature of prejudice*. Cambridge, MA: Addison-Wesley.

Balci, O. (1994). Validation, verification, and testing techniques throughout the life cycle of a simulation study. *Annals of Operations Research, 53*, 121–173.

Crooks, A. T., & Castle, C. (2012). The integration of agent-based modelling and geographical information for geospatial simulation. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 219–252). Dordrecht: Springer.

Crooks, A. T., & Heppenstall, A. J. (2012). Introduction to agent-based modelling. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 85–105). Dordrecht: Springer.

Epstein, J. M. (2006b). *Generative social science*. Princeton: Princeton University Press.

Epstein, J. M., & Axtell, R. (1996). *Growing artificial societies: social science from the bottom up*. Cambridge, MA: MIT Press.

Friedman-Hill, E. (2003). *Jesss in action: Rule-based systems in Java*. Greenwich: Maning.

Garson, D. G. (1998). *Neural networks: An introductory guide for social scientists*. London: Sage Publications.

Gilbert, N. (2008b). *Agent-based models*. London: Sage.

Gilbert, N., Pyka, A., Ahrweiler, P. (2001). Innovation networks: A simulation approach. *Journal of Artificial Societies and Social Simulation, 4*(3). Available at: http://jasss.soc.surrey.ac.uk/4/3/8.html

Gilbert, N. et al. (2006). Emerging artificial societies through learning. *Journal of Artificial Societies and Social Simulation, 9*(2). Available at: http://jasss.soc.surrey.ac.uk/9/2/9.html

Hutchins, E., & Hazlehurst, B. (1995). How to invent a lexicon: The development of sharedsymbols in interaction. In N. Gilbert & R. Conte (Eds.), *Artificial societies: The computer simulation of social life* (pp. 157–189). London: UCL Press.

Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology, 25*(6), 881–919.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence, 33*(1), 1–64.

Macy, M., & Willer, R. (2002). From factors to actors: Computational sociology and agent-based modeling. *Annual Review of Sociology, 28*, 143–166.

Massey, D. S., & Denton, N. A. (1988). The dimensions of residential segregation. *Social Forces, 67*(2), 281–315.

Ngo, T. A., & See, L. M. (2012). Calibration and validation of agent-based models of land cover change. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 181–196). Dordrecht: Springer.

Parry, H. R., & Bithnell, M. (2012). Large scale agent-based modelling: A review and guidelines for model scaling. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 525–542). Dordrecht: Springer.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology, 1*, 143–186.

Scott, J. (2000). *Social network analysis* (2nd ed.). London: Sage.

Sherif, M. (1966). *In common predicament; social psychology of intergroup conflict and cooperation*. Boston: Houghton Mifflin.

Sidanius, J., Pratto, F., van Laar, C., & Levin, S. (2004). Social dominance theory: Its agenda and method. *Political Psychology, 25*(6), 845–880.

Sun, R. (2006). The CLARION cognitive architecture: Extending cognitive modeling to social simulation. In R. Sun (Ed.), *Cognition and multi-agent interaction: From cognitive modeling to social simulation* (pp. 79–99). New York: Cambridge University Press.

Tajfel, H. (1981). *Human groups and social categories: Studies in social psychology*. Cambridge: Cambridge University Press.

Taatgen, N., Lbiere, C., & Anderson, J. (2006). Modeling paradigms in ACT-R. In R. Sun (Ed.), *Cognition and multi-agent interaction: From cognitive modeling to social simulation* (pp. 28–51). Cambridge: Cambridge University Press.

Terna, P. (1997). A laboratory for agent based computational economics. In R. Conte, R. Hegselmann, & P. Terna (Eds.), *Simulating social phenomena* (pp. 77–88). Berlin: Springer.

Wilensky, U. (1997a). *NetLogo traffic basic model*. Evanston: Center for Connected Learning and Computer-Based Modeling, Northwestern University. Available at: http://ccl.northwestern.edu/netlogo/models/TrafficBasic

Wilensky, U. (1997b). *NetLogo segregation model*. Evanston: Center for Connected Learning and Computer-Based Modeling, Northwestern University. Available at: http://ccl.northwestern.edu/netlogo/models/Segregation

Wilensky, U. (1999). *NetLogo*. Evanston: Center for Connected Learning and Computer- based Modeling, Northwestern University.

Wray, R. E., & Jones, R. M. (2006). Considering Soar as an agent architecture. In R. Sun (Ed.), *Cognition and multi-agent interaction: From cognitive modeling to social simulation* (pp. 53–78). New York: Cambridge University Press.

Ye, M., & Carley, K. M. (1995). Radar-Soar: Towards an artificial organization composed of intelligent agents. *Journal of Mathematical Sociology, 20*(2–3), 219–246.

Chapter 9
Modelling Human Behaviour in Agent-Based Models

William G. Kennedy

Abstract The modelling of human behaviour is not at all obvious. First, humans are not random. Second, humans are diverse in their knowledge and abilities. Third, besides being controlled by rational decision-making, human behaviour is also emotional. This chapter attempts to present principles driving human behaviour and reviews current approaches to modelling human behaviour.

9.1 Introduction

The behaviour of humans as individuals, in small groups, and in societies is the subject of several fields of research because it has such an important role in many aspects of daily life. However, incorporating human behaviour into Agent-Based Models (ABMs) is a real challenge, primarily because of the short history of our scientific observation of human behaviour, but there is hope. This chapter discusses the challenges of modelling human behaviour, presents and critiques the major approaches available along with some basic principles of human behaviour before providing information on how to integrate human behaviour into ABMs. The chapter starts with how not to model human behaviour.

9.2 How Not to Model Human Behaviour

To start, humans are not random. They (we) are strange and wonderful. Their behaviour may be unexpected or inconsistent (i.e., noisy), but it is not random. As an example, here is a simple demonstration. An easy question will be presented below

W.G. Kennedy (✉)
Krasnow Institute for Advanced Study, George Mason University, Fairfax, VA, USA
e-mail: wkennedy@gmu.edu

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_9,

and you may take hundreds of milliseconds to answer, but do answer. The question is: “Pick a number between one and four.” Have an answer?

The most common response is “three” and there is a secondary effect of this task: people feel a need to explain why they chose whatever answer they did. The second most common answer is “two”. Very few people decide to respond with either “one” or “four”. Sadly, there is not a serious study of this behaviour but undocumented sources suggest that the response statistics are close to 50% for “three”, 30% for “two” and about 10% for the other two answers.

The common explanation for the selection of “three” is that it was the most “interesting” number in the range. There is also a small number of people who are compelled to answer outside of the range, with fractions, or irrational numbers. These are rare occurrences. Similar results are obtained when the task is to pick a number between 1 and 20. The similarity is that people pick their most interesting number. For this range, the most common response is 17, occurring about 40% of the time, well above a “rationally”, “logically” expected 5%. Other primes are also favored as answers because they too are interesting.

This behaviour is interesting. The decision-making process should be simple, but it certainly does not appear to be a simple random selection among equally likely options. What this shows us is that people cannot even be random when they want to be. Further, if this task had been modeled as a uniform random distribution among equally likely choices, it would have been very different from actual behaviour.

Modeling human choices as uniform random distributions is making a very serious claim about human behaviour. It is saying that all choices are equally likely even when we know nothing about how people actually decide. It also assumes people have no preferences, do not consider the consequences of their actions, have no memory of previous choices, and can be more consistent than the data shows. Modeling human behaviour requires some data or some experience. Luckily, modelers are human and should know better.

9.3 Levels of Modelling Human Behaviour

The first question in an effort to model human behaviour is at what level the behaviour is to be modelled. The choices are basically at the individual level, at some small grouping of individuals, such as a household, and as a society. Modelling of a society can be done statistically, i.e., without dealing with individuals within the society. They could be inanimate particles because there is no effort to represent their decision-making process, only to describe what they have done. Small groups are typically modelled as if they were individuals and the science behind modelling individuals applies to small groups as well. This chapter addresses the modelling of individuals.

9.4 The Science Behind Modelling Human Behaviour

There appears to be at least two levels of sophistication in social organisms, "sociobiology" as E.O. Wilson termed it (Wilson 1975/2000). Social organisms such as slime moulds and social insects seem to be totally driven by inherited instincts that fully define their reactions to environmental stimuli. Social mammals, on the other hand, appear to have some degree of general problem-solving capabilities, such as a Theory of Mind, or in other words, their own model of other agents. This general capability results in the social behaviour of at least mammals being far more complicated than seems possible from a fixed set of inherited instincts. Humans and a majority of the great apes have many traits and resulting behaviours in common – see Wilson (1978/2004).

The study of human behaviour is as old as social primates themselves. A large part of social behaviour is the internal modelling of others for the purpose of knowing how to get along with them successfully. Prehistoric oral traditions have taught us how people supposedly behaved and the consequences of that behaviour (Stone 2011). The scientific study of how humans behave began less than 150 years ago with the advent of psychology as a modern scientific field – see James (1892/2001). The work is progressing, but due to the nature and complexity of the human mind, progress could be said to be slow.

In the mid-1950s, a cognitive revolution resulted in the research in behaviour changing from explaining all behaviour as simple stimulus-response associations to applying a new theory. The new theory was that behaviour could be explained in computational terms, but not simply via a "computer metaphor", i.e., literally like a computer, but a "computational theory of mind". This meant that the mind could be explained "using some of the same principles" as computers (Pinker 2002, p. 32).

One of the early concepts that has been both useful and distracting, is the metaphor of the brain as a computer (Newell and Simon 1972). It has been useful in providing a framework to understand the mind in terms of inputs, processes, and outputs. This reductionist approach has led to advances in understanding the modular organization of the mind and the brain (Anderson 2007). However, our focus on the von Neumann computer architecture, i.e., a separate memory and processor, which operate serially, has resulted in a symbol vs. connections debate (Anderson 2007). Neural network approaches to modelling cognition is an ongoing research area, but such systems are difficult to build and it has been difficult to make steady, incremental progress.

The pursuit of modelling or replicating human behaviour has developed two camps: Artificial Intelligence and Cognitive Science. The work in AI is aimed at replicating the intelligent behaviour of humans and surpassing human intelligence when possible, as in mathematics from arithmetic to calculus. However, most AI researchers have little interest in replicating the all too human errors or unintelligent behaviour observed in nature. On the more psychological side, researchers in

Cognitive Science seek to understand human cognition in all of its forms, rational as well as emotional, intuitive, and erroneous. Both approaches have developed methods and techniques that can be useful in modelling human behaviour.

Focusing on the rational and analytic side of human cognition has generated the largest amount of research in this area and significant progress had been made, e.g. see Kahneman (2003). There has been far less research on other behavioural drivers such as intuition or emotions, but research is growing in this area – see Damasio (1994).

9.5 Basic Principles

In this section, a set of basic principles of human behaviour is provided. These principles are focused on the causes of human social behaviour, not the behaviour of individuals alone or over very short periods.

9.5.1 Humans as Information Processors

Humans process sensory information about the environment, their own current status, and their remembered history to decide what actions to take. However, their environmental sensors are limited in type to the traditional five senses (touch, sight, hearing, taste, and smell). Humans can also sense temperature, internals (kinesthetic or proprioception), pain, balance, and acceleration. Each has a range and a minimum sensitivity and duration threshold.

Humans also have diverse personality traits. These are characteristics that effect the thoughts, behaviour and emotions that they are born with, which seem to be relatively constant over a life span, and that are a large part of individual differences. Traits are intended to be relatively independent and seem to have normal distributions with large populations. There are two taxonomies of personality traits known as a three-factor model (Eysenck 1967/2006) and a five-factor model (McCrae and Costa 1987). Both share two traits: extraversion (sociability) and neuroticism (tendency toward emotional behaviour). Other potentially important traits associated with social behaviour include agreeableness, risk avoidance, and impulsivity.

Taken together, humans as information processing systems have a limited informational input bandwidth, limited memory, and limited processing capability. However, because humans have language, their information sources can be very wide, and with written language, they can have memories spanning centuries.

9.5.2 Human Motivations

A very highly cited 1943 paper on human motivation provided an organization of human motivations into a “Hierarchy of Needs” (Maslow 1943). This ordering is not rigid but has survived intact over the years. Maslow proposed that humans’ first need is to meet their basic physiological requirements. After these are adequately met, the next priority is for safety and security. When these are adequately addressed, the next priority is the social needs of friendship, family, and sexual intimacy. The last two layers deal with external esteem and self-actualization. This hierarchy is useful in ordering potentially competing priorities of agents representing humans in ABMs.

9.5.3 Humans Behaving Rationally

Human behaviour is commonly thought of as being rational. Rational Choice Theory (Coleman 1990) is based on the presumption that humans behave in ways to maximize their benefits or minimize their costs, and in either case, follow logical processes. This approach typically assumes all possible actions are known, all agents have perfect knowledge of the environment, and that the preferences of agents are well behaved, i.e., have necessary ordering and transitivity properties. Tempering this approach is the idea that agents have "bounded rationality", i.e., have limited information, limited cognitive abilities, and limited time to make decisions (Simon 1996). In addition, there may be limitations as to how many variables humans can process and how mathematically sophisticated the evaluation of those variables are in order to determine their rational behaviour. Although many forms of knowledge representation are possible, the representation of human knowledge is generally accepted to be in two basic forms: declarative knowledge of facts and procedural knowledge typically represented in IF-THEN rules (Newell 1990; Anderson 2007). Rational behaviour also includes learning of declarative knowledge, and new procedural knowledge in some cases. How long knowledge is retained varies from systems that never forget knowledge to systems that have very little memory for either form of knowledge. Clearly, systems of human behaviour need to have some memory, but how much and how formally it is modelled depends on the purpose of the system. Therefore, a rationally behaving model needs to be able to represent knowledge, learn, remember new knowledge, and apply that knowledge to determining the behaviour of the agent.

9.5.4 Humans Behaving Emotionally/Intuitively/Unconsciously

In addition to being rational beings, humans have other factors that affect their behaviour. These include emotional, intuitive, or unconscious decision making processes. The representation of human behaviour in ABMs may need to include these other decision-making processes. Research in emotions and the effect of emotions on decision-making is taken in this discussion as the leading representative of the non-traditionally rational decision-making processes.

There is evidence of a common set of basic emotions: interest, joy, happiness, sadness, anger, disgust, and fear (Izard 2007). These emotions are considered evolutionarily very old and have neurobiological bases. They are generally infrequent, short lived, and do not directly affect cognition. However, emotions can lead to longer-term moods and result in complex behaviour.

There have been many studies of emotions but the relation of emotion to cognition, and therefore to behaviour is a highly debated topic in psychology (e.g., LeDoux 1995). Whether emotions are modifiers of the rational decision-making process or a separate mental process is not yet settled. Kahnemann (2003) discusses a System 1 and System 2 approach to dual cognitive processes. The predominant theory of emotion is Appraisal Theory (Scherer 1999).

The appraisal theory poses that there are a fixed set of dimensions of factors needed to determine the emotional status of an individual. However, there is wide variance of thought on what the dimensions are. Progress is being made and repeatable results are starting to produce interesting results (Scherer 1999).

Although it may seem natural to presume humans behave to maximize their expected emotions, the effect of emotions on decision-making can be more richly discussed (Loewenstein and Lerner 2003). Emotions can alter rational decision-making by distorting the agent's perceptions of the environment and the likelihood of future evaluations. Loewenstein and Lerner (2003) offer two limitations concerning the impact of emotions on decision-making. First, some behaviour is not the result of decision-making and can be the result of emotional drivers directly. Second, the impact of emotions on decision-making cannot be easily classified as improving or degrading the rational decision-making process.

9.5.5 Humans Behaving Socially

As social beings, the behaviour of individuals is shaped by input from others in two basic ways. First, humans have a Theory of Mind by which they imagine what others have as their goals and what they are thinking and feeling (Dunbar 2004). Second, human behaviour is influenced by and combines with the behaviour of others (Latané 1981; Friedkin and Johnsen 1999; Surowiecki 2005; Kennedy and Eberhart 2001).

A Theory of Mind supports the transference of information based on establishing and sharing common concepts among agents, i.e., language. The exchange of information and goods and services among groups of agents then provides for the development of culture and economies within and among societies.

Latané proposed a formulation of social influence based on experiments where a group attempts to influence a human subject (Latané 1981). The relationship he found was of the form:

$$I = s\,N^{t} \tag{9.1}$$

where I is the influence in terms of the percentage conforming or imitating behaviour in the subject, s is a constant associated with the circumstances, N is the number of others involved, and t is a factor less than one and often near one half. However, this influence also inhibits action by, in a sense, distributing the social responsibility to act such that a social inhabitation to act by bystanders has been found (Latané 1981). Extending the study of influence, Friedkin and Johnsen (1999) reported on the influence of a group's members on each other and the result can be that the group settles on the group's mean, a compromise different from the mean, on the position of an influential member of the group, or may not form a consensus.

Groups can also develop results greater than those of any of the individuals. Groups of diverse people independently making evaluations with an appropriate method of bringing their results together can have this kind of result (Surowiecki 2005). He explored conditions that resulted in good collective results and found that they result from the differences in the evaluations among group members, not compromises or achieving consensus. This appears to be another outgrowth of social influences, which can lead to conformity, a lack of independence, and then poor results. For example, he reports that in a crowd, due to diversity, there will be some willing to riot, some who would never riot, and many that will decide based on social influences.

This section has attempted to identify the basic principles of human behaviour. They are intended to be the causes of human social behaviour, not the behaviour of individuals. Of course, this is incomplete, possibly wrong, and the subject of much research. The next section addresses current approaches in applying this knowledge to modelling human behaviour.

9.6 Current Approaches

Although this book is about ABMs, within an ABM, the representation of the cognition driving a modelled human's behaviour can have its own internal architecture. A cognitive architecture (Newell 1990) is the structure and functionality that is unchanging throughout the simulation and supports the cognitive model that drives behaviour. There are several cognitive approaches to consider. For presentation here they are grouped as: (1) *ad hoc* direct and custom coding of behaviours mathematically in the simulation's programming language; (2) conceptual frameworks to be implemented within the target system; and (3) research-quality tools for modelling the cognitive functioning of an individual at the millisecond scale.

9.6.1 Mathematical Approaches

Mathematical approaches to modelling human behaviour are methods that produce agent behaviour through the use of mathematical simplifications. First among these, and the most severe simplification, is the use of random number generators to select between predefined possible choices. The fallacies of this approach were addressed at the beginning of this chapter, and includes that people are not random, that random number generation is not a replacement for unknown quantities, and that using a random number generator is making very strong and very wrong claims about human behaviour.

Better than relying on random number generators would be to directly code threshold-based rules. These are of the form that when an environmental parameter passes a threshold, a specific human behaviour would result. This would provide simple behaviour, but they would be explainable and could approximate human behaviour. The parameter could be transformed so that the action is taken when the transformation of the parameter is above, below, or between thresholds.

Using a threshold is equivalent to comparing two values in that the difference in the two values can be compared to a threshold. For example, if the intent is to compare function1 with function2, this is the same as comparing (function1 – function2) and a threshold value of 0. For instructional purposes, all sample rules presented here are in the form of a function compared to a threshold. Variables and functions are descriptively named between "<>" and actions are in italics.

As an example:

IF<hunger>is below<hungerThreshold1>THEN *agent-dies*.
IF<hunger>is above<hungerThreshold2>THEN *address-another-goal*.
IF<hunger>is between<hungerThreshold1>and<hungerThreshold2>
THEN *search-for-food*.

Another mathematical approach is the use of multi-dimensional functions of parameters rather than comparing a single environmental parameter to a threshold. Here, several parameters are combined to define a modelled human's behaviour. The major weakness in this approach is that available data does not validate humans as pure optimizing agents.

IF<evaluation of<hunger>&<thirst>>is above thresholdHT
THEN *focus-on-safety-issues*.

Finally, Dynamic Modeling (Hannon and Ruth 1994) represents human decision-making as "stocks and flows" or, in a sense, as a hydraulic system with pipes, tanks, valves, and pumps. The representational sophistication of this modelling approach is that the rate of change of a variable can be a function of its own magnitude. Such a model uses differential equations to describe relationships in the model. The hydraulic theory of emotion can be traced back to René Decartes (1596–1650) (Evans 2001). An example is:

IF<anger>is above<ventThreshold>THEN *act-to-vent-anger*.

These mathematical approaches to modelling human behaviour rely on a simplification of the perception, reasoning, and actions important to the purpose of the model. For many models, the vast majority of the human behaviour is not of interest to the model and the behaviour of interest can be reasonably well specified. If more general behaviour is important to the modelling effort, a more general approach may be appropriate.

9.6.2 Conceptual Frameworks

Conceptual frameworks are approaches to modelling human decision-making using more abstract concepts than mathematical transformations of environmental parameters. They involve concepts such as beliefs, desires, and intentions (BDI), emotional state and social status (PECS), and "fast and frugal" decision hierarchies. Three conceptual frameworks will be addressed.

The first approach is based on beliefs, desires, and intentions (BDI) (Rao and Georgeff 1995). The BDI approach is a theoretical framework based on the idea that human behaviour can be modelled by representing an individual's beliefs, desires, and intentions. Beliefs are the individual's knowledge about the world, i.e., the world as they perceive it to be. Desires are the individual's motivation, i.e., its goals. Intentions are the agent's deliberative states. A BDI implementation develops a decision tree and this complete decision tree is transformed into a possible worlds model from which a deliberation process decides the best course of action. The BDI framework is very general and can be realized in many ways. Its weakness is that it is so general that it provides little more than a conceptual framework for thinking about how to model the human cognition behind behaviour. The next framework is more specific and provides more guidance for implementing a model of human behaviour within an agent-based system.

The second framework involves physical, emotional, cognitive, and social factors (PECS) affecting behavioural decisions (Schmidt 2002). This framework includes a representation of the human mind, specifically perception and behaviours, and mathematical representations of physiology, emotion, cognition, and social status. Within cognition are mathematical transformations for a self-model, an environmental model, memory for behaviour protocols, planning, and reflection. The declared purpose of the PECS framework is to replace the BDI framework, and it is more specific and implemented. The PECS framework can represent simple stimulus-response behaviours and more complex behaviours that involve the determination of drives, needs, and desires and their transformation into motives. Motives, depending on their intensity, are state variables that indirectly determine behaviour. Advantages of this framework are that behaviours can be explained in terms of their causes in a reasonably plausible manner. Two challenges for this framework are the internal parameters for the mathematical transformations of environmental parameters into the internal state variables and the combination, prioritization, and integration of the various motives into the selected behaviour.

The third framework is called "fast and frugal" and was developed by analyzing data on human decisions. Gigerenzer (2007) reported on the analysis of how intensive care units make decisions about whether a patient is having a heart attack and how two judges evaluate court cases and make decisions on whether to grant bail for defendants. The analysis in both cases identified three sequential questions that could be answered by environmental variables, and the use of these "fast and frugal" trees performed very well compared to human decision-making. In the medical case, the decision tree developed for a U.S. hospital performed better than the heart disease predictive instrument or physicians, and the decision trees explained 92% of the two UK magistrates' decisions (Gigerenzer 2007). The design of these rules in these trees is not aimed at identifying all the variables to justify implementation of a particular behaviour, but an attractive characteristic of this framework important to ABMs is that these decision trees are inexpensive computationally and should scale up well to large numbers of agents.

These three frameworks are different approaches to modelling human behaviour at a level of rigor between the pure mathematical representations and full, research quality models of human cognition. The third level, research-quality models are tools intended for use usually in representing the cognitive decision-making of individuals.

9.6.3 Cognitive Architectures

A third approach is to use research tools developed for a purpose different from agent-based modelling for social simulation. Their purpose is research into abstract or theoretical cognition on the one hand and understanding human cognition on the other. This section discusses Soar, ACT-R, and other architectures. These are architectures in the sense that the basic system is unchanging throughout the use of the system. Cognitive models of specific tasks are implemented within these cognitive architectures. Such a cognitive model can be used to drive the human behaviour of an ABM.

Soar (Lehman et al. 2006) is an Artificial Intelligence system originally based on matching human performance in problem-solving tasks at a symbolic level of granularity and is the basis of Newell's proposal for Unified Theories of Cognition (Newell 1990). As an AI system, its purpose is to meet or exceed human performance on a wide variety of tasks. The Soar system could be considered to be an implementation of a BDI architecture in that it maintains an internal representation of the world, is always working to solve a goal, and has available internal state variables. Soar has a long history of modelling human behaviour framed as problem solving in research settings and for commercial customers. Although a stand-alone system, Soar has been connected to several other environments including games. A Soar model consists of a collection of rules written as text that uses environmental or internal variables and either changed internal variables or takes an action that changes the environment. The system, which includes demonstration models, is available at no cost from the Soar website, http://sitemaker.umich.edu/soar/home. There is also a Java based version being developed at http://code.google.com/p/jsoar/. There is an active Soar community, it offers training on using Soar, and 40–60 members meet annually.

ACT-R (Anderson and Lebiere 1998; Anderson et al. 2004), which most recently stands for Atomic Components of Thought-Rational, has been used in basic research in cognition for many years. ACT-R provides architecture assumptions based on both symbolic and sub-symbolic representations of knowledge. Over the years, ACT-R has evolved into a comprehensive cognitive architecture demonstrating successful models of many cognitive phenomena and it has been linked to the functional regions of the brain (Anderson 2007). Successful here means closely matching human performance data. However, ACT-R is focused on relatively low-level cognitive phenomena operating over very short time periods. It does not support higher-level concepts such as beliefs, desires, or intentions explicitly. An ACT-R model consists of a collection of declarative facts and rules written as text that uses environmental or internal variables and either changed internal variables or initiates actions in the environment. ACT-R is also available at no cost and has an active community supporting it. Courses on using ACT-R are offered in Europe and the United States annually and the community meetings of 40–60 people also occur approximately annually. Their home page is: http://act-r.psy.cmu.edu/. There is also a Java version of ACT-R in development and use: http://jactr.org/, which has been connected to and operates a mobile robot – see http://www.nrl.navy.mil/aic/iss/.

There are other cognitive architectures used in research. Several are reviewed in the National Research Council report (Zacharias et al. 2008). However, none of these other symbolic architectures have the wide acceptance and active community that Soar and ACT-R have.

9.7 Challenges in Modelling Human Behaviour

There are at least three challenges in the efforts to model human behaviour in agent-based systems: understanding humans, data, and validation & verification. As should be obvious, although human behaviour has been noticed for thousands of years and scientifically studied for a couple of hundreds of years, there is still much unknown. The genetic, historical, and current environmental factors affecting the behaviour of such diverse agents as humans may appear incomprehensible, but progress is being made and will continue. Research continues to develop data on how people behave under certain circumstances and this is replacing the poor default of assuming that human behaviour is random and unknowable. However, data for many or most behaviours of interest to the ABM community may not yet exist. The lack of data makes validation and verification of models of human behaviour difficult, at best. However, as humans are the ones constructing ABMs of human behaviour, hopefully, some knowledge, some generally accepted practices, and a good dose of common sense will result in good models of human behaviour.

Additional Resources

While research and the practice of modelling human behaviour continues, there are sources supporting this effort. The U.S. Air Force asked the U.S. National Research Council to provide “advice on planning for future organizational modelling research”

(Zacharias et al. 2008, p. 1). The resulting report provides an excellent review of the state of the art, although a criticism is that it does not adequately address work outside the U.S.

Current research and results in agent-based modelling is presented in scientific conferences held regularly. In the United States, the Behaviour Representation in Modeling and Simulation Society meets annually to present and discuss current work. Their website is: http://brimsconference.org/. In Europe, the European Council for Modelling and Simulation meets annually and their web site is: http://www.scs-europe.net/.

References

Anderson, J. R. (2007). *How can the human mind occur in the physical Universe?* Oxford: Oxford University Press.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah: Lawrence Erlbaum Associates.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglas, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of mind. *Psychological Review, 111*(4), 1036–1060.

Coleman, J. S. (1990). *Foundations of social theory*. Cambridge, Massachusetts, USA.

Damasio, A. (1994). *Decartes' error: Emotion, reason and the human brain*. New York: Penguin.

Dunbar, R. (2004). *The human story*. London: Faber & Faber.

Evans, D. (2001). *Emotion: The science of sentiment*. New York: Oxford University Press.

Eysenck, H. J. (1967/2006). *The biological basis of personality*. Springfield: Charles C Thomas.

Friedkin, N. E., & Johnsen, E. C. (1999). Social influence networks and opinion change. *Advances in Group Processes, 16*, 1–29.

Gigerenzer, G. (2007). *Gut feelings: The intelligence of the unconscious*. New York: Penguin.

Hannon, B., & Ruth, M. (1994). *Dynamic modeling*. New York: Springer.

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science, 2*(3), 260–280.

James, W. (1892/2001). *Psychology, the briefer course*. Mineola: Dover Publications.

Kahneman, D. (2003). A perspective on judgment and choice. *American Psychologist, 58*, 697–720.

Kennedy, J., & Eberhart, R. (2001). *Swarm intelligence*. Waltham: Morgan Kaufmann Academic Press.

Latané, B. (1981). The psychology of social impact. *American Psychologist, 36*(4), 343–356.

LeDoux, J. E. (1995). EMOTION: Clues from the brain. *Annual Review of Psychology, 46*, 209–235.

Lehman, J. F., Laird, J., & Rosenbloom, P. (2006). A gentle introduction to Soar, an architecture for human cognition: 2006 update. URL: http://ai.eecs.umich.edu/soar/sitemaker/docs/misc/GentleIntroduction-2006.pdf

Loewenstein, G. F., & Lerner, J. S. (2003). The role of affect in decision making. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective science* (pp. 619–642). Oxford: Oxford University Press.

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review, 50*, 370–396.

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*(1), 81–90.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs: Prentice-Hall.

Pinker, S. (2002). *The blank slate*. New York: Viking Penguin.

Rao, A. S., & Georgeff, M. P. (1995). *BDI-agents: From theory to practice*. In *Proceedings of the First International Conference on Multi-Agents-Systems*. Menlo: AAAI Press.

Scherer, K. R. (1999). Appraisal theory. In T. Dalgleish & M. Power (Eds.), *Handbook of cognition and emotion* (pp. 637–663). New York: Wiley.

Schmidt, B. (2002). *Modelling of human behaviour the PECS reference model*. In A. Verbraeck & W. Krug (Eds.) *Proceedings of the 14th European Simulation Symposium*, SCS Europe BVBA.

Simon, H. (1996). *The sciences of the artificial* (3rd ed.). Cambridge: MIT Press.

Stone, R. (2011). Red in tooth and claw among the Literati. *Science, 332*, 654–657.

Surowiecki, J. (2005). *The wisdom of crowds*. New York: Anchor Books.

Wilson, E. O. (1975). *Sociobiology: The new synthesis*. Cambridge, MA: The Belknap Press.

Wilson, E. O. (1978/2004). *On human nature*. Cambridge, MA: Harvard University Press.

Zacharias, G. L., MacMillan, J., & Van Hemel, S. B. (Eds.) (2008) *Behavioral modeling and simulation: From individuals to societies*. Committee on Organizational Modeling from Individuals to Societies. Washington, DC: The National Academy Press.

Chapter 10
Calibration and Validation of Agent-Based Models of Land Cover Change

The An Ngo and Linda See

Abstract This chapter considers two important issues in the development of agent-based models, i.e. calibration and validation. These terms are defined and framed into a step-by-step process. Each step is then explained in further detail and illustrated using an agent-based model of shifting cultivation developed by Ngo (2009) as part of his PhD research project. Although the process of model validation presented here is applicable to agent-based models in general, some of the finer details are more relevant to agent-based models of land use and land cover change.

10.1 Introduction

Model validation is a process for determining if a model is able to produce valid and robust results such that they can serve as the basis for decision makers (Berger et al. 2001). The validation process provides the information needed to assess how well the model approximates the real world system and meets the original objectives of the model development. Before the outputs of a model are validated, there is a calibration process whereby the model parameters are determined using real world data. Together both calibration and validation represent one of the seven challenges of agent-based modeling (Crooks et al. 2007). One of the main reason for this

The An Ngo (✉)
Hanoi University of Agriculture, Trau Quy, Gia Lam, Hanoi, Vietnam
e-mail: ntan@hua.edu.vn

L. See
IIASA, Laxenburg, A-2361 Austria
e-mail: see@iiasa.ac.at

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_10, © Springer Science+Business Media B.V. 2012

challenge is that the concepts related to validation are still being debated, and conflicts remain in the way that validation terminologies are used (Carley 1996; Crooks et al. 2007; Troitzsch 2004). Moreover, the different techniques for validation are quite varied, which has led to a confusing situation for modellers. Therefore, it is important to have a systematic approach to the overall validation process, and one that is integrated throughout the development phase of an agent-based model (ABM). This chapter attempts to provide such an approach to ABM validation.

Numerous publications have been devoted to reviewing different validation methods for ABMs (Berger et al. 2001; Carley 1996; Klügl 2008; Parker et al. 2002; Troitzsch 2004; Windrum et al. 2007). Among these, several types of validation are mentioned, e.g. empirical validation, statistical validation, conceptual validation, internal validation, operational validation, external validation, structural validation and process validation. However, Zeigler (1976) provides a good characterization of these methods into three main types:

- **Replicative validation:** where model outputs are compared to data acquired from the real world;
- **Predictive validation:** where the model is able to predict behaviour that it has not seen before, e.g. that which might come from theories or which might occur in the future; and
- **Structural validation:** where the model not only reproduces the observed system behaviour, but truly reflects the way in which the real system operates to produce this behaviour.

In this chapter, the focus is on structural validation, which in broad terms consists of the following four processes as defined below (Carley 1996; Klügl 2008):

- **Face Validation:** is often applied at the early phase of a simulation study under the umbrella of conceptual validation. This technique consists of at least three methodological elements:
  - **Animation assessment:** involves observations of the animation of the overall simulated system or individual agents and follows their particular behaviours.
  - **Immersive assessment:** monitors the dynamics of a particular agent during the model run.
  - **Output assessment:** establishes that the outputs fall within an acceptable range of real values and that the trends are consistent across the different simulations.
- **Sensitivity Analysis:** assesses the effect of the different parameters and their values on particular behaviours or overall model outputs.
- **Calibration:** is the process of identifying the range of values for the parameters and tuning the model to fit real data. This is conducted by treating the overall model as a black box and using efficient optimisation methods for finding the optimal parameter settings.
- **Output Validation:** involves graphically and statistically matching the model's predictions against a set of real data.

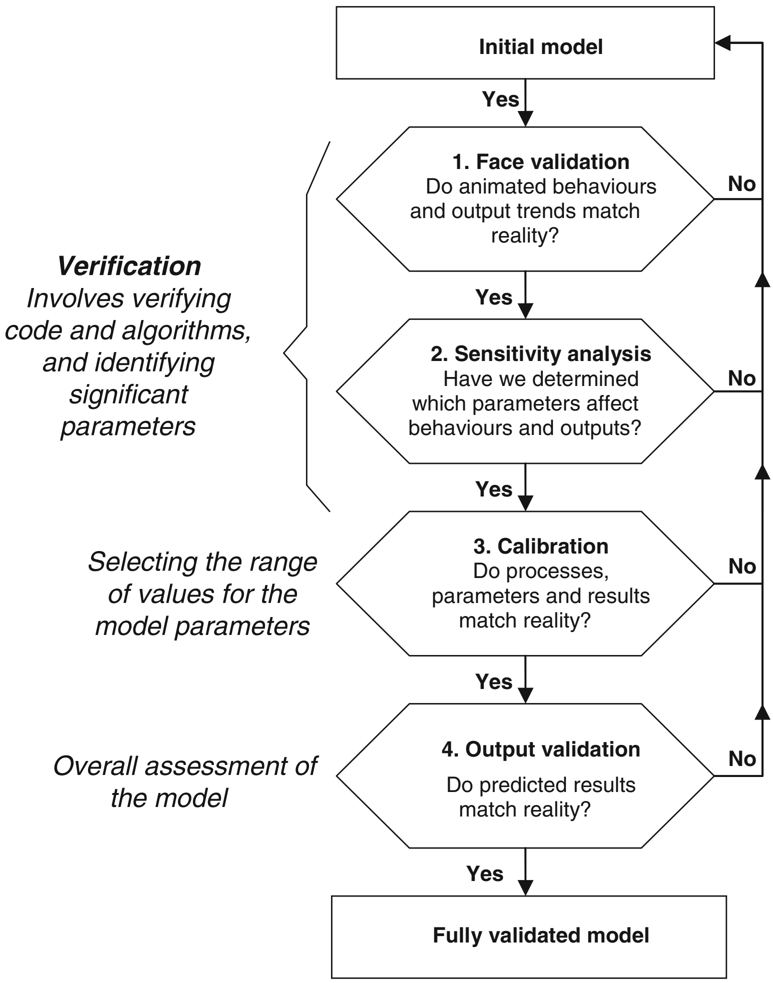


Fig. 10.1 General validation process of an ABM

Face validation and sensitivity analysis are sometimes collectively referred to as verification (Parker et al. 2002). The different processes above are often carried out iteratively in a step-by-step process as illustrated in Fig. 10.1.

A model is able to generate reliable and valid results within its experimental frame only if these validation processes are wholly implemented. However, there are very few examples of where comprehensive system validation has been applied to ABMs. For land use and land cover change modeling in particular, many studies have only concentrated on *output validation* (e.g. Castella and Verburg 2007; Jepsen et al. 2006; Le 2005; Wada et al. 2007) whereas the other steps mentioned above have not been treated explicitly. Therefore, the results may not truly reflect the way the system operates as per the definition of structural validation provided earlier.

The rest of the chapter discusses each of the stages in the validation process (Fig. 10.1) in more detail, providing examples from an ABM of shifting cultivation (SCM) as described in Ngo (2009) and Ngo et al. (2012).

10.2 Verification of ABMs

Verification is the process whereby the logic of the model is tested for acceptability and validity. Basically the model is checked to see if it behaves as it should. Crooks et al. (2007, p. 10) refer to this as testing the "inner validity" of the model. Verification often involves examining processes within the model and then comparing the model outputs graphically or statistically against the real data. However, the level of detail needed for verification is less than that required for calibration (Carley 1996).

As defined previously in Fig. 10.1, model verification consists of face validation together with the sensitivity analysis. Face validation is conducted to ensure that the processes and initial outcomes of the model are reasonable and plausible within the basic theoretical framework of the real system. Sensitivity analysis, on the other hand, is applied to examine the effect of the model parameters on the outcome of the model. Parameters with no significant effect are then removed from the model to make it more coherent and easier to operate. The sensitivity analysis is, therefore, necessary in the pilot phase of complicated simulation studies as the parameters that are identified as being important are those that will require calibration or identification using optimisation or some other means.

10.2.1 Face Validation

Face validation should be applied to several aspects of the model in its early development phase. The dynamic attributes of the agents can be analysed visually across many iterations of the model. All behaviours such as those used for identifying the relationships between agents, and the automatic updating of related parameters are checked for consistency and accuracy. These processes are essentially the *animation* and *immersive* assessments referred to in Sect. 10.1, which can be undertaken in a visual and qualitative way.

A simple example of visual validation is demonstrated in Fig. 10.2, which has been conducted for the SCM of Ngo (2009). Figure 10.2 shows the results of the dynamic monitoring of a random household agent with their relatives over time. As time increases (on an annual time step), the household characteristics of the agent are updated gradually from a state when the household was young to when the first partitioning occurs and the first son marries, forming a new household. Replacement by the second son then takes place when the head of household agent dies to form a new household. Visual analyses like these were used to determine whether the SCM (Ngo 2009) was able to produce acceptable results when simulating real human relationships in a shifting cultivation system.

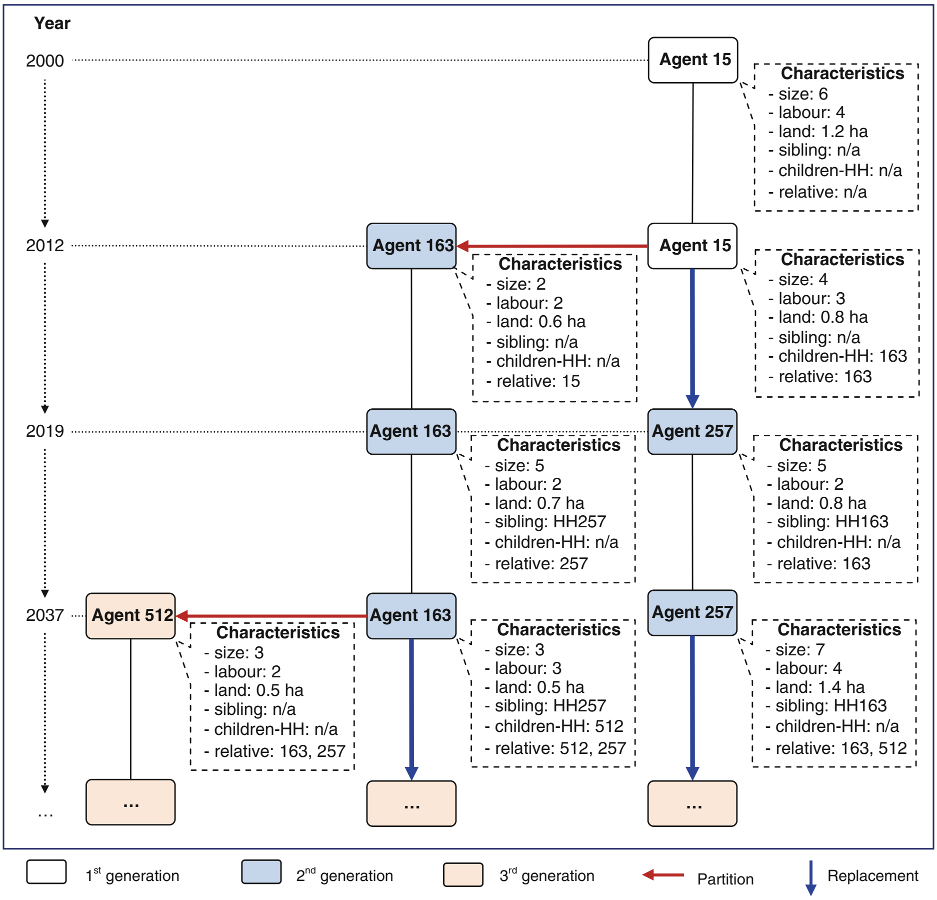


Fig. 10.2 Dynamic monitoring of selected household agents over time in the shifting cultivation model (Ngo 2009)

The second part of the face validation process relates to output assessment in order to ensure that the simulated results fall within an acceptable range of real values across the simulations. The simulated results might include the important parameter values which were used to describe an agent's characteristics. The analyses can be conducted as follows. Firstly, the model is run several times (where all inputs are held constant) in order to generate the initial outputs related to the characteristics of the agents. The number of runs should be sufficiently large as to be statistically significant (e.g. 30). These data were then analysed visually to ensure that they fell within the range that corresponds to the real world (based on a comparison with survey data obtained from fieldwork).

A statistical comparison between the data from the simulated runs and the real data is shown in Table 10.1. In terms of the statistical distribution, it is important to check for Standard Errors (SE) and compare the mean values of the simulated results with real world values to ensure that the model can provide consistent results. Once

Table 10.1 Household data from the model simulation and the survey data collected in 2007 (Ngo 2009)

	Model outputs in 2007			Survey data in 2007		
	Mean	SE	SE (%)	Lower bound	Upper bound	Mean
Age of household heads (year)	40	0.149	0.37	39	45	42
Household size (# of people)	6.27	0.054	0.89	6.17	7.54	6.86
Household labour (# of labourers)	3.70	0.047	1.27	3.31	4.34	3.83
Land per capita (ha)	0.18	0.000	0.00	0.18	0.21	0.19

the simulated results appear to be consistent (e.g. SE < 5%), their mean values can be then compared with the ranges of the real data, which is often indicated by the lower and upper bounds in statistical terms.

The simulated data in Table 10.1 shows that the model output results have SEs of less than 5% compared to the mean values, indicating that the results are consistent and can therefore be compared with the findings from the survey. The mean values of the model outputs fall within the upper and lower bounds of the survey data, which confirms that the SCM can produce household characteristics that are similar to the survey data.

Another assessment of the output within the face validation framework is to check how consistently the model can produce the same or similar outcomes between the different model runs. There are several ways to do this but the *Test for Homogeneity of Variances* (Winer 1971) is one possible approach. In practice, we might measure the variances of the simulated results for several time steps (i.e. t, t+1, t+2, t+3, t+n) with several replications. If the hypothesis is accepted, i.e. the variations between model runs are similar, then the model would pass this test.

Regarding ABMs related to land use and land cover change analysis, it is also important to compare output values from model runs produced at different scales. Since the level of detail is reduced at lower resolutions, there will most likely be some difference between the model outputs run at varying scales. However, if this difference is not statistically significant, then the model could be run at the coarser scale to reduce the running time of the model. This reduction in computational time could be very significant if the model is applied to a large area.

10.2.2 Sensitivity Analysis

In an ABM context, sensitivity analysis is often carried out to examine the effect of input parameters and their values on model behaviours and model outputs. This analysis is essential for selecting significant parameters for the simulation before the model is calibrated or used in scenario analysis. A common approach is to modify only one parameter at a time, leaving the other parameter values constant (Happe 2005). However, this approach is not as easily applicable to agent-based systems (Manson 2002) and sensitivity analysis has often been undertaken in an unstructured way (Kleijnen et al. 2003). In order to avoid oversimplification of the underlying model due to leaving out possible interactions between input parameters, Kleijnen et al. (2003) and Happe (2005) have suggested that the sensitivity analysis should be conducted systematically by applying the statistical techniques of Design of Experiments (DOE) and metamodelling (Box et al. 1978; Kleijnen and Van Groenendaal 1992).

The suitability of DOE techniques in the context of ABMs has been recognised previously as it can help to determine the importance of the input parameters and also provide information about model behaviour and the logic employed in the programme (Happe 2005; Kleijnen et al. 2003). In DOE terminology, model input parameters are called factors, and model output measures are referred as to responses. A full factorial design consists of i factors, with an assumption that each factor takes j levels and therefore involves $n = i^{j}$ factor setting combinations. This means that n simulations are required to determine the effect of i factors. However, this procedure can only be applied to a small number of factors because the computation time increases exponentially with each additional factor and each additional factor level (or categories in which each factor is divided). It is obvious that alternative methods are therefore necessary to undertake a sensitivity analysis of the model if the number of factors is large.

To deal with the computational problem due to the large number of factors, Bettonvil and Kleijnen (1997) proposed the Sequential Bifurcation (SB) technique which is essentially a method to determine the most important factors among those that affect the performance of the system.

SB operates with three assumptions: (*i*) the importance of factors to the model performance can be approximated as a first-order polynomial; (*ii*) the sign of each factor effect is known; and (*iii*) errors in the simulation model are assumed to be zero. The overall procedure can be described as follows. Firstly, the analysed parameters are converted to binary variables with values of 0 or 1, which correspond to low and high simulation outputs, respectively. The simplest approximation of the simulation model output y is a first-order polynomial of the standardised variables $(x_1,\ldots,x_j,\ldots,x_K)$, which has main effects β_j and overall mean β_0, and can be expressed as:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_j x_j + \cdots + \beta_K x_K \tag{10.1}$$

The manner of the variable standardisation mentioned above implies that all the main effects in (10.1) are non-negative: $\beta_j \geq 0$. In terms of DOE, the standardised variables also indicate that the combination of experimental factors relates to the switch-on (1) and switch-off (0) of the equation's elements. To deal with the interaction between factors, i.e. the dependence of a specific factor on the levels of other factors, (10.1) can be approximated as:

$$y = \beta_0 + \sum_{j=1}^{K} \beta_j x_j + \sum_{j=1}^{K-1} \sum_{j=j+1}^{K} \beta_{j,j} x_j x_j \tag{10.2}$$

where $\beta_{j,j'}$ is the two factor interaction effect between factor j and j'.

Secondly, SB is operated in an iterative procedure where the next factor is selected based on the outputs of previous factor combinations that have already been simulated. The procedure might contain several stages, depending on the lower limit of the effect level defined by the users. The first stage always estimates the simulated results from the two extreme factor combinations, namely y_0 (all factors low) and y_K (all factors high). If $y_0 < y_K$, then the sum of all the individual main effects is important and the second stage of SB is entered. SB then splits the factors into two subsets of equal size and continues the estimation process for each subgroup, which is the same as that described in the first stage, and the procedure continues in an iterative manner. SB terminates when the effect level (i.e. $y_j - y_0$) reaches the lower effect limit defined by the user.

More detailed instructions on how to apply the SB technique can be found in Bettonvil and Kleijnen (1997) and Ngo (2009). In general, the effective level of the factor found by the SB indicates its sensitivity. The factors that are identified by the SB as having little importance or were less effective should be eliminated from the model. The remaining factors or model parameters will then need to be calibrated if unknown *a priori*. In the SCM of Ngo (2009), sensitivity analysis was used to eliminate a number of variables from the model, leaving a subset for calibration.

10.3 Model Calibration

Once the sensitivity analysis is completed, the next stage in validation (Fig. 10.1) is calibration of the model. The calibration process is conducted to identify suitable values for the model parameters in order to obtain the best fit with the real world. This process, therefore, involves the optimisation of the parameters. There are many different optimisation methods available (Fletcher 2000) but a genetic algorithm (GA) is particularly well suited for implementing this task. A GA has novel properties such as being able to undertake a parallel search through a large solution space (Holland 1992). GAs have also been used to calibrate other ABMs in the past (Heppenstall et al. 2007; Rogers and Tessin 2004).

10.3.1 The Principle of Parameter Optimisation Using GAs

A GA applies the principle of "survival of the fittest" from the field of genetics to a population of competing individuals or solutions within a given environment called the search space (Soman et al. 2008). The procedures involved in a GA are similar to the process that occurs in genetics where the parameters in the GA play the role of chromosomes; the range of data is the genotype; while the results of the model runs are the phenotype. The general steps in a GA are illustrated in Fig. 10.3.

The GA starts with a randomly generated number of solution samples which is collectively called the population, which is the first generation of the species.

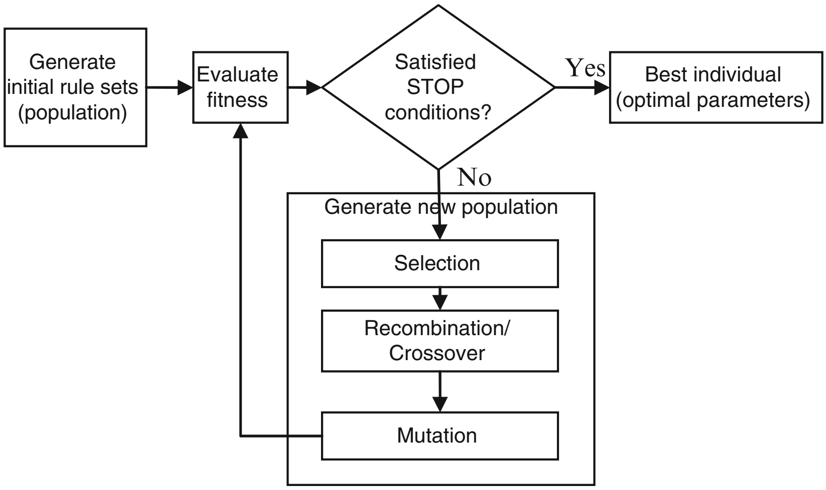


Fig. 10.3 The general steps in a genetic algorithm

A single solution or individual in the population is the combination of parameters with particular values. The solution is therefore equivalent to a natural chromosome with a specific genotype. The next step is the evaluation of fitness using the objective function specified by the user. If there is any individual with a fitness value that satisfies the threshold condition, the programme is terminated and the best individual will be the best solution. Otherwise the GA will operate in a loop creating new generations or populations. Within the loop, individuals (i.e. chromosomes) with higher fitness values are given a higher probability of mating with each other, so as to produce offspring that may better fit the environment.

Several selection methods for selecting the best fit individuals are available such as roulette wheel, tournament, rank and elitism (Mitchell 1996). The most popular method is tournament selection, which is not only suitable for a medium and small population size but also provides marginally better accuracy compared to the roulette wheel selection (Al-Ahmadi et al. 2009). The tournament selection chooses the best fit individuals from several random groups iteratively. For example, if a total of 35 best fit individuals must be selected out of a population of 50 members, the tournament will firstly select a random group (e.g. a group of three random members); within this group, a best fit individual will be the first selected member. These selection processes continue with the next random group to choose the second member until the 35th member is reached. All selected individuals are then entered into the recombination or crossover step which replaces the old chromosomes with the new ones. In the crossover phase, two selected individuals from two random tournament groups perform crossover with a certain number of gene exchanges.

The process of selection and recombination do not inject new genes, so the solution can converge to a local optimum (Soman et al. 2008). The process of mutation, which prevents GAs from premature convergence to a local optimum, is performed to achieve local perturbation by randomly replacing the parameter values with new ones. The frequency of the replacement and the level of perturbation (i.e. the number of parameter values that are replaced) is defined by the mutation rate. Selection, recombination and mutation are then applied to each generation iteratively until an optimal solution is reached. The condition to be satisfied could be reached after a maximum number of generations or if there is observed stability in statistics such as the mean and/or variance of the population fitness values from a generation (Soman et al. 2008). The optimisation programme ends when the terminated conditions are matched and the optimal solution is reported.

10.3.2 Measurement of the Fitness of a GA

There are several techniques for measuring fitness and errors in the simulation model such as the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), Median Absolute Percentage Error (MdAPE), the Relative Operating Characteristic (ROC), a Confusion Matrix (CM), the Kappa Index of Agreement (KIA), Fractal Analysis (Mandelbrot 1983) and Multiple Resolution Goodness-of-fit (MRG). These techniques and goodness-of-fit statistics measure different aspects of the model performance, and may therefore be suited to different objectives. The selection of which evaluation measures to use depends upon the purpose of the validation and the characteristics of the measures, i.e. what the different measures are intended to show.

With respect to the GA, the RMSE is the most commonly used fitness or error measure (Chatfield 1992) because it indicates the magnitude of error rather than relative error percentages (Armstrong and Collopy 1992). This statistic measures the squared differences between the simulated or predicted values and the observed or reference values:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(x_{1,i} - x_{2,i}\right)^2}{n}} \tag{10.3}$$

where $x_{1i} - x_{2,i}$ is the difference between variable i from data source 1 (i.e. the simulated result) and data source 2 (i.e. the reference or observed data); and n is the total number of variables. The RMSE provides a global measure of performance that aggregates all individual differences into a single measure of predictive power.

Other measures of evaluation such as the ROC and the MRG are more suited to evaluation of the model outputs once the model is calibrated so are described in more detail in Sect. 10.4.

10.3.3 Interpreting Calibration Results from the GA

In practice, a GA does not produce a single unique set of parameters but a range of solutions that sit on a Pareto front (Madsen et al. 2002; Yapo et al. 1998). This means that the GA operations will produce a range of different parameter combinations

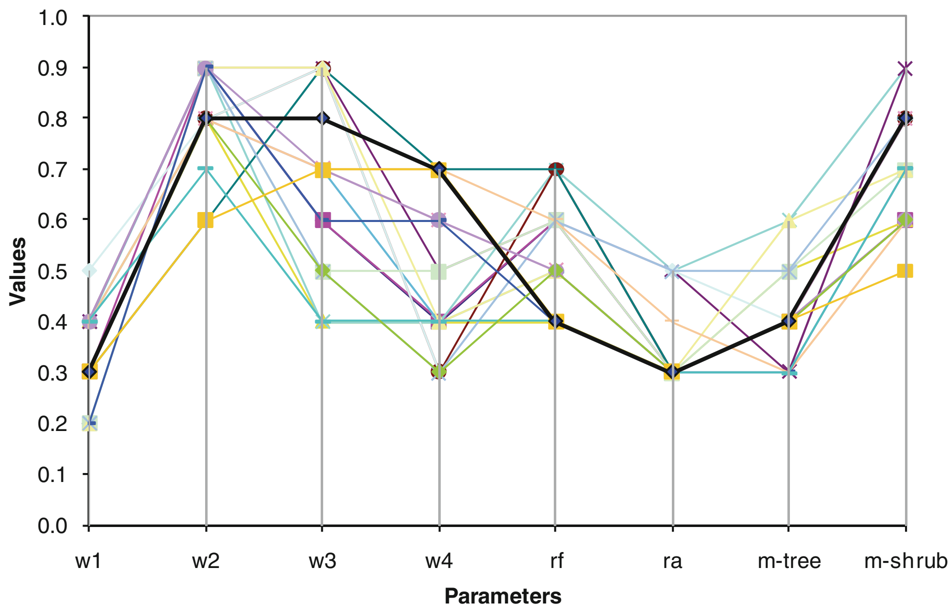


Fig. 10.4 The calibrated parameters provided by 30 GAs (Ngo 2009)

that can give acceptable solutions, rather than generating a single solution. An example of the optimised parameter set from the SCM of Ngo (2009) is shown in Fig. 10.4.

Each line in Fig. 10.4 represents a solution that consists of values for eight calibrated parameters. For each parameter there is a range of possible solutions, indicating the error in the values found by the GAs produced in several runs. Therefore, the additional step that needs to be done is to check for the standard errors for each parameter from all runs. If the errors are not high and the relationships between the values of the parameters and real conditions are reasonable, the solution will be potentially accepted.

As explained above, all parameter combination sets provided by the GA are potentially acceptable solutions. In addition, each parameter clusters around a central value, suggesting that there is a global optimum for the multiple objectives. However, later analyses using the ABM such as validation of model outputs and scenario analyses will require a consistent set of parameters. The way of selecting a set of parameters for further analysis depends strongly on the purpose of the modeller. An acceptable way could be to run the model several times with different parameter sets provided by the GA and then compare the output(s) that is considered as important or significant by the modeller. The parameter set that yielded the highest average fitness value compared to the real data is one method for selection. For example, the bold line in Fig. 10.4 is a parameter set that provided highest fitness values for land cover and was therefore selected as the best solution for the SCM (Ngo 2009).

10.4 Validation of Model Outputs

The final stage in the validation process (Fig. 10.1) is validation of the ABM outputs. This is the most important process in model development because it ensures that the model has the right behaviour for the right reasons (Klügl 2008; Qudrat-Ullah 2005; Troitzsch 2004). Validation of the model outputs is concerned with how well they represent real world behaviour and they are, therefore, compared with actual observations (Parker et al. 2002).

The measurement techniques that determine how the model outputs match the real data are varied. However, the Relative Operating Characteristic (ROC) and the Multiple Resolution Goodness-of-fit (MRG) are two good measures for validating ABM model outputs. These two measures are explained in more detail below.

10.4.1 Relative Operating Characteristic (ROC)

The ROC is used to evaluate the performance of a classification or prediction scheme by identifying where instances fall in a certain class or group (Beck and Shultz 1986). The classification is based on the value of a particular variable in which the boundary between classes must be determined by a threshold or cut-off value. An example would be the prediction of illegal cultivation measured by the SCM (Ngo 2009), where the threshold value used to predict whether or not a household would cultivate illegally in the protected forest is a value between 0 and 1. The result is therefore a two-class prediction, labelled either as positive (illegal) (*p*) or negative (not illegal) (*n*). There are four possible outcomes from a binary predictor: *true positive, false positive, true negative and false negative*. A *true positive* occurs when both the prediction and the actual value are *p*; *false positive* when the prediction is *p* but the actual value is *n*; *true negative* when the predicted value is *n* and the actual value is also *n*; and false negative when the predicted value is *n* while the actual value is *p*. The four outcomes can be formulated in a two by two confusion matrix or contingency table as shown in Fig. 10.5 (Fawcett 2003). Definitions of precision, accuracy and specificity are also provided.

The ROC evaluation is based on the ROC curve, which is a graphical representation of the relationship between the sensitivity or tp-rate and the specificity or 1 – fp-rate of a test over all possible thresholds (Beck and Shultz 1986). A ROC curve involves plotting the sensitivity on the y-axis and 1-specificity on the x-axis as shown in Fig. 10.6.

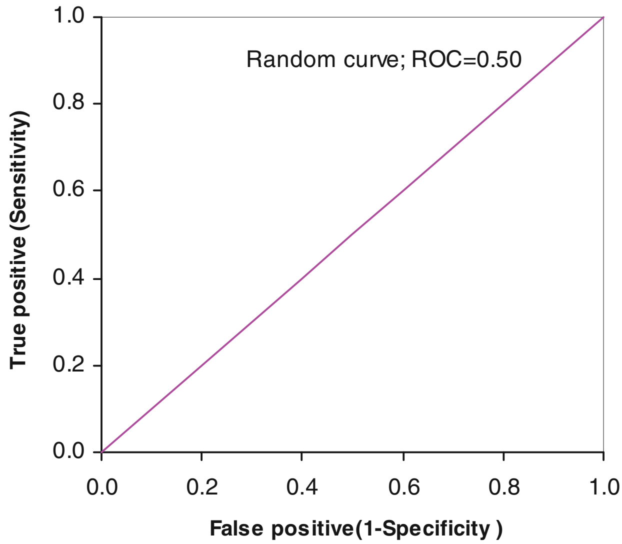
This graphical ROC approach makes it relatively easy to grasp the interrelationships between the sensitivity and the specificity of a particular measurement. In addition, the area under the ROC curve provides a measure of the ability to correctly classify or predict those households with and without illegal cultivation. The

		True class	
		P	*N*
Predicted outcome	*P*	True Positives (*TP*)	False Positives (*FP*)
	N	False Negatives (*FN*)	True Negatives (*TN*)
	Total	*P*	*N*

tp-rate = TP/N = sensitivity;
fp-rate = FP/N;
precision = TP/(TP + FP);
accuracy = TP+TN/(P+N);
specificity = TN/(FP+TN)
= 1 – fp- rate

Fig. 10.5 The confusion matrix to calculate the ROC (Adapted from Fawcett 2003)

Fig. 10.6 A basic ROC curve (Adapted from Fawcett 2003)



ROC area under the curve (AUC) would reach a value of 1.0 for a perfect test, while the AUC would reduce to 0.5 if a test is no better than random (Fawcett 2003).

The ROC has been proposed as a method for land cover change validation (Pontius and Schneider 2001). However, it is less useful in terms of capturing the spatial arrangement of the model outputs in relation to the real world results (Pontius and Schneider 2001). Thus, in the case of the SCM (Ngo 2009), the ROC is more useful for validating the number of illegal cultivators than the area of illegal cultivation predicted by the SCM (Ngo 2009).

10.4.2 *Multiple Resolution Goodness-of-Fit (MRG)*

Multiple resolution goodness-of-fit (MRG) has been proposed for measuring the spatial patterns of the model output at several resolutions. This measurement is especially relevant when validating the spatial outputs of ABMs that model land cover and land use change (Turner et al. 1989).

The MRG procedure is expressed in (10.4), which measures the fit at a particular sampling window size (F_w), which is then aggregated for all samples (Costanza 1989):

$$F_w = \frac{\sum_{s=1}^{t_w} \left[1 - \frac{\sum_{i=1}^{p} \left| a_{1i} - a_{2i} \right|}{2w^2} \right]_s}{t_w} \tag{10.4}$$

where F_w is the fit for the sampling window size w, a_{ki} is the number of cells of category i in the image k within the sampling window, p is the number of different categories in the sampling window, s is the sampling window of dimension w by w which moves across the image one cell at a time, and t_w is the total number of sampling windows in the image of window size w.

The fit for each sampling window is calculated as 1 minus the number of cells that would need to change in order that each category has the same number of cells in the sampling window irrespective of where they appear in the image.

The weighted average of all the fits, F_t, over all window sizes is then calculated to determine the overall degree of fit between the two maps as follows:

$$F_t = \frac{\sum_{w=1}^{n} F_w e^{-k(w-1)}}{\sum_{w=1}^{n} e^{-k(w-1)}} \tag{10.5}$$

where F_w is defined above in (10.4) and k is a constant. When $k=0$, all window sizes have the same weight while for $k=1$, only the smaller windows sizes are important. For the purpose of matching the spatial pattern of land use, a value of k of 0.1 gives an 'adequate' amount of weight to the larger window sizes (Costanza 1989).

The MRG is a much more suitable way of assessing the fitness of the spatial outputs compared to the more conventional methods used in ABM model output validation such as a confusion matrix or kappa statistic calculated at a single resolution only. The Kappa test, for example, can be used to measure the fit between two land cover maps based on a pixel-by-pixel comparison, but it ignores the relationships between one measured pixel and its neighbours. Hence, it will only tell us whether the total number of pixels in each land cover category is significantly different between the two maps, and not say anything about the accuracy of their spatial arrangement (Costanza 1989). The MRG, however, captures the details of the spatial and temporal patterns in the data. More details on the application of MRG can be found in Costanza (1989). The use of the MRG in validating the model outputs of the SCM can be found in Ngo (2009).

10.5 Summary

Calibration and validation are crucial stages in the development of ABMs yet remain a key challenge (Crooks et al. 2007). This chapter has defined these terms and presented the process as a series of steps that should be followed when building a model. Although the process is generic to ABMs in general, particular attention was given to ABMs of land use and land cover change, especially in terms of the measures for evaluating the output of the model. More specifically, examples from the calibration and validation of the SCM of Ngo (2009) were provided to illustrate the process. It should be noted that this represents only one view of the calibration and validation process based on experience gained through building an ABM of shifting cultivation. There are clearly a range of methods available that could be used in or adapted to any part of the calibration and validation process, e.g. different methods of parameter optimization, different measures of evaluating performance, etc. Until more guidance is provided in the literature, calibration and validation will remain a key challenge.

References

Al-Ahmadi, K., Heppenstall, A., Hogg, J., & See, L. (2009). A Fuzzy Cellular Automata Urban Growth Model (FCAUGM) for the city of Riyadh, Saudi Arabia. Part 1: Model structure and validation. *Applied Spatial Analysis and Policy, 2*, 65–83.

Armstrong, J. S., & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting, 8*, 69–80.

Beck, J. R., & Shultz, E. (1986). The use of relative operating characteristic (ROC) curves in test performance evaluation. *Archives of Pathology & Laboratory Medicine, 110*, 13–20.

Berger, T., Goodchild, M., Janssen, M.A., Manson, S. M., Najlis, R., & Parker, D. C. (2001). Methodological considerations for agent-based modelling of land-use and land-cover change. In D. C. Parker, T. Berger, & S. M. Manson (Eds.), *Agent-based models of land-use and land-cover change*. *Report and Review of an International Workshop*. Irvine.

Bettonvil, B., & Kleijnen, J. P. C. (1997). Searching for important factors in simulation models with many factors: Sequential bifurcation. *European Journal of Operational Research, 96*, 180–194.

Box, G. E. P., Hunter, J. S., & Hunter, W. G. (1978). *Statistics for experimenters: An introduction to design, data analysis, and model building*. New York: Wiley.

Carley, K. M. (1996). *Validating computational models* (Working Paper). Pittsburgh: Carnegie Mellon University.

Castella, J.-C., & Verburg, P. H. (2007). Combination of process-oriented and pattern-oriented models of land-use change in a mountain area of Vietnam. *Ecological Modelling, 202*, 410–420.

Chatfield, C. (1992). A commentary on error measures. *International Journal of Forecasting, 8*, 100–102.

Costanza, R. (1989). Model goodness of fit: A multiple resolution procedure. *Ecological Modelling, 47*, 199–215.

Crooks, A. T., Castle, C. J. E., & Batty, M. (2007). *Key challenges in agent-based modelling for geo-spatial simulation* (CASA Working Paper 121). URL: http://www.casa.ucl.ac.uk/publications/workingPaperDetail.asp?ID=121

Fawcett, T. (2003). *ROC graphs: Notes and practical considerations for researchers* (HP Labs Technical Report HPL-2003-4).

Fletcher, R. (2000). *Practical methods of optimization* (2nd ed.). Chichester: Wiley.

Happe, K. (2005). *Agent-based modelling and sensitivity analysis by experimental design and metamodelling: An application to modelling regional structural change*. European Association of Agricultural Economists, Copenhagen.

Heppenstall, A. J., Evans, A. J., & Birkin, M. H. (2007). Genetic algorithm optimisation of an agent-based model for simulating a retail market. *Environment and Planning B: Planning and Design, 34*, 1051–1070.

Holland, J. H. (1992). *Adaptation in natural and artificial systems*. Cambridge, MA: MIT Press.

Jepsen, M. R., Leisz, S., Rasmussen, K., Jakobsen, J., Moller-Jensen, L., & Christiansen, L. (2006). Agent-based modelling of shifting cultivation field patterns, Vietnam. *International Journal of Geographical Information Science, 20*, 1067–1085.

Kleijnen, J. P. C., & Van Groenendaal, W. (1992). *Simulation: A statistical perspective*. Chichester: Wiley.

Kleijnen, J. P. C., Sanchez, S. M., Lucas, T. W., & Cioppa, T. M. (2003). *A user's guide to the brave new world of designing simulation experiments*. Tilburg: Tilburg University, Center for Economic Research.

Klügl, F. (2008). A validation methodology for agent-based simulations. In *Proceedings of the 2008 ACM Symposium on Applied Computing – Advances in Computer Simulation.* ACM. New York, NY, USA, 39–43.

Le, Q. B. (2005). *Multi-agent system for simulation of land-use and land cover change: A theoretical framework and its first implementation for an upland watershed in the central coast of Vietnam*. Gottingen: ZEF.

Madsen, H., Wilson, G., & Ammentorp, H. C. (2002). Comparison of different automated strategies for calibration of rainfall-runoff models. *Journal of Hydrology, 261*, 48–59.

Mandelbrot, B. B. (1983). *The fractal geometry of nature*. Oxford: WH Freeman and Co.

Manson, S. M. (2002). Validation and verification of multi-agent systems. In M. A. Janssen (Ed.), *Complexity and ecosystem management – The theory and practice of multi-agent systems*. Cheltenham: Elgar.

Mitchell, M. (1996). *An introduction to genetic algorithms*. Cambridge, MA: MIT Press.

Ngo, T. A. (2009). *Simulating spatial patterns of shifting cultivation – A village case study from the uplands of Vietnam*. Unpublished PhD thesis, School of Geography, University of Leeds, Leeds.

Ngo, T. A., Drake, F., & See, L. (2012). An agent-based modelling application of shifting cultivation. In A. J. Heppenstall, A. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models for geographical systems* (pp. 611–627). Dordrecht: Springer.

Parker, D. C., Manson, S. M., Janssen, M. A., Hoffmann, M. J., & Deadman, P. (2002). Multi-agent systems for the simulation of land-use and land-cover change: A review. *Annals of the Association of American Geographers, 93*, 314–337.

Pontius, R. G., & Schneider, L. C. (2001). Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. *Agriculture, Ecosystems and Environment, 85*, 239–248.

Qudrat-Ullah, H. (2005). Structure validation of system dynamics and agent-based simulation models. In Y. Merkuryev, R. Zobel, & E. Kerckhoffs (Eds.), *Proceedings of the 19th European Conference on Modelling and Simulation* (pp. 481–485). Riga: ECMS.

Rogers, A., & von Tessin, P. (2004, May). Multi-objective calibration for agent-based models. In: *Agent-Based Simulation 5*, Lisbon.

Soman, S., Misgna, G., Kraft, S., Lant, C., & Beaulieu, J. (2008). An agent-based model of multi-functional agricultural landscape using genetic algorithms. In *American Agricultural Economics Association Annual Meeting*, Orlando.

Troitzsch, K. G. (2004). Validating simulation models. In *Proceedings of the 18th European Simulation Multi-conference*. SCS Europe, Magdeburg.

Turner, M. G., Costanza, R., & Sklar, F. H. (1989). Methods to evaluate the performance of spatial simulation models. *Ecological Modelling, 48*, 1–18.

Wada, Y., Rajan, K. S., & Shibasaki, R. (2007). Modelling the spatial distribution of shifting cultivation in Luangprabang, Lao PDR. *Environment and Planning B: Planning and Design, 34*, 261–278.

Windrum, P. F., Fagiolo, G., & Moneta, A. (2007). Empirical validation of agent-based models: Alternatives and prospects. *Journal of Artificial Societies and Social Simulation*, *10*(2), 8. URL: http://jasss.soc.surrey.ac.uk/10/2/8.html

Winer, B. J. (1971). *Statistical principles in experimental design*. New York: McGraw-Hill.

Yapo, P. O., Gupta, H. V., & Sorooshian, S. (1998). Multi-objective global optimization for hydrologic models. *Journal of Hydrology, 204*, 83–97.

Zeigler, B. P. (1976). *Theory of modelling and simulation*. New York: Wiley.

Chapter 11
Networks in Agent-Based Social Simulation

Shah Jamal Alam and Armando Geller

Abstract Computational social science and in particular agent-based social simulation continue to gain momentum in the academic community. Social network analysis enjoys even more popularity. They both have much in common. In agent-based models, individual interactions are simulated to generate social patterns of all kinds, including relationships that can then be analyzed by social network analysis. This chapter describes and discusses the role of agent-based modeling in the generative-analytical part of this symbiosis. More precisely, we look at what concepts are used, how they are used (implemented), and what kind of validation procedures can be applied.

11.1 Introduction

Agent-based modeling and network analysis enjoy a symbiotic relationship in the field of computational social science. The former is a method of computationally representing individual interactions from which social patterns emerge; the latter is a method that affords (dynamic) structural analysis of (socio-) structural patterns. The renowned anthropologist Clyde Mitchell stated that the starting point of any analysis should be the actual relationships in which people are involved (Mitchell 1989, pp. 77–79). What he did not think of, interestingly, is to analyze, other than by observational and descriptive means, how these relationships form.

S.J. Alam
School of Public Health, University of Michigan, Ann Arbor, MI, USA
e-mail: sjalam@umich.edu

A. Geller (✉)
Krasnow Institute for Advanced Study, George Mason University, Fairfax, VA, USA
e-mail: ageller1@gmu.edu

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_11,

Today, agent-based models (ABMs) are mostly implemented as object-oriented computer programs. They consist of autonomous agents that can be perceived as computer programs themselves. In principal, agents have three features: they behave and interact according to a given set of rules, possess cognitive capabilities to process information, and constitute their own environment (cf. Cederman 2001; Ferber 1999). Empirically seen, the key question is how the design of agent behavior and cognition is informed. Standard research practices suggest that the agent design process can rely on qualitative data (Alam et al. 2010; Hoffer 2006), experimental data (Barreteau et al. 2001), and empirically validated theoretical knowledge (Cioffi-Revilla and Osman 2009).

At this point it should be evident that we favor an empirical approach over a pure Popperian procedure. The importance of this statement lies in the fact that agent interactions as defined by agent behavior are tantamount to what is called in social network analysis 're-wiring', i.e., according to which rules (algorithms) do the different nodes in a network get connected with each other. ABMs claiming to have relevance for the social sciences should assume plausible behavior at the individual level. ABMs are considered non-black-box models (Boudon 1998). Should ABMs serve as social network generators, then one requirement is that they can explain how the network came about. Hence, from an epistemological perspective, the model needs to exhibit construct-valid mechanisms and processes.

The kinds of networks that can be generated and represented by agent-based social simulations are manifold. They can range from networks with only a few vertices and edges to complicated networks in which agents are embedded in several different layers, so called multiplex networks (Granovetter 1985). Networks generated by ABMs can represent social, geographical, and even cognitive (semantic network) spaces. In their capacity as thematic maps, networks can be used to elucidate such concepts as exchange, power, or identity. Paired with social simulation, these questions can be further explored insofar as agent-based modeling enables the study of the underlying agent behavior, and social mechanisms and processes (Hedström 2005). This is a powerful combination.

Agent-based social simulations are usually analyzed based on hypotheses. One way of testing the hypotheses is observing time-series charts for a number of measures. In analyzing agent-based social networks, an important issue is to understand the role of social processes in constraining the dynamics of the generated networks. The purpose of agent-based social networks is to explore the simulated data trajectories and to understand the modeled phenomena. This is different compared to stochastic models for dynamic social networks (Snijders et al. 2010), where existing longitudinal data are used for model fitting and parameter estimation.

When generating social networks by means of agent-based modeling, two concepts are in the foreground: the processes that bring about the network and the structure this network has. Process and structure are interdependent processes. How agents behave is, of course, influenced by how they are connected to others; that is how they are embedded in society. To this a third dimension is added in agent-based modeling. Agents are usually placed on some kind of surface.

The focus of this chapter is to describe and discuss the symbiosis of agent-based social modeling and social network analysis. We shall look at how model topologies affect network topologies and provide an overview of different social network generation processes. How networks are implemented in ABMs and how agent-based social networks may be analyzed are also discussed.

11.2 Social and Physical Space in ABMs

In this section, we discuss physical and social neighborhoods in agent-based social simulation models. Agent-based simulation models of social phenomena date back to the mid 1980s. As Axelrod (1997) argues, the goal of this modeling approach has been to break simplistic assumptions required for mathematical tractability, e.g. homogeneity, ignoring interaction. With the advent of multiagent models, social simulation benefited from it most as these models provided the provisions of simulating the social behavior of autonomous individuals and the interactions between them. ABMs have been accredited, in most cases, as suitable for decentralized scenarios, especially when individual interactions lead to the emergence of collective patterns, like in the case of complex social networks.

11.2.1 Representing Physical Neighborhoods

Agent-based modeling affords taking geographical space into account in a straightforward manner. This is true for abstract spaces as well as for detailed Geographic Information Systems (GIS) referenced spaces. Perhaps the most commonly used topologies in agent-based modeling are the von Neumann or Moore neighborhoods on a plain or a toroid surface. Other possible topologies are, for example, irregular, hexagonal grids or vector-based (Crooks et al. 2008). Differences in topologies lead to differences in network generation processes and resulting network structures (c.f. Flache and Hegselmann 2001). The reason for this is that model topologies limit agents not only in their movement, but also in the manner by which they perceive information and interact with other agents. The underlying assumption is that space is important and matters in everyday (artificial) life, affecting both the individual's behavior and society as a whole. Choice of a topology depends very much on the modeler's needs. The focus of the discussion here is on the effect that different model topologies have on network evolution processes and network structures. In other words, how are dynamic social networks coupled to model space? Note that this question is distinct from questions of how space is represented in networks, which we discuss below.

As Bailey and Gatrell (1995, p. 4) explain, "spatial data analysis is involved when data are spatially located and explicit consideration is given to the possible importance of their spatial arrangement in the analysis or interpretation of results."

Spatial analysis, for example, that is based on GIS techniques, highlights the importance, provided it exists, of neighborhood influences, if any, in the actors' behavior caused by the spatial-context. Schensul et al. (1999) have thoroughly covered the issues involving spatial mapping of data; we report a few of the most relevant points. For any social networks, the atomic units are usually the individuals. In gathering data about individuals, it is quite useful to identify the general spatio-temporal constraints that limit most individuals' movements and interaction in the region. Typically, in spatially explicit models, agents may include stakeholders, land owners, farmers, public institutions, and policy or decision-making agencies. As Brown (2006) explains, the behavior of such agents may vary from being triggered by some external stimulus or coping with certain stresses to being goal-oriented.

11.2.2 Networks from Embedded Social Mechanisms and Processes

Earlier in this section we discussed the importance of signifying boundaries and neighborhoods. Modeling a social network requires identifying the spaces in which the agents exist and are related. All relations among real entities exist and are constrained through physical spaces. More importantly, case-studies involving land use change, distribution and utilization of physical resources are modeled spatially explicitly *per se*.

Social networks are generated through social mechanisms and processes, i.e., agents that are embedded in society and that interact with each other produce them. It has become more and more accepted in the social sciences that the agents' (e.g., humans, primates, ants) behavior does not follow a linear pattern, but is non-linear in its own right. Social complexity, according to Moss (2008, p. 2), is a "condition whereby social behavior cannot be understood simply as a scaled-up replication of the behavior of the individuals comprising the society". The interplay of social processes as an outcome of socially embedded individuals gives rise to the social behavior, which, as Moss (2008, p. 3) explains, "cannot be forecast on the basis of individuals' characteristics and predictions alone". The macro-phenomena resulting from such micro-level interactions are often complex in nature. We understand complexity as a "type of condition in which agent behavior and social interaction combine to generate macro-level outcomes that could not be predicted from knowledge of the behavior and nature of interactions alone, and result in sporadic volatile episodes, the timing, magnitude, duration and outcomes of which are themselves unpredictable" (Geller and Moss 2008, p. 322). By contrast, in the study of so called complex networks, the notion of complexity is related to network structures (both local and global) and characteristics that are not statistically significant in a random network (Newman 2004; Wasserman and Faust 1994). We are aware of other definitions of both complexity and complex networks (see Edmonds (1999) for a review), but those given should suffice for the purpose of this chapter.

Self-organized criticality (SOC) addresses the local mechanisms and processes that drive the emergence of complex systems. It can be interpreted as the response of a slowly driven system such that the outcome of the system's behavior is limited by the order of the magnitude of its size, thus, leading to the scale-free property (see below). Following Jensen (1998), one may explain SOC as the development of emergent patterns due to interactions among meta-stable agents, so that at some critical state, the result of interactions affects the entire system such that all members of the system influence each other. For the rest of the period, any local distortions resulting from agents' interactions in their neighborhoods remain confined locally. Systems governed by SOC leave characteristic traces in the data they produce. The data conforms not to the assumption underlying standard statistical methods, namely that the mean and standard deviation of the distribution of the data are known and stable. Consequentially, the conditions for standard statistical hypothesis testing and regression techniques are not satisfied anymore, and there are cases where variance is infinite (Barthélémy 2006). However, of more importance to us in the present context is the fact that investigating such signatures provides useful guidance for the analysis of social simulations (Moss 2002). Leptokurtosis in a distribution of relative changes can be a reflection of episodes of volatility that are themselves unpredictable (Moss 2002). That is, unpredictable clustering of volatility and the corresponding extreme events are identifiably complex features of time series. Conversely, finding leptokurtosis in time series data would naturally incline us to look for extreme events. A vital implication of such approaches is that it is practically impossible to predict the outcomes to the system from simple stimuli (Jensen 1998).

ABMs – not only of social systems – can represent such properties. This is an important assumption that needs to be taken into account when modeling networks with an agent-based approach, for agent behavior and interaction – as understood in SOC – will affect the kind of networks that emerge. With this in mind, we now present an overview of characteristic complex network topologies associated with complexity concepts that an agent-based modeler has to expect when running a simulation. Presented will be also measures appropriate for the analysis of agent-based social simulation generated networks.

Modeling dynamic social networks where agents communicate with each other and build relations over time requires the introduction of "social" spaces that go beyond the physically situated agents. Such agents can be called "socially embedded" (Edmonds 2006; Granovetter 1973), i.e., an agent's behavior is fairly influenced by the network of social relations that it is part of. Physical resources and interaction with the environment do not fulfill the demand for capturing the social interactions that may influence, for example, a farmer's decision to plant a certain type of crop, or use of their land. Social spaces and the agents' interactions may either be constrained by a local neighborhood, or could be global (i.e., each agent may be directly related to any other agent in the space). In the former case, the sociability of agents depends on the spatial neighborhoods, and thus, according to Edmonds (2006), the physical space is used as a proxy for social space.

Not many social network models exploit combining the social and physical spaces, which is pivotal for analyzing the underlying complexity and for which ABMs are well suited as they support modeling the spatial neighborhood as well as agents' cognition in building relations. Hence, symbiosis of the two "spaces" remains an active area of research.

11.2.3 Types of Complex Networks

The term complex networks is used as an umbrella term for the size, similarity of structure, and dynamics in real and simulated networks (for two comprehensive articles on the issue see Newman (2004) and Fortunato (2009)). Cross-disciplinary research, especially in the last decade, has resulted in identifying characteristic network types and their statistical properties. Network structures are either modeled phenomenologically or they emerge from agents' local interaction (for an older, but relevant review concerning networks for ABMs see Amblard (2002)). We briefly look at three commonly occurring network structures in agent-based social simulation: random graphs, small world and scale-free networks. Regular lattice networks are used in cellular automata models – a lattice is a graph where vertices are placed on a grid and are connected to the neighboring vertices only.

An early attempt to study the behavior of complex networks dates back to Erdős and Rényi's (1959) seminal work on random graph theory. The basic Erdős-Rényi (ER) model requires connecting N nodes through n edges chosen randomly such that the resulting network is from a space of equally likely graphs, where N is the size of the network. Several nodes can have the same degree in a random graph. Given a high wiring probability $p,$ the diameter of random graphs increases logarithmically with the growth of the graph. The ER graph also predicts the appearance of subgraph structures and the emergence of a unique giant component.

Random networks are to social network data what the Gaussian distribution is to statistical data; it is neither very likely to find random network structures in real world data nor very realistic to assume that real world networks are of a random nature. Firstly, people do not behave randomly. Secondly, societies are complex systems. Randomness is diametrically opposed to this idea. It is, however, worthwhile to consider random networks as a useful concept in agent-based social network modeling since they constitute a test case. The networks generated by the simulation, and which are meant to represent an identified (real world) target system, should be significantly different with regard to certain key metrics from the corresponding ER network.

In 1998 Watts and Strogatz (1998) presented the Watts-Strogatz (WS) model, which interpolated a small world graph as an intermediate of a purely random and a regular graph. They showed that as the length of the shortest path between two nodes tends towards *O(ln(N))*, which is small, a random graph exhibits the so-called small-world effect. That is, a WS network is characterized by short average path length (L) and a high clustering coefficient (C) compared to an Erdős-Rényi graph of the same size and density. This property displayed by small-world networks has been observed in a number of social systems, including friendship, co-worker, and conflict networks.

Informally, a high C supports the ideas that the "friend of my friend is my friend" and that the neighbors of a node are more likely to be linked to each other than in a random network. More generally, small-world type networks should be of interest to us because they exhibit properties which are "sufficiently well connected to admit rich structure, yet each element is confined to operate within a local environment that encompasses only a tiny fraction of the entire system" (Watts 1999, p. 499). This specification of the micro-level processes leading to the emergence of small-world networks is closely related to the idea of SOC and complex systems.

Albert and Barabási (2002) argued that simply using ER or WS models does not capture the important aspects of real-world networks. The Barabàsi-Albert (BA) model is a special case of the stochastic model proposed by Herbert Simon (Simon 1955) for generating a class of highly skewed distributions, including the power-law distribution. The number of starting vertices is fixed and the chances of a vertex being linked to another are equally likely. Instead, real-world networks evolve over time and exhibit a feature that is called preferential attachment. Albert and Barabási address these issues by introducing network growth. The network starts off with a small number of connected vertices. New vertices are added to the network one at a time and are linked to existing vertices. Then they introduce the idea of preferential attachment, meaning the probability that a new vertex is connected to an existing vertex depends upon the connectivity of the vertex, where k is the degree of the i-th vertex in the existing network. The network evolves into a scale-variant such that the degree distribution follows a power law.

11.3 Incorporating Networks into Agent-Based Simulation Models

Unlike physical systems, social processes are modeled descriptively and validated qualitatively. The evidence is gathered through fieldwork. An individual's relations and actions are driven by their position and other factors affecting the system. Where the actions are constrained by both the endogenous and exogenous factors, one may find episodic volatility in the observed time series (Moss and Edmonds 2005). Next, we discuss some of the issues concerning social network data collection and incorporating them into ABMs. We then give some examples of ABMs of social networks.

11.3.1 Data for Networks in Agent-Based Social Simulation

Acquiring data on social networks is a challenging task for fieldwork researchers depending upon socio-cultural and socio-political aspects of their research and resource constraints. Schensul et al. (1999) identify features for data that are of use for the description of social networks:

- Identification of network actor;
- Definitions of and rules to define group members by the people;
- Inclusion and exclusion rules defining social network boundaries;
- Familial and sexual relationships, (if any), within groups.

Network boundaries constitute the edges of networks and are defined by rules for entry and exit from groups as well as by other cultural patterns of participation that differentiate one group from another. An important facet of a community is the existence of so-called community organizations, which operate within the perimeters of the community. Such organizations can be characterized as being informal or institutionalized. Physical neighbors can be described in terms of land use and segregation of sub-regions. The social aspect of neighborhood is based, for example, on the "local social interactions, social class, ethnic and radical origins, life cycle characteristics of the population, length of residence, and place of work" (Schensul et al. 1999). The concept of locality is embedded in its definition; hence a community can be identified as sharing social characteristics or a community space, where social interactions are likely to take place.

Social network data can be derived from census data, third-party surveys and various forms of quantitative data (e.g., Eubank et al. 2004; Bearman et al. 2004; Geller and Moss 2008).

Social network data may also be extracted from existing databases such as e-mail correspondence within an organization or social interactions among individuals in online communities. On the other hand, it is very difficult to conduct fully-fledged surveys for acquiring social network data in distant, stressed or conflict-torn regions such as Yemen or Afghanistan. Knowledge elicitation techniques based on participatory approaches (Barreteau et al. 2001; Pahl-Wostl and Hare 2004) may be used to model the behavior and social interaction of relevant actors through an iterative process involving data collection, validation and scenario exploration.

11.3.2 Implementing Networks in Agent-Based Models

A social network is a graph where actors (e.g., individuals, households, firms) are represented as vertices and an existing relation between any two nodes represented as an edge between them. Multiple relations among agents embedded in space are represented as a two-mode sociomatrix, a hypergraph or a bipartite graph, where one representation can equivalently be mapped to another (Wasserman and Faust 1994). Bipartite graphs are useful for simultaneous analysis of both actors and the affiliations

(Degenne and Forsé 1999). Typically, a graph, i.e. a social network, is implemented as an adjacency matrix or a doubly-linked list besides others. Choice of a suitable data structure for manipulating social networks may depend upon the structure of the underlying social network, e.g., single or multiple relations; directed/undirected; weighted edges, etc. Two of the most widely adopted data formats used for social networks are GraphML (Brandes et al. 2004) and DyNetML (Tsvetovat 2005), both based on XML. Both support directed, undirected, and mixed graphs; hypergraphs; hierarchical graphs; and store nodes and edges attributes, for example agents' characteristics or type or strength of edges (see Tsvetovat (2005) for a comparison of commonly used social network data formats). Another well known data format is Pajek's .net format for rich social network data (de Nooy et al. 2005).

Several simulation toolkits and software exist with built-in data structures and operations for analyzing and visualizing social networks. Widely-used software includes Pajek (de Nooy et al. 2005), ORA (Carley et al. 2007), StOCNET (Boer et al. 2006) and UCINet (Borgatti et al. 2004) (for a list of social network analysis software, see for example Wikipedia's entry under "Social network analysis software"). Several agent-based modeling platforms provide functionality for implementing and analyzing networks at runtime. These include RePast 3.1/Simphony (North and Macal 2007), MASON (Luke et al. 2005), NetLogo (Wilensky and Rand in press) and Swarm (Minar et al. 1996). Most of them intentionally provide only limited support for network analysis measures such as the basic centrality measures and community detection algorithms (Nikolai and Madey 2009). Dedicated network modeling and analysis libraries such as the Java Universal Network/Graph library (JUNG) (O'Madadhain et al. 2005); the R Project packages statnet, sna, and igraph are to be used for more computationally-extensive handling of network data generated by ABMs. Social network analysis software and APIs provide an interface to read/write social network data in data formats such as GraphML or DyNetML. For a detailed discussion on the integration of GIS and agent-based modeling, see Crooks and Castle (2012).

11.3.3 Some Examples of Spatially-Explicit Agent-Based Social Simulation Models

In this section, we present a selection of relevant work dealing with implementations of social networks in ABMs.

11.3.3.1 Land Use Models

Central to landscape modeling, such as land use, land cover, habitat conservation and farming, is the identification of community space and distinct regions (Brown 2006; Parker 2005). For instance, Krebs et al. (2007) developed a spatially explicit ABM of a water irrigation system in the Odra River Valley in Poland. In their model, farmers' decisions to maintain the irrigation water canal depend on the relative location of their land (up- or downstream), how they perceive their physical neighbors, and the underlying social network. For a recent review on land-use from an agent-based modeling perspective, see Matthews et al. (2007) and Crooks (2010). Becu et al. (2003) modeled the impact of upstream management in Thailand and explored several scenarios concerning land managers' collective action given their characteristic and social interaction (Ziervogel et al. 2006).

FEARLUS is an established modeling framework designed for the assessment of land use change scenarios (Polhill et al. 2008). Built upon the Swarm modeling platform (Minar et al. 1996), it supports a variety of agent-based modeling techniques and extensions such as a biophysical component, land trade and the effects of climatic variability on land parcels. The FEARLUS simulation begins with the land parcels assigned to land managers. At each annual cycle, managers select the land use of their land parcels based on the available selection strategies. They decide to harvest based on the expected yield for a particular year, select land parcels for sale or to clear off deficits, or decide to retire, allowing new land managers to enter the system. FEARLUS incorporates social and physical neighborhoods. Social neighborhood and spatial distribution are both used by agents, representing farmers or land owners, to observe each other and decide what action to take. Further information on FEARLUS and how the physical and the embedded social neighborhoods are implemented can be found online at http://www.macaulay.ac.uk/fearlus.

11.3.3.2 Neighborhood and Segregation Models

Edmonds (2006) extended the Schelling (1971) segregation model by adding an explicit social structure in the form of a friendship network to the agent neighborhoods which are defined by their spatial location on a regular grid. The friendship network is assigned randomly at the start based on the preference parameters: number of friends, neighborhood bias, and bias for racial similarity. Edmonds thus changes the motivation for switching the neighborhood. Instead of intolerance based on race, as implemented by Schelling, fear as a result of personal insecurity makes people leave for another neighborhood. Fear is a function of security related incidences and spreads through the friendship network. Communication of fear depends on the density of the social network on the other hand. At the same time friendship networks are not necessarily in the geographical vicinity of an agent. An agent can thus be attracted away to where its friends live. As a result, social and physical space becomes disjointed.

In their model of neighborhood change, Bruch and Mare (2006) used a variety of choice functions to introduce heterogeneity in individuals' preferences, thereby relaxing several of Schelling's (1971) assumptions. They utilized real data from several US cities where the population was divided into multiple racial and ethnic types. They demonstrate that the choice of the utility function can significantly affect the observed patterns of segregation and neighborhood change. Crooks (2010) studied residential segregation using a spatially-explicit ABM using vector GIS. The model takes into account socioeconomic and geographical data where agents represent households with preferences for a neighborhood depending upon their properties. Crook's model is initialized with available aggregate census data of the wards in London (UK). Werth and Moss (2007) modeled migration under socio-economic stress in the Sahel region in North Africa. They used an abstract spatial representation of the region, where household decisions to migrate to another location depend upon their existing social and kinship ties with other households in the neighborhood, in addition to their available food status. Rakowski et al. (2010) studied contact patterns among individuals in a transportation model in Poland.

11.3.3.3 Propagation Models

Spatial and social propinquity can be key determinants in the spread of infectious diseases depending upon their infectiousness and the required level of intimacy for transmission. For instance, sexual transmission of HIV or transmission by sharing injection needles may be driven by the social and physical proximity among potential sex or needle-sharing partners. Diseases like smallpox may be transmitted when individuals happen to be in the same location where an infected person is present. Spread of airborne infections with high infectivity such as influenza, depends upon the migratory or activity patterns in a given population.

EpiSims is a large-scale disease propagation ABM capable of simulating millions of agents based on real data (Eubank et al. 2004). Locations in EpiSims represent a physical place, for example an office or a school building, where individuals get into contact with each other provided that they are in the same location at the same time given their preferences and shared activities. During the simulation, a dynamic contact network is developed by recording the amount of time each individual shares with each other person. The duration of contact between infected and susceptible persons determines the spatially-distributed spread of the infectious disease (Stroud et al. 2007). Yang and Atkinson (2008) developed an ABM of the transmission of airborne infectious diseases using activity bundles, where individual contacts are driven by social activities or physical proximity or both. Huang et al. (2004) modeled the spread of the SARS epidemic by using a small-world social network whereas the individuals' activity spaces were modeled upon a two-dimensional cellular automata. Dunham (2005) demonstrated an implementation of the spread of three viruses using a spatially-explicit agent-based epidemiological model developed in MASON. Huang et al. (2010) propose a four-layer architecture for network-based epidemic simulation comprised of individuals' social interaction, passive connections between individuals and locations, use of abstract geographical mapping to reflect the neighborhood, and the use of demographic or geographic data.

11.3.3.4 Miscellaneous Models

In addition to the models presented above, there are many other examples of ABMs of social networks. In many of these models, the mechanisms generating the social networks have been empirically derived. This stands in stark contrast to modeling exercises where the authenticity of social network generating mechanisms is less of a concern, such as in statistical mechanics. The purpose of many of these models is an explanatory one. Companion (sometimes also called participatory) modeling and role-playing games are certainly at the forefront of an explanatory modeling agenda. The primary objective of companion modeling is to understand complex environments through stakeholder participation, affording to validate model assumptions and to make informed policy recommendations (Barreteau et al. 2001, 2003). Companion modeling stresses that no *a priori* hypotheses are made about the target system. Priority is thus assigned to evidence gathering during fieldwork. Similarly, role-playing games incorporate a special function in the understanding and validation of ABMs. The idea is to consider role-playing games as "living" multi-agent systems in which players are the agents and the set of roles is the rule base. Through rule design in collaboration with the players an understanding of the complexity of the system to be modeled is developed.

It should become clear at this point, that integrating social networks in ABMs goes beyond mere measurement of social network metrics at the aggregate level, but includes a thorough study of the processes underlying network generation, i.e., the structural-dynamic consequences of the actual relationships in which people are involved as mentioned by Mitchell (1989). Geller and Moss (2008) developed a model of power structures in Afghanistan. Barthélémy (2006) modeled water consumption, where a household was represented as the smallest unit in the modeled community space. Alam and Meyer (2010) studied dynamic sexual networks based on a village in the Limpopo Valley case where neighborhood and kinship networks serve as *safety-nets* at times of socioeconomic stress for the households.

Pujol et al. (2005) have modeled the evolution of complex networks from local social exchange, simulating networks with similar characteristics as scale-free and small-world networks. They show that properties characterizing complex networks emerge from the local interactions of the agents, imperfect knowledge and sociologically plausible behavior. Jin et al. (2001) demonstrated how a small-world friendship network may be evolved from simple probabilistic rules. The forest fire model by Leskovec et al. (2005) is another example of a generative process that represents networks phenomenologically with heavy-tailed distributions and shrinking diameters.

So far we have only talked about extra-individual networks. But networks do not only exist between agents; they exist also as mappings of organization beyond social structure. "Structure exists not only as sets of ties between actors but as networks among cognitive and cultural entities and study of these entities by means of network analysis is just as important as study of interpersonal relations" (Tsvetovat 2005, p. 111). The utilization of networks in agent-design and in particular in the agent reasoning processes hence becomes obvious. In this respect the concept of semantic networks offers particular usefulness, for it expresses, in the most general way, relations of meanings between concepts in terms of nodes and links. Semantic networks are thus often used for the representation of knowledge; knowledge that bears – represented as a semantic network – some form of content-related domain specificity (DiMaggio 1997). It is beyond the scope of this chapter to pursue this route any further. We would nevertheless like to make clear that we see great potential in the use of semantic networks in the modeling of socio-culturally grounded cognitive and action selection processes.

11.4 Analyzing Social Networks in Agent-Based Social Simulation

The choice of suitable measures for agent-based social networks depends upon our understanding of the phenomenon under study. Analyzing social networks in and generated by agent-based social simulation does not impose new requirements for social network analysis. Metrics such as: geodesic distance; average path length; network density; reachability; clustering; centrality and centralization and their meaning, continue to be useful in that they characterize network topologies and process-based complexity. However, most of the analysis will have to deal with dynamic social networks. Considering only *a priori* and ex post snapshots of networks is not helpful in identifying network measures for agent-based simulations. Applying graph-theoretic measures over a network snapshot may increase the risk of losing the context of a particular agent's position in the network (Borgatti et al. 2006; Carley 2003; Edmonds and Chattoe 2005). In complex systems, it is hard to anticipate how emerging patterns result from interactions at the micro-level. It could be thus misleading to apply measures on a single snapshot of the network. Carley's dynamic network analysis introduces the meta-matrix, a scheme for coping with the problem of multiple relations and co-evolution of both agents (entities, vertices) and their dynamically changing edges (Carley 2003). This approach is further supplemented by combining social network analysis with cognitive science and multiagent systems, the idea being that change in one network may affect change in another. Edmonds and Chattoe (2005) suggest a scheme that makes use of agent-based social simulation in order to find better means for abstraction.

Again, networks in ABMs are dynamic in nature and ties may be added or removed between agents during a simulation run. The network evolves with changes in the agent population, i.e., the agents that participate in a given (social) network. Consequently, the time-series measures of the simulated social network changes as the network evolves. Therein the focus can lie on standard statistical metrics, such as skewness and kurtosis of the absolute relative differences of network measures, such as changes in the clustering coefficient over time. Since we deal with a complex system, we would expect these measures to be indicators for volatile episodes in the time series (Moss and Edmonds 2005). Of course, we would not expect the time-series to be normally distributed and exhibit heteroskedasticity. But in general, for dynamic networks, where the population of participating agents in a network changes over time, we should also look for the stability (or change) of network measures over one or multiple simulation runs. The choice of measures is therefore important when comparing networks of varying sizes within and/or across simulation runs. See McCulloch (2009) and Alam et al. (2009) for methods of detecting patterns in dynamic social networks.

Networks sharing similar global characteristics can nevertheless differ in terms of their local structures. Identifying subgraph structures and their properties have been studied extensively in social network analysis, particularly with regard to triads as building blocks of a network (c.f. Wasserman and Faust 1994). Milo et al. (2002) introduced the concept of local structures as "motifs" that are statistically significant in comparison to local structures in a random network. Hales and Arteconi (2008) provide a good example of applying motif analysis in an ABM of a peer-to-peer network.

Closely related to motif analysis are endeavors to identify communities in networks. Fortunato (2009) and Mucha et al. (2010) provide a good overview of community detection algorithms for static and longitudinal networks. Without going into the details, the problem is twofold: Firstly, from a socio-scientific point of view, the non-trivial issue of solving the boundary specification problem needs to be solved for a given network. Second, the algorithm for dealing with boundary specification issues needs to be fast, since iterating over the whole network at each time step is computationally expensive.

Agent-based social simulations should be cross-validated (Moss and Edmonds 2005). That is, the model output should be compared against the model's target system data. This comparison can happen at the aggregated level (e.g., statistical signatures of time-series data) or it can happen at a qualitative level, informing on social mechanisms that are assumed to drive the social network. For example, as Watts (1999) reports, small-world structures are likely to be present in many real social networks. Geller and Moss (2008) report a small-world-like structure for Afghan power structures.

11.5 Conclusions

Social and physical networks are important with respect to modeling systems that require both socio-cultural as well as geographical information. However, spatial ABMs incorporating social networks are few. On the other hand, social spaces in the form of friendship, kinship and other socio-cultural networks are often modeled in ABMs without any explicit reference to physical spatial representation or constraints. Some of the examples cited in this chapter show how physical and social space can be coupled together for the purpose of understanding complex social systems. Social networks in ABMs may emerge as a result of agent interaction, which can be contextualized or abstract. On the other hand, incorporating physical networks such as a neighborhood, road networks, etc. is important when understanding the dynamics of urban planning and growth, irrigation systems and road transport. We also discussed in this chapter issues related to data collection for social networks as well as the technical aspects of incorporating networks in ABMs.

References

Alam, S. J., & Meyer, R. (2010). Comparing two sexual mixing schemes for modelling the spread of HIV/AIDS. In K. Takadama, G. Deffuant, & C. Cioffi-Rivella (Eds.), *Proceedings of the second world congress on social simulation* (pp. 33–45). Berlin: Springer.

Alam, S. J., Edmonds, B., & Meyer, R. (2009). Identifying structural changes in networks generated from agent-based social simulation models'. In A. Ghose, G. Governatori, & R. Sadananda (Eds.), *Agent computing and multi-agent systems* (pp. 298–307). Berlin: Springer.

Alam, S. J., Geller, A., Meyer, R., & Werth, B. (2010). Modelling contextualized reasoning in complex societies with endorsements. *Journal of Artificial Societies and Social Simulation 13*(4), 6. http://jasss.soc.surrey.ac.uk/13/4/6.html

Albert, R., & Barabási, A. (2002). Statistical mechanics of complex networks. *Review of Modern Physics, 74*(1), 47–97.

Amblard, F. (2002). Which ties to choose? A survey of social networks models for agent-based social simulations. In *Proceedings of the 2002 SCS International Conference on Artificial Intelligence, Simulation and Planning in High Autonomy Systems* (pp. 253–258), Lisbon, Portugal, April 2002.

Axelrod, R. (1997). Advancing the art of simulation in the social science. In R. Conte, R. Hegselmann, & P. Terna (Eds.), *Simulating social phenomena* (pp. 21–40). Berlin: Springer.

Bailey, T. C., & Gatrell, A. C. (1995). *Interactive spatial data analysis*. Burnt Mill, Essex: Longman Scientific & Technical.

Barreteau, O., Bousquet F., & Attonaty J. M. (2001). Role-playing games for opening the black box of multi-agent systems: Method and lessons of its application to Senegal River Valley irrigated systems. *Journal of Artificial Societies and Social Simulation*, *4*(2). http://jasss.soc.surrey.ac.uk/4/2/5.html

Barreteau, O. et al. (2003). Our companion modelling approach. *Journal of Artificial Societies and Social Simulation*, *6*(2). http://jasss.soc.surrey.ac.uk/6/2/1.html

Barthélémy, O. T. (2006). *Untangling scenario components with agent based modelling: An example of social simulations of water demand forecasts*. Unpublished PhD thesis, Centre for Policy Modelling, Manchester Metropolitan University, Manchester.

Bearman, P. S., Moody, J., & Stovel, K. (2004). Chains of affection: The structure of adolescent romantic and sexual networks. *The American Journal of Sociology, 110*(1), 44–91.

Becu, N., Perez, P., Walker, A., Barreteau, O., & Le Page, C. (2003). Agent based simulation of a small catchment water management in northern Thailand: Description of the CATCHSCAPE model. *Ecological Modelling, 170*(2–3), 319–331.

Boer, P., Huisman, M., Snijders, T. A. B., Steglich, C. H., Wichers, L. H. Y., & Zeggelink, E. P. H. (2006). *StOCNET: An open software system for the advanced statistical analysis of social networks, version 1.7*. Groningen: ICS SciencePlus.

Borgatti, S., Everett, M., & Freeman, L. (2004). *UCINET: Software for social network analysis*. analytictech.com/ucinet

Borgatti, S. P., Carley, K., & Krackhardt, D. (2006). On the robustness of centrality measures under conditions of imperfect data. *Social Networks, 28*(2), 124–136.

Boudon, R. (1998). Social mechanisms without black boxes. In P. Hedström & R. Swedberg (Eds.), *Social mechanisms: An analytical approach to social theory*. Cambridge: Cambridge University Press.

Brandes, U., Eiglsperger, M., Kaufmann, M., & Lerner, J. (2004). *The GraphML file format.* graphml.graphdrawing.org

Brown, D. G. (2006). Agent-based models. In H. Geist (Ed.), *The earth's changing land: An encyclopedia of land-use and land-cover change* (pp. 7–13). Westport: Greenwood Publishing Group.

Bruch, E. E., & Mare, R. D. (2006). Neighborhood choice and neighborhood change. *The American Journal of Sociology, 112*(4), 667–709.

Carley, K. (2003). Dynamic network analysis. In R. Breiger, K. Carley, & P. Pattison (Eds.), *Dynamic social network modeling and analysis: Workshop summary and papers*. Washington, DC: The National Academies Press.

Carley, K. M., Columbus, D., Reno, M., Reminga, J., & Moon, I.-C. (2007). *ORA user's guide 2007* (Technical Report, CMU-ISRI-07-115). Carnegie Mellon University, School of Computer Science, Institute for Software Research.

Cederman, L.-E. (2001). Agent-based modeling in political science. *The Political Methodologist, 10*(1), 16–22.

Cioffi-Revilla, C., & Osman, H. (2009). *A theoretical model of individual radicalization: Understanding complexity in terrorism and insurgency'*. Poster prepared for the 2009 annual European Social Simulation Association conference, University of Surry, Guildford.

Crooks, A. T. (2010). Constructing and implementing an agent-based model of residential segregation through vector GIS. *International Journal of Geographical Information Science, 24*(5), 661–675.

Crooks, A. T., & Castle, C. (2012). The integration of agent-based modelling and geographical information for geospatial simulation. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 219–252). Dordrecht: Springer.

Crooks, A. T., Castle, C. J. E., & Batty, M. (2008). Key challenges in agent-based modelling for geo-spatial simulation. *Computers, Environment and Urban Systems, 32*(6), 417–430.

De Nooy, W., Mrvar, A., & Batagelj, V. (2005). *Exploratory social network analysis with Pajek*. Cambridge: Cambridge University Press.

Degenne, A., & Forsé, M. (1999). *Introducing social networks* (A. Borgess, Trans.). London: Sage.

DiMaggio, P. (1997). Culture and cognition. *Annual Review of Sociology, 23*, 263–287.

Dunham, J. B. (2005). An agent-based spatially explicit epidemiological model in MASON'. *Journal of Artificial Societies and Social Simulation*, *9*(1). http://jasss.soc.surrey.ac.uk/9/1/3.html

Edmonds, B. (1999). *Capturing social embeddedness: A constructivist approach, adaptive behavior* (Vol. 7, pp. 323–348). London: Sage.

Edmonds, B. (2006). How are physical and social spaces related? – Cognitive agents as the necessary "glue". In F. C. Billari, T. Fent, A. Prskawetz, & J. Scheffran (Eds.), *Agent-based computational modelling: Applications in demography, social, economic and environmental sciences* (pp. 195–214). Berlin: Springer.

Edmonds, B., & Chattoe, E. (2005). When simple measures fail: Characterising social networks using simulation. In *Social Network Analysis: Advances and Empirical Applications Forum*, Oxford. <http://cfpm.org/cpmrep158.html>

Erdös, P., & Rényi, A. (1959). On random graphs I. *Publicationes Mathematicae, 6*, 290–297.

Eubank, S., Guclu, H., Kumar, A., Marathe, M. V., Srinivasan, A., Toroczkai, Z., & Wang, N. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature, 429*, 180–184.

Ferber, J. (1999). *Multi-agent systems: An introduction to distributed artificial intelligence*. London: Addison-Wesley.

Flache, A., & Hegselmann, R. (2001). Do irregular grids make a difference? Relaxing the spatial regularity assumption in cellular models of social dynamics. *Journal of Artificial Societies and Social Simulation*, *4*(4). http://jasss.soc.surrey.ac.uk/4/4/6/html

Fortunato, S. (2009). Community detection in graphs, *Physics Reports.* http://dx.doi.org/10.1016/j.physrep.2009.11.002

Geller, A., & Moss, S. (2008). Growing *qawm*: An evidence-driven declarative model of Afghan power structures. *Advances in Complex Systems, 11*(2), 321–335.

Granovetter, M. S. (1973). The strength of weak ties. *The American Journal of Sociology, 78*(6), 1360–1380.

Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *The American Journal of Sociology, 91*(3), 481–510.

Hales, D., & Arteconi, S. (2008). Motifs in evolving cooperative networks look like protein structure networks. *Networks and Heterogeneous Media, 3*(2), 239–249.

Hedström, P. (2005). *Dissecting the social: On the principles of analytical sociology*. Cambridge: University of Cambridge Press.

Hoffer, L. (2006). *Junkie business: The evolution and operation of a heroin dealing network*. Belmont: Thomson Wadsworth Publishing.

Huang, C.-Y., Sun, C.-T., Hsiehm J.-L., & Lin, H. (2004). Simulating SARS: Small-world epidemiological modeling and public health policy assessments. *Journal of Artificial Societies and Social Simulation*, *7*(4). http://jasss.soc.surrey.ac.uk/7/4/2.html

Huang, C.-Y., Tsai, Y.-S., & Wen, T.-H. (2010). Network-based simulation architecture for studying epidemic dynamics. *Simulation, 86*(5-6), 351–368. (May 2010) doi=10.1177/0037549709340733 http://dx.doi.org/10.1177/0037549709340733.

Jensen, H. (1998). *Self-organized criticality: Emergent complex behavior in physical and biological systems*. Cambridge: University of Cambridge Press.

Jin, E. M., Girvan, M., & Newman, M. E. J. (2001). Structure of growing social networks. *Physical Review E, 64*(4).

Krebs, F., Elbers, M., & Ernst, A. (2007). Modelling social and economic influences on the decision making of farmers in the Odra Region. In F. Amblard (Ed.), *Proceeding of the 2007 Annual European Social Simulation Association*, Toulouse.

Leskovec, J., Kleinberg, J., & Faloutsos, C. (2005). Graphs over time: Densification laws, shrinking diameters and possible explanations. In *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining,* Chicago.

Luke, S., Cioffi-Revilla, C., Panait, L., Sullivan, K., & Balan, G. (2005). MASON: A multi-agent simulation environment. *Simulation, 81*(7), 517–527.

Matthews, R. B., Gilbert, N. G., Roach, A., Polhill, J. G., & Gotts, N. M. (2007). Agent-based land-use models: A review of applications. *Landscape Ecology, 22*, 1447–1459.

McCulloh, I. (2009). *Detecting changes in a dynamic social network*. Unpublished PhD thesis, Carnegie Mellon University, Pittsburgh.

Milo, R., Shen-Orr, S., Itzkovitz, S., et al. (2002). Network motifs: Simple building blocks of complex networks. *Science, 298*, 824–827.

Minar N., Burkhart R., Langton C., & Askenazi M. (1996). *The Swarm simulation system: a toolkit for building multiagent simulations* (Working paper 96-06-042). Santa Fe: Santa Fe Institute.

Mitchell, C. (1989). Ethnography and network analysis. In T. Schweizer (Ed.), *Netzwerkanalyse: Ethnologische Perspektiven*. Berlin: Reimer.

Moss, S. (2002). Policy analysis from first principles. *Proceedings of the National Academy of Sciences, 99*(suppl. 3), 7267–7274.

Moss, S. (2008). Simplicity, generality and truth in social modelling part 1: Epistemology. In T. Takama, G. Deffuant, & C. Cioffi-Rivella (Eds.), *Proceedings of the second world congress on social simulation*. Berlin: Springer.

Moss, S., & Edmonds, B. (2005). Sociology and simulation: Statistical and qualitative cross-validation. *The American Journal of Sociology, 110*(4), 1095–1131.

Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., & Onnela, J.-P. (2010). Community structure in time-dependent, multiscale, and multiplex networks, *328*(5980), 876–878. doi=10.1126/science.1184819 http://arxiv.org/abs/0911.1824

Newman, M. E. J. (2004). The structure and function of complex networks. *SIAM Review, 45*, 167–256.

Nikolai, C., & Madey, G. (2009). Tools of the trade: A survey of various agent-based modeling platforms. *Journal of Artificial Societies and Social Simulation*, *12*(2). http://www.sciencemag.org/content/328/5980/876; http://www.sciencemag.org/content/328/5980/876

North, M. J., & Macal, C. M. (2007). *Managing business complexity: Discovering strategic solutions with agent-based modelling and simulation*. New York: Oxford University Press.

O'Madadhain, J., Fisher, D., Smyth, P., White, S., & Boey, Y.-B. (2005). Analysis and visualization of network data using JUNG. *Journal of Statistical Software, VV*(II), 1–35.

Pahl-Wostl, C., & Hare, M. (2004). Processes of social learning in integrated resources management. *Journal of Community & Applied Social Psychology, 14*(3), 193–206.

Parker, D. C. (2005). Integration of geographic information systems and agent-based models of land use: Challenges and prospects. In D. J. Maguire, M. F. Goodchild, & M. Batty (Eds.), *GIS, Spatial analysis and modeling* (pp. 403–422). Redlands: ESRI Press.

Polhill, J. G., Parker, D. C., & Gotts, N. M. (2008). Effects of land markets on competition between imitators and innovators: Results from FEARLUS-ELMM. In B. Edmonds, C. Hernandez, & K. Troitzsch (Eds.), *Social simulation: Technologies, advances and new discoveries* (pp. 81–97). Hershey: GI Global.

Pujol, J. M., Flache, A., Delgado, J., & Sangüesa, R. (2005). How can social networks ever become complex? Modelling the emergence of complex networks from local social exchanges'. *Journal of Artificial Societies and Social Simulation*, *8*(4). http://jasss.soc.surrey.ac.uk/8/4/12.html

Rakowski, F., Gruziel, M., Krych, M., & Radomski, J. P. (2010). Large scale daily contacts and mobility model – An individual-based countrywide simulation study for Poland. *Journal of Artificial Societies and Social Simulation, 13*(1), 13. http://jasss.soc.surrey.ac.uk/13/1/13.html

Schelling, T. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology, 1*(2), 143–186.

Schensul, J. J., LeCompte, M., Trotter, P., & Singer, M. (1999). *Mapping social networks, spatial data & hidden populations*. Thousand Oaks: Sage.

Simon, H. (1955). On a class of skew distribution functions. *Biometrika, 42*(3–4), 425–440.

Snijders, T. A. B., van de Bunt, G. G., & Steglich, C. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks, 32*(1), 44–60.

Stroud, P., Del Valle, S., Sydoriak, S., Riese, J., & Mniszewski, S. (2007). Spatial dynamics of pandemic influenza in a massive artificial society. *Journal of Artificial Societies and Social Simulation*, *10*(4). http://jasss.soc.surrey.ac.uk/10/4/9.html

Tsvetovat, M. (2005). *Social structure simulation and inference using artificial intelligence techniques*. Unpublished PhD thesis, Carnegie Mellon University, Pittsburgh.

Wasserman, S., & Faust, K. (1994). *Social network analysis: methods and applications*. Cambridge: Cambridge University Press.

Watts, D. J. (1999). Networks, dynamics, and the small-world phenomenon. *The American Journal of Sociology, 13*(2), 493–527.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature, 393*, 440–442.

Werth, B., & Moss, S. (2007). Modelling migration in the Sahel: An alternative to cost-benefit analysis. In S. Takahashi, D. Sallach, & J. Rouchier (Eds.), *Advancing social simulation: The first world congress* (pp. 331–342). Berlin: Springer.

Wilensky, U., & Rand, W. (in press). An introduction to agent-based modeling: Modeling natural, social and engineered complex systems with NetLogo. Cambridge, MA: MIT Press.

Yang, Y., & Atkinson, P. M. (2008). Individual space – Time activity-based model: A model for the simulation of airborne infectious-disease transmission by activity-bundle simulation. *Environment and Planning B, 35*(1), 80–99.

Ziervogel, G., Bharwani, S., & Downing, T. E. (2006). Adapting to climate variability: Pumpkins, people and policy. *Natural Resource Forum, 30*(4), 294–305.

Part III
Methods, Techniques and Tools for the Design and Construction of Agent-Based Models

Chapter 12
The Integration of Agent-Based Modelling and Geographical Information for Geospatial Simulation

Andrew T. Crooks and Christian J.E. Castle

Abstract Within this chapter we focus on the integration of Geographical Information System (GIS) and Agent-based modelling (ABM) and review a selection of toolkits which allow for such integration. Moreover, we identify current capabilities of modelling within a GIS and methods of coupling and integrating GIS with agent-based models. We then introduce suggested guidelines for developing geospatial simulations with ABM toolkits and offer practical guidelines for choosing a simulation/modelling system before providing a review of a number of simulation/modelling systems that allow for the creation of geospatial agent based models along with the identification of a number references for further information.

12.1 Introduction

The Agent-Based modelling (ABM) paradigm is developing into a powerful tool in many disciplines as seen in Crooks and Heppenstall (2012), Johansson and Kretz (2012) and Harland and Heppenstall (2012), but also in a other disciplines such as archaeology (Axtell et al. 2002), economics (Tesfatsion and Judd 2006), health (Epstein 2009), geography (Batty 2005) and computational social science more generally (see Cioffi-Revilla 2010 for a discussion). Such models allow researchers

A.T. Crooks (✉)
Department of Computational Social Science, George Mason University, Research 1,
4400 University Drive, Fairfax, VA 22030, USA
e-mail: acrooks2@gmu.edu

C.J.E. Castle
Mott MacDonald Group, Mott MacDonald House, 8-10 Sydenham Road, Croydon,
Surrey CR0 2EE, UK
e-mail: Christian.Castle@mottmac.com

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_12,

to explore how through the interaction of many individuals more emergent phenomena arise. Moreover, it allows for practitioners to build models of complex social phenomenon by simulating the interactions of the many actors in such systems. Thus gaining insights that will lead to greater understanding and, in some cases, better management of the behaviour of complex social systems. The intention of this chapter is to the outline how one can develop geospatial agent-based models (i.e. that model spatially explicit geographic phenomena – where the nature of the features and movement that is represented varies over the Earth's surface). Essentially, geospatial models depend on the location of the features or phenomena being modelled, such that if one or more of those locations change, the results of the model change (Wegener 2000). Geographical Information Systems (GIS) are a particularly useful medium for representing model input and output of a geospatial nature. However, GIS are not well suited to dynamic modelling (Goodchild 2005; Maguire 2005) as will be discussed in Sect. 12.2. Consequently, Sect. 12.2.2 explores the opportunity of linking (through coupling or integration/embedding) a GIS with a simulation/modelling system purposely built for the task at hand (Sect. 12.3), and therefore better suited to supporting the requirements of ABM.

12.2 Modelling Within GIS: Current Capabilities

It can be difficult to comprehend how GIS technology, built essentially for handling maps and "map-related ideas", can be adapted to the needs of dynamic simulation modelling; especially when it is not even perceived as an optimal platform for modelling (Goodchild 2005). Particular criticisms of GIS with respect to modelling is their ability to handle time (Langran 1992; Peuquet 2005– see Sect. 12.2.1), the representation of continuous variation (Longley et al. 2005), and most have only rudimentary modelling capabilities (Maguire 2005). Nevertheless, there are several good reasons to justify why the use, or linkage of GIS with simulation/modelling systems (see Sect. 12.2.2), is an effective means of modelling when spatial and temporal analysis is necessary.

Current commercial and public domain GIS software systems all contain numerous tools for acquiring, pre-processing, and transforming data. Their use in modelling includes data management, format conversion, projection change, re-sampling, raster-vector conversion, etc. GIS also include excellent tools for visualisation/mapping, rendering, querying, and analysing model results, as well as assessing the accuracies and uncertainties associated with inputs and outputs.

Typically, all of the capabilities described above are accessible via end-user graphical and command line interfaces. However, these capabilities have recently become accessible through application programming interfaces (APIs), via software libraries. The exposure of APIs was a significant recent improvement in terms of GIS and spatial modelling, as external programmers now have access to the underlying software components upon which GIS software vendors base their end-user versions of systems. This is perhaps the most pertinent enhancement, as many of the techniques used in GIS analysis are potentially far more robust if they can be linked with an extensive toolkit of methods for simulation; an issue which is addressed at greater length later in Sect. 12.2.2. GIS vendors have invited this situation as it allows GIS to be extended and customised for use in new application areas, thus expanding the market potential of their systems.

Alternatively, a model can be expressed as a sequence of GIS commands executed by a script (Maguire 2005). Recently in GIS there has been a move to use industry-standard low-level programming languages (e.g. Java, C++, and Visual Basic), and scripting languages (e.g. Python, VBScript, and Jscript), rather than proprietary, home grown scripting languages (e.g. ESRI's Arc Macro Language, AML, or Avenue). Interoperability standards such as the Microsoft. Net framework facilitate this process by allowing compliant packages to be called from the same script.

In addition to scripts, graphical flowcharts can be used to express sequences of operations that define a model. Longley et al. (2005) note that one of the first graphic platforms for conceptualising and implementing spatial models was probably the ERDAS IMAGINE software, which allows the user to build complex modelling sequences from primitive operations. ESRI is another GIS vendor that provides an environment that allows models to be authored and executed in a graphical environment: ModelBuilder within ArcGIS 9.x, which superseded Spatial Modeller within ArcView 3.

In principle, graphic-model building can be used for dynamic modelling via an iterative process, where the output of one time step becomes the input for the next. However, this method posses two dilemmas: (1) the GIS will not have been designed for an iterative process, requiring the user to re-enter the data at the beginning of each time step, and; (2) the time required to run a model could be considerable. The former of these problems can be overcome with scripting languages (e.g. Python in ArcGIS); both can potentially be overcome by integrating the GIS with a simulation/modelling system better equipped for the task at hand. Before exploring the possibilities of linking GIS and simulation/modelling systems (Sect. 12.2.2), the following section of this chapter evaluates the capability of GIS to handle space-time information, which computer simulations generate in volume, and has always been a limitation.

12.2.1 Representing Time and Change Within GIS

The subject of time within GIS has received a considerable amount of attention. Heywood et al. (2006) comments that ideally, GIS would be able to represent temporal change using methods that explicitly represent spatial change, as well as different states through time. Furthermore, methods allowing direct manipulation and comparison of simulated or observational data in a temporal and spatial dimensions should be catered for. In reality, two main challenges for the integration of time within GIS exist: (1) continuous data over a period of time are rarely available for an entity or system of interests; (2) data models and structures able to record, store, and visualise information about an object in different temporal states are still in their infancy (Heywood et al. 2006). In the context of this chapter, the former challenge is less of a constraint since an agent-based computer simulation is capable of generating an abundance of data over a continuous period of time, while much progress has been made on the later issue. The following discussion outlines issues related to the representation of time and change, as well as approaches for incorporating space-time information within GIS.

The basic objective of any temporal database is to record change over time, where change can be thought of as an event or collection of events. An event might be a change in state of one or more locations, entities, or both. Changes that might affect an event can be distinguished in terms of their temporal pattern; Peuquet (2005) has suggested four types: (1) continuous – events occurring throughout some period of time; (2) majorative – events occurring most of the time; (3) sporadic – events occurring some of the time, and; (4) unique – events that only occur once. The distribution of events within these temporal patterns can also be very complex (e.g. chaotic, cyclic, or steady state), complicated further as change, to some extent, is always occurring at various rates as well (e.g. from sudden to gradual). Hence, duration and frequency are important descriptive characteristics within this taxonomy of temporal patterns.

There are three approaches for capturing space-time information within a GIS: (1) location-based; (2) time-based, and; (3) entity-based. The only method of viewing a data model within existing GIS, as a space-time representation, is as a temporal series of spatially-registered 'snapshots' (Peuquet 2005). Invariably this approach employs a raster data model, although vector has also been used, with only a single information type stored (e.g. elevation, density, precipitation, etc.) for each cell at any one point in time. Information for the entire layer is stored for each time step, regardless of whether change has occurred since the previous step. There are several criticisms of this approach. Firstly, the data volume increases enormously, because redundant data is stored in consecutive snapshots. The state of a spatial entity can only be retrieved by querying cells of adjacent snapshots, because information is stored implicitly between each time step. Finally, the exact point when change has occurred cannot be determined. Langran (1992) has proposed a modification of this approach. The temporal-raster (or grid) approach allows multiple values to be stored for each pixel. A new value, and the time at which change occurred for each pixel is stored, which can result in a variable number of records for each cell. Recording the time at which change has occurred allows for values to be sorted by time. The most recent value for each cell can therefore be retrieved, which represents the present state of the system. The obvious advantage to this approach is the reduction of redundant data stored for each cell.

Peuquet and Duan (1995) have proposed a time-based approach to storing space-time information within a GIS, where change is stored as a sequence of events through time. Time is stored in increasing order from an initial point, with the temporal interval correlating to successive events. An event is recorded at the time when the amount of accumulated change is considered significant, or by another domain-specific rule. This type of representation has the advantage of facilitating time-based queries, and the addition of a new event is straight forward as it can simply be added to the end of the timeline. Furthermore, in terms of modelling an important capacity of any model is the ability to represent alternative versions of the same reality. The concept of representing multiple realities over time is called branching. Branching allows various model simulation runs to be compared, or simulation results to be compared to observed data. The time-based approach facilitates the branching of time in order to represent alternative or parallel sequences of events resulting from specific scenarios, because it is strictly an ordinal timeline.

Finally, several entity-based space-time models have been proposed. Conceptually these models extend the topological vector approach (e.g. coverage model); tracking changes in the geometry of entities incrementally through time. The amendment vector model was the first of this type, and extended frameworks have been proposed subsequently. Besides maintaining the integrity of entities and their changing topology, these approaches are able to represent asynchronous changes to entity geometries. However, the space-time topology of these vectors becomes increasingly complex as amendments accumulate through time. In addition, aspatial entity attributes can change over time. To record aspatial changes, a separate relational database is often used. However, if change occurs at a different rate between the spatial and aspatial aspects of an entity, maintaining the identity of individual entities becomes difficult, especially when entities split or merge.

Object-oriented data models have transformed the entity-based storage of space-time information within GIS (Zeiler 1999), and have become mainstream within commercial GIS (e.g. the geodatabase structure with ArcGIS). They have grown increasingly more sophisticated, catering for a powerful modelling environment. The object-oriented data model approach provides a cohesive representation that allows the identity of objects, as well as complex interrelationships to be maintained through time. Specifically, temporal and location behaviour can be assigned as an attribute of features rather than the space itself, which has the distinct advantage of allowing objects to be updated asynchronously. Despite the advantages of the object-oriented data model, Reitsma and Albrecht (2006) observe that, to date, no data model or data structure allows the representation of processes (i.e. recording a process that has changed the state of an object within a model).[1] Consequently, queries about where a process is occurring at an instant of time cannot be expressed with these current approaches. Notwithstanding, object-oriented data models are the canonical approach to the storage of space-time data generated by agent-based models, and their visualisation within GIS, given their complementarities. Nevertheless, the visualisation of agent-based models within GIS is still limited to a temporal series of snapshots.

[1] However this is an active research topic and holds much promise with respect to creating geospatial agent-based models (see Torrens 2009 for a more detailed discussion).

12.2.2 Linkage – Coupling Versus Integration/Embedding

Models implemented as direct extensions of an underlying GIS, through either graphic model-building or scripts, generally make two assumptions: (1) all operations required by the model are available in the GIS (or in another system called by the model); and, (2) the GIS provides sufficient performance to handle the execution of the model (Longley et al. 2005). In reality, a GIS will often fail to provide adequate performance, especially with very large datasets and a large number of iterations, because it has not been designed as a simulation/modelling engine. This one-size-fits-all approach inherent in GIS provides limited applicability, and attention has subsequently been devoted to linking, either through coupling or integration/embedding, GIS with simulation/modelling systems more directly suited to users needs. General classifications have been produced by numerous authors (e.g. Maguire 1995; Bernard and Krüger 2000; Westervelt 2002; Goodchild 2005; Longley et al. 2005; Maguire 2005). Several of their definitions now overlap as technological advance has blurred the boundaries of their classifications, whist some definitions are convoluted because terminology has been used interchangeably or sometimes inappropriately (e.g. coupling, linkage or integration). Nevertheless, categorisation of these techniques is possible, and a brief description of each is developed below, in an attempt to clarify the situation. This is followed by a critique of these different approaches, with a view to identifying an appropriate method for developing geospatial agent-based models.

In situations where GIS and simulation/modelling systems already exist (e.g. as commercial products), or the cost of rebuilding the functionality of one system into another is too great, the systems can be coupled together (Maguire 2005). Coupling can therefore be broadly defined as the linkage of two stand-alone systems by data transfer. Three types of coupling are distinguishable, although these are only a subset of the much larger fields of enterprise application integration (Linthicum 2000) and software interoperability (Sondheim et al. 2005). The attributes of each approach cascaded along the coupling continuum, from loose to tight/close (Table 12.1 summaries the competing objectives of the different coupling approaches; greyed boxes are considered more desirable characteristics – adapted from Westervelt 2002):

1. **Loose Coupling.** A loose connection usually involves the asynchronous operation of functions within each system, with data exchanged between systems in the form of files. For example, the GIS might be used to prepare inputs, which are then passed to the simulation/modelling system, where after execution the results of the model are returned to the GIS for display and analysis. This approach requires the GIS and simulation/modelling system to understand the same data format; if no common format is available an additional piece of software will be required to convert formats in both directions. Occasionally, specific new programmes must be developed to perform format modifications;
2. **Moderate Coupling.** Essentially this category encapsulates techniques between loose and tight/close coupling. For example, Westervelt (2002) advocates remote procedure calls and shared database access links between the GIS and

Table 12.1 Comparison of coupling approaches (Adapted from Westervelt 2002)

Objective and explanation	Loose	Moderate	Close/tight
Integration Speed: The programmer time involved in linking the programmes	Fast	Medium	Slow
Programmer Expertise: Required level of software development expertise	Low	High	Medium
Multiple Authorship Avoidance: In some instances it might be necessary for the programmer to modify the original software product. Any alteration reduces the ownership responsibility. Major alterations could totally sever this link, resulting in limited or no support by the original author(s)	High	Medium	Low
Execution Speed: How rapidly does the integrated software execute?	Slow	Medium	Fast
Simultaneous Execution: Can components of the system run simultaneously and communicate with one another? Can the components operate on separate platforms?	Low	Low	High
Debugging: How difficult is it to locate execution errors in the linked system?	Easy	Moderate	Hard

simulation/modelling system, allowing indirect communication between the systems. Inevitably, this reduces the execution speed of the integrated system, and decreases the ability to simultaneously execute components belonging to the different software; and,

3. **Tight or Close Coupling.** This type of linkage is characterised by the simultaneous operation of systems allowing direct inter-system communication during the programme execution. For example, standards such as Microsoft's COM and .NET allow a single script to invoke commands from both systems (Ungerer and Goodchild 2002). A variant of this approach allows inter-system communication by different processes that may be run on one of more networked computers (i.e. distributed processing).

Coupling has often been the preferred approach for linking GIS and simulation/modelling systems. However, this has tended to result in very specialised and isolated solutions, which have prevented the standardisation of general and generic linkage. An alternative to coupling is to embed or to integrate the required functionality of either the GIS or simulation/modelling system within the dominant system using its underlying programming language (Maguire 2005). The final system is either referred to as GIS-centric or modelling-centric depending on which system is dominant. In both instances, the GIS tools or modelling capabilities can be executed by calling functions from the dominant system, usually through a graphical user interface (GUI). Compared to coupling, an embedded or integrated system will appear seamless to a user (Maguire 1995). However, in the past integration has been based on existing closed and monolithic GIS and simulation systems, which poses a risk of designing systems that are also closed, monolithic, and therefore costly (Fedra 1996).

Interest in modelling-centric systems has increased considerably over recent years, predominately due to the development of simulation/modelling toolkits with scripting capabilities that do not require advanced computer programming skills (Gilbert and Bankes 2002). Often the simulation/modelling toolkit can access GIS functions, such as data management and visualisation capabilities, from a GIS software library. For example, the RepastJ (see Sect. 12.3.3.3) toolkit exploits functions from GeoTools (a Java GIS software library) for importing and exporting data, Java Topology Suite (JTS) for data manipulation, and OpenMap for visualisation. The toolkit itself maintains the agents and environment (i.e. their attributes), using identity relationships for communication between the different systems. Functions available from GIS software libraries reduce the development time of a model, and are likely to be more efficient because they have been developed over many years with attention to efficiency. Additionally, the use of standard GIS tools for spatial analysis improves functional transparency of a model, as it makes use of well known and understood algorithms. Alternatively, spatial data management and analysis functions can be developed within the modelling toolkit, although this strategy imposes huge costs, in terms of time to programme the model, and time required to frequently update spatial data or use spatial analysis functions within the model.

Conversely, the GIS-centric approach is an attractive alternative; not least because the large user-base of some GIS expands the potential user-base for the final model. Analogous to the modelling-centric approach, GIS-centric integration can be carried out using software libraries of simulation/modelling functions accessed through the GIS interface. There are many examples of simulation/modelling systems integrated within commercial GIS, including: the Consequences Assessment Tool Set (2011, CATS) system, designed for emergency response planning; the Hazard Prediction and Assessment Capability (2004, HPAC) system, for predicting the effect of hazardous material releases into the atmosphere; the NatureServe Vista (2011) system, for land use and conservation planners.

Brown et al. (2005) propose an alternative approach which straddles both the GIS-centric and modelling-centric frameworks. Rather than providing functionality within one system, the middleware-based approach manages connections between systems, allowing a model to make use of the functionality available within the GIS or the simulation/modelling toolkit most appropriate for a given task. Thus, the middleware approach allows the simulation/modelling toolkit to handle the identity and relationship of, and between agents and their environment. Conversely, the GIS would manage spatial features, as well as temporal and topological relationships of the model. Essentially, the simulation/modelling toolkit handles what it is designed for (i.e. implementing the model), while the GIS can be used to run the model, and visualise the output. An example of this approach is the ABM extension within ArcGIS (referred to as Agent Analyst), which allows users to create, edit, and run RepastPy models from within ArcGIS (Redlands Institute 2010). However, it is the opinion of the authors that only a dichotomy of integration classifications exists. A GIS is either integrated into a simulation/modelling toolkit, or vice versa. The definition of the middleware approach is essentially tight coupling (see above).

12.3 Developing Geospatial Simulations with Agent-Based Modelling Toolkits

The process of building an agent-based model begins with a conceptual model, where basic questions or goals, elements of the system (e.g. agent attributes, rules of agent interaction and behaviour, the model environment, etc.), and the measurable outcomes of interest are identified (Brown 2006). It is important to 'ground' a model during the conceptualisation process (i.e. establish whether simplifications made during the design process do not seriously detract from the credibility and likelihood that the model will provide important insights; Carley 1996). It is usual for a modeller to set forth a claim as to why the proposed model is reasonable. This claim will be enhanced if the applicability of the model is not over stated, and by defining the models limitations and scope. Grounding can be reinforced by demonstrating that other researchers have made similar or identical assumptions in their models, and by justifying how a proposed model will be of benefit in relation to pre-existing models.

Conceptualising the fundamental aspects of an agent-based model (i.e. one or more agents interacting within an environment), juxtaposed with the distinction between explanatory vs. predictive purposes of a model suggests a fourfold typology of agent and environment types (Table 12.2). Couclelis (2001) classifies agents and their environment as either being designed (i.e. explanatory) or analysed (i.e. predictive – empirically grounded). If designed, agents are endowed with attributes and behaviours that represent (often simplified) conditions for testing specific hypotheses about general cases. Analysed agents are intended to accurately mimic real-world entities, based on empirical data or ad hoc values that are realistic substitutes for observed processes. Similarly, the environment that agents are situated within can be designed (i.e. provided with characteristics that are simplified to focus on specific agent attributes), or analysed (i.e. represent a real-world location).

The boundary between designed and analyzed is not always distinct, especially when ad hoc data are employed. Subtle but profound differences, both practical and conceptual, exist between the design or analysis approach of developing agents and their environment. A major difference in practical terms is that designing something provides direct (partial or total) control over the outcome, whereas there can only be hope that something has been analyzed correctly (Couclelis 2001). Table 12.2 provides further details to consider when developing agents and their environment; including a brief description of the model, the purpose and intent of the model (see Parker et al. 2001), verification and validation strategies used to assess the model outputs (see Parker et al. 2001; Crooks et al. 2008), and appropriate software for the development of a model (see Sect. 12.3.2).

Once a model has been conceptualised, it must be formalised into a specification which can be developed into a computer programme (Grimm and Railsback 2012 and Abdou et al. 2012 offer constructive advice on this); if the model is required to be run as a computer simulation. The process of formalisation involves being precise about what an identified theory relating to a phenomena of interest means,

Table 12.2 Description, purpose/intent, verification and validation strategies, and appropriate development tools for agent-based models incorporating designed or analysed agents/environments (Adapted from Berger and Parker 2001)

		Agent	
		Designed	Analysed
Environment	Designed	Model description – Abstract Purpose/intent – Discovery of new relationships – Existence proof Verification and validation strategy – Theoretical comparison – Replication Appropriate development tools – Easy to implement simulation/modelling system Example model – Filatova et al. (2009)	Model description – Experimental Purpose/intent – Role-playing games among stakeholders – Laboratory experiments Verification and validation strategy – Repetitions – Adequacy of design Appropriate development tools – Flexible simulation/modelling systems with well developed user interfaces Example model – Mooij et al. (2002)
	Analysed	Model description – Historical Purpose/intent – Explanation Verification and validation strategy – Qualitative: goodness of fit Appropriate development tools – Advanced simulation/modelling systems linked with GIS Example model – Mathevet et al. (2003)	Model description – Empirical Purpose/intent – Explanation – Projection – Scenario analysis Verification and validation strategy – Quantitative: goodness of fit Appropriate development tools – Low-level programming languages Example model – Jackson et al. (2008)

making sure that it is complete and coherent. There are several reasons why computer simulation is more appropriate for formalising social science theories than mathematics, which has often been used in the social sciences (Gilbert and Troitzsch 2005). First, programming languages are more expressive and less abstract than most mathematical techniques. Second, a computer simulation can deal more easily with parallel process and processes without well defined order or actions than systems of mathematical equations. Third, a computer model can include heterogeneous agents (e.g. pedestrians with varying degrees of knowledge about a building layout), while this is usually relatively difficult using mathematics. Finally, computer programmes are (or can easily be made to be) modular, so that major changes can be made to one part of the model without requiring large changes in other parts of the programme, an ability which mathematical systems often lack.

The object-oriented paradigm provides a very suitable medium for the development of agent-based models. In particular, it provides the aforementioned modularity useful for developing a computer simulation. It is not the intention of this chapter to outline the fundamental object-oriented concepts, this has been achieved by numerous others (refer to Booch (1994) for a seminal discussion and Armstrong (2006) for a useful evaluation and clarification of key object-oriented notions).

At the time of writing, there are many simulation/modelling systems available to assist the development stage of ABM. The majority of these simulation/modelling systems are programmed, and/or require the user to develop their model in an object-oriented language. The subsequent section of this chapter identifies some of the simulation/modelling systems available for ABM, highlighting key questions that should be considered for a user to determine an appropriate system for their needs.

12.3.1 Types of Simulation/Modelling Systems for Agent-Based Modelling

In general, two types of simulation/modelling systems are available to develop agent-based models: toolkits or software.[2] Based on this dichotomy, toolkits are simulation/modelling systems that provide a conceptual framework for organising and designing agent-based models. They provide appropriate libraries[3] of software functionality that include pre-defined routines/functions specifically designed for ABM. However, the object-oriented paradigm allows the integration of additional functionality from libraries not provided by the simulation/modelling toolkit, extending the capabilities of these toolkits. Of particular interest to this chapter is the integration of functionality from GIS software libraries (e.g. OpenMap, GeoTools, ESRI's ArcGIS, etc.), which provide ABM toolkits with greater data management and spatial analytical capabilities required for geospatial modelling (see Sect. 12.2).

The development of agent-based models can be greatly facilitated by the utilisation of simulation/modelling toolkits. They provide reliable templates for the design, implementation and visualisation of agent-based models, allowing modellers to focus on research (i.e. building models), rather than building fundamental tools necessary to run a computer simulation (see Tobias and Hofmann 2004; Railsback et al. 2006). In particular, the use of toolkits can reduce the burden modellers face programming parts

[2] An agent-based model could be programmed completely from scratch using a low-level programming language if a modeller has sufficient programming knowledge and experience; see below for disadvantages of this approach.

[3] A collection of programming classes grouped together, termed packages (i.e. classes with similar purpose).

of a simulation that are not content-specific (e.g. GUI, data import-export, visualisation/display of the model). It also increases the reliability and efficiency of the model, because complex parts have been created and optimised by professional developers, as standardised simulation/modelling functions. Unsurprisingly, there are limitations of using simulation/modelling systems to develop agent-based models, for example: a substantial amount of effort is required to understand how to design and implement a model in some toolkits; the programming code of demonstration models or models produced by other researchers can be difficult to understand or apply to another purpose; a modeller will have to learn or already have an understanding of the programming language required to use the toolkit; and finally the desired/required functionality may not be present, although additional tools might be available from the user community or from other software libraries. Benenson et al. (2005) also note that toolkit users are accompanied by the fear of discovering that a particular function cannot be used, will conflict, or is incompatible with another part of the model late in the development process.

Probably the earliest and most prominent toolkit was SWARM, although many other toolkits now exist. At the time of writing there are more than 100 toolkits available for ABM (see AgentLink 2007; SwarmWiki 2010; Nikolai and Madey 2009; Tesfatsion 2010; Wikipedia 2010 for comprehensive listings). However, variation between toolkits can be considerable. For example, their purpose (some toolkits have different design objectives e.g. Artificial Intelligence (AI) rather than social science focus, or network opposed to raster or vector model environments), level of development (e.g. some models are no longer supported or have ceased development), and modelling capabilities (e.g. the number of agents that can be modelled, degree of interaction between agents) can vary. A review of all toolkits currently available is beyond the scope of this chapter. However, we identify a selection of noteworthy simulation/modelling toolkits (e.g. Swarm, MASON, Repast, AnyLogic), highlighting their purpose and capabilities, as well as resources providing further information.

In addition to toolkits, software is available for developing agent-based models, which can simplify the implementation process. For example, simulation/modelling software often negates the need to develop an agent-based model via a low-level a programming language (e.g. Java, C++, Visual Basic, etc.). In particular, software for ABM is useful for the rapid development of basic or prototype models. However, modellers using software are restricted to the design framework advocated by the software. For instance, some ABM software will only have limited environments (e.g. raster only) in which to model, or agent neighbourhoods might be restricted in size (e.g. von Neumann or Moore). Furthermore, a modeller will be constrained to the functionality provided by the software (unlike ABM toolkits modellers will be unable to extend or integrate additional tools), especially if the toolkit is written in its own programming language (e.g. NetLogo). Section 12.3.3 identifies a selection of noteworthy software for the development of agent-based models; StarLogo, its derivative NetLogo, and AgentSheets.

12.3.2 Guidelines for Choosing a Simulation/Modelling System

Ideally, a modeller would have comprehensive practical experience in a range of modelling/simulation systems before choosing which system to use for a modelling endeavour. Unfortunately, this is not usually feasible. For this reason several authors (Najlis et al. 2001; Gilbert and Bankes 2002; Serenko and Detlor 2002; Tobias and Hofmann 2004; Dugdale 2004; Rixon et al. 2005; Robertson 2005; Andrade et al. 2008; Berryman 2008; Liebert et al. 2008; Nikolai and Madey 2009) have gained practical experience and/or have surveyed several systems, identifying key criteria that should be considered before making a decision. General criteria include, but are not limited to: ease of developing the model/using the system; size of the community using the system; availability of help or support (most probably from the user community); size of the community familiar with the programming language in which the system is implemented (if a programming language is necessary to implement the model); is the system still maintained and/or updated; availability of demonstration or template models; technical and how-to documentation, etc. Criteria relating specifically to a systems modelling functionality include: number of agents that can be modelled; degree of interaction between agents; ability to represent multiple organisational/hierarchical levels of agents; variety of model environments available (network, raster, and vector); possible topological relationship between agents; management of spatial relationships between agents, and agents with their environment; mechanisms for scheduling and sequencing events, etc. These criteria will be weighted differently depending on a modeller's personal preferences and abilities (e.g. the specification of the model to be developed, programming experience/knowledge, etc.).

Another important distinction separating simulation/modelling systems is there licensing policy; open source, shareware/freeware, or proprietary. Open source simulation/modelling systems constitute toolkits or software whose source code is published and made available to the public, enabling anyone to copy, modify and redistribute the system without paying royalties or fees. A key advantage of open source simulation/modelling systems relates to the transparency of their inner workings. The user can explore the source code, permitting the modification, extension and correction of the system if necessary. This is particularly useful for verifying a model (see Crooks et al. 2008). The predominant open source simulation/modelling systems are toolkits (e.g. MASON, Repast, Swarm, etc.). The distinction between an open source simulation/modelling system and a shareware/freeware system is subtle. There is no one accepted definition of the term shareware/freeware, but the expression is commonly used to describe a system that can be redistributed but not modified, primarily because the source code is unavailable. Consequently, shareware/freeware systems (e.g. StarLogo, NetLogo, etc.) do not have the same flexibility, extendibility or potential for verification (in relation to access to their source code), as open source systems. Similarly, shareware/freeware systems tend to be toolkits,

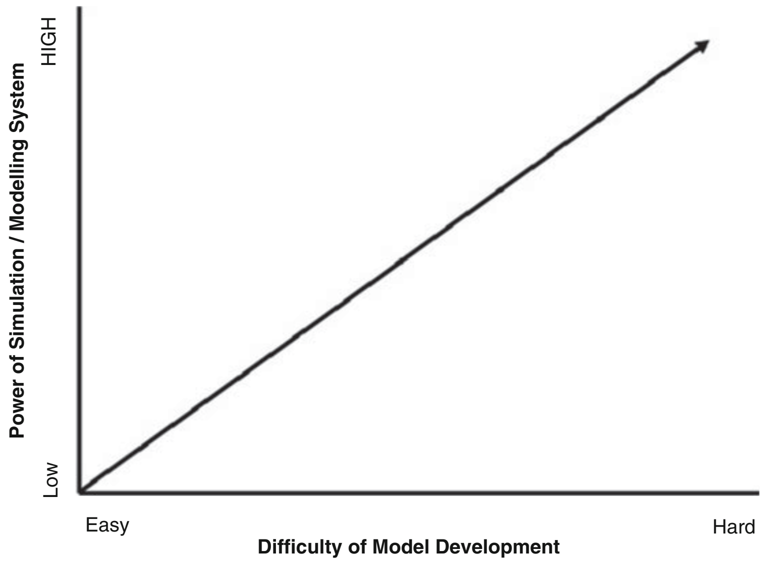


Fig. 12.1 Balance between power versus difficulty of developing a model with a simulation/modelling system

rather than software.[4] Finally, proprietary simulation/modelling systems are available for developing agent-based models. Proprietary systems are mainly software, developed by an organisation who exercises control over its distribution and use; most require a licence at a financial cost to the user. These systems have the advantage of being professionally designed and built for a specific use, and are often relatively simple to use. However, they often lack the community support found with open source or shareware/freeware systems. Moreover, since access to their source code is prohibited, a model developed with proprietary software is essentially black box. A modeller will therefore, to some extent, be left unsure about the inner validity of a model constructed with a proprietary system. This situation is compounded when the output of a model is emergent or unexpected.

Striking a balance between the aforementioned criteria is difficult. Unfortunately, while identifying a suitable system for the development of an agent-based model, too much time can often be expended trying to find this balance. This balance can be perceived as a trade off between the difficulty of developing a model (e.g. in terms of time required to programme the model, understand how to develop a model with a specific system, or acquiring experience and knowledge of a programming language if required, etc.), versus the modelling power provided by the simulation/modelling system (e.g. modelling capabilities and functionality, Fig. 12.1). The key

[4]Other shareware/freeware systems used for the creation of spatial agent-based models include OBEUS (Benenson et al. 2006) and CORMAS (Bousquet et al. 1998). These systems are not reviewed in this chapter for space requirements.

is striking a 'personal' balance between these criteria. For example, those more accustomed to programming may prefer the functionality and flexibility of a simulation/modelling toolkit. However, modellers that only wish to develop a basic or prototype model quickly and easily, possibly with little or no programming skills may prefer to use simulation/modelling software (see Railsback et al. 2006).

12.3.3 Simulation/Modelling Systems for Agent-Based Modelling

This section provides key criteria pertaining to a selection of simulation/modelling systems available for the development of agent-based models (the rationale for each criterion was described in Sect. 12.3.2). Although there are many systems available for developing agent-based models, this chapter reviews seven, separated into three categories of licensing policy (1) open source (Swarm, MASON and Repast); (2) shareware/freeware (StarLogo and NetLogo); and (3) proprietary systems (AgentSheets and AnyLogic). These systems were chosen because they fulfilled the (majority of the) following criteria, they are: maintained and still being developed; widely used and supported by a strong user community; accompanied by a variety of demonstration models and in some instances the model's programming script or source code is available; and finally they are capable of developing spatially explicit models, possibly via the integration of GIS functionality. Tables 12.3–12.5 tabularise information of each system for comparison purposes; categorised by their licensing policy (adapted from Najlis et al. 2001 and Parker 2001). The reminder of this section provides further information about each system, identifying examples of geospatial models that have been developed with the system. A caveat must be noted at this point, the information provided within this section is accurate at the time of publication. However, the systems reviewed are constantly being updated, thus modellers are advised to check each systems website to obtain up to date information.

12.3.3.1 Swarm

Swarm (Table 12.3) is an open source simulation/modelling system designed specifically for the development of multi-agent simulations of complex adaptive systems (Swarm 2010); although agent-based models can easily be develop using Swarm as well. Inspired by artificial life, Swarm was designed to study biological systems; attempting to infer mechanisms observable in biological phenomena (Minar et al. 1996). In addition to modelling biological systems (e.g. Railsback and Harvey 2002), Swarm has been used to develop models for anthropological, computer science, ecological, economic, geographical, and political science purposes. Useful examples of spatially explicit models include: the simulation of pedestrians in the urban centres (Schelhorn et al. 1999 and Haklay et al. 2001); and the examination of crowd congestion at London's Notting Hill carnival (Batty et al. 2003).

Table 12.3 Comparison of open source simulation/modelling systems (Adapted from Najlis et al. 2001 and Parker 2001)

Open source simulation/modelling systems	Swarm	MASON	Repast
Developers	Santa Fe Institute/SWARM Development Group, USA	Evolutionary Computation Laboratory and Center for Social Complexity, George Mason University, USA	University of Chicago, Department of Social Science Research Computing, USA
Date of inception	1996	2003	2000
Website	http://www.swarm.org	http://cs.gmu.edu/~eclab/projects/mason	http://repast.sourceforge.net
E-mail list	http://www.swarm.org/mailman/listinfo	https://listserv.gmu.edu/archives/mason-interest-l	https://lists.sourceforge.net/lists/listinfo/repast-interest
Implementation language	Objective-C/Java	Java	Java/Python/Microsoft.Net
Operating system	Windows, UNIX, Linux, Mac OSX	Windows, UNIX, Linux, Mac OSX	Windows, UNIX, Linux, Mac OSX
Required programming experience	Strong	Strong	Strong
Integrated GIS functionality	Yes (e.g. Kenge GIS library for Raster data: http://www.gis.usu.edu/swarm)	Yes	Yes (Repast 3 simulations can also be run within ArcGIS through an extension called Agent Analyst).
Integrated charting/graphing/statistics	Yes (e.g. R and S-plus statistical packages)	Yes (e.g. wrappers for JFreeChart)	Yes (e.g. Colt statistical package, and basic Repast functionality for simple network statistics) While in RepastS there are automated connections to R and VisAD
Availability of demonstration models	Yes	Yes	Yes
Source code of demonstration models	Yes	Yes	Yes

Tutorials/how-to documentation	Yes	Yes	Yes
Additional information	Minar et al. (1996) Contributed user example models: http://www.swarm.org/index.php/Swarm:_contributed_code	GeoMason website: http://www.cs.gmu.edu/~eclab/projects/mason/extensions/geomason/ Example publications: Luke et al. (2004), Kennedy et al. (2010) Useful weblog: http://www.gisagents.blogspot.com	Agent Analyst Extension (http://www.institute.redlands.edu/agentanalyst) Useful weblogs: http://www.gisagents.blogspot.com http://crimesim.blogspot.com/

Table 12.4 Comparison of shareware/freeware simulation/modelling systems (Adapted from Najlis et al. 2001 and Parker 2001)

Shareware/freeware simulation/modelling systems		
System name	StarLogo	NetLogo
Developers	Media Laboratory, Massachusetts Institute of Technology, USA	Centre for Connected Learning and Computer-Based Modelling, Northwestern University, USA
Date of inception	Early 1990s, Java based version 2000	1999
Website	http://education.mit.edu/starlogo/	http://ccl.northwestern.edu/netlogo
E-mail list	http://education.mit.edu/pipermail/starlogo-users	None
Implementation language	Proprietary scripting	Proprietary scripting
Operating system	Windows, UNIX, Linux, Mac OSX	Windows, UNIX, Linux, Mac OSX
Required programming experience	Basic	Basic
Integrated GIS functionality	None	Yes
Integrated charting/graphing/statistics	Yes	Yes
Availability of demonstration models	Yes	Yes
Source code of demonstration models	Yes	Yes
Tutorials/how-to Documentation	Yes	Yes
Additional information	OpenStarLogo website: http://education.mit.edu/openstarlogo/	http://groups.yahoo.com/group/netlogo-users http://ccl.northwestern.edu/netlogo/docs/gis.html http://backspaces.net/wiki/NetLogo_Bag_of_Tricks

Najlis et al. (2001) identify the steep learning curve of Swarm as a significant factor to consider before choosing this system to develop an agent-based model; although this should be less of a problem for a modeller with strong programming skills.

12.3.3.2 MASON

MASON (Multi Agent Simulation Of Neighbourhood – Table 12.3) is developed by the Evolutionary Computation Laboratory (ECLab) and the Centre for Social Complexity at George Mason University (see Luke et al. 2005). Currently MASON provides much of the same functionality as Repast, for example, dynamically charting

Table 12.5 Comparison of proprietary simulation/modelling systems (Adapted from Najlis et al. 2001 and Parker 2001)

Proprietary simulation/modelling systems		
	AgentSheets	AnyLogic
Developers	AgentSheets Inc., USA	XJ Technologies, Russia
Date of inception	1991	Unknown
Website	http://www.agentsheets.com	http://www.xjtek.com
E-mail list	None	None
Implementation language	Proprietary scripting	Proprietary scripting
Operating system	Windows, UNIX, Linux, Mac OSX	Windows, UNIX, Linux, Mac OSX
Required programming experience	None – Basic	Moderate
Integrated GIS functionality	None	None
Integrated charting/graphing/statistics	Yes	Yes
Availability of demonstration models	Yes http://repast.sourceforge.net/examples/index.html	Yes http://repast.sourceforge.net/examples/index.html
Source code of demonstration models	N/A	N/A
Tutorials/how-to documentation	Yes	Yes
Additional information	Carvalho 2000 and Repenning et al. 2000	http://www.xjtek.com/support/forums/general

(e.g. histograms, line graphs, etc.) and model output during a simulation. A recent addition to MASON is GeoMASON (2010) which allows GIS vector data to be imported/exported. In addition MASON also supports the use of raster data in the creation of geospatial agent-based models (e.g. Kennedy et al. 2010) as shown in Fig. 12.2.

MASON has a growing set of technical documents and well commented Javadocs and a user group which is actively supports its e-mail list. MASONs how-to documentation, demonstration models (e.g. the seminal heat bugs example, network models, etc.), and several publications detailing the implementation and/or application of MASON are available for a prospective modeller to evaluate the system further (MASON 2010). Examples of spatially explicit models utilizing MASONs GIS functionally include exploring conflict between herdsmen and farmers in East Africa (Kennedy et al. 2010), pastoralists in Inner Asia (Cioffi-Revilla et al. 2010), residential dynamics in Arlington County, Virginia (Hailegiorgis 2010) and understanding the Afghan drug industry (Łatek et al. 2010).

12.3.3.3 Repast

Repast (Recursive Porous Agent Simulation Toolkit – Table 12.3) was originally developed at the University of Chicago, and is currently maintained by Argonne

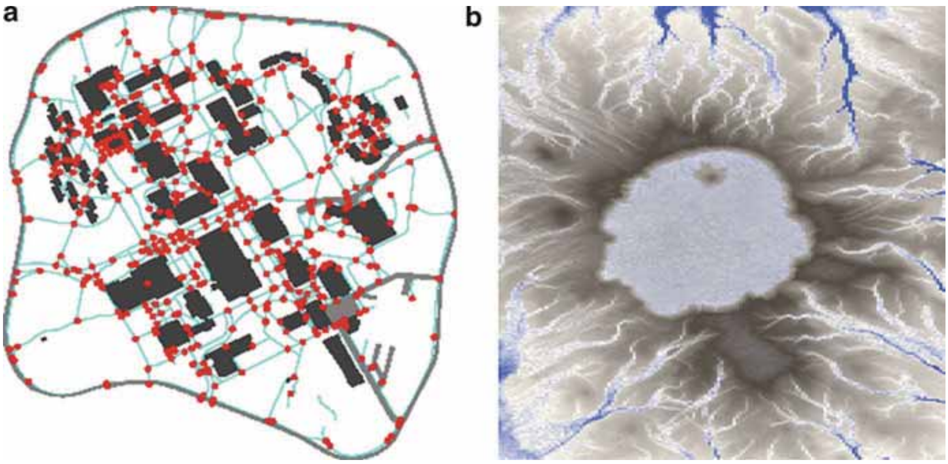


Fig. 12.2 Examples of raster and vector agent-based models in MASON. (**a**) Agents are *red* points which move around the footpaths (*Blue Lines*). (**b**) A rainfall model where agents are *blue* and flow down the Terrain (Built from a Digital Elevation Model)

National Laboratory and managed by the Repast Organisation for Architecture and Development (ROAD). Earlier incarnations of Repast catered for the implementation of models in three programming languages: Python (RepastPy); Java (RepastJ and Repast Simphony); and Microsoft.Net (Repast.Net). RepastPy allows basic models to be developed by modellers with limited programming experience via a 'point-and-click' GUI (Collier and North 2005). RepastPy models can subsequently be exported/converted into Java for further development in RepastJ. Repast.Net and RepastJ allow for more advanced models to be developed (Vos 2005), because more complex functionality can be programmed into a model. Agent Analyst is an ABM extension that allows users to create, edit, and run Repast models from within ArcGIS (Redlands Institute 2010). For further information of earlier versions of Repast, readers are referred to Crooks (2007). Repast has a relatively large user group and an actively supported e-mail list, as well as extensive how-to documentation and demonstration models available from the system website.

Whilst still being maintained RepastJ, Repast.Net and RepastPy have now reached maturity and are no longer being developed. They have been superseded by Repast Simphony (RepastS), which provides all the core functionality of RepastJ or Repast.Net, although limited to implementation in Java. For a comparison of RepastS and previous versions readers are referred to North and Macal (2009). RepastS was initially released in late 2006 and now provides the same GIS functionality of previous versions. The main improvements with RepastS over Repast 3.0 is a new optional GUI point-and-click environment for model development that generates Java classes, however models can still be coded manually. Secondly a improved runtime GUI, the GUI can now be used to build displays (both in 2 and 3D) or charts, output data, interrogate agents, and interface with other programs (like R for statistics) via a point-and-click interface at run time. This means that these tasks are

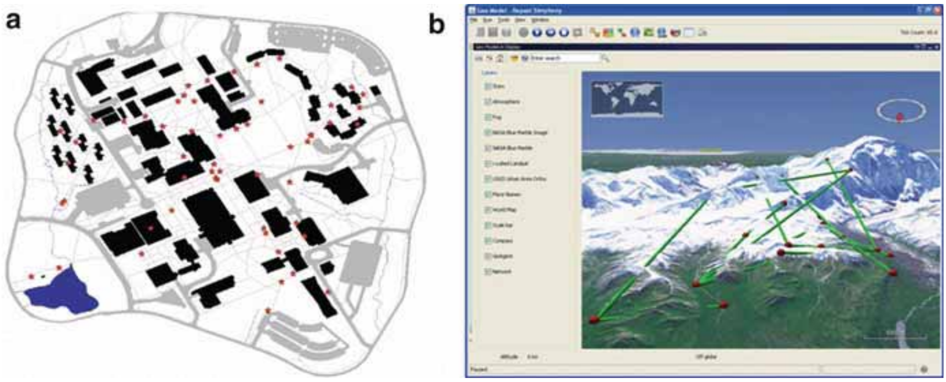


Fig. 12.3 Examples of vector agent-based models in RepastS. (**a**) Agents (*Red Dots*) moving about on footpaths (*Grey Lines*). (**b**) An agent-based model overlaid on NASA world wind (Source: Repast 2011)

done more quickly after the model has been built and compiled, and do not feature in the underlying code at all, unlike previous Repast implementations.

The Repast development team have provided a series of articles regarding RepastS. The architecture and core functionality are introduced by North et al. (2005a), and the development environment is discussed by Howe et al. (2006). The storage, display and behaviour/interaction of agents, as well as features for data analysis (i.e. via the integration of the R statistics package) and presentation of models within Repast S are outlined by North et al. (2005b). Tatara et al. (2006) provide a detailed discussion outlining how-to develop a "simple wolf-sheep predation" model; illustrating RepastS modelling capabilities. In relation to the integration of GIS functionality the reader is referred to the tutorials by Malleson, (2008) which demonstrates how to create a virtual city via the importation of shapefiles, create agents and then move the agents around a road network (this tutorial was used for the creation of Fig. 12.3a). Furthermore, within RepastS it is possible to embed spatially explicit agent-based models directly into a 3D GIS display. For this RepastS provides methods to directly visualise agent-based models to NASA's (2010) virtual globe – World Wind. This new interactive 3D GIS display allows one to visualise agents with satellite imagery, elevated terrain and other scientific datasets as shown in Fig. 12.3b. RepastS also supports the importation of NetLogo (see Sect. 12.3.3.5) models into the Repast framework via ReLogo (Ozik 2010). Such functionality aims to allow for rapid prototyping of agent-based models by first building simple agent-based models in NetLogo and once satisfied allowing one to migrate and extend them in RepastS. Not only does RepastS provide tools for the conversion of simple models from NetLogo, it also supports high performance distributed computing, via Repast for High Performance Computing (Repast HPC, see Collier 2010).

Useful examples of spatially explicit models created using Repast include the studying of segregation, and residential and firm location (Crooks 2006, 2010), residential dynamics (Jackson et al. 2008) crime (Malleson et al. 2010) and the evacuation of pedestrians from within an underground station (Castle 2007).

12.3.3.4 StarLogo

StarLogo (Table 12.4) is an shareware/freeware modelling system developed at the Media Laboratory, Massachusetts Institute of Technology (MIT) It has undergone some change, the original StarLogo modelling system has been released as an open source project (see OpenStarLogo 2010) however, it is still included in this section as the new version, StarLogo TNG (The New Generation) is still shareware/freeware. StarLogo TNG moves StarLogo from the 2D to the 3D realm through the use OpenGL graphics API and aims to lower the barrier for programming agent-based models through the use of a drag and drop programming graphical interface. Modellers can drag commands from a set of model building blocks (a block based graphical language) rather than creating models using the StarLogo syntax thus allowing for rapid model development. StarLogo TNG uses OpenGL for displaying the models at run time therefore providing a 3D display termed ‘SpaceLand’. The terrain within such models is editable and can be manually shaped. Agents can also be programmed to move in x, y and z directions.

StarLogo lacks the same flexibility offered by open source systems, since modellers are constrained to functionality provided by the system. Despite this limitation, StarLogo is very easy to use, notably for people with very little programming experience. Dynamic charting functionality of model output during a simulation is provided. In addition, a number of demonstration models and detailed how-to documentation relating to these models is supplied with StarLogo, and many more are available to download from the World Wide Web (WWW). While StarLogo does not support GIS per se, it does allow one to import GIFs, therefore allow pixels to be converted into patches. Batty et al. (1998) used this approach to examine visitor movement within London’s British Tate Gallery, specifically how changes in room configuration can affect movement between exhibits.

12.3.3.5 NetLogo

NetLogo (originally named StarLogoT – Table 12.4) is a variant of StarLogo, originally developed at the Centre for Connected Learning and Computer-Based Modelling at Northwestern University, to allow StarLogo models to be developed on computers using the Macintosh operating system. It is now possible to create StarLogo models on a computer using a Macintosh operating system, thus the critically distinction between the two simulation/modelling systems is that NetLogo is specifically designed for the deployment of models over the internet (NetLogo 2010). Initially both NetLogo and StarLogo only provided functionality to import image files, which can be used to define the environments within which agents are

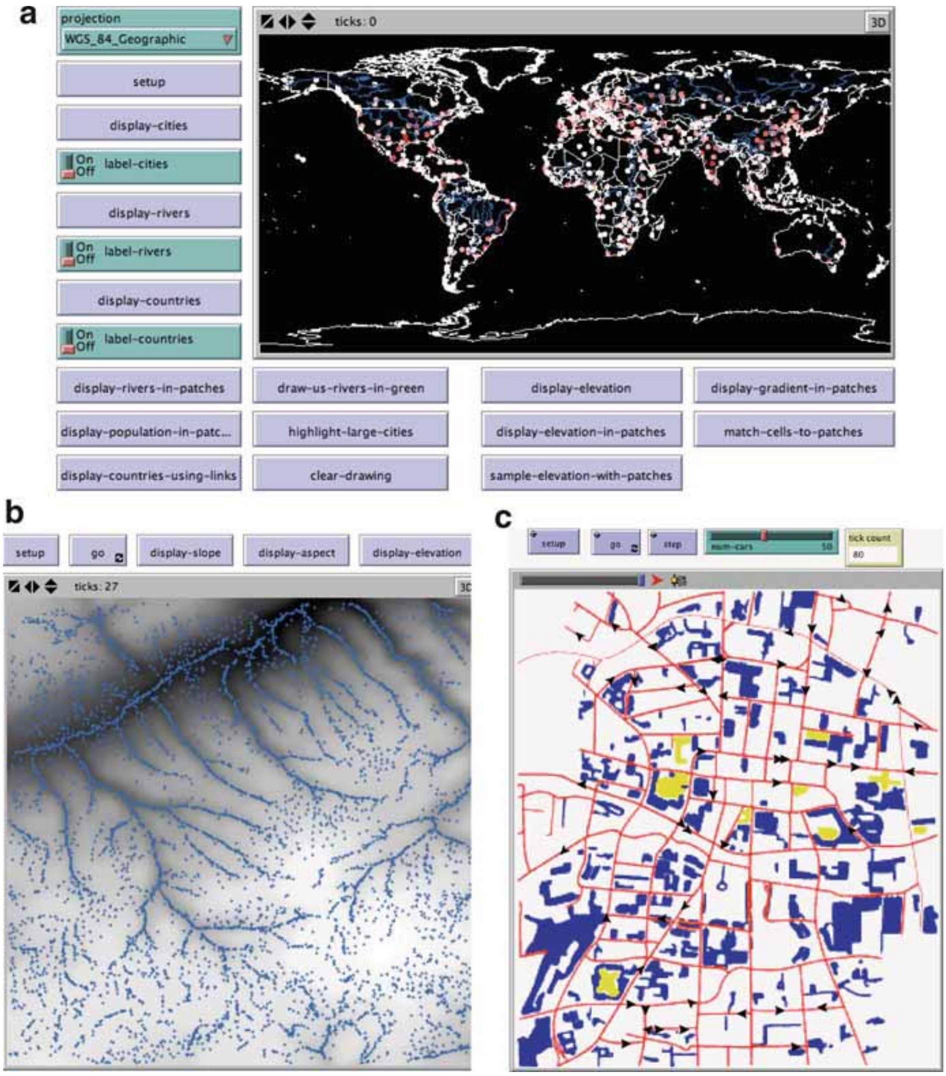


Fig. 12.4 Example of GIS integration in NetLogo. (**a**) Demonstration model of using *point*, *line* and *polygon* shapefiles for creating a landscape. (**b**) NetLogo's gradient example and (**c**) the cruising model where cars move along the roads (Red lines) (Source: NetLogo 2010)

located, thus facilitating the development of spatial models (Fig. 12.4). However, within NetLogo it is now possible to import both raster (in the form of .asc files) and vector data (shapefiles). This new ability opens up a range of possibilities for the easy creation of spatial agent based models. For example, for the studying of surface erosion (Wilensky 2006) as shown in Fig. 12.4b.

The NetLogo installation comes with two demonstration models highlighting this functionality. For vector data, four different GIS datasets: a point file of world cities, a polyline file of world rivers, a polygon file of countries (however there is no way to distinguish if the polygon has holes in it) are imported into a NetLogo model and converted into patches as shown in Fig. 12.4a. For the raster example, a raster file of surface elevation is loaded into a NetLogo model to demonstrate the possibilities of working with spatial data as shown in Fig. 12.4b. In this example, Agents follow the surface to lower elevations. Such functionality potentially lowers the barrier between coupling agent-based models and GIS to none expert programmers. For example, the gradient example presented above could be used to model process that relies on cost surfaces such as emergency evacuation of buildings (see Crooks et al. 2008, for an example). As with StarLogo TNG (Sect. 12.3.3.4), models within NetLogo can be viewed in a 3D environment however unlike StarLogo TNG it is only the agents that appear in 3D while the surface remains a 2D plane.

NetLogo has been used to develop applications in disciplines varying from biology and physics to the social sciences. Extensive how-to documentation/tutorials and demonstration models are available from the system website, and functionality can be extended through APIs, although the source code for the system is currently unavailable. Useful examples of spatially explicit models created using NetLogo include the study of gentrification (Torrens and Nara 2007), residential housing demand (Fontaine and Rounsevell 2009) and the emergence of settlement patterns (Graham and Steiner 2006) and the reimplementation of Axtell et al. (2002) artificial Anasazi model by Janssen (2009).

12.3.3.6 AgentSheets

AgentSheets (Table 12.5) is a proprietary simulation/modelling system that allows modellers with limited programming experience to develop an agent-based model, because models are developed through a GUI (Repenning et al. 2000). A number of demonstration models are available from the system website. For example, Sustainopolis is a simulation analogous to the computer game SimCity; exploring pollution dispersal within a city (Fig. 12.5). Furthermore, AgentSheets can be linked to real time information over the internet (Repenning and Ioannidou 2004). For example, AgentSheets has been used in conjunction with real time weather feeds and used to make mountain biking recommendations in Boulder County. Within the model, agents represent locations that are possible candidates for biking featuring real time, web accessible weather sensors. This information is then used by the biker to reach a decision on where to go biking. Carvalho (2000) has used AgentSheets extensively to teach undergraduate students. He comments that it is easy to use the system to develop models quickly and provides students with hands-on experience of ABM without the need to learn a programming language. However, he also found that models created with AgentSheets were limited in their sophistication (notably in terms of the complexity of representation of agent behaviour and interaction). Furthermore, agents are limited to movement within a two-dimensional cell-based environment.

Fig. 12.5 The Sustainopolis model developed in AgentSheets (2010)

12.3.3.7 AnyLogic

AnyLogic (Table 12.5) incorporates a range of functionality for the development of agent-based models. For example, models can dynamically read and write data to spreadsheets or databases during a simulation run, as well as dynamically chart model output. Furthermore, external programmes can be initiated from within an AnyLogic model for dynamic communication of information, and vice versa. However, AnyLogic models can only be created on Microsoft operating systems, although a simulation can be run on any Java-enabled operating system once compiled (e.g. a Macintosh operating system). The AnyLogic website notes that models have been developed for a diverse range of applications including: the study of social, urban (Fig. 12.6) and ecosystem dynamics (e.g. a predator-prey system); planning of healthcare schemes (e.g. the impact of safe syringe usage on HIV diffusion); computer and telecommunication networks (e.g. the placement of cellular phone base stations); and the location of emergency services and call centres. Further information pertaining AnyLogic modelling applications can be found in Parinov (2007), these include imitating the functioning of a emergency department in a large hospital. However, the source code of these examples and/or documentation of these

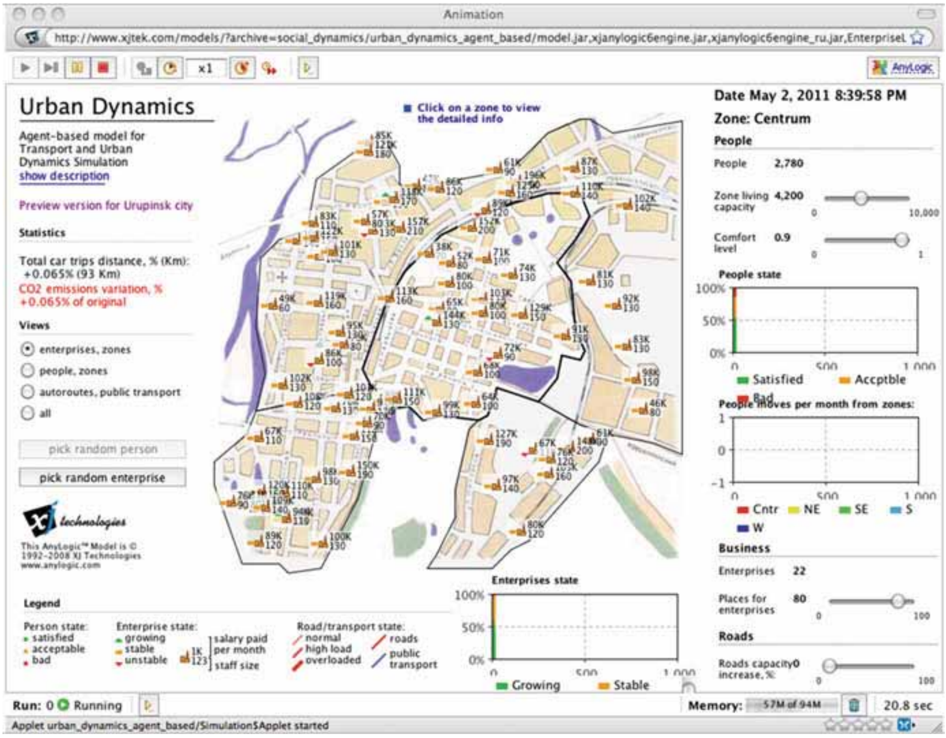


Fig. 12.6 An urban and transport dynamics model developed in AnyLogic (2010)

models is unavailable. Example applications utilizing AnyLogic for spatial agent-based modelling include: Makarov et al. (2008) who studied traffic jams in Moscow and explored different scenarios for reducing such events either by road pricing or new road building. Johnson and Sieber (2009) used AnyLogic to explore tourism in Nova Scotia, while Pint et al. (2010) used AnyLogic to explore organised crime in Rio's favelas.

12.4 Summary

This chapter has reviewed the current capabilities of modelling within a GIS and suggests that agent-based modellers interested in developing geospatial models involving many (possibly tens of thousands) interacting agents with complex behaviours and interactions between themselves, and their environment should consider either GIS-centric or modelling-centric integration. Moreover, we have discussed considerations one should take when thinking about utilizing an agent-based simulation/modelling system. Furthermore, we have outlined a selection of simulation/modelling systems which can be used for the creation of geospatial agent-based models along with providing examples of applications.

Each of simulation/modelling systems discussed within this chapter can be positioned within the continuum illustrated in Fig. 12.1 (power versus difficulty of developing a model with a simulation/modelling system). However, the exact location of each system is very subjective (i.e. dependant upon a modeller's knowledge and experience of ABM in general, and each simulation/modelling system in particular). The information presented within this chapter is aimed at providing the reader with a selection of useful criteria to assess the seven simulation/modelling systems presented, allowing each system to be (approximately) located within this continuum based on the readers own knowledge and experience. That is not to say that the selection criteria cannot be utilized for other simulation/modelling systems and once a candidate system(s) has been identified the reader will need to investigate the potential suitable of the system(s) further.

However, it needs to be noted that while such tools exist, integrating GIS data for ABM is still a difficult process (Gilbert 2007) and many considerations are needed such as what data is needed, how should the data be utilised, how should agents interact with the data, etc. Nevertheless, such systems lower the entry level needed to create geospatial agent-based models and thus allowing a greater number of social scientists to create geospatial agent-based models. One note of caution however is needed, that is there is still a computational challenge when it comes to the creation of geospatial agent-based models with thousands of agents operating and interacting with raster or vector features (see Kennedy et al. 2009 for a discussion) but over time this should be reduced with increased computational power.

References

Abdou, M., Hamill, L., & Gilbert, N. (2012). Designing and building an agent-based model. In A.J. Heppenstall, A.T. Crooks, L.M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 141–166). Dordrecht: Springer.

AgentLink. (2007). Software. Available at http://eprints.agentlink.org/view/type/software.html. Accessed on 12 July 2010.

AgentSheets. (2010). AgentSheets: The Sustainopolis model. Available at http://www.agentsheets.com/Applets/sustainopolis/. Accessed on 12 July 2010.

Andrade, P., Monteiro, A., & Câmara, G. (2008). Entities and relations for agent-based modelling of complex spatial systems. *BWSS 2008: 1st Brazilian Workshop on Social Simulation*, Bahia.

AnyLogic. (2010). AnyLogic: Urban dynamics agent based model. Available at http://www.xjtek.com/anylogic/demo_models/48/. Accessed on 12 July 2010.

Armstrong, D. J. (2006). The quarks of object-oriented development. *Communication of the ACM, 49*(2), 123–128.

Axtell, R., Epstein, J. M., Dean, J. S., Gumerman, G. J., Swedlund, A. C., Harburger, J., Chakravarty, S., Hammond, R., Parker, J., & Parker, M. (2002). Population growth and collapse in a multi-agent model of the Kayenta Anasazi in Long House Valley. *Proceedings of the National Academy of Sciences of the United States of America (PNAS), 99*(3), 7275–7279.

Batty, M. (2005). *Cities and complexity: Understanding cities with cellular automata, agent-based models, and fractals*. Cambridge, MA: The MIT Press.

Batty, M., Conroy, R., Hillier, B., Jiang, B., Desyllas, J., Mottram, C., Penn, A., Smith, A., & Turner, A. (1998). *The virtual tate*, Working paper 5. London: Centre for Advanced Spatial Analysis (University College London).

Batty, M., Desyllas, J., & Duxbury, E. (2003). Safety in numbers? Modelling crowds and designing control for the Notting Hill carnival. *Urban Studies, 40*(8), 1573–1590.
Benenson, I., Aronovich, S., & Noam, S. (2005). Let's talk objects: Generic methodology for urban high-resolution simulation. *Computers, Environment and Urban Systems, 29*(4), 425–453.
Benenson, I., Birfur, S., & Kharbash, V. (2006). Geographic automata systems and the OBEUS software for their implementation. In J. Portugali (Ed.), *Complex artificial environments: Simulation, cognition and VR in the study and planning of cities* (pp. 137–153). Berlin: Springer.
Berger, T., & Parker, D. C. (2001). Examples of specific research: Introduction to standard descriptions of projects'. In D. C. Parker, T. Berger, & S. M. Manson (Eds.), *Meeting the Challenge of Complexity* (pp. 48–54). *Proceedings of a Special Workshop on Land-Use/Land-Cover Change*, Irvine.
Bernard, L., & Krüger, T. (2000). Integration of GIS and spatio-temporal simulation models: Interoperable components for different simulation strategies. *Transactions in GIS, 4*(3), 197–215.
Berryman, M. (2008). *Review of software platforms for agent based models*. Edinburgh: Defence Science and Technology Organisation, DSTO-GD-0532.
Booch, G. (1994). *Object-oriented analysis and design with applications*. Redwood City: Benjamin/Cummings.
Bousquet, F., Bakam, I., Proton, H., & Le Page, C. (1998). CORMAS: Common-pool resources and multi-agent systems. In A. P. Pobil, J. Mira, & A. Moonis (Eds.), *Lecture notes in artificial intelligence 1416* (pp. 826–838). Berlin: Springer.
Brown, D. G. (2006). Agent-based models. In H. Geist (Ed.), *The Earth's changing land: An encyclopaedia of land-use and land-cover change* (pp. 7–13). Westport: Greenwood Publishing Group.
Brown, D. G., Riolo, R., Robinson, D. T., North, M. J., & Rand, W. (2005). Spatial process and data models: Toward integration of agent-based models and GIS. *Journal of Geographical Systems, 7*(1), 25–47.
Carley, K. M. (1996). *Validating computational models*, Working Paper. Pittsburgh: Carnegie Mellon University.
Carvalho, J. (2000). Using agentSheets to teach simulation to undergraduate students. *Journal of Artificial Societies and Social Simulation, 3*(3). Available at http://jasss.soc.surrey.ac.uk/3/3/forum/2.html
Castle, C. J. E. (2007). *Agent-based modelling of pedestrian evacuation: A study of London's king's cross underground station.* Ph.D thesis, University College London, London.
Cioffi-Revilla, C. (2010). Computational social science. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(3), 259–271.
Cioffi-Revilla, C., Rogers, J. D., & Latek, M. (2010). The MASON householdWorlds of pastoral nomad societies. In K. Takadama, C. Cioffi-Revilla, & G. Deffaunt (Eds.), *Simulating interacting agents and social phenomena: The second world congress in social simulation*. Berlin: Springer.
Collier, N. T. (2010). *Repast HPC manual.* Available at http://repast.sourceforge.net/docs.html
Collier, N. T., & North, M. J. (2005). Repast for Python Scripting. *Annual Conference of the North American Association for Computational Social and Organizational Science (NAACSOS)*, Notre Dame. Available at http://www.casos.cs.cmu.edu/events/conferences/2005/2005_proceedings/Collier.pdf
Consequences Assessment Tool Set. (2011). Incident manager and first responder tools. Available at http://www.saic.com/products/security/cats/. Accessed on 12 Apr 2011.
Couclelis, H. (2001). Why I no longer work with agents: A challenge for ABMs of human-environment interactions. In D. C. Parker, T. Berger, & S. M. Manson (Eds.), *Meeting the challenge of complexity* (pp. 14–16). *Proceedings of a Special Workshop on Land-Use/Land-Cover Change*, Irvine.
Crooks, A. T. (2006). Exploring cities using agent-based models and GIS. In D. Sallach, C. M. Macal, & M. J. North (Eds.), *Proceedings of the Agent 2006 Conference on Social Agents:*

Results and Prospects, University of Chicago and Argonne National Laboratory, Chicago. Available at http://agent2007.anl.gov/2006procpdf/Agent_2006.pdf

Crooks, A. T. (2007). *The repast simulation/modelling system for geospatial simulation*, Working Paper 123. London: Centre for Advanced Spatial Analysis (University College London).

Crooks, A. T. (2010). Constructing and implementing an agent-based model of residential segregation through vector GIS. *International Journal of GIS, 24*(5), 661–675.

Crooks, A. T., Castle, C. J. E., & Batty, M. (2008). Key challenges in agent-based modelling for geo-spatial simulation. *Computer, Environment and Urban Systems, 32*(6), 417–430.

Crooks, A.T., & Heppenstall, A.J. (2012), Introduction to agent-based modeling. In A.J. Heppenstall, A.T. Crooks, L.M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 85–105). Dordrecht: Springer.

Dugdale, J. (2004). An evaluation of seven software simulation tools for use in the social sciences. Available at http://www.irit.fr/COSI/training/evaluationoftools/Evaluation-Of-Simulation-Tools.htm. Accessed on 12 July 2010.

Epstein, J. M. (2009). Modelling to contain pandemics. *Nature, 460*, 687.

Fedra, K. (1996). Distributed models and embedded GIS: Integration strategies and case studies. In M. F. Goodchild, L. T. Steyaert, & B. O. Parks (Eds.), *GIS and environmental modelling: Progress and research issues* (pp. 413–417). Fort Collins: GIS World Books.

Filatova, T., Parker, D., & van der Veen, A. (2009). Agent-based urban land markets: Agent's pricing behavior, land prices and urban land use change. *Journal of Artificial Societies and Social Simulation, 12*(1). Available at http://jasss.soc.surrey.ac.uk/12/1/3.html

Fontaine, C. M., & Rounsevell, M. (2009). An agent-based approach to model future residential pressure on a regional landscape. *Landscape Ecology, 24*(9), 1237–1254.

GeoMASON. (2010). GIS integration for MASON. Available at https://mason.dev.java.net/source/browse/mason/contrib/geomason/. Accessed on 14 July 2010.

Gilbert, N. (2007). *Agent-based models*. London: Sage.

Gilbert, N., & Bankes, S. (2002). Platforms and methods for agent-based modelling. *Proceeding of the National Academy of Sciences of the USA, 99*(3), 7197–7198.

Gilbert, N., & Troitzsch, K. G. (2005). *Simulation for the social scientist*. Milton Keynes: Open University Press.

Goodchild, M. F. (2005). GIS, spatial analysis, and modelling overview. In D. J. Maguire, M. Batty, & M. F. Goodchild (Eds.), *GIS, spatial analysis and modelling* (pp. 1–18). Redlands: ESRI Press.

Graham, S., & Steiner, J. (2006). Travellersim: Growing settlement structures and territories with agent-based modelling. In J. T. Clark & E. M. Hagemeister (Eds.), Digital discovery: Exploring new frontiers in human heritage. *Proceedings of the 34th Conference on Computer Applications and Quantitative Methods in Archaeology*, Fargo.

Grimm, V., & Railsback, S. F. (2012). Designing, formulating and communicating agent-based models. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.) *Agent-based models of geographical systems* (pp. 361–377). Dordrecht: Springer.

Hailegiorgis, A. B. (2010). Changing residence in the city: An agent based model of residential mobility in Arlington County. *The Association of American Geographers (AAG) Annual Meeting*, Washington, DC, 14–18 Apr 2010.

Haklay, M., O'Sullivan, D., Thurstain-Goodwin, M., & Schelhorn, T. (2001). "So go downtown": Simulating pedestrian movement in town centres. *Environment and Planning B, 28*(3), 343–359.

Harland, K., & Heppenstall, A. J. (2012). Using agent-based models for education planning: Is the UK education system agent-based. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 481–497). Dordrecht: Springer.

Hazard Prediction and Assessment Capability. (2004). Available at http://computing.ornl.gov/cse_home/hpac.shtml. Accessed on 12 Apr 2011.

Heywood, I., Cornelius, S., & Carver, S. (2006). *An introduction to geographical information systems* (3rd ed.). Harlow: Pearson Education.

Howe, T. R., Collier, N. T., North, M. J., Parker, M. T., & Vos, J. R. (2006). Containing agents: Contexts, projections, and agents. In D. Sallach, C. M. Macal, & M. J. North (Eds.), *Proceedings of the Agent 2006 Conference on Social Agents: Results and Prospects*, University of Chicago and Argonne National Laboratory, Chicago. Available at http://agent2007.anl.gov/2006procpdf/Agent_2006.pdf

Jackson, J., Forest, B., & Sengupta, R. (2008). Agent-based simulation of urban residential dynamics and land rent change in a gentrifying area of Boston. *Transactions in GIS, 12*(4), 475–491.

Janssen, M. A. (2009). Understanding artificial Anasazi. *Journal of Artificial Societies and Social Simulation, 12*(4). Available at http://jasss.soc.surrey.ac.uk/12/4/13.html

Johnasson, A., & Kretz, T. (2012). Applied pedestrian modeling. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 451–462). Dordrecht: Springer.

Johnson, P. A., & Sieber, R. E. (2009). Agent-based modeling: A dynamic scenario planning approach to tourism PSS. In S. Geertman & J. Stillwell (Eds.), *Planning support systems: Best practices and new methods* (pp. 211–226). Berlin: Springer.

Kennedy, R. C., Lane, K. E., Fuentes, A., Hollocher, H., & Madey, G. (2009). Spatially aware agents: An effective and efficient use of GIS data within an agent-based model. *Agent-Directed Simulation Symposium (ADS'09)*, San Diego, 22–27 Mar 2009.

Kennedy, W. B., Hailegiorgis, A. B., Rouleau, M., Bassett, J. K., Coletti, M., Balan, G. C., & Gulden, T. (2010). An agent-based model of conflict in East Africa and the effect of watering holes. *Behavior Representation in Modeling and Simulation (BRiMS) Conference*, Charleston.

Langran, G. (1992). *Time in geographic information systems*. London: Taylor & Francis.

Łatek, M. M., Rizi, S. M. M., & Geller, A. (2010). Persistence in the political economy of conflict: The case of the Afghan Drug Industry. *Complex Adaptive Systems: Resilience, Robustness, and Evolvability Workshop, Fall Symposium of Association for the Advancement of Artificial Intelligence*, 11–13 Nov 2010, Arlington. Available at http://www.css.gmu.edu/projects/irregularwarfare/

Liebert, K. D., Earnest, D. C., & Tolk, A. (2008). Using GIS vector data to build virtual environments for agent based models. *Proceedings of the 2008 Spring simulation multiconference* (pp. 45–52), Ottawa.

Linthicum, D. S. (2000). *Enterprise application integration*. Boston: Addison-Wesley.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2005). *Geographical information systems and science* (2nd ed.). New York: Wiley.

Luke, S., Cioffi-Revilla, C., Panait, L., & Sullivan, K. (2004). MASON: A new multi-agent simulation toolkit. *SwarmFest 2004, Eighth Annual Swarm Users/Researchers Conference*, University of Michigan, Ann Arbor. Available at http://www.cscs.umich.edu/swarmfest04/Program/PapersSlides/seanLuke-SwarmFest04-040507-2100.pdf

Luke, S., Cioffi-Revilla, C., Panait, L., Sullivan, K., & Balan, G. (2005). MASON: A multi-agent simulation environment. *Simulation, 81*(7), 517–527.

Maguire, D. J. (1995). Implementing spatial analysis and GIS applications for business and service planning. In P. A. Longley & G. Clarke (Eds.), *GIS for business and service planning* (pp. 171–191). Cambridge: GeoInformation International.

Maguire, D. J. (2005). Towards a GIS platform for spatial analysis and modelling. In D. J. Maguire, M. Batty, & M. F. Goodchild (Eds.), *GIS, spatial analysis and modelling*. Redlands: ESRI Press.

Makarov, V. L., Bakhtizin, A. R., & Zhitkov, V. A. (2008). Agent-based model for traffic jams in Moscow. *ABS2 Agent Based Spatial Simulation International Workshop*, 24–25 Nov 2008, Paris. Available at http://s4.csregistry.org/ABS2

Malleson, N. (2008). RepastCity – A demo virtual city. Available at http://portal.ncess.ac.uk/access/wiki/site/mass/repastcity.html. Accessed on 21 Jan 2009.

Malleson, N., Heppenstall, A., & See, L. (2010). Crime reduction through simulation: An agent-based model of burglary. *Computers, Environment and Urban Systems, 34*(3), 236–250.

MASON. (2010). Multi agent simulation of neighbourhood. Available at http://cs.gmu.edu/~eclab/projects/mason/. Accessed on 12 July 2010.

Mathevet, R., Bousquet, F., Le Page, C., & Antona, M. (2003). Agent-based simulations of interactions between duck population, farming decisions and leasing of hunting rights in the Camargue (Southern France). *Ecological Modelling, 165*(2–3), 107–126.

Minar, N., Burkhart, R., Langton, C. and Askenazi, M. (1996). The swarm simulation system: A toolkit for building multi-agent simulations. Available at http://www.santafe.edu/media/workingpapers/96-06-042.pdf. Accessed on 12 July 2010.

Mooij, W. M., Bennetts, R. E., Kitchens, W. M., & DeAngelis, D. L. (2002). Exploring the effect of drought extent and interval on the Florida snail kite: Interplay between spatial and temporal scales. *Ecological Modelling, 149*(1–2), 25–39.

Najlis, R., Janssen, M. A., & Parker, D. C. (2001). Software tools and communication issues. In D. C. Parker, T. Berger, & S. M. Manson (Eds.), *Meeting the Challenge of Complexity: Proceedings of a Special Workshop on Land-Use/Land-Cover Change* (pp. 17–30), Irvine. Available at http://www.csiss.org/resources/maslucc/ABM-LUCC.pdf

NASA. (2010). World wind. Available at http://worldwind.arc.nasa.gov/java/. Accessed on 15 July 2010.

NatureServe Vista. (2011). Decision support for better planning, Available at http://www.natureserve.org/prodServices/vista/overview.jsp. Accessed on 12 Apr 2011.

NetLogo. (2010). NetLogo: Multi-agent programmable modelling environment. Available at http://ccl.northwestern.edu/netlogo/. Accessed on 12 July 2010.

Nikolai, C., & Madey, G. (2009). Tools of the trade: A survey of various agent based modeling platforms. *Journal of Artificial Societies and Social Simulation, 12*(2). Available at http://jasss.soc.surrey.ac.uk/12/2/2.html

North, M. J., & Macal, C. M. (2009). Foundations of and recent advances in artificial life modeling with Repast 3 and Repast Simphony. In A. Adamatzky & M. Komosinski (Eds.), *Artificial life models in software* (2nd ed., pp. 37–60). New York: Springer.

North, M. J., Howe, T. R., Collier, N. T., & Vos, J. R. (2005a). The Repast Simphony development environment. In C. M. Macal & M. J. North (Eds.), *Proceedings of the Agent 2005 Conference on Generative Social Processes, Models, and Mechanisms*, Chicago. Available at http://agent2007.anl.gov/Agent_2005.pdf

North, M. J., Howe, T. R., Collier, N. T., & Vos, J. R. (2005b). The Repast Simphony runtime system. In C. M. Macal, M. J. North, & D. Sallach (Eds.), *Proceedings of the Agent 2005 Conference on Generative Social Processes, Models, and Mechanisms*, Chicago. Available at http://agent2007.anl.gov/Agent_2005.pdf

OpenStarLogo. (2010). OpenStarLogo homepage. Available at http://education.mit.edu/openstarlogo/. Accessed on 12 July 2010.

Ozik, J. (2010). *RELOGO getting started guide.* Available at http://repast.sourceforge.net/docs/ReLogoGettingStarted.pdf

Parinov, S. I. (2007). New possibilities for simulations of socio-economic systems. *Artificial Societies, 2*(3–4): 23–53. Available at http://www.artsoc.ru/html_e/journal.htm

Parker, D. C. (2001). Object-orientated packages for agent-based modelling. Available at http://mason.gmu.edu/~dparker3/spat_abm/lectures/lecture2_tables.pdf Accessed on 12 June 2010.

Parker, D. C., Berger, T., & Manson, S. M. (2001). *Proceedings of an International Workshop on Agent-Based Models of Land-Use and Land-Cover Change,* Irvine. Available at http://www.csiss.org/maslucc/ABM-LUCC.htm

Peuquet, D. J. (2005). Time in GIS and geographical databases. In P. A. Longley, M. F. Goodchild, D. J. Maguire, & D. W. Rhind (Eds.), *Geographical information systems: Principles, techniques, management and applications (Abridged Edition)* (pp. 91–103). Hoboken: Wiley.

Peuquet, D. J., & Duan, N. (1995). An event-based spatio-temporal data model (ESTDM) for temporal analysis of geographic data. *International Journal of Geographical Information Systems, 9*(1), 2–24.

Pint, B., Crooks, A. T., & Geller, A. (2010). An agent-based model of organized crime: Favelas and the drug trade. *2nd Brazilian Workshop on Social Simulation*, Sao Bernardo do Campo.

Railsback, S. F., & Harvey, B. C. (2002). Analysis of habitat selection rules using an individual-based model. *Ecology, 83*(7), 1817–1830.

Railsback, S. F., Lytinen, S. L., & Jackson, S. K. (2006). Agent-based simulation platforms: Review and development recommendations. *Simulation, 82*(9), 609–623.

Redlands Institute. (2010). What is agent analyst? Available at http://www.institute.redlands.edu/agentanalyst. Accessed on 12 July 2010.

Reitsma, F., & Albrecht, J. (2006). A process-oriented data model. In J. Drummond, R. Billen, D. Forrest, & E. João (Eds.), *Dynamic and mobile GIS: Investigating change in space and time* (pp. 77–87). London: Taylor & Francis.

Repast. (2011). Recursive porous agent simulation toolkit. Available at http://repast.sourceforge.net/. Accessed on 7 Mar 2011.

Repenning, A., & Ioannidou, A. (2004). Agent-based end-user development. *Communications of the Association of Computing Machinery, 47*(9), 43–46.

Repenning, A., Ioannidou, A., & Zola, J. (2000). AgentSheets: End-user programmable simulations. *Journal of Artificial Societies and Social Simulation, 3*(3). Available at http://jasss.soc.surrey.ac.uk/3/3/forum/1.html.

Rixon, A., Moglia, M., & Burn, S. (2005). Bottom-up approaches to building agent-based models: Discussing the need for a platform. *Proceedings of the Joint Conference on Multi-Agent Modelling for Environmental Management*, Bourg-St-Maurice.

Robertson, D. A. (2005). Agent-based modeling toolkits. *Academy of Management Learning and Education, 4*(4), 525–527.

Schelhorn, T., O'Sullivan, D., Haklay, M., & Thurstain-Goodwin, M. (1999). *STREETS: An agent-based pedestrian model*, Working paper. London: Centre for Advanced Spatial Analysis (University College London).

Serenko, A., & Detlor, B. (2002). *Agent toolkits: A general overview of the market and an assessment of instructor satisfaction with utilizing toolkits in the classroom,* Working Paper 455. Hamilton: McMaster University (School of Business). Available at http://merc.mcmaster.ca/wpapers/wpaper455.html

Sondheim, M., Gardels, K., & Buehler, K. (2005). GIS interoperability. In P. A. Longley, M. F. Goodchild, D. J. Maguire, & D. W. Rhind (Eds.), *Geographical information systems: Principles, techniques, management and applications (Abridged Edition)* (pp. 347–358). Hoboken: Wiley.

Swarm. (2010). Swarm: A platform for agent-based models. Available at http://www.swarm.org/. Accessed on 12 July 2010.

SwarmWiki. (2010). Tools for Agent-Based Modelling. Available at http://www.swarm.org/wiki/Tools_for_Agent-Based_Modelling. Accessed on 12 July 2010.

Tatara, E., North, M. J., Howe, T. R., Collier, N. T., & Vos, J. R. (2006). An introduction to repast simphony modelling using a simple predator-prey example. In D. Sallach, C. M. Macal, & M. J. North (Eds.), *Proceedings of the Agent 2006 Conference on Social Agents: Results and Prospects*, University of Chicago and Argonne National Laboratory, Chicago. Available at http://agent2007.anl.gov/2006procpdf/Agent_2006.pdf

Tesfatsion, L. (2010). General software and toolkits. Available at http://www.econ.iastate.edu/tesfatsi/acecode.htm. Accessed on 12 July 2010.

Tesfatsion, L., & Judd, K. L. (Eds.). (2006). *Handbook of computational economics: Agent-based computational economics* (Vol. 2). Amsterdam: North-Holland.

Tobias, R., & Hofmann, C. (2004). Evaluation of free java-libraries for social-scientific agent based simulation. *Journal of Artificial Societies and Social Simulation, 7*(1). Available at http://jasss.soc.surrey.ac.uk/7/1/6.html

Torrens, P. M. (2009). *Process models and next-generation geographic information technology* (GIS best practices: Essays on geography and GIS, Vol. 2, pp. 63–75). Redlands: ESRI Press.

Torrens, P. M., & Nara, A. (2007). Modelling gentrification dynamics: A hybrid approach. *Computers, Environment and Urban Systems, 31*(3), 337–361.

Ungerer, M. J., & Goodchild, M. F. (2002). Integrating spatial data analysis and GIS: A new implementation using component object model (COM). *International Journal of Geographic Information Science, 16*(1), 41–53.

Vos, J. R. (2005). Repast .NET: The Repast framework implemented in the .NET. *Annual Conference of the North American Association for Computational Social and Organizational Science (NAACSOS)*, Notre Dame. Available at http://www.casos.cs.cmu.edu/events/conferences/2005/2005_proceedings/Vos.pdf

Wegener, M. (2000). Spatial models and GIS. In A. S. Fotheringham & M. Wegener (Eds.), *Spatial models and GIS: New potential and new models* (pp. 3–20). London: Taylor & Francis.

Westervelt, J. D. (2002). Geographic information systems and agent-based modelling. In H. R. Gimblett (Ed.), *Integrating geographic information systems and agent-based modelling techniques for simulating social and ecological processes* (pp. 83–104). Oxford: Oxford University Press.

Wikipedia. (2010). Comparison of agent-based modeling software. Available at http://en.wikipedia.org/wiki/Comparison_of_agent-based_modeling_software. Accessed on 12 July 2010.

Wilensky, U. (2006). NetLogo grand canyon model. Available at http://ccl.northwestern.edu/netlogo/models/GrandCanyon. Accessed on 15 July 2010.

Zeiler, M. (1999). *Modelling our world: The ESRI guide to geodatabase design*. Redlands: ESRI Press.

Chapter 13
Space in Agent-Based Models

Kiril Stanilov

Abstract The chapter offers an overview of the issues related to the integration and representation of space in agent-based models (ABMs), with a focus on those models that can be considered spatially explicit. Key aspects of space in ABM are highlighted, related to: the role of space as an attribute of agents and the environment; as an interaction component; as a determinant of issues of scale; and as a tool for communicating and validating model outcomes. The chapter reviews the issues and challenges arising from the difficulties of integrating space in agent-based modeling. It outlines the emerging trend towards improving the level of realism in representing space, which can lead not only to an enhanced comprehension of model design and outcomes, but to an enhanced theoretical and empirical grounding of the entire field of agent-based modelling.

13.1 Introduction

One of the main characteristics of agent-based systems is that the interactions of the modeled agents do not take place in a vacuum, but are situated within structures that both condition agents' behavior and are in turn influenced by it (Epstein and Axtell 1996). These interaction structures can be physical or social environments, or networks that encode geographic or other feature-based differences (Riolo et al. 2001). Consequently, a key advantage of ABMs is their ability to integrate these two components – agents and their environment – through systematic specification of interdependencies and feedbacks (Parker et al. 2003).

K. Stanilov (✉)
Centre for Applied Spatial Analysis (CASA), University College London, London, UK
e-mail: kiril.stanilov@ucl.ac.uk

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_13,

It should be noted, however, that traditionally the emphasis in agent-based modeling has been clearly placed on the development of agents and their behavior at the expense of less sophisticated representations of space and spatial relationships (Brown et al. 2005). Many ABMs, in fact, consider spatial relationships as a marginal issue, or at least treat space as a feature of the model that becomes relevant only at the macro scale. Examples of such models include investigations of social and cultural phenomena such as investment management, dynamics of labor markets, shifts in consumer behavior, or spread of technological innovations. In contrast, another string of ABMs, more tightly related to investigations of geographic phenomena, considers space as an integral component of their system. These models, referred in the literature as spatially explicit ABMs, include a diverse group of studies ranging from explorations of urban growth and natural resource management to agricultural economics and archaeology. Commonly, these models try to establish explicit links between environmental characteristics and agent behavior (Benenson and Torrens 2004).

The discussion offered in this chapter on issues related to space and its representation in ABMs is centered on those models that can be considered spatially explicit. The review of the literature on which this paper is based is far from balanced as it relies heavily on examples from the field of urban modeling. This is due partially to the author's background, but more importantly to the fact that in urban modeling the consideration of space is inevitably explicit (Berger et al. 2002). The proliferation of spatially explicit ABMs in the last 10 years is particularly impressive in the area of land use analysis where such models have become popular tools for understanding land-use systems (Polhill et al. 2001; Deadman et al. 2004). Here ABMs are considered particularly well suited for representing complex spatial interactions under heterogeneous conditions (Parker et al. 2003).

The discussion of space offered on the following pages is structured into two parts. The first one provides an overview of the general concepts of space and its integration within agent-based modeling. The key aspects of space in ABM are highlighted related to: the role of space as an attribute of agents and the environment; as an interaction component; as a determinant of issues of scale; and as a tool for communicating and validating model outcomes. A further discussion in this section addresses the various ways in which space is represented in the ABM world. The second part of the chapter reviews the issues and challenges arising from the difficulties of integrating space in agent-based modeling. The most promising venues towards a better representation of space are outlined, reviewing the shift from cell-based to object-based applications. The chapter concludes by sketching the contours of an emerging trend aimed to move the theory and practice of ABM beyond the grid-vs-vector debate, offering some new prospects for the integration of space within agent-based modeling frameworks.

13.2 The Concept of Space in ABM

13.2.1 The Integration of Space in Modeling Frameworks

This section outlines several aspects of space critical for its integration within spatially explicit agent-based modeling systems.

13.2.1.1 Space as an Attribute

An apparent role of space in ABMs that try to incorporate the significance of spatial phenomena in the simulation of social processes is the function of space as an attribute of a model's components – both of the environment and of the agents that operate within it.

The spatial characteristics of the environment could be represented with various levels of detail (this topic is discussed in more detail later in this section), but at a minimum, the model environment could be described as a non-differentiated plane with geographic or relative coordinates on which the actions of the agents take place. In such models the environment influences the agents' interactions simply by measures of distance and direction (Castle and Crooks 2006). In models that represent the physical characteristics of the environment with a greater level of sophistication, the agents respond to attributes of the landscape such as physical barriers, soil types, infrastructure, or aesthetic qualities by adopting their behavior to the features of the modeled environment.

Space as a characteristic of agents in ABMs is a more flexible concept. The agents could be spatially explicit or they could be implicit (meaning that their precise spatial location is not essential for the operation of the model). In addition, spatially explicit agents could be static (tied to a specific location in the environment) or dynamically situated (free to move within the environment either with or without predefined constraints).

13.2.1.2 Types of Space-Agent Interactions

Due to the wide variety of details with which both the environment and the agents within an ABM could be specified, the nature of the interactions between them could be rather complex. First, it is possible for an agent to be associated with only one spatial feature in a one-to-one relationship. A typical example of such a relationship is a household and its place of habitation in a simple residential location model or a local government and its jurisdiction in an urban growth management simulation. An agent, however, could be associated with more than one spatial feature in a one-to-many relationship. Examples of such cases are models in which households are linked with their places of residence, work, shopping, entertainment, etc.

In addition to the level of connectivity, there are two ways in which environment-agent interactions could be constructed: as a simple unidirectional relationship in which the environment is affected by the behavior of the agents (or vice versa), or as a multidirectional cycle of interactions and feedbacks between the two. Examples of models integrating space in a simple one-way causal environment-agent relationship are relatively few. In such models the environment is the only factor governing agent behavior. The agents adopt strategies that allow them to react to a heterogeneous environment given their goals and actions (Parker et al. 2003). Alternatively, the causal relationship could be pointed the other way by modeling changes in the environment as a result of the agents' behavior. Examples here include studies of deforestation due to agricultural practices, fragmentation of the natural habitat due to urban sprawl, etc.

In reality, the interactions between humans and their environment are always more complex, never confined to a single unidirectional link – a fact which is recognized by the majority of agent based modelers. A good example of the complexity of environment-agent interaction is urban gentrification, where a chain of events dynamically transforms both the actors and the environment. In this process, agents are drawn to urban areas due to specific locational or environmental characteristics; they engage in interactions with other actors in the local property market thus changing its dynamics; as a result the environment is changing; this in turn draws new actors to the scene affecting further the dynamics of the process. Another good example of modeling the complexity of environment-agent interactions is the SLUCE model of residential location at the urban fringe (Rand et al. 2002; Brown et al. 2005). Here residents make decisions about where to locate based on a combination of environmental factors including density, distance to service centers, and the aesthetic quality of the landscape. New service centers locate near recent residential development, influencing, in turn, the behavior of future homebuyers. A main challenge for the models exploring the complexity of environment-agent linkages is to separate the effects of endogenous interactions from spatially correlated exogenous landscape features (Irwin and Bockstael 2002).

13.2.1.3 Space and Scale

Scale is another important aspect of the task of integrating space in ABM frameworks. The issues of scale become relevant in the construction of the model in two distinct ways linked to the determination of the spatial extent and the spatial resolution of the data used (Goodchild 2001). First, in terms of the spatial extent of the modeled area, studies have demonstrated that changes in spatial extent have a significant impact on the outcomes of spatial analysis (Saura and Millan 2001). This fact highlights the need to capture processes at the scale at which they operate. This principle of scale-dependency is also particularly important in determining the level of spatial resolution, or the level of detail captured in the model (Lam and Quattrochi 1992). A coarse granularity of the data tends to iron out both spatial heterogeneity and spatial dynamics (Batty 2005). The issues of spatial aggregation are particularly relevant for ABMs that try to capture emergent behavior (Goodchild 2001). The "modifiable areal unit problem" (MAUP) and associated issues of ecological fallacy (Openshaw 1983) loom large in all models based on assumptions that larger units are representative of smaller units. While this does not seem to be an issue with the specification of agents, which is commonly done at the level of individuals and households, finding the proper level of representation of environmental characteristics and processes presents significant methodological difficulties. The use of a very fine data resolution, on the other hand, has been found to produce patterns that are overly fragmented (Menard and Marceau 2005; Chen and Mynett 2003). Finally, making the integration of space in ABM an even more challenging task, is the recognition of the fact that an individual agent is likely influenced by, and in turn influences, processes occurring at multiple spatial scales (Batty 2005; Parker et al. 2003).

The consideration of scale also becomes pertinent in ABM through the definition of neighborhoods of interaction. In the classic cellular automata (CA) conceptualizations on which the majority of ABM environments are based, neighborhoods are defined on the principle of spatial proximity. Here the magnitude of interaction is described as a distance decay function following Tobler's law, which postulates that near things are more related than distant things (Tobler 1970). While the size of the neighborhoods in many CA and ABMs is predetermined by a fixed (and in many cases somewhat arbitrary) radius, a relatively small number of studies have carried out systematic analysis of the impact of this critical neighborhood parameter. A recent study of residential segregation, for instance, has emphasized the importance of scale over the shape of neighborhoods, which in this case is interpreted as the field of the agents' vision (Fossett and Dietrich 2009). Other studies have proposed more refined techniques of neighborhood definition taking into account different spatial scales relevant for the modeled interactions (Batty et al. 1999; Vancheri et al. 2008). In recognition of the larger spatial scale at which neighborhood interactions operate, some scholars have introduced the concept of domains – large scale spatial ensembles representing a group of neighborhoods populated by agents of homogeneous characteristics – devising algorithms for the identification of emerging domains and techniques for following their evolution (Benenson et al. 2005).

It should be noted, however, that operational ABMs of larger scale systems such as metropolitan areas are still quite rare (e.g. Benenson et al. 2002; Mathevet et al. 2003), pointing to a lack of studies taking on the challenge of modeling processes that operate on multiple spatial scales (e.g., from the level of individual parcels and neighborhoods to the scale of urban regions). The task of simulating large-scale dynamics based on detailed representation of micro-scale processes poses many new challenges in terms of computational algorithms, data organization, and model architecture (Ettema et al. 2007).

13.2.1.4 Space as a Tool of Validation and Communication

Another important aspect of integrating space in ABM is its utility as a powerful tool of communication and validation of model outcomes. These two areas – communication and validation – have been identified as key challenges for the future development of the ABM field (Crooks et al. 2008) having received so far only scant coverage in the professional literature. This fact is somewhat surprising, considering that in many cases a comparison between model outcomes and real data along their spatial characteristics is the ultimate form of model validation. Yet one needs to be aware that location-specific estimates based solely on landscape metrics may not be as useful as having model outcomes reproduce realistic patterns, or as Mandelbrot simply put it – they must "look right" (Mandelbrot 1983). And while ABMs have the potential of being more easily comprehended by the general public due to the fact that they simulate "real world" behavior based on simple rules, quite often the outcomes of these models are not immediately transparent for a wide range of potential users who happen to lack the appropriate technical background for interpreting the results. In this sense, the visualization of model outcomes through maps and other types of commonly used spatially referenced information can serve as a great medium of communicating a model's results, reaching effectively a wider range of users (Axtell 2000).

13.2.2 The Representation of Space

The level of detail with which the environment is described in spatially explicit ABMs depends primarily on the type and the purpose of the model. Thus while in theoretical interaction models environmental characteristics are traditionally simplified (Irwin and Bockstael 2002), in models that are based on real-world locations the representation of landscape heterogeneity is a critical feature of a model's design. These two approaches have been referred to in the literature as *designed* (in the case of the more abstract theoretical models) and *analyzed* (in the case of applied inductive studies) (Parker et al. 2003). It should be noted that the distinction between the two approaches in not always clear-cut, with a substantial number of models straddling the boundary between abstract and more realistic representations. At the same time, since the early days of ABM, there has been a gradual yet noticeable trend towards more detailed representations of socio-spatial systems (Epstein and Axtell 1996). This could be explained by the natural course of the evolution of the field striving for higher fidelity of the modeled reality on one hand; and the increasing pressure to develop tools that are geared towards end-users and other stakeholders on the other (Matthews et al. 2007).

In general, the majority of spatially explicit ABMs rely on a regular cell framework used as a basis for representing the environment (Barros 2003; Batty et al. 2003). This concept of spatial organization is borrowed directly from the field of CA due to the kinship between the two modeling techniques in the analysis of related socio-spatial phenomena. With the conceptual linkages between ABM and CA being so tight, often CA models are re-interpreted as ABMs by attributing anthropomorphic state variables to cells (Torrens and Benenson 2005), using transition rules as proxies to decision making (Parker et al. 2003). However, regardless of these attempts, an important distinction remains. While CA can be described entirely through the interaction of spatial phenomena, they do not provide support for typical actor-based processes (Ligtenberg et al. 2001). As a result, CA models rely on a fixed interaction topology whereas the interactions in ABMs can be changed dynamically since they are defined at the level of mobile agents (Brown 2005).

The general model formulation, based on CA populated by agents migrating between cells, seems to be a natural process of merging ABM and CA by building on the strengths of each modeling approach (Portugali et al. 1994; Portugali and Benenson 1997). The ability of such systems to separate out the influence of actors, institutions, and the environment have been enthusiastically embraced more specifically in urban high-resolution modeling (Parker et al. 2003; Manson 2006). Here, the urban environment is represented in two layers, one for the city's infrastructure (immobile), and the other for migrating human individuals (mobile) (Portugali 2000; Polhill et al. 2001). Correspondingly, in many land-change models, agents choose cells from a gridded landscape for their productive utility, either for agriculture or home building (Evans and Manson 2007).

A key conceptual dilemma in the construction of model environments in ABMs is in the choice of selecting the best way to represent the environment's critical properties. Choosing between raster vs. vector-based representations is not always an easy decision to make. While raster-based structures are best fitted to capture continuous field data, vectors are best suited to depict the properties of discrete objects. Since the natural and built environments are composed of both, the question is which way would be most appropriate for capturing the essence of the modeled spatial phenomena. Traditionally, the prevailing practice in both CA and ABMs has been to favor a rigid partitioning of space into regular cells, and there are several factors that have solidified this choice. Some of the main reasons include the conceptual foundations of CA theory and its grounding in cell space; the prevailing availability of remote sensing data in raster formats; the advantages of using the functionality of raster-based GIS data preparation and analysis in model development; and the computational efficiency of working with regular grids (Stanilov 2009).

Deviations from the practice of using a rectangular tessellation of space in CA and ABMs have included experimentation with hexagonal grids (Phipps 1989; Sanders et al. 1997), yet it has been recognized that in order to make models applicable in the arena of public policy, modelers need to move away from abstract cellular representations in order to incorporate the detailed geography of the real places (Xie and Batty 2003). While the literature has long suggested the integration of irregular structures in microsimulation (Couclelis 1985), only recently have ABMs begun to use real-world spatial data (Brown et al. 2005). Early attempts have considered non-uniform partitions of urban space, accounting exclusively for infrastructure units (Erickson and Lloyd-Jones 1997; Semboloni 2000). One of the first ABMs to use real-world geographic features was developed in the field of natural resource management, in a simulation of the recreational use in a state park in Arizona (Gimblett et al. 2002). More recent work based on the integration of parcel-level data has included the development of custom-built model environments such as MABEL (Alexandridis and Pijanowski 2007), but most common has become the use of hybrid raster-vector environments in which vector-based features are used to calculate spatial attributes of raster-based cells such as calculating accessibility of cells based on the distance to the road network (Brown et al. 2008).

The evolution of the grid vs. vector dilemma within the field of agent-based modeling is discussed in more detail in the following section which offers a summary of the main challenges related to the integration of space within ABMs.

13.3 Issues and Challenges

One of the main challenges for agent-based modeling is to move both practice and theory from the arena of experimental and hypothetical applications towards empirically-based research (Berger and Schreinemachers 2006; Janssen and Ostrom 2007). This process entails a transition from abstract towards more realistic representations of the environment (Torrens and O'Sullivan 2001). While CA and agent-based systems have been introduced in the modeling world with the intent to infuse it with a recognition of the finer scale on which spatial relationships operate in both the natural and the built environments, these models, in their majority, continue to be based on highly restrictive assumptions related to the integration and representation of space. This situation has been primarily a function of the limitations imposed by the direct utilization of the generic spatial constructs underlying CA theory, rather than the application of empirical or theoretical knowledge on how systems function in space (Torrens and Benenson 2005).

13.3.1 From Cells to Objects

The deficiencies of employing a rigid tessellation of space as a basis of ABM environments stem from the fact that pixel-based cellular dynamics seldom match spatial phenomena (Xie and Batty 2003). To begin with, many linear features of both the natural and the built environment (rivers, infrastructure, etc.) do not lend themselves to be easily represented in a grid format that engenders the proper integration of network elements in the specification of spatial interactions (Benenson et al. 2005). Additional problems arise with the depiction of entities and agents that are either larger or smaller than a single grid cell. The representation of entities larger than the size of the basic modular unit calls for aggregation of cells based on a unique shared attribute describing the identity of the depicted object. The grouping of cells on this principle, however, creates conceptual and computational problems challenging the basic premises on which cell-based structures operate. In cases when cells are larger than the spatial entities on which they are superimposed, the issue of cell heterogeneity presents significant methodological challenges. The problems created by such a tessellation could be as difficult to address as the MAUP in aggregate models where the boundaries are at least drawn with the idea of maintaining a certain level of area homogeneity.

There have been several attempts to increase the fidelity of the model outcomes by fine-tuning the size of the grid cells of the lattice underlying CA and ABM environments. While common sense logic would suggest that smaller cell sizes increase data resolution, thus leading to more accurate results, in some cases the findings of sensitivity analysis indicate that a coarser resolution can generate more realistic spatial patterns (Jenerette and Wu 2001). Support for this claim has been provided by studies concluding that using the finest resolution does not provide the best results (Menard and Marceau 2005; Chen and Mynett 2003). Overall, there appears to be a general agreement shared in the field that the choice of cell size has considerable impact on simulation results (Kocabas and Dragicevic 2006), and that one needs to perform a systematic sensitivity analysis to determine the optimal cell size for a particular model (Jantz and Goetz 2005). This task, however, takes a significant amount of resources and ultimately the selection of cell sizes in many projects is determined somewhat arbitrary, mostly relying on previous studies.

The problems stemming from the application of abstract rectangular grids as a spatial framework for modeling are compounded further by the use of rigid raster cells for defining the spatial extent of neighborhoods of influence. While the utilization of a universal nondiscriminatory grid might be appropriate in modeling environmental processes where influence is mainly determined by proximity (e.g., the spread of brushfires) in urban environments spatial relationships tend to be much more complex in their dimensions and magnitude of interaction.

The field of CA/ABM abounds with experimentation aimed at optimizing the definition of neighborhoods (much more so than with studies questioning the applicability of raster lattices). In the majority of cases, this has included experiments with extending the radius of influence beyond the traditional von Neumann and Moore neighborhoods (White et al. 1997). Some have used hierarchical neighborhoods defined on a neighborhood, regional, or global level. Others have proposed to define neighborhood interactions based on empirical analysis derived from neighborhood characteristics by calculating over- or under-representation of particular parameters (e.g., land use class) relative to their representation in the entire study area (Verburg et al. 2004).

The issues associated with the application of rigid grid lattices has spurred a strand of research exploring the utility of alternative conceptualizations of spatial structures underlying model environments, including the Voronoi model of spatial tessellation (Shi and Pang 2000; Flache and Hegselmann 2001) and the use of graph-based CA (O'Sullivan 2001; Torrens and Benenson 2005). This path of exploration has drawn its own share of critics, pointing to the fact that Voronoi polygons do not correspond to real-world entities, but are generated automatically for simplicity of computation.

Recently, attempts to link closer the tessellation of space to real world entities have been emphasized in the field of urban modeling with several studies employing parcels as the basic unit of spatial organization (Stevens and Dragicevic 2007; Alexandridis and Pijanowski 2007). The use of parcel-based cells in urban ABMs offers several avenues for refining the definition of neighborhoods and transition rules that are not available in the conventional raster-based modeling environment. The utilization of a cadastral–based lattice provides an opportunity to incorporate important parameters of spatial interaction that cannot be accounted for in the traditional grid-based models. Such systems of structuring the modeled environment can be linked to the following methodological advantages (Stanilov 2009):

- An environment in which cells are based on parcel boundaries allows for the integration of cell size as a factor of spatial interaction, reflecting the fact that smaller parcels exert a smaller impact on neighboring cells and vice versa.
- Parcel-based cells can account for variations in the magnitude of cell interaction that are due not only to differences in the size of neighboring cells but in their mutual orientation as well. Such relationships are captured by the length of their shared boundaries.
- Parcel-based cells also have the advantage of being homogeneous in terms of their land use. This allows for a more precise definition of land use interactions, thus eliminating the problems associated with cell heterogeneity.
- Parcel-based cells can take cognizance of variations in the intensity of development better than nondescript raster cells. The use of parcel boundaries can capture, for instance, the fact that a large parcel with a small building footprint can have less of an impact on its neighboring cells than an intensely developed smaller parcel.

The use of cadastral property lines as a basis for creating the underlying lattice of a model environment is of fundamental importance for capturing the essence of urban form generation. Research in urban morphology has consistently stressed the essential role that land ownership patterns play in setting up the spatial configuration of urban environment. Parcel boundaries, although not physical entities per se, outline the basic spatial framework within which the urban landscape is constituted (Conzen 1960). The use of historic cadastral boundaries makes particular sense in the context of modeling the growth of the urban periphery where the pre-urban cadastre has set the basic framework within which the pieces of urban development are distributed.

The integration of parcel data in ABM indicates a new direction for the development of the field marked by the transition from raster to vector-based data and from cells to objects as descriptors of both agents and their environment. Indeed, some of the most exciting and promising theoretical advances in ABM in recent years have been related to experimentations with the object-oriented data modeling approach. Such developments have been driven by the similarity in abstraction shared between the agent-based and object-oriented paradigms (Castle and Crooks 2006). The fact that most ABMs use object-oriented programming languages, such as C++, Java, or Objective-C, points naturally to conceptualizations describing the environment as a collection of spatially discrete objects (Benenson and Torrens 2004). The possibilities for the effective implementation of the object-based approach seem to be most frequently recognized in the development of high-resolution simulations of urban dynamics.

One of the most conceptually advanced systems of this type is the Object-Based Environment for Urban Simulation (OBEUS) (Torrens and Benenson 2005; Benenson et al. 2005). Here discrete objects directly represent real-world urban entities and both agents and features are treated as individual automata situated in space through a set of geo-referencing rules. The model distinguishes between fixed objects (described with the coordinates of their vertices, edges, centroids, minimal bounding rectangles, etc.) and non-fixed urban objects identified by pointing to one or several fixed ones. Neighborhoods are defined by Voronoi coverages constructed on the base of centroids, and by interaction rules which allow neighborhoods to be varied in space or time in the course of the simulation. Such object-based models have the added advantage in their ability to assign temporal and location behavior as an attribute of features rather than space itself, allowing objects to be updated asynchronously (Castle and Crooks 2006).

In spite of the numerous advantages of employing an object-based modeling framework, there have been a relatively limited number of cases embracing this approach in the field of ABM. The reluctance to venture into this territory is related to several factors. First, compared to models based on raster data, vector-based structures require significant computational resources and object-based programming knowledge. In addition, the departure from traditional cellular-based space representations leads to several conceptual problems (Castle and Crooks 2006). A major obstacle is that, while the neighborhood relationship between identical cells in a CA-based model do not vary, in an object-based vector model the magnitude of the neighborhood interactions is impacted by the spatial attributes of the objects (Benenson et al. 2005), which makes them conceptually and procedurally difficult to model.

Another problem in object-based modeling arises from the challenge of dynamically updating connected or adjacent features whose shapes change over time (Miller 1999). In such cases the space-time topology of objects' vectors becomes increasingly complex as amendments accumulate during the simulation runs (Castle and Crooks 2006). Of particular interest in urban modeling, for instance, are the processes of parcel subdivision or amalgamation which underline the morphogenetic processes of growth. Yet, due to the issues outlined above, these processes have not found adequate representation in ABMs so far. The few attempts to incorporate dynamic repartitioning of space rely on rather mechanically construed Voronoi tessellation algorithms (Semboloni 2000; Benenson et al. 2005) that do not bear much resemblance to the complex patterns generated by the processes of land subdivision. In spite of the growing number of experiments with the object-based approach, moving forward from agents with fixed vector boundaries remains to this day a seemingly insurmountable challenge in ABM (Hamman et al. 2007).

13.3.2 Beyond the "Grid vs. Vector" Debate

Another interesting area of development within ABMs, situated outside the territory of the grid vs. vector and cells vs. objects debate, is composed of a recent group of studies concerned with the integration of urban form characteristics that have been previously overlooked. An early example of such an attempt is the ILUTE project (Miller et al. 2004) in which the built environment is described by the type and amount of floorspace, while transition rules incorporate the age of development as well as local and global vacancy rates. Similar attributes of the built environment are used in another detailed land use change model, which adds to the spatial parameters the amount of land surface covered by buildings, thus identifying spatial resources available in each cell for further development (Vancheri et al. 2008).

A further effort to capture key features of the built environment in an ABM structure is aimed at incorporating representation of physical design elements. The DEED model (Brown et al. 2008) locates residential agents using a utility calculation that considers the landscape characteristics associated with a range of subdivision types. Each of the four types is defined on the basis of observed land-cover proportions and patterns, street patterns, and lot sizes. The characteristics of subdivision design are also incorporated in a high-resolution data model which evaluates how different subdivision designs might influence development under varying population growth rates and buyer preferences (Stevens and Dragicevic 2007).

A logical step in the progression towards higher levels of sophistication with which the environment is represented in ABMs is the incorporation of the third dimension of space. The field of ABM has traditionally been dominated by two dimensional approaches, with very few experiments venturing into 3D space (Dibble and Feldman 2004; Thorp et al. 2006). Most of these projects are conceptual developments creating hypothetical environments such as CityDev, which offers an interactive multi-agent simulation model of city development organized spatially in cubic cells (Semboloni et al. 2004). A few studies, however, have tried to incorporate 3D features into models simulating the development of real urban environments. Of particular interest among these examples is the quality of views offered within a given landscape. In such studies, viewshed analysis is used to describe the degree of visibility as a determining factor for residential location (Yin and Muller 2007).

An interesting venue of exploration within the ABM world is the use of 3D environments for the purposes of visualization (see Patel and Hudson-Smith 2012 for an overview of visualizing ABM outputs). One of the first illustrations of such capabilities utilized a combination of Repast software libraries and GIS layers (Dibble and Feldman 2004), allowing the movement and interaction of agents to be followed in real-time 3D networks. The system has been used to model a number of socio-spatial phenomena including the transmission of infectious diseases, the dynamics of civil violence, and the coordination of social networks. Latest attempts to develop further conceptually the application of 3D visualization include the idea of moving ABM simulation environments from individual workstations to collaborative geographic space using Second Life as a platform for the dissemination of geographic content (Crooks and Hudson-Smith 2008). Such experiments underscore the great potential for the development of the field charted by the advancement of the concepts of space within ABMs.

13.4 Conclusions

The primary strength of ABMs is as a testing ground for a variety of theoretical assumptions and concepts about human behavior. As a result of this concentration on behavior-driven social processes, ABMs tend to be traditionally less concerned with realistic representation of the physical environment. Therefore, they are rarely used as predictive models for real-world sites where the concern is that they can be overly fitted to existing data, thus losing their power of generalization or ability to explore alternative systems.

As the field of ABMs develops and matures, it has faced the need to refine its underlying theoretical concepts, including the role played by the environment in conditioning the interactions of agents. Research has highlighted the point that dynamic behavior-based processes can be significantly impacted by even small changes in underlying spatial structures (O'Sullivan 2001). This has directed the attention of agent-based modelers towards new paths for better integration and representation of the spatial aspects of the modeled environment.

The most numerous group of such studies have been constrained within a general effort to refine CA-based structures, which continue to be utilized as an underlying environment for the majority of ABMs. These efforts have included the employment of higher resolution data, larger areal extents, and experiments with alternative methods of grid tessellation. An interesting departure from the dominant tradition is based on the work of a relatively small but growing number of researchers who have tried to break away from the bind of CA constructs by experimenting with environments defined by vectors and objects. This approach holds the promise of producing very interesting results, especially in view of the natural affinity between the agent-based and object-oriented paradigms. The third stream of innovations in the integration of space in ABMs is built on the idea of achieving a richer representation of the spatial characteristics of the environment through the inclusion of features that have been previously overlooked but which might have a critical importance for the dynamics of the modeled phenomena. An important conceptual leap forward here is the inclusion of the third dimension of space which opens up exciting opportunities for exploration of model parameters and the visualization of simulated phenomena.

All of these new avenues of exploration present new challenges for the development of the field of agent-based modelling. Many of the conceptual and technical considerations related to the integration of space are pushing the field forward as modellers are charged to apply forward thinking, which should not be confined by the limitations of the tools and concepts in currency today. This chapter has presented the argument that improving the level of realism in representing space can lead not only to an enhanced comprehension of model design and outcomes, but to an enhanced theoretical and empirical grounding of the entire field of agent-based modelling. It appears that this new decade will be a critical time for meeting these goals.

References

Alexandridis, K., & Pijanowski, B. C. (2007). Assessing multiagent parcelization performance in the MABEL simulation model using Monte Carlo replication experiments. *Environment and Planning B: Planning and Design, 34*(2), 223–244.

Axtell, R. L. (2000). *Why agents? On the varied motivations for agent computing in the social sciences, center on social and economic dynamics* (Working Paper 17). Washington ,DC: The Brookings Institute.

Barros, J. (2003). Simulating urban dynamics in Latin American cities. In *Proceedings of the 7th International Conference on Geocomputation*, University of Southampton, Southampton.

Batty, M. (2005). Agents, cells, and cities: New representational models for simulating multiscale urban dynamics. *Environment and Planning A, 37*, 1373–1394.

Batty, M., Xie, Y., & Sun, Z. (1999). Modeling urban dynamics through GIS-based cellular automata. *Computers, Environment and Urban Systems, 23*, 205–233.

Batty, M., Desyllas, J., & Duxbury, E. (2003). The discrete dynamics of small-scale spatial events: Agent-based models of mobility in carnivals and street parades. *International Journal of Geographic Information Science, 17*(7), 673–697.

Benenson, I., & Torrens, P. M. (2004). Geosimulation: Object-based modeling of urban phenomena. *Computers, Environment and Urban Systems, 28*, 1–8.

Benenson, I., Omer, I., & Hatna, E. (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., Aronovich, S., & Noam, S. (2005). Let's talk objects: Generic methodology for urban high-resolution simulation. *Computers, Environment and Urban Systems, 29*(4), 425–453.

Berger, T., & Schreinemachers, P. (2006). Creating agents and landscapes for multiagent systems from random samples. *Ecology and Society, 11*(2), 19.

Berger, T., Couclelis, H., Manson, S. M., & Parker, D. C. (2002). Agent based models of LUCC. In D. C. Parker, T. Berger & S. M. Manson (Eds.), *Agent-based Models of Land Use and Land Cover Change,* (*LUCC*) (Report Series No. 6) (pp. 1–2). LUCC Focus 1 Office, Indiana University, Bloomington.

Brown, D. G. (2005). Agent-based models. In H. Geist (Ed.), *The earth's changing land: An encyclopedia of land-use and land-cover change*. Westport: Greenwood Publishing Group.

Brown, D. G., Riolo, R., Robinson, D. T., North, M., & Rand, W. (2005). Spatial process and data models: Toward integration of agent-based models and GIS. *Journal of Geographic Systems, 7*(1), 25–47.

Brown, D. G. R., An, L., Nassauer, J. I., Zellner, M., Rand, W., Riolo, R., Page, S. E., Low, B., & Wang, Z. (2008). Exurbia from the bottom-up: Confronting empirical challenges to characterizing a complex system. *Geoforum, 39*(2), 805–818.

Castle, C., & Crooks, A. T. (2006). *Principles and concepts of agent-based modelling for developing geospatial simulations* (Working Paper 110). London: CASA.

Chen, Q., & Mynett, A. E. (2003). Effects of cell size and configuration in cellular automata based prey–predator modeling. *Simulation Modelling Practice and Theory, 11*, 609–625.

Conzen, M. R. G. (1960). *Alnwick, Northumberland: A study in town plan analysis* (Publication No. 27). London: Institute of British Geographers.

Couclelis, H. (1985). Cellular worlds: A framework for modeling micro-macro dynamics. *Environment and Planning A, 17*, 585–596.

Crooks, A. T., & Hudson-Smith, A. (2008). Techniques and tools for three dimensional visualisation and communication of spatial agent-based models. In *Proceedings from Agent-based Spatial Simulation Workshop*. Paris: ISC-PIF.

Crooks, A., Castle, C., & Batty, M. (2008). Key challenges in agent-based modelling for geo-spatial simulation. *Computers, Environment and Urban Systems, 32*, 417–430.

Deadman, P. J., Robinson, D. T., Moran, E., & Brondizio, E. (2004). Effects of colonist household structure on land use change in the Amazon Rainforest: An agent based simulation approach. *Environment and Planning B: Planning and Design, 31*, 693–709.

Dibble, C., & Feldman, P. G. (2004). The GeoGraph 3D computational laboratory: Network and terrain landscapes for repast. *Journal of Artificial Societies and Social Simulation*, *7*(1). Available at: http://jasss.soc.surrey.ac.uk/7/1/7.html

Epstein, J., & Axtell, R. (1996). *Growing artificial societies: Social science from the bottom up*. Cambridge: MIT Press.

Erickson, B., & Lloyd-Jones, T. (1997). Experiments with settlement aggregation models. *Environment and Planning B: Planning and Design, 24*(6), 903–928.

Ettema, D., de Jong, K., Timmermans, H., & Bakema, A. (2007). PUMA: Multi-agent modelling of urban systems. In *45th Congress of the European Regional Science Association.* Amsterdam: Vrije Universiteit.

Evans, T. P., & Manson, S. (2007). Space, complexity, and agent-based modeling – Editorial. *Environment and Planning B: Planning and Design, 34*(2), 196–199.

Flache, A., & Hegselmann, R. (2001). Do irregular grids make a difference? Relaxing the spatial regularity assumption, in cellular models of social dynamics. *Journal of Artificial Societies and Social Simulation*, *4*(4). Available at: http://www.soc.surrey.ac.uk/JASSS/4/4/6.html

Fossett, M., & Dietrich, D. R. (2009). Effects of city size, shape, and form, and neighborhood size and shape in agent-based models of residential segregation: Are Schelling-style preference effects robust? *Environment and Planning B: Planning and Design, 36*, 149–169.

Gimblett, H. R., Richards, M. T., & Itami, R. M. (2002). Simulating wildland recreation use and conflicting spatial interactions using rule-driven intelligent agents. In H. R. Gimblett (Ed.), *Integrating geographic information systems and agent-based modeling techniques for simulating social and ecological processes* (pp. 211–243). Oxford: Oxford University Press.

Goodchild, M. (2001). Issues in spatially explicit modeling. In D. Parker, T. Berger & S. M. Manson (Eds.), *Agent-based models of land-use and land-cover change* (pp. 13–17). Irvine.

Hamman, Y., Moore, A., & Whigham, P. (2007). The dynamic geometry of geographical vector agents. *Computers, Environment and Urban Systems, 31*(5), 502–519.

Irwin, E., & Bockstael, N. (2002). Interacting agents, spatial externalities, and the evolution of residential land-use patterns. *Journal of Economic Geography, 2*(1), 31–54.

Janssen, M. A., & Ostrom, E. (2007). Empirically based agent-based modeling. *Ecology and Society, 11*(2), 37.

Jantz, C. A., & Goetz, S. J. (2005). Analysis of scale dependencies in an urban land-use-change model. *International Journal of Geographical Information Science, 19*(2), 217–241.

Jenerette, G. D., & Wu, J. (2001). Analysis and simulation of land-use change in the central Arizona – Phoenix region, USA. *Landscape Ecology* (16), 611–626.

Kocabas, V., & Dragicevic, S. (2006). Assessing cellular automata model behaviour using a sensitivity analysis approach. *Computers, Environment and Urban Systems, 30*(6), 921–953.

Lam, N., & Quattrochi, D. A. (1992). On the issues of scale, resolution, and fractal analysis in the mapping sciences. *The Professional Geographer, 44*, 88–98.

Ligtenberg, A., Bregt, A. K., & von Lammeren, R. (2001). Multi-actor-based land use modelling: Spatial planning using agents. *Landscape and Urban Planning, 56*, 21–33.

Mandelbrot, B. B. (1983). *The fractal geometry of nature*. New York: W.H. Freeman.

Manson, S. (2006). Land use in the southern Yucatan peninsular region of Mexico: Scenarios of population and institutional change. *Computers, Environment and Urban Systems, 30*, 230–253.

Mathevet, R., Bousquet, F., Le Page, C., & Antona, M. (2003). Agent-based simulations of interactions between duck population, farming decisions and leasing of hunting rights in the Camargue (Southern France). *Ecological Modelling, 165*, 107–126.

Matthews, R. B., Gilbert, N. G., Roach, A., Polhill, J. G., & Gotts, N. M. (2007). Agent-based land-use models: A review of applications. *Landscape Ecology, 22*, 1447–1459.

Menard, A., & Marceau, D. J. (2005). Exploration of spatial scale sensitivity in geographic cellular automata. *Environment and Planning B: Planning and Design, 32*, 693–714.

Miller, H. J. (1999). Measuring space-time accessibility benefits within transportation networks: Basic theory and computation procedures. *Geographical Analysis, 31*(2), 187–213.

Miller, E. J., Hunt, J. D., Abraham, J. E., & Salvini, P. A. (2004). Microsimulating urban systems. *Computers, Environment and Urban Systems, 28*(1–2), 9–44.

Openshaw, S. (1983). *The modifiable areal unit problem* (CATMOG 38). Norwich: GeoBooks.

O'Sullivan, D. (2001). Graph-cellular automata: A generalised discrete urban and regional model. *Environment and Planning B Planning and Design, 28*, 687–705.

Parker, D. C., Manson, S. M., Jansen, M. A., Hoffmann, M. J., & Deadman, P. (2003). Multi-agent systems for the simulation of land-use and land-cover change: A review. *Annals of the Association of American Geographers, 93*(2), 314–337.

Patel, A., & Hudson-Smith, A. (2012). Agent-tools, techniques and methods for macro and microscopic simulation. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 379–407). Dordrecht: Springer.

Phipps, M. (1989). Dynamical behavior of cellular automata under the constraint of neighborhood coherence. *Geographical Analysis, 21*, 197–215.

Polhill, J. G., Gotts, N. M., & Law, A. N. R. (2001). Imitative versus nonimitative strategies in a land use simulation. *Cybernetics and Systems, 32*(1–2), 285–307.

Portugali, J. (2000). *Self-organization and the city*. Berlin: Springer.

Portugali, J., & Benenson, I. (1997). Human agents between local and global forces in a self-organizing city. In F. Schweitzer (Ed.), *Self-organization of complex structures: From individual to collective dynamics* (pp. 537–546). London: Gordon & Breach.

Portugali, J., Benenson, I., & Omer, I. (1994). Sociospatial residential dynamics: Stability and instability within a self-organizing city. *Geographical Analysis, 26*(4), 321–340.

Rand, W., Zellner M., Page, S. E., Riolo, R., Brown, D. G., Fernandez, L. E. (2002). The complex interaction of agents and environments: an example in urban sprawl. In *Proceedings of Agent 2002*. Chicago: Argonne National Laboratory.

Riolo, R. L., Axelrod, R., & Cohen, M. D. (2001). Evolution of cooperation without reciprocity. *Nature, 414*, 441–443.

Sanders, L., Pumain, D., Mathian, H., Guerin-Pace, F., & Bura, S. (1997). SIMPOP: A multiagent system for the study of urbanism. *Environment and Planning B: Planning and Design, 24*, 287–305.

Saura, S., & Millan, M. (2001). Sensitivity of landscape pattern metrics to map spatial extent. *Photogrammetric Engineering and Remote Sensing, 67*(9), 1027–1036.

Semboloni, F. (2000). The growth of an urban cluster into a dynamic self-modifying spatial pattern. *Environment and Planning B: Planning and Design, 27*(4), 549–564.

Semboloni, F., Assfalg, J., Armeni, S., Gianassi, R., & Marsoni, F. (2004). CityDev, an interactive multi-agents urban model on the web. *Computers, Environment and Urban Systems, 28*(1–2), 45–64.

Shi, W., & Pang, M. Y. C. (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.

Stanilov, K. (2009). Capturing urban form patterns and processes: Insights from the field of urban morphology. In *Conference Presentation, S4 European Spatial Analysis Network*. London: UCL.

Stevens, D., & Dragicevic, S. (2007). A GIS-based irregular cellular automata model of land-use change. *Environment and Planning B: Planning and Design, 34*(4), 708–724.

Thorp, J., Guerin, S., Wimberly, F., Rossbach, M., Densmore, O., Agar, M., & Roberts, D. (2006). Agent-based modelling of wildfire evacuation. In D. Sallach, C. M. Macal & M. J. North (Eds.), *Proceedings of the Agent 2006 Conference on Social Agents: Results and Prospects*. Chicago: University of Chicago and Argonne National Laboratory. Available at: http://agent2007.anl.gov/2006procpdf/Agent_2006.pdf

Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*(2), 234–240.

Torrens, P. M., & Benenson, I. (2005). Geographic automata systems. *International Journal of Geographical Information Science, 19*(4), 385–412.

Torrens, P. M., & O'Sullivan, D. (2001). Cellular automata and urban simulation: Where do we go from here? *Environment and Planning B: Planning and Design, 28*(2), 163–168.

Vancheri, A., Giordano, P., Andrey, D., & Albeverio, S. (2008). Urban growth processes joining cellular automata and multiagent systems. Part 1: Theory and models. *Environment and Planning B: Planning and Design, 35*(4), 723–739.

Verburg, P. H., de Nijs, T. C. M., van Eck, J. R., Visser, H., & de Jong, K. (2004). A method to analyse neighbourhood characteristics of land use patterns. *Computers, Environment and Urban Systems, 28*(6), 667–690.

White, R., Engelen, G., & Uljee, I. (1997). The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. *Environment and Planning B: Planning and Design, 24*(3), 323–343.

Xie, Y., & Batty, M. (2003). *Integrated urban evolutionary modeling* (Working Paper 68). London: CASA.

Yin, L., & Muller, B. (2007). Residential location and the biophysical environment: Exurban development agents in a heterogeneous landscape. *Environment and Planning B: Planning and Design, 34*, 279–295.

Chapter 14
Large Scale Agent-Based Modelling: A Review and Guidelines for Model Scaling

Hazel R. Parry and Mike Bithell

Abstract This chapter provides a review and examples of approaches to model scaling when constructing large agent-based models. A comparison is made between an aggregate 'super-individual' approach, as run on a single processor machine, and two different approaches to parallelisation of agent models run on multi-core hardware. Super-individuals provide a straightforward solution without much alteration of the model formulation and result in large improvements in model efficiency (speed and memory use). However, there are significant challenges to using a super-individual approach when relating super-individuals to individuals in time and space. Parallel computing approaches accept the requirement for large amounts of memory or CPU and attempt to solve the problem by distributing the calculation over many computational units. This requires some modification of the model software and algorithms to distribute the model components across multiple computational cores. This can be achieved in a number of different ways, two of which we illustrate further for the case of spatial models, an 'agent-parallel' and an 'environment-parallel' approach. However, the success of such approaches may also be affected by the complexity of the model (such as multiple agent types and agent interactions), as we illustrate by adding a predator to our example simulation. Between these two parallelisation approaches to the case study, the environment-parallel version of the model, written in C++ instead of Java, proved more efficient and successful at handling parallel processing of complex agent interactions. In conclusion, we use our experiences of creating large agent-based simulations to provide some general guidelines for best practice in agent-based model scaling.

H.R. Parry (✉)
Department of Entomology, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Canberra, Australia
e-mail: Hazel.Parry@csiro.au

M. Bithell
Department of Geography, University of Cambridge, Cambridge, UK
e-mail: mike.bithell@geog.cam.ac.uk

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_14,

14.1 Introduction

In agent-based simulation (ABS), the term 'large scale' refers not just to a simulation that contains many agents, but also refers to the problem of managing the complexity of the simulation (Parry 2009). Another term also used for such simulations is 'Massively Multi-agent Systems (MMAS)' or 'Massive Agent-based Systems (MABS)' (Ishida et al. 2005; Jamali et al. 2008), the term 'Massive' being used in the general computing sense where it implies extremely large numbers (i.e. millions) of agents.

Resource limitations in ABS may be encountered as the modeller adds more agents to investigate whole system behaviour, as the modeller adds complexity to each agent in the form of rules and parameters, or when the modeller wishes to examine the response of an agent in a more realistic and complex environment. Haefner (1992, pp. 156–157) had the foresight nearly 20 years ago to identify aspects of ecological individual-based models that would benefit from advanced computing: multi-species models; models of large numbers of individuals within a population; models with greater realism in the behavioural and physiological mechanisms of movement; and models of individuals with 'additional individual states' (e.g. genetic variation). The introduction of a spatial dimension also adds complexity and puts demands on computing resources, yet many agent-based models (ABMs) are spatial.

In this chapter we focus on spatial ABMs. We compare the aggregate 'super-individual' approach as run on a single processor machine with two different approaches to parallelisation of agent models run on multi-core hardware, using Message-Passing Interface (MPI) libraries to achieve communication between cores. We use a model of insect population dynamics to provide specific examples of each approach. We point out the potential pitfalls that arise from aggregation of individuals in a spatial context and from communication complications that arise when moving from serial to parallel code. The advantages and disadvantages of each approach for speeding up computation and managing memory use will be discussed.

14.2 Review of Large-Scale Modelling Techniques

A number of methodologies have arisen to deal with the problem of 'large scale' simulations in the agent-based literature in a number of disciplines, ranging from molecular physics, social science, telecommunications and ecology, to military research. Some of these methods are given in Table 14.1. This chapter focuses on the last two entries in the table, as the most common types of solution found in the literature: (1) model software restructuring; (2) computer hardware and software programming solutions, including the use of vector computers, Graphics Processing Units (GPUs) and parallel computing.

Table 14.1 Potential solutions to implement when faced with a 'large scale' ABM (Adapted from Parry 2009)

Solution	Pro	Con
Reduce the number of agents, or level of agent complexity, in order for model to run on existing hardware	No reprogramming of model	Assumes dynamics of a smaller or less complex system are sufficiently identical to larger systems, or that there is a simple scaling relationship deducible from the reduced model
Revert to a population-based modelling approach	Could potentially handle any number of individuals	Lose insights from agent approach. Effects of diversity in agent population lost. Emergent properties from simulation of non-linear interactions at agent level difficult to capture. Construction of entirely new model (not agent-based)
Invest in a larger or faster serial machine	No reprogramming of model	High cost. CPU speeds limited to gains of only a few percent (CPU speeds no longer increasing with Moore's law). Most gain likely for large memory problems, but again maximum machine memory is limited. Multi-threading or parallelism would increase the utility of this approach (see last entry in the table)
Run the model on a vector computer	Potentially more efficient as more calculations may be performed in a given time	High cost. Vector hardware not easy to obtain (although Graphics Processing Units (GPU) may compensate this somewhat – see below). This approach works more efficiently with SIMD (see glossary), possibly not so suitable for ABMs with heterogeneous model processes
Super-individuals (model software restructuring)	Relatively simple solution, keeping model formulation similar	Restructuring of model. Aggregation can change dynamics. Potentially inappropriate in a spatial context (Parry and Evans 2008)
Invest in a large scale computer network and reprogram the model in parallel	Makes available high levels of memory and processing power	High cost (although lowering with advent of multi-core and GPU computing). Advanced computing skills required for reprogramming of model software. Algorithms need to be modified to cope with out-of-order execution on different cores. Communication efficiency between cores becomes important. Solutions required are problem dependent

14.3 Model Software Restructuring: 'Super-individuals'

A relatively simple option is to implement an aggregation of the individual agents into 'super-agents', such as the 'super-individual' approach in ecological modelling (Scheffer et al. 1995). Other terms coined for this approach in ecology are the 'Lagrangian Ensemble' method (Woods and Barkmann 1994; Woods 2005) and 'generalised individuals' (Metz and de Roos 1992). A similar approach has been termed 'agent compression' in social science (Wendel and Dibble 2007), which is derived from an earlier ecological paper (Stage et al. 1993). In many ways these approaches are analogous to the concept of 'cohorts', which has been used for a long time in entomological modelling (e.g. Barlow and Dixon 1980; Ramachandramurthi et al. 1997). There are a number of examples of the super-individual method in relation to ABMs in a wide range of literature, with examples in ecology (Schuler 2005; Parry and Evans 2008) and social science (epidemiology) (Dibble et al. 2007; Rao et al. 2009). The basic concept of this approach is shown in Fig. 14.1.

The challenge to using a super-individual approach is relating super-individuals to individuals in time and space (Parry and Evans 2008). Some solutions to managing super-individuals spatially have been proposed, e.g. to maintain a constant number of super-individuals within a spatial unit or cell, so that individuals migrate from one super-individual in one cell to become part of a super-individual in another cell. However, these solutions still affect model behaviour and it comes down to a 'trade-off between error and computing costs' (Hellweger 2008, pp 148). This approach is still likely to have some limitations when behaviour at low densities is important and there is a strong spatial effect on the individuals.

Recent work has proposed a dynamic approach to the creation of super-individuals (Wendel and Dibble 2007). Compression algorithms are applied to homogenous super-individuals to selectively compress their attributes. The algorithm can maintain the integrity of the original data; however, it can be an advantage for the algorithm to combine similar pieces of information to produce a more compact representation. The result is super-individuals that contain varying numbers of similar or identical individuals, from just a single individual to many, depending on the uniqueness of the individuals. The attributes of the individuals contained within the super-individual are monitored over time, so that if individuals differentiate themselves from the group (e.g. they change spatial location, perhaps to another spatial cell), they are extracted from the super-individual and become separate individuals. If the attributes of the uncontained agent now match another super-individual, they may join that super-individual (e.g. they are added to a super-individual at their new spatial location). Although there is some computing overhead for this 'dynamic agent compression', it has been show that it may give some efficiency gain over an individual-based model whilst promising to preserve heterogeneity as necessary (Wendel and Dibble 2007). In general, the fewer unique agents in the simulation the more effective this approach will be.

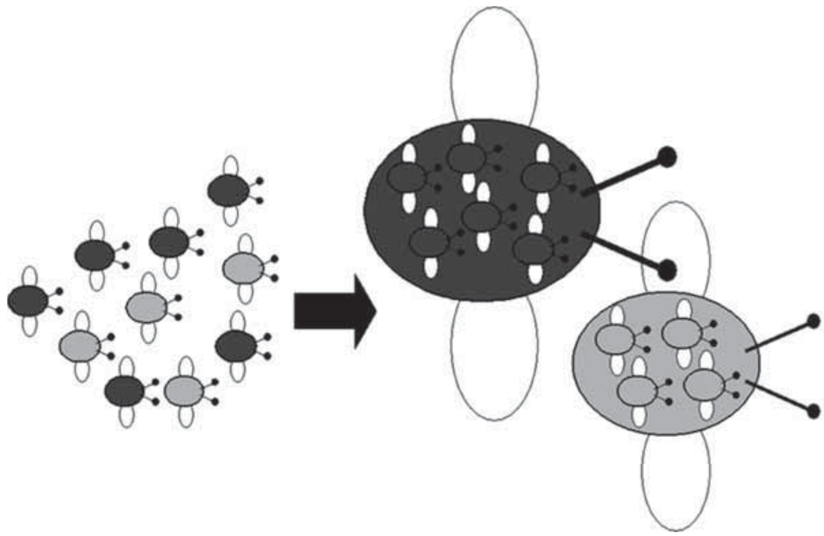


Fig. 14.1 'Super-agents': grouping of individuals into single objects that represent the collective (Taken from Parry and Evans 2008)

14.4 Parallel Computing

Instead of attempting to reduce the computational load by agent-aggregation, parallel approaches accept the requirement for large amounts of memory or CPU and attempt to solve the problem by distributing the calculation over many computational units. One sense in which such distribution can be used is for parameter-space exploration or Monte-Carlo simulations, in which many runs of a small serial (i.e. single-CPU) code may be required. In such cases, efficient use of computer clusters can be made by running identical copies of the code on many separate machines using solutions such as CONDOR (http://www.cs.wisc.edu/condor). While these are in a sense 'large-scale' and make good use of multi-core or distributed computer resources on heterogeneous hardware, here we discuss the use of parallel computing to address the issue of models that require significant resources even for a single model run.

Reprogramming a model in parallel is challenging. Despite this, over the last 10 years or so it has become a popular solution for agent-based modellers in many different fields of research. These range from ecology (Lorek and Sonnenschein 1995; Abbott et al. 1997; Wang et al. 2004, 2005, 2006a, b; Immanuel et al. 2005; Parry et al. 2006a) and biology (Castiglione et al. 1997; Da-Jun et al. 2004) to social and economic science (Massaioli et al. 2005; Takeuchi 2005) and computer science (Popov et al. 2003), including artificial intelligence and robotics (Bokma et al. 1994; Bouzid et al. 2001). In the early 1990s, work in the field of molecular-dynamics (MD) simulations proved parallel platforms to be highly successful in enabling large-scale MD simulation of up to 131 million particles – equivalent to very simple 'agents' (Lomdahl et al. 1993). Today the same code has been tested and used to simulate up to 320 billion atoms on the BlueGene/L architecture containing 131,072 IBM PowerPC440 processors (Kadau et al. 2006). ABS in ecology and social science tend to comprise more complex agents. Therefore, distributed execution resources and timelines must be managed, full encapsulation of agents must be enforced, and tight control over message-based multi-agent interactions is necessary (Gasser et al. 2005). ABMs can vary in complexity, but most tend to be complex, especially in the key model elements of spatial structure and agent heterogeneity.

14.4.1 Multi-core Architectures

'Parallel computing' encompasses a wide range of computer architectures, where the common factor is that the system consists of a number of interconnected 'cores' (processing units), which may perform simultaneous calculations on different data (Wilkinson and Allen 2004). These calculations may be the same or different, depending upon whether a 'Single Instruction Multiple Data' (SIMD) or 'Multiple Instruction Multiple data' (MIMD) approach is implemented (see glossary). Large-scale shared-memory vector processing machines operating via SIMD are now something of a rarity (although individual processors will usually use such methods internally). On the other hand, desktop machines now typically have multi-core processors (with each core essentially acting as a separate CPU), and large-scale high performance computer (HPC) clusters built from such machines with fast low-latency network inter-connects allow the same code to be tested on a desktop and then deployed to a larger system with little or no modification. As there is no longer a trend toward increasing individual CPU speeds, increases in computing power are mostly coming from higher numbers of cores per chip, so that building parallel applications will be a necessary part of exploiting hardware improvements. By designing models that exploit local desktop parallelism and scale to HPC machines, one can not only benefit from desktop speed improvements but also thoroughly test parallelization before making larger runs on more expensive systems. In practice MPI-based applications fulfil this role well, but alternative architectures are beginning to compete with this approach.

14.4.2 Graphics Processing Units (GPUs)

Recent advances in the power of Graphics Processing Units (GPU) now make it easier for modellers to take advantage of data-parallel computer architectures on desktop machines (Lysenko and D'Souza 2008). Multi-core graphics cards can be used not just for display purposes, but also for more general numerical computing tasks (sometimes referred to as GPGPU (General Purpose GPU)). The need for high levels of inter-agent communication and agent movement can make it difficult for cluster-based parallel computing to be efficient, an issue that may be addressed by tighter communication within a GPU as these devices have been designed with very high memory bandwidth (although this comes at the cost of higher memory latency).

Essentially GPUs are similar to vector computers (see glossary). The structure of agent simulations (often with asynchronous updating and heterogeneous data types) could mean that running a simulation on a vector computer may make little difference to the simulation performance. This is because an ABM typically has few elements that could take advantage of SIMD: rarely the same value will be added (or subtracted) to a large number of data points (Nichols et al. 2008). In particular, vector processors are less successful when a program does not have a regular structure, and they do not scale to arbitrarily large problems (the upper limit on the speed of a vector program will be some multiple of the speed of the CPU (Pacheco 1997)). GPUs offer some advantage over vector processors – their operation is single process multiple data (SPMD) rather than SIMD, so that all processing units need not be executing that same instruction as in a SIMD system (Kirk and Hwu 2010). Although it is difficult to keep the advantages of object-oriented code in a GPU environment, there can be considerable benefits in terms of speed.

The architecture of GPUs is rather different from traditional cluster systems. Groups of stream processors are arranged with their own local shared memory, plus access to global memory that resides on the GPU. To make use of this, data must be copied from the CPU-accessible memory into the graphics card. Then the data can be processed by invoking one of a number of 'Kernel functions' that run on the GPU. Lysenko and D'Souza (2008) reformulated two ABMs (Sugar-scape and Stupid Model) to operate on a GPU by the use of large, multi-dimensional arrays to contain the complete state of an agent. Kernels were programmed to run update functions on these arrays. A different kernel was created for each update function, which operated one at a time on the dataset. Some careful coding was required when handling mobile agents (see below), but good performance was obtained for models with a few millions of agents on a domain of up to 2,048 × 2,048 cells. However, their approach required explicit use of the graphics card's texture maps and pixel colour values – such technical details make it awkward for the general programmer to easily access and exploit hardware of this type. Since that time, further developments have made it more straightforward to use GPUs for general computation with the advent of better hardware and libraries designed for the purpose such as NVIDIA's CUDA (http://developer.nvidia.com/object/cuda.html). These libraries relieve the programmer of some of the previous awkwardness involved in converting code for use on a GPU, although awareness of the hardware layout is still required in order to get good performance. Other similar libraries such as Apple's openCL (Khronos 2010), Intel Ct and Microsoft Direct Compute also exist, but as of the time of writing, seem to be in a less advanced state of development. These latter libraries also seek to incorporate some level of hardware independence and are therefore likely to be somewhat more involved to code with than CUDA (Kirk and Hwu 2010). Object-oriented Molecular Dynamics (MD) code already exists that can exploit the CUDA library (Stone et al. 2007), so that the prospect for making individual-based or agent-based code that exploits these libraries in the future would seem to be good. Typically for MD codes, a 240 core GPU seems to be able to deliver similar performance to a 32 core CPU cluster (see for example http://codeblue.umich.edu/hoomd-blue/benchmarks.html). Simulations of cell-level biological systems using FLAME (Richmond et al. 2009a, b), a finite-state machine agent architecture designed specifically to exploit parallel hardware, seem to bear out the potential for simulation speedup that a GPU can offer. However, problems with very large memory requirements may still be challenging for these architectures (as of the time of writing the largest GPU memories are of order 4 GB). The solution in such cases is likely to require running on multiple GPUs, possibly distributed over many independent nodes, with the result that the message passing techniques described below will still be needed.

14.4.3 Challenges of Parallel Computing

Several key challenges arise when implementing an ABM in parallel, which may affect the increase in performance achieved. These include load balancing between cores, synchronising events to ensure causality, monitoring of the distributed simulation state, managing communication between nodes and dynamic resource allocation (Timm and Pawlaszczyk 2005). Good load balancing and inter-node communication with event synchronisation are central to the development of an efficient parallel simulation, a full discussion of which is in Parry (2009). Notable examples of load balancing strategies can be found in Pacheco (1997), including 'block mapping' and 'cyclic mapping' (see glossary).

A further major hurdle is that many (perhaps most) ABMs are constructed with the aid of agent toolkits such as RePast or NetLogo. These toolkits may not be able to handle this conversion to another program representation (particularly an issue for GPU). Recently, Minson and Theodoropoulos (2008) have used a High Level Architecture (HLA) to distribute the RePast Toolkit for a small number of highly computationally intensive agents over up to 32 cores with significant improvements in performance. Rao et al. (2009) express reservations about the general availability of such HLAs, however. In the examples that follow, we show an instance of RePast parallelised using a library (MPIJava[1]) that adds external Message Passing Interface (MPI)[2] calls to Java, but use of this library required extensive restructuring of the original model code, as it was originally designed for serial execution. Since this work was carried out, a facility for making MPI-parallel models using C++ has been added to RePast. Conversion of existing Java code to C++ is usually fairly straightforward, (we will use an alternative C++ library later in this chapter) but the algorithmic considerations regarding the changes needed to ensure correct functioning of parallel code discussed below are still relevant.

[1]Message Passing Interface for Java (MPIJava) http://www.hpjava.org/mpiJava.html is no longer available for download online. It has been super-ceded by MPJ-Express http://mpj-express.org/

[2]See glossary for definition of MPI

14.4.4 Approaches to Agent Parallelism

Parallel agent modelling requires that agent computation is distributed in a way that allows model updates to be carried out on many computational cores simultaneously. This can be achieved in a number of different ways, two of which we will illustrate further for the case of spatial models. In both cases the idea is to send the whole data-structure involved with each agent out to processor cores for updating. In the 'agent parallel' approach, this is done without reference to any spatial structure, but is needed for carrying out update tasks. The 'environment parallel' approach instead divides up the spatial domain between cores and carries the agents associated with each spatial unit along with the spatial sub-division.

14.4.4.1 The 'Agent-Parallel' Approach

This approach focuses on the agents and divides them between the cores, which keep track of the individual agents' properties and spatial location. Thus, each core must keep up-to-date information on the complete environment and surrounding agents. Communication with other cores is necessary to update the actual agent densities for a given location as a result of movement, birth and death. This form of parallelisation is similar to 'functional decomposition' (Foster 1995), which divides various model processes or calculations, though not necessarily agents, between cores. The advantage is that load balancing is more straightforward, as cores can be loaded with agents symmetrically so that each core bears as nearly as possible an equal share of the computation. However, since the spatial data are not included in this process, an extra overhead is implied in ensuring that spatially localized agent interactions are dealt with consistently, as co-location on a core does not guarantee co-location in space.

Examples from ecology:

- Aphids and hoverflies (Parry and Evans 2008), the example used in this chapter.
- Schools of fish (Lorek and Sonnenschein 1995) – includes an extension where fish are dynamically redistributed according to their neighbourhood to improve efficiency.
- Trees (one processor per tree) (Host et al. 2008).
- Landscape vegetation model (functional decomposition) (Cornwell et al. 2001).
- Daphnia, distributing individuals between processors as cohorts or ecotypes, similar to super-individuals (Ramachandramurthi et al. 1997; Nichols et al. 2008).

Examples from social science:

- Financial markets (Massaioli et al. 2005).
- Crowd simulation (Lozano et al. 2007).

14.4.4.2 The 'Environment-Parallel' Approach

This approach divides the geographical space between cores. The parallelisation focuses on a point in space (e.g. a grid cell), which is assigned to each core. The core then keeps track of all agent activity within that space. This has also been termed 'geometric' or 'domain' decomposition (Foster 1995). Local spatial interactions between agents are now likely also to be local to a single core, with potentially easier co-ordination of agent updates. However, when the agents are highly mobile, or when the density of agents varies spatially over time, balancing the load between cores becomes more of an issue, as the allocation of agents to cores must be re-calculated at intervals that depend upon the model dynamics.

Examples from ecology:

- Parallel individual-based modeling of everglades deer ecology (Abbott et al. 1997).
- Design and implementation of a parallel fish model for South Florida (Wang et al. 2004).
- Fire simulation (Wu et al. 1996).
- Forest modelling (Chave 1999).

Examples from social science:

- Parallel implementation of the TRANSIMS micro-simulation model (Nagel and Rickert 2001).
- Abstract agent model 'StupidModel' (Lysenko and D'Souza 2008).
- Traffic simulation (Dupuis and Chopard 2001).
- Disaster Mitigation (Takeuchi 2005).

14.5 Model Software Restructuring Example: Spatial Super-Individuals

This example uses a spatially-explicit individual-based aphid model detailed in (Parry 2006; Parry et al. 2006b); see also Sect. 14.6.1. Turning the individuals in this simulation into 'super-individuals' involved only a small alteration of the model's structure; for details see Parry and Evans (2008). A variable was added to record the number of individuals that all super-individuals actually represent. Equations that were dependent on density (such as morphology determination) were altered so that the density values were related to the real number of individuals in the simulation, not the number of super-individuals.

Movement of super-individuals followed the same rules as that of individuals; however, this produced spatial clustering of the populations. The model was tested by Parry and Evans (2008) using varying populations of individuals (100, 1,000, 10,000 and 100,000 and 500,000 individuals) represented by varying numbers of super-individuals. A brief summary of the findings in this paper follow.

The super-individual model runs on a cellular landscape of 50×50 25m cells, with the initial population of apterous adult aphids initiated at the central cell.

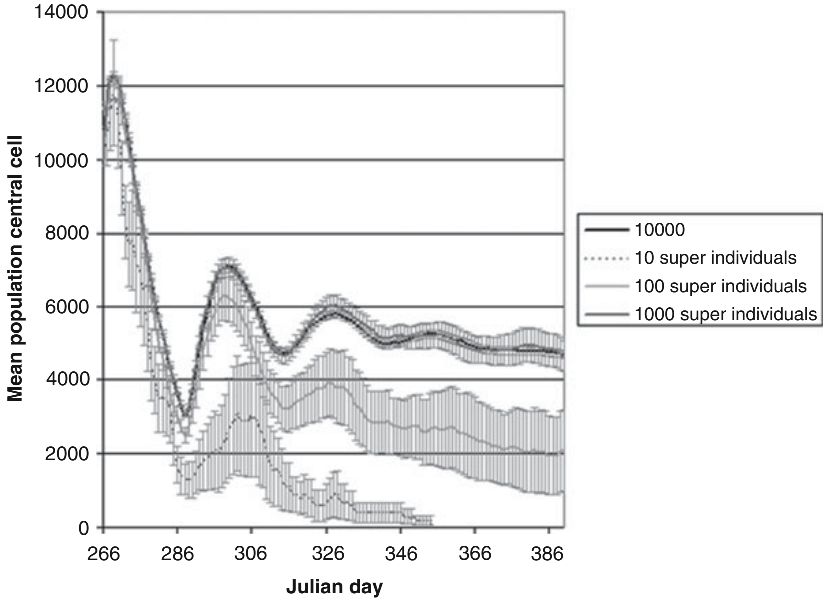


Fig. 14.2 10,000 individuals: comparison between individual-based simulation, 1,000 super-individual simulation (each represents 10 individuals), 100 super-individual simulation (each represents 100 individuals) and 10 super-individual simulation (each represents 1,000 individuals), showing 95% confidence limits derived from the standard error (Taken from Parry and Evans 2008)

14.5.1 Temporal Results

The temporal comparison of super-individuals (representing 10,000 individuals) given in Parry and Evans (2008) is shown in Fig. 14.2. The results for 1,000 super-individuals (scale factor ten individuals per super-individual) are the only results that fall within the 95% confidence limits of the original model for the duration of the simulation period. This is due to excessive discretization of mortality in the model for the super-individuals. Therefore, super-individuals composed of large numbers of individuals as shown here with low scale factors may be the only acceptable way to use this approach, in this case.

14.5.2 Spatial Results

The spatial results given in Parry and Evans (2008) are summarised in Fig. 14.3. Clustering is evident in the spatial distribution. The super-individuals are contained in fewer cells, closer to the origin, than the individual-based simulation for all instances of super-individuals, even those with a low scale factor. Thus, it is an important consideration for spatially-explicit models to test super-individual scaling approaches spatially as well as temporally, as temporal testing will not show the more sensitive spatial errors.

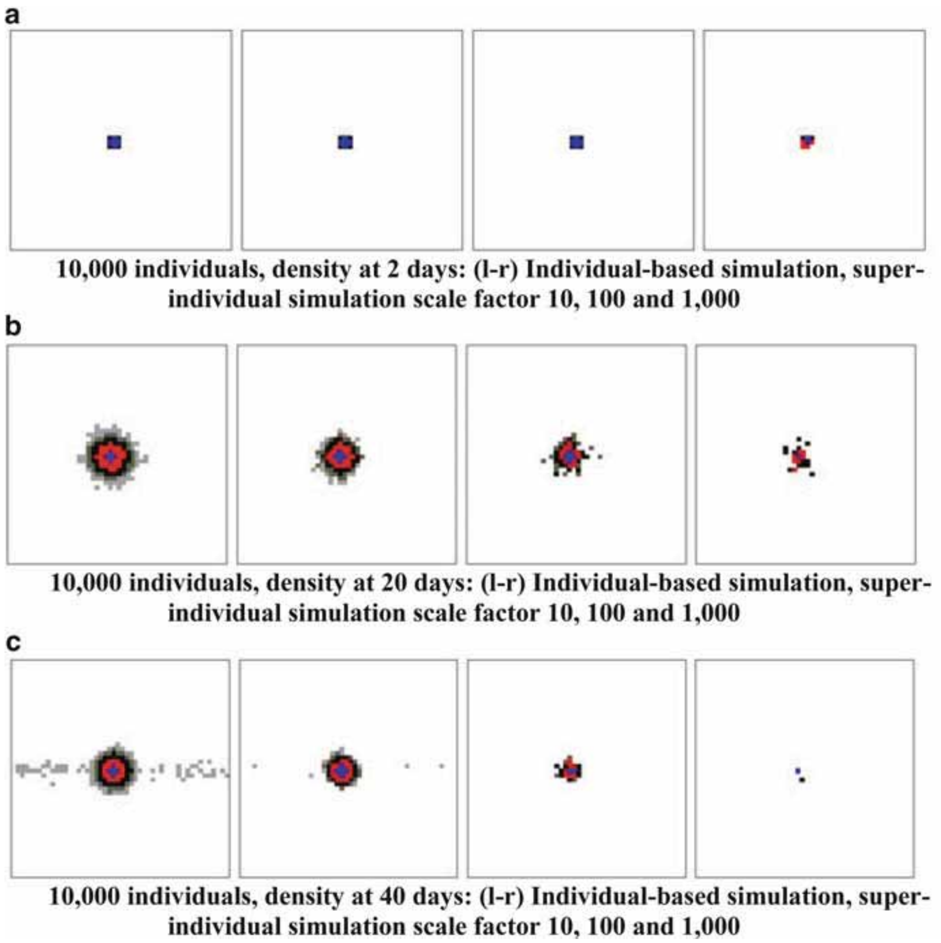


Fig. 14.3 Spatial density distributions for individual-based versus super-individual simulations (10,000 aphids) at (**a**) 2 days (**b**) 20 days and (**c**) 40 days. The distribution further from the central cell is influenced by the constant westerly wind direction to result in a linear movement pattern (Taken from Parry and Evans 2008)

14.6 Parallel Computing Examples: 'Agent-Parallel' and 'Environment-Parallel' Approaches

14.6.1 Example of the Use of an Agent-Parallel Approach

This example uses a spatial predator–prey (hoverfly-aphid) model to show how an agent-parallel model can be established. The model was constructed with the RePast 2.0 agent-based software development toolkit for Java (http://repast.sourceforge.net/). The example illustrates how spatial interactions between predators and prey can lead to difficulties in reproducing the results from serial code.

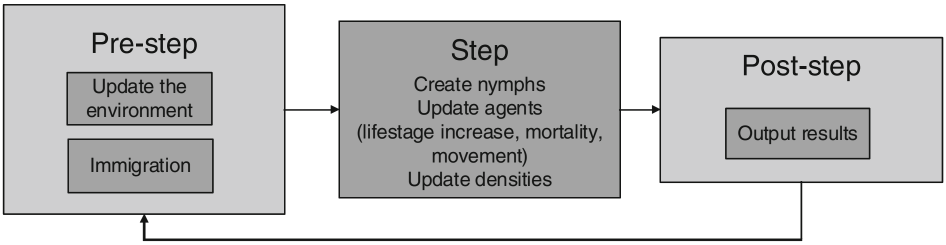


Fig. 14.4 Simplified flow chart for the aphid model

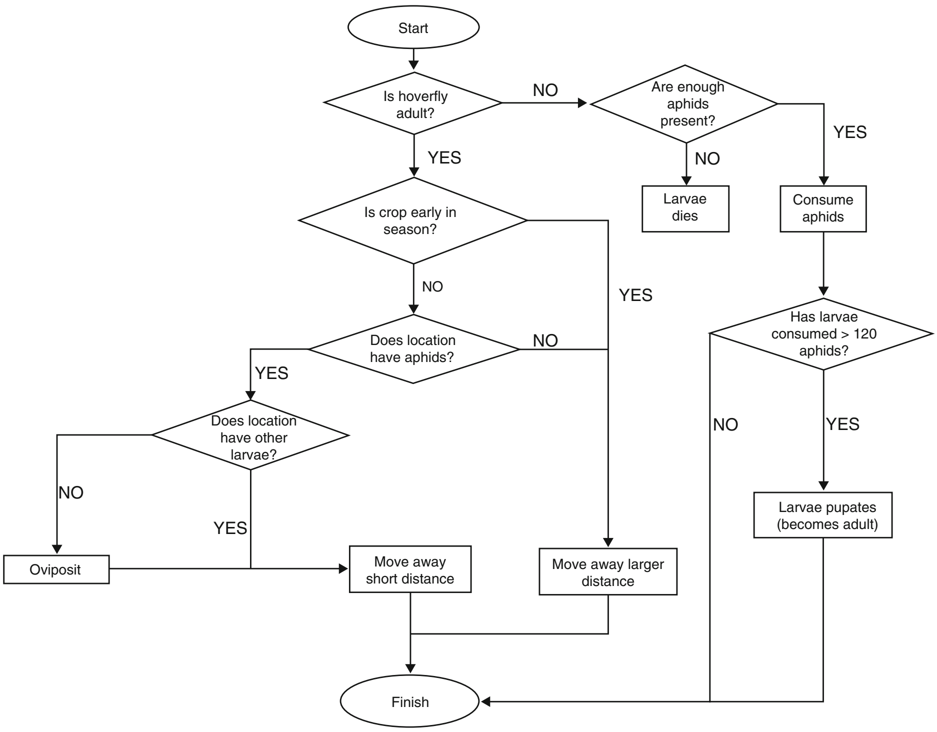


Fig. 14.5 Flowchart of the syrphid model

The basic overall structure of the model system is similar to the structure used by Tenhumberg (2004), which refers to two interacting sub-models for syrphid larvae and aphids. The model describes the population lifecycle of an aphid, *Rhopalosiphum padi*. However, in the individual-based model presented here, the movement of adult female syrphids across the landscape is also modelled. This includes spatial as well as temporal population dynamics within a field. Full details of the aphid sub-model can be found elsewhere (Parry 2006; Parry et al. 2006b), with a highly simplified model flow diagram shown in Fig. 14.4.

The basic rules followed in the syrphid model are given in Fig. 14.5, with more detail on the rules used in the hoverfly model given in the Appendix, as this sub-model is unpublished elsewhere. The two sub-models (aphids and hoverflies) are

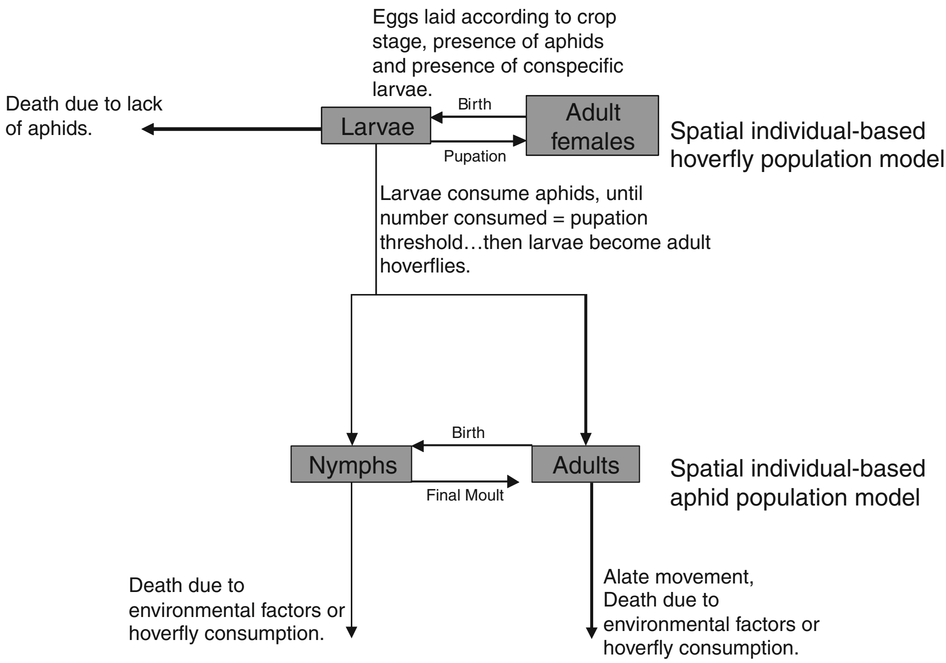


Fig. 14.6 Key processes of the hoverfly-aphid model

connected to one another, by the consumption of aphids by hoverfly larvae. The relationship between the two models is simplified in Fig. 14.6.

The simple model landscape is as shown later in this chapter, two rectangular fields split by a central margin (see Fig. 14.11). The space is divided into a set of square cells, each of area 1 m^2. The model is initiated with one apterous adult aphid in each field cell and one female adult hoverfly in each cell at the field margin.

In order to parallelise the model to distribute the agents to different cores in a cluster, a Message Passing Interface (see glossary) for Java was used http://www.hpjava.org/mpiJava.html (no longer available for download, see footnote 1), run on a Beowulf cluster (see glossary). At each time step, agents are updated on the worker cores (see Fig. 14.7), as the control core maintains global insect density and aphid consumption information and controls the simulation flow.

Testing just the aphid model, simple tests of the parallel code versus the original model (without hoverfly larvae) showed the parallel model to replicate the original serial model accurately.

However, when hoverfly larvae were introduced, the parallel implementation did not replicate the original, non-parallel version. The added complexity of including predators gave rise to two major problems. The most complex element of the model to program was the interaction between the hoverflies and the aphids (i.e. aphid consumption). This involved additional message passing, as the hoverfly might attempt to consume aphids allocated to another processor (although in the same cell geographically). Therefore, consumption for each cell had to be totalled on the control core and then messages passed to each core to instruct the core to remove a

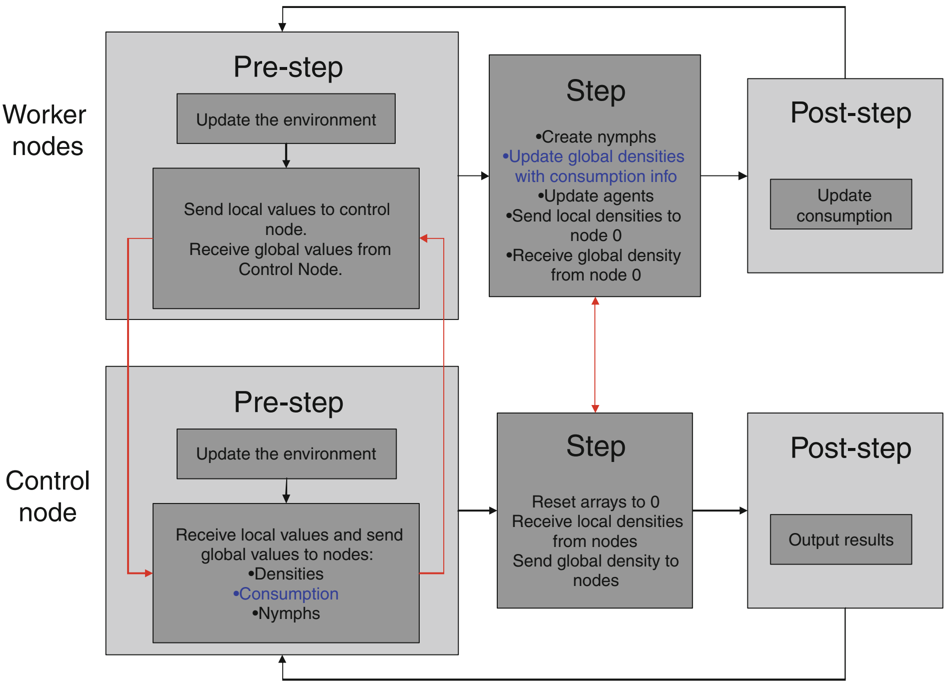


Fig. 14.7 Parallel model flow chart where blue text indicates interaction between the two sub-models and red arrows indicate interaction between the control core and the worker cores

given number of aphids in each cell. However, as these messages are only sent once per iteration, it was possible for more than one hoverfly larvae to consume the same aphid (as the hoverfly larvae would only have information from the previous model iteration on the total aphid densities within the cell, and would be unaware if an aphid had been consumed by another hoverfly larva on another core).

The result was that, occasionally, the total calculated consumption of aphids per iteration per cell was greater than the total density of aphids per cell in that iteration. A simple fix was added to recalculate the total consumption, so that when the total aphid consumption was greater than the total aphid density, the consumption was reduced to the total aphid density. However, the problem still remained, and it gave rise to lower aphid populations in the parallel model than in the non-parallel model, as shown by Fig. 14.8.

In addition, more hoverflies were born into a cell than should be. During the same iteration, different female hoverflies on different processors may perceive a cell to have no larvae present, and then both lay in that cell. However, the model rules state that once larvae are present in a cell, no more larvae should be laid there. The result is likely to be higher numbers of larvae throughout the simulation, as shown in Fig. 14.9. This also acts to reduce the aphid population below that of the non-parallel simulation.

The knock-on effect is that, although higher populations of larvae are present in the non-parallel model due to the artificial reduction in the aphid population and

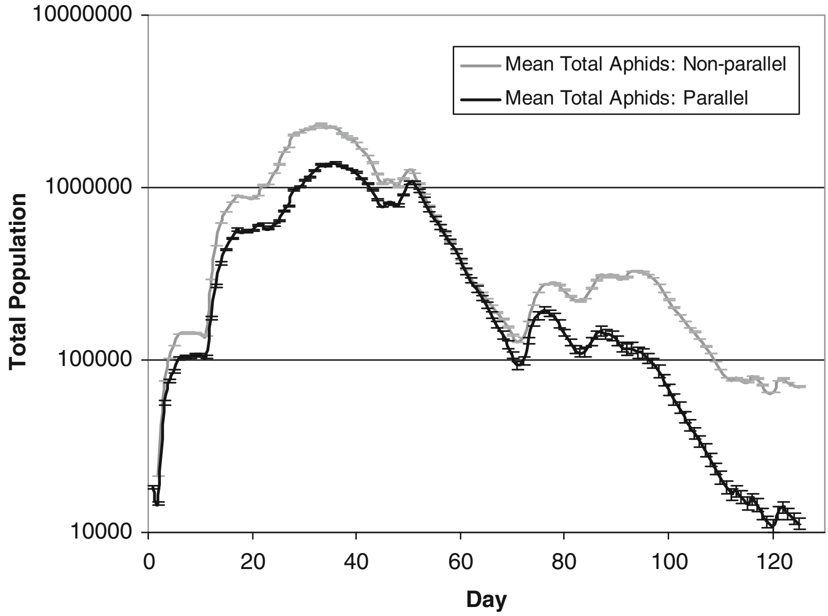


Fig. 14.8 Comparison of the temporal dynamics of the total population of aphids between parallel and non-parallel simulation implementations (error bars show standard error)

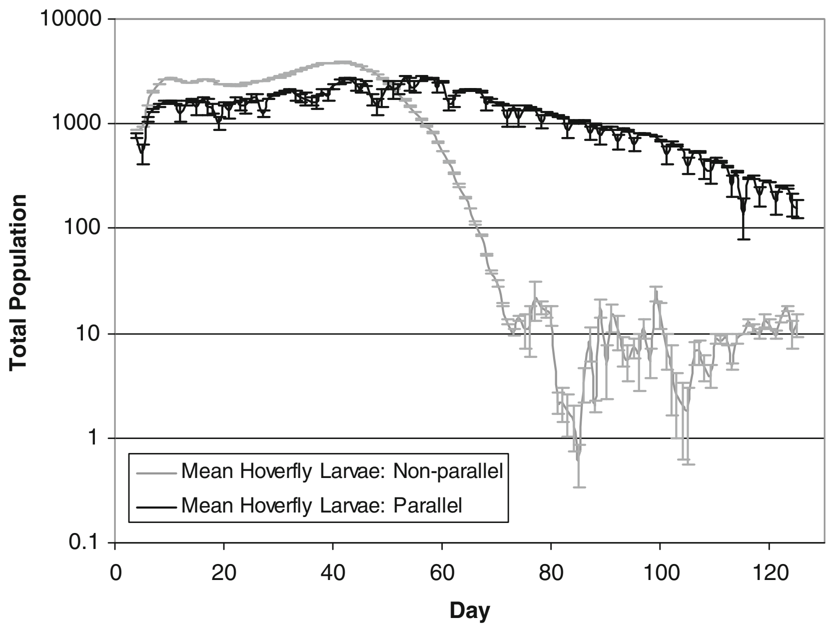


Fig. 14.9 Comparison of the temporal dynamics of the total population of hoverfly larvae between parallel and non-parallel simulation implementations

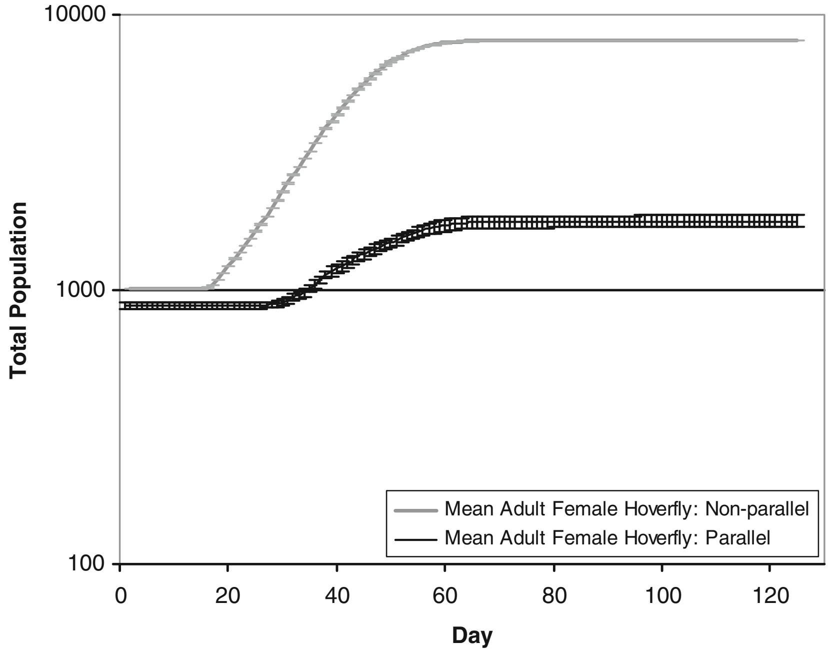


Fig. 14.10 Comparison of the temporal dynamics of the total population of adult female hoverfly between parallel and non-parallel simulation implementations (no mortality)

artificial increase in the larvae population, these larvae are less likely to reach adulthood as there are not enough aphids to consume so that they undergo the transition to adulthood in the model before dying (a combination of higher competition due to the higher larvae density and lower aphid populations due to the higher consumption rate) (Fig. 14.10).

These problems are not experienced in the non-parallel model, as it is straightforward to re-set the number of hoverfly larvae present within a cell during a time-step so that further hoverfly larvae are not introduced mid-iteration, and the consumption of aphids does not conflict as information on the number of aphids present can also be updated easily mid-iteration.

Such programming issues need to be resolved before the agent-parallel model can be used further in scenario development. However, the comparisons provide a valuable insight into the difficulties that may arise when simulating increasingly complex ABMs in parallel. One possible solution may be the use of ‘ghost’ agents, as done by Nichols et al. (2008). However, until tested with this particular model, it is uncertain if this would fully resolve the issues. More generally, this indicates that as the complexity of an ABM increases, it may be more efficient to distribute the model environment (as described in the next section), rather than the agents, so that local agents may interact directly and update parameters within a single model iteration.

14.6.2 Example of the Use of an Environment-Parallel Approach

The environment-parallel approach is essentially a form of domain-decomposition in which spatial units are passed out for processing by remote cores, rather than individual agents. Two challenges are: firstly, to efficiently distribute the environment across cores so as to keep the processor load as even as possible and secondly, how to handle the interaction between, and movement of, the agents.

For the hoverfly-aphid model described here, handling interactions is relatively simple – the landscape (see Fig. 14.11) is divided into a regular cellular grid, which is used to organise the search process by which hoverflies discover their prey. Note that this particle-in-cell approach need not constrain the actual spatial locations of agents, which may still take on values to a much higher level of precision than cell locations (c.f. Bithell and Macmillan (2007)) – the cells can simply act as agent containers. Since the hoverfly larvae are relatively immobile their search process is approximated as involving only the cell that they currently occupy (as opposed to having to search nearby cells – this introduces further complication as noted below). Cells can then be handed off to remote cores, for processing of all parts of the model that do not involve movement beyond cell boundaries (egg-laying by hoverfly adults, predation by larvae, progression of larvae to adult hoverfly, production of young by aphids, calculation of movement by either type of insect) during the first part of the model timestep. Since all cells are independent at this point, this results in a high degree of efficiency in the use of the distributed cores (provided that the cell distribution gives equal numbers of insects per core) whilst also resolving the issues arising in the agent-parallel methodology described above.

For the current simulation, cells are 1 m^2 – this means that typical movement per timestep (1 day) exceeds the cell size (see the Appendix) – insect movement may therefore necessitate transfer of agents from their current core to a remote core upon which their new cell is located. At the end of the above predation timestep, therefore, all the cells are synchronized across cores (to ensure that the same stage of calculation has been reached) and then a communication step is performed to move agents to their correct new locations (see Fig. 14.12). As this communication step is relatively expensive, it reduces the level of speedup achievable somewhat.

In order to implement the above scheme, the model was re-cast into C++, so that advantage could be taken of an existing data-parallel formulation (the graphcode library – Standish and Madina 2008), in which the MPI-parallel part of the code is encapsulated in the formulation of the model grid, along with a utility program (named classdesc) that allows packing and unpacking of arbitrarily structured agents for transfer between cores, making it possible to define the agent dynamics independent of the details of the MPI libraries.

The serial model, when re-coded into C++, produces essentially identical results (barring very small variations introduced by the use of random number generators) to the original Java version. The parallel version of the code in this case shows negligible differences from the serial version. The re-coding of the model into C++ might be expected to have efficiency gains before any parallelisation of the model (as shown for a similar individual-based model of a plant-aphid-disease system by

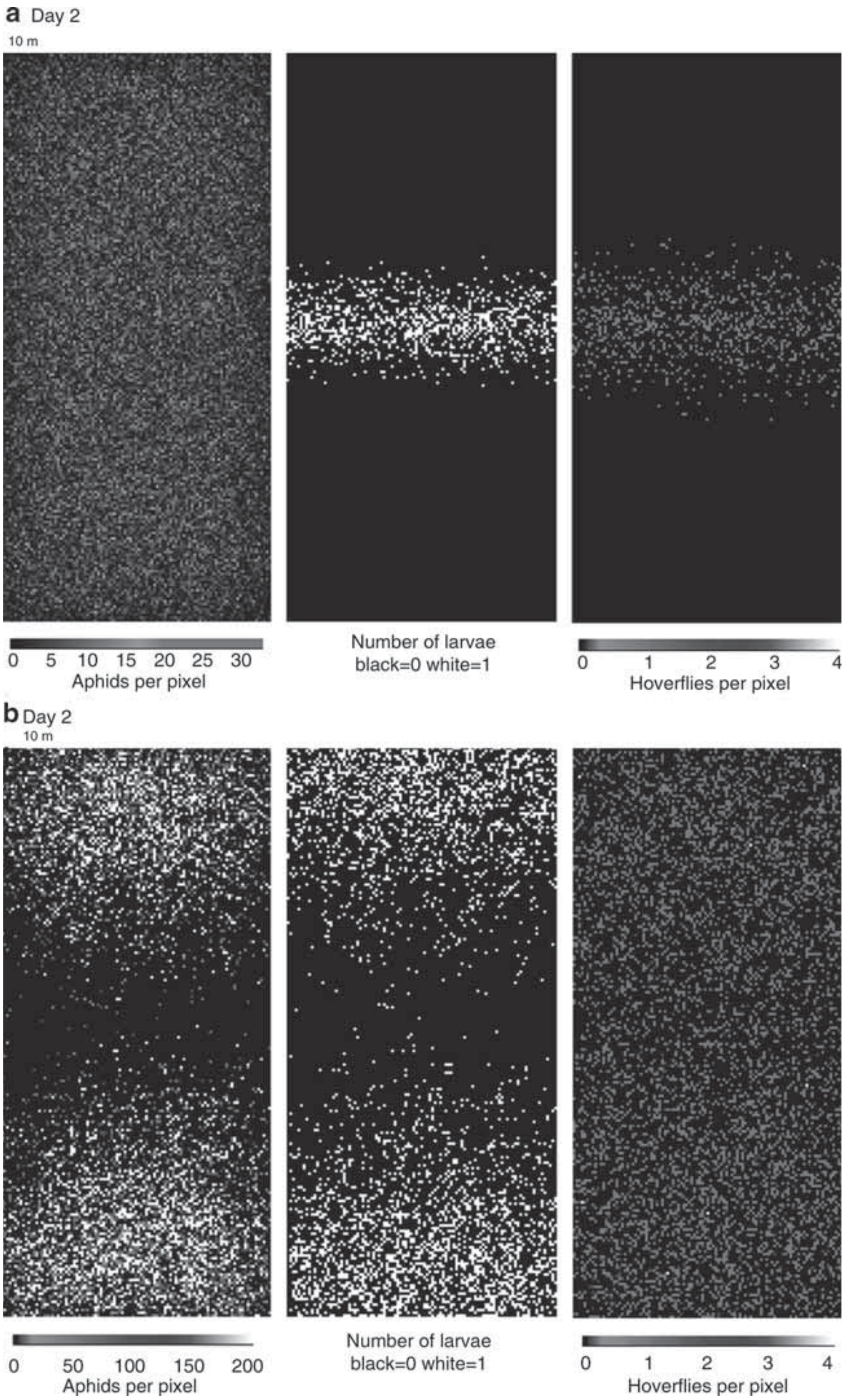


Fig. 14.11 Snapshots of spatial distributions of aphids, hoverfly larvae and hoverfly adults showing spatial distribution over a 100 m × 200 m domain

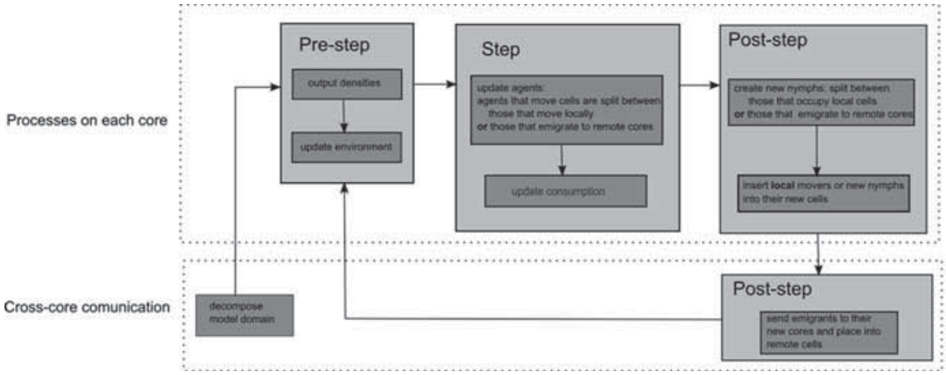


Fig. 14.12 Schematic to show the sequencing of the environment-parallel model. Note that here there is no distinction between workers and control – all cores are treated equally and all run the same set of processes

Barnes and Hopkins (2003)). However, at least for the current implementation, using Java openjdk 1.6.0 and gnu C++ 4.3.2, runtimes of the serial version of the code in the two languages proved to be comparable. The parallel versions of the two implementations are not compared as the Java simulation had significant errors introduced by the parallelisation, as discussed in the preceding sections. An analysis of the speed-up of the Java model (when simulating aphids only) is given later in this chapter, which also draws comparisons with the speed of the super-individual model implementation and the efficiency of the C++ environment-parallel model.

While the environment-parallel version of the model successfully reproduced the results of the serial code, the example presented so far has two simplifications that in practice side-step two of the more awkward issues that need to be addressed in creating parallel agent code – namely (a) domain decomposition is performed only once at the start of the run, where in principle it should be a dynamic process that is adaptive depending on agent density, in order to ensure a balanced load and (b) the interaction between agents takes place only within a single cell, thereby limiting the necessary processes to a single core. We discuss each of these in the following sections.

(a) Balancing loads in the spatially decomposed case
When the density of agents does not vary significantly across the spatial domain (or the density is uniform but the internal computation within each agent is not spatially variable), then the decomposition of the domain can be achieved at the start of the run by allocating equal area blocks of cells to different processors; see e.g. Abbott et al. (1997). However, where there are mobile agents, the density of occupation of the domain need not be uniform either spatially or temporally. Figure 14.11 shows two snapshots from the run of the aphid-hoverfly model – one at day 2 and the other after 45 days. Note that initially the aphids are completely uniformly distributed, but hoverflies and larvae are concentrated near the middle of the domain. However, once significant predation has taken place, aphids are

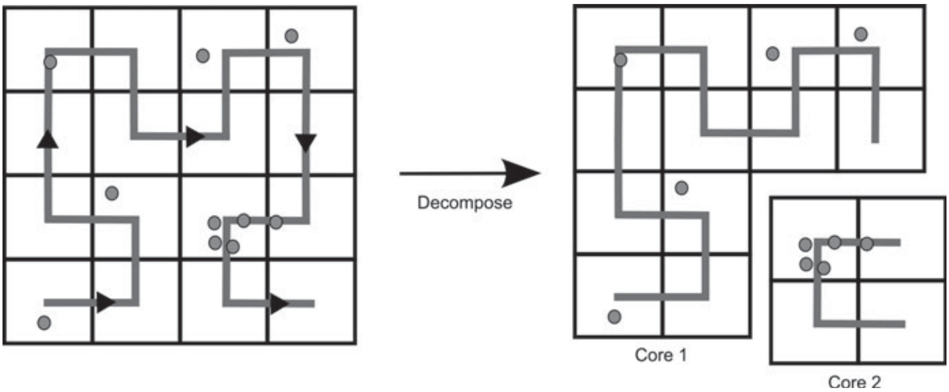


Fig. 14.13 Spatial domain decomposition using a Peano-Hilbert space filling curve. A self-similar path is drawn connecting all the cells in the grid. The path is then traversed (as shown by the arrows), counting up the computational load, and the grid is then segmented along sections of the curve so that equal loads can be distributed to each core (here load is assumed proportional to the number of agents, shown as red dots)

almost entirely excluded from the domain centre, with a similar distribution to the larvae, whereas the hoverfly adults are almost uniformly spread. Since the aphids constitute the bulk of the computational load, a simple block decomposition of the domain with cores being allocated horizontal strips of cells across the domain from top to bottom would lead to cores near the domain centre spending much of their time idle compared to those nearer the upper and lower boundaries.

Since the evolution of the density is not necessarily known from the start of the run, a re-allocation of the cell-to-core mapping should be recomputed automatically as the run proceeds. In practice this is not always a simple thing to do efficiently. Standish and Madina (2008) use the parallel graph partitioning library PARMETIS (http://glaros/dtc/umn.edu/gkhome/metis/parmetis/overview). Other methodologies exist based on space filling curves, e.g. Springel (2005) – see Fig. 14.13. The latter has the advantage of being straightforward to code directly, but unlike PARMETIS, does not explicitly take into account communication overhead, and has the added disadvantage of requiring a domain that can be easily mapped by a self similar structure (e.g. in the example shown, the grid has to have a number of cells in each dimension that is a power of 2), making irregular regions with complex boundaries more difficult to handle.

In addition, any domain re-partitioning implies an overhead in re-building the allocation of cells to processor cores. How often this needs to be done and whether it is worth the time is problem dependent. For example, the C++ version of the example code on a 200×100 m domain runs 124 days on 32 cores in just 7 s. A much larger domain or a larger number of days would likely be required before load-balancing the code would provide a practical benefit.

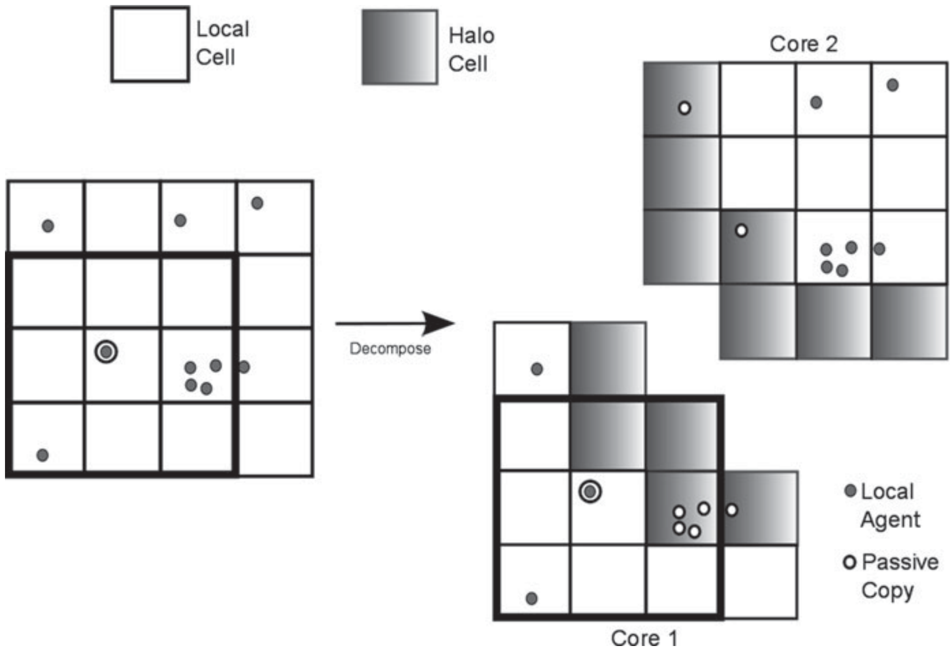


Fig. 14.14 Domain decomposition where agents interact with others outside their own local cell. The circled agent interacts with those in its own cell, but also those in the eight-member neighbourhood outlined by the blue square. On decomposition, part of this neighbourhood lies on a remote core. A halo region is therefore defined around the boundary of each decomposed part of the grid, into which passive copies of the appropriate remote cells can be placed. Locally active agents can then examine these copies in order to make decisions about interaction with the remotely stored agents. In this case, the circled agent can see one active agent on its own core, and 4 passive copies that are active on core 2. Agent copies in the halo cells are updated whenever their corresponding active counterparts on a remote core are changed

(b) Dealing with non-local agent interactions

As mentioned above, we can overcome the problem of predators on different cores accessing the same prey by using the environment-parallel approach when the predators do not look beyond their own local cell. However, once a region of interaction exists that extends across many cells, the problem of co-ordinating agent actions on different cores re-surfaces. Indeed the typical particle-in-cell code uses at least a four or eight cell interaction region about a central cell; see e.g. Bithell and Macmillan (2007). Once the spatial domain is split across cores, such interaction regions also get subdivided. Typically the first level required to deal with this problem is to maintain a 'halo' or 'ghost' region on each core, in which copies of the boundary cells that lie on a neighbouring core, together with *passive* copies of their contained agents, are kept on the local machine (Fig. 14.14).

This allows any independently computable symmetrical or uni-directional interactions to be accounted for immediately (examples would be molecular, smooth particle hydrodynamic or discrete element models, where forces encountered between interacting particles are equal and opposite, or are possibly

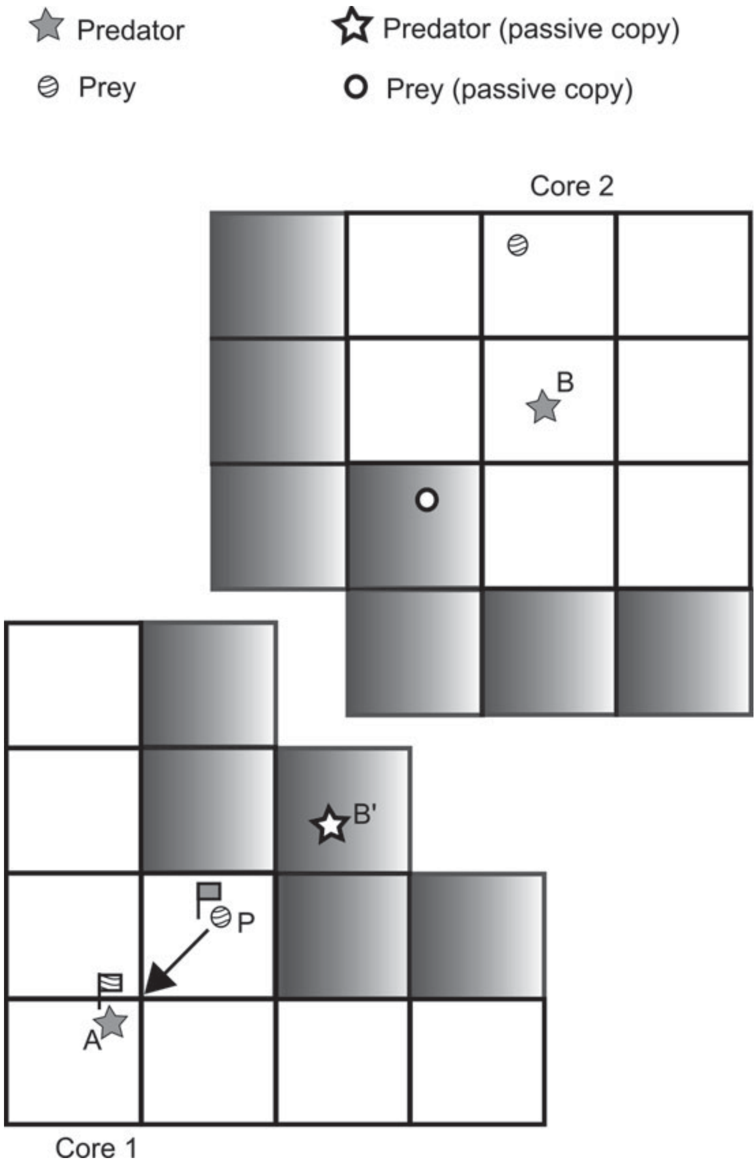


Fig. 14.15 Predator–prey interaction taking place across cores. *Prey P* can see both *predator A* and the passive copy *B'*. *Predator A*, however, only knows about *P*, and not about *B'*. Active *predator B* on core 2 can see two prey, one of which is the passive copy of *P*. Predators and prey need to set and communicate flags to ensure consistency of action (see text). Once flags are consistent (flag on *P* labels it with *A*, flag on *A* labels it with *P*), then prey can be consumed as indicated by the *arrow*

determined by an external field, or disease models where contact with infectives leads to susceptibles acquiring disease, but the nature of the interaction is uni-directional, with no feedback to the infecting agent). Update of the passive agent copies can be performed at the end of each timestep as required. However, for typical ecological or social simulations, this is unlikely to be sufficient. Figure 14.15 illustrates a typical case. Here agent A is a predator that can see only the prey (P) on its own core. Agent B can see a prey on its own core, but also the passive copy of the prey visible to agent A. Suppose both A and B choose to attack prey P. Since the passive copy at first knows nothing of the attack of agent A, potentially A and B could independently attempt to consume the whole of P, leading to over-counting of the available prey. Any solution of this problem must additionally take account of the fact that the order of execution on different cores cannot be guaranteed.

Lysenko and D'Souza (2008) encountered a similar problem in the allocation of single-occupancy spatial cells in their implementation of Stupid Model (Railsback et al. 2005) – they overcame this using a two-pass method in which the agents initially attempted to place a flag in the cell they wish to occupy – a pre-allocated priority allowed agents to compute independently which would succeed – and on a second pass, those agents with highest priority got to occupy the cells of their choice. However, in general, it will not be known a priori which agent should have priority over others, requiring some form of conflict resolution to be performed: in the predator-prey case a competition between predators needs to ensue, and the outcome of this may not be known ahead of time. Mellott et al. (1999) discuss such a case in their implementation of deer predation by panthers, an extension of the earlier work by Abbott et al. (1997). In essence, a further layer of communication is needed in order to ensure consistency between the cores. Looking back at Fig. 14.15, we can envisage a three-pass algorithm in which the initial exchange is for each predator to mark itself with a flag indicating their interest in prey P. This flag is then copied across to the passive copy of the predator (in this case B') on the neighbouring core. Prey P then examines predators that are within range and runs a conflict resolution process (which may involve a more or less elaborate chase sequence involving A and B') to resolve the winner of A and B', setting a flag on itself with the identity of the winner. This flag can then also be copied across cores, and the predators can compare the flag on P with their own identity in order to find the outcome. Clearly this kind of algorithm may need to be extended in the case of more complex predator strategies (hunting as groups, for example) or more complex cognitive agents able to take account of a more extensive view of their surroundings and the available options for attack or escape. Again the result would seem to be that a general algorithm for dealing with this kind of parallel consistency issue is unlikely to be possible – the necessary solution is dictated by the problem at hand.

14.7 Potential Efficiency Gains

This section firstly compares the super-individual model with a parallel implementation of the aphid model only, described in Parry and Evans (2008). The aphid-only model parallelised well using the agent-parallel method as it lacked the complexity of the hoverfly interactions. This shows how parallelisation and super-individuals can both help deal with increasing numbers of agents.

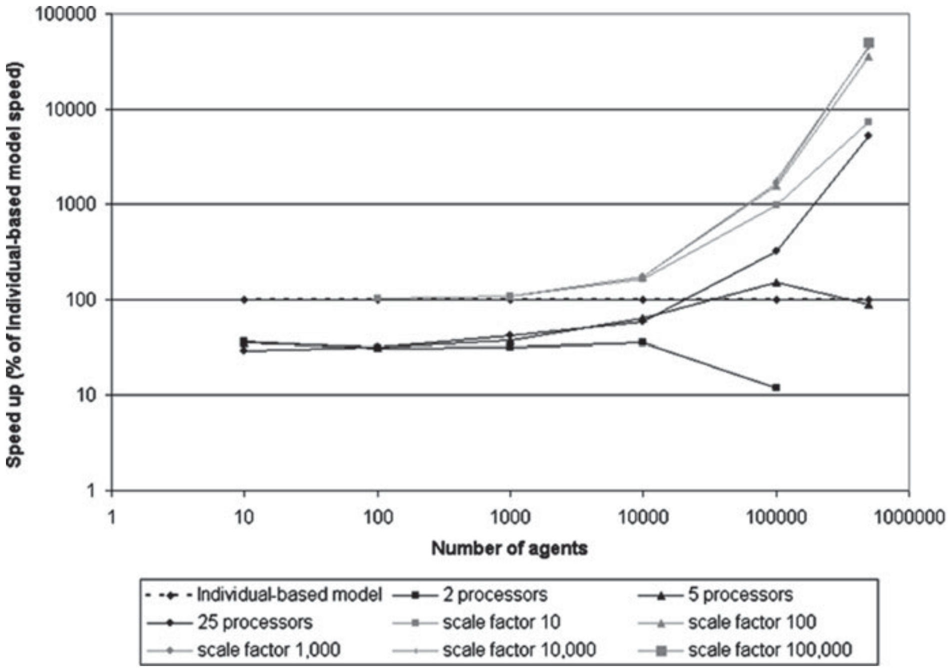


Fig. 14.16 Plot of the percentage speed up from the individual-based (non-parallel) model against number of agents modelled: comparison between super-individuals of scale factor 10, 100, 1,000, 10,000, 100,000 and 500,000

The second part of the section presents the efficiency gains in terms of memory and speed with increasing numbers of processors for the environment-parallel version of the aphid-hoverfly model, to illustrate how efficient this method has been in parallelising this more complex model.

14.7.1 Model Speed and Increasing Numbers of Agents

Super-individuals always improve the model speed with increasing numbers of agents (Fig. 14.16). This improvement is linear (shown here on a log-log scale). The speed improvement is enormous for the largest simulations: 500,000 individuals simulated with super-individuals using a scale factor of 100,000 increases the model speed by over 500 times. However, it was shown above that only large simulations with a low scale factor (10–100) may benefit from the super-individual approach. Thus for these scale factors, an improvement in model speed of approximately 10,000–30,000% (100–300 times) the original speed would result for simulations of 100,000–500,000 individuals.

For the agent-parallel implementation, adding more processors does not necessarily increase the model speed. Figure 14.16 shows that for simulations run on two cores (one control core, one worker core) the simulation takes longer to run in parallel compared to the non-parallel model. Message passing time delay and the modified structure of the code are responsible. As the number of cores used increases, the speed improvement depends on the number of agents simulated. The largest improvement in comparison to the non-parallel model is when more than 500,000 agents are run across 25 cores, where model speed does scale linearly as the number of individuals increases. However, the parallel model is slower than the serial code for fewer than about 30,000 individuals. When only five cores are used, the relationship is more complex: for 100,000 agents, five cores are faster than the non-parallel model, but for 500,000, the non-parallel model is faster. This is perhaps due to the balance between communication time increasing as the number of cores increases versus the decrease in time expected by increasing the number of cores. Overall, these results seem to suggest that when memory is sufficient on a single processor, it is unlikely to be efficient to parallelise the code unless the number of individuals is sufficiently large.

14.7.2 Model Memory Use and Increasing Numbers of Agents

The individual-based model has a linear increase in the memory used as agent numbers increase (shown here on a log-log scale, Fig. 14.17).

Super-individuals always reduce the memory requirements of the simulation (Fig. 14.17). The relationship between the number of (real) individuals in the simulation and the memory used is linear for each scale factor (number of individuals represented by each super-individual). The memory requirement for a simulation of super-individuals has a similar memory requirement to that of an individual-based simulation with the same number of agents as super-individuals. For simulations of 100,000 agents, this can reduce the memory requirement to less than 10% of the memory required for the individual-based simulation with a scale factor of 10,000. For simulations of 500,000 agents, this may be reduced to around 1% with the same scale factor. Thus, when large scale factors are used and as agent numbers increase, there is very little extra demand on memory.

The mean maximum memory usage by each worker core in the agent-parallel simulations is significantly lower than the non-parallel model, for simulations using more than two cores (Fig. 14.17). The relationship between the number of agents in the simulation and the memory used is linear for each number of cores. The two core simulation used more memory on the worker core than the non-parallel model when the simulation had 100,000 agents or above. This is probably due to the memory saved due to the separation of the visualization of the output onto the control core being over-ridden by the slight additional memory requirements introduced by the density calculations. However, when 5 and 25 cores are used, the memory requirements on each core are very much reduced, below that of the super-individual approach in some cases. The super-individual approach uses the least memory for 500,000 individuals, apart from when only a scale factor of 10 is used (after which the 25 core parallel simulation is more memory efficient).

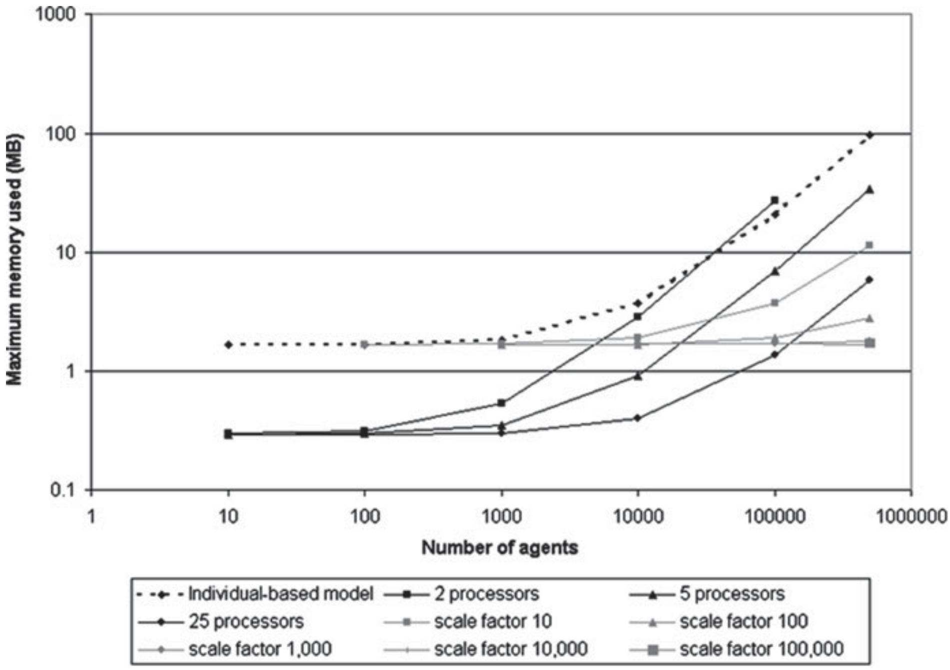


Fig. 14.17 Plot of the mean maximum memory used in a simulation run against number of agents for the model, for different scale factors for super-individuals

14.7.3 Hoverfly-Aphid Model Environment-Parallel Programming Efficiency

The C++ programmed environment-parallel version of the hoverfly-aphid model was run on a dedicated cluster at CSIRO Black Mountain, Canberra. Each node in this network has 28x dual 3.2 GHz Xeon, with 2 or 4 Gbytes per node.

The speed-up of the model approximates a power law for up to 32 cores in comparison to the non-parallel serial model code run on a single processor (Fig. 14.18). At 64 processors the speed-up drops, probably due to the overhead required for each processor to run the model and the time taken for processors to communicate now exceeding the time take for the distributed model to run (at 32 processors the model takes less than 7 s to run) – if tested with a longer or larger (more agents) run of the model, 64 processors would perhaps continue to show increased efficiency as this would remove the effect of this overhead. In terms of memory, the parallel model uses much less memory per processor than the serial implementation, again approximately following a power-law decay up to 32 processors (Fig. 14.19). Overall, of the two parallel approaches, the environment-parallel version of the model, written in C++ instead of Java, proved more efficient and successful at handling parallel processing of complex agent interactions in this case study.

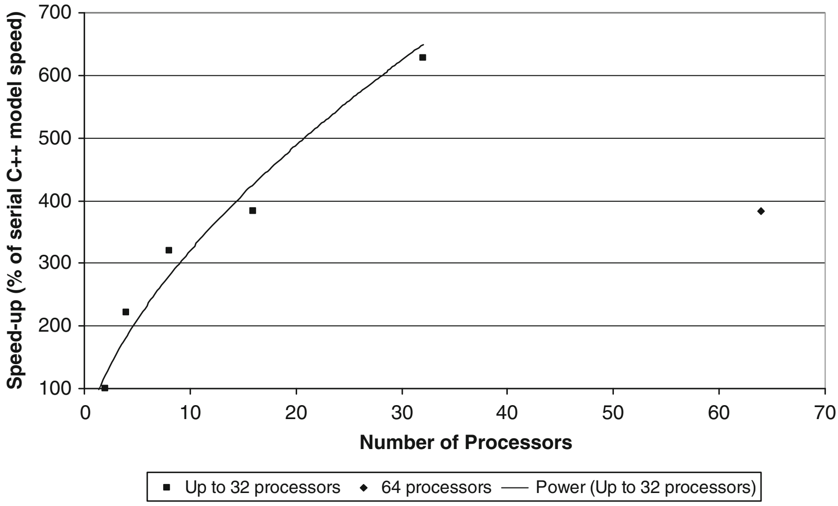


Fig. 14.18 Environment-parallel hoverfly-aphid model: percentage speed-up from the individual-based (non-parallel) model against number of processors. Under 32 processors, this approximates a power law relationship, as shown

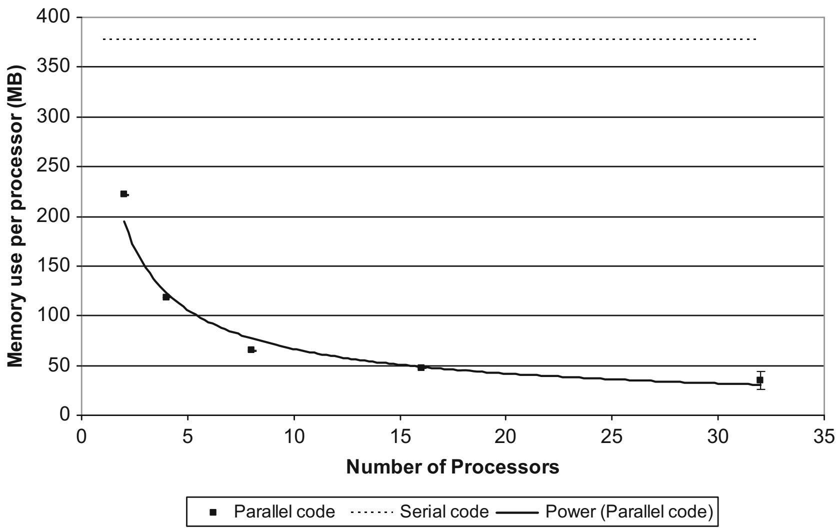


Fig. 14.19 Environment-parallel hoverfly-aphid model: Plot of the mean maximum memory used per processor in a simulation run against number of processors

14.8 Guidelines for Agent-Based Model Scaling

There is no standard method for the development of ABMs, although there are a number of agent modelling toolkits and recently some design protocols have arisen, e.g. Gilbert (2007), Grimm et al. (2006) and Grimm and Railsback (2012). Therefore, as stated in Parry (2009), there is no standard method with which a large scale ABM can be best developed. Instead, Parry (2009) puts forward some key questions to consider at this stage of model development, from Parry (2009, pp 152):

1. What program design do you already have and what is the limitation of this design?
   (a) What is the memory footprint for any existing implementation?
   (b) What are your current run times?
2. What are your scaling requirements?
   (a) How much do you need to scale now?
   (b) How far do you need to scale eventually?
   (c) How soon do you need to do it?
3. How simple is your model and how is it structured?
4. What are your agent complexities?
5. What are your output requirements?

The answers to these questions will help to determine the kind of solution you might seek to the problems of scale. By initially investigating the 'bottlenecks' in your model, you will be able to understand whether it is memory availability or processor speed that is limiting your model. If simple adjustments to your model code are insufficient to solve this, other solutions will then need to be sought. Perhaps a hardware upgrade may be sufficient, but if anything other than moderate scaling is required a more drastic but longer term solution might be necessary.

Question 3 is important to help decide which method may be optimal to scale up the model. Model complexity, agent interaction and spatial model environments will all pose challenges to the use of any method presented here. Some suggestions are made in this chapter as to how best to use some popular solutions when scaling a complex model. However, this cannot be exhaustive and a great deal of experimentation, creativity and development of solutions appropriate to the individual model is likely to be necessary.

Model outputs may also pose limits on the model, in terms of memory for data storage or the way that the output is handled (which may become critical as the model is scaled up). This should be considered when scaling-up an ABM and altering the model structure.

14.8.1 A Protocol

In relation to the key considerations highlighted above, a simple protocol for developing a large scale ABS was defined by Parry (2009, pp 153):

1. Optimise existing code.
2. Clearly identify scaling requirements (both for now and in the future).

3. Consider simple solutions first (e.g. a hardware upgrade).
4. Consider more challenging solutions.
5. Evaluate the suitability of the chosen scaling solution on a simplified version of the model before implementing it on the full model.

The main scaling solution to implement (e.g. from Table 14.1) is defined by the requirements of the model. Implementation of more challenging solutions should be done in stages, where perhaps a simplified version of the model is implemented on a larger scale using some of the techniques described here. Also, as demonstrated here, it is best to initially test the model with numbers lower than perhaps required for realism, to allow for faster run times when testing and experimenting with different approaches. Agent simulation development should originate with a local, flexible 'prototype', and then as the model development progresses and stabilises larger scale implementations can be experimented with (Gasser et al. 2005). For complex solutions, such as parallel computing, a simplified model is often necessary to experiment with large numbers. Improvements to model efficiency are not necessarily linear and optimal solutions tend to be model specific. Thus solutions demonstrated here will work for some ABMs but perhaps not so well for others. A key point, however, is to devise a set of test cases against which the code modifications can be validated at every stage. Although this should be a standard part of any software development programme, it becomes even more vital in developing parallel solutions, where subtle issues to do with timing of agent updates and access to data across cores can lead to difficult debugging problems.

Acknowledgements Much of this work was undertaken whilst the lead author was at the Food and Environment Research Agency, UK and the School of Geography at the University of Leeds, UK, funded by Defra. Many thanks to the Advanced Scientific Computing team at CSIRO for their assistance in running the models on the cluster, particularly Aaron McDonagh and Tim Ho, and also to Dr Andrew Evans at the University of Leeds for assistance in the early stages of this work.

Appendix: Rules for Hoverfly Sub-Model

Development

Development of hoverflies is highly simplified, and birth and death is minimised (see below). The only development that occurs in the model is the transition of larvae to adults. In this, there is a 50% probability that the hoverfly will be female (determined at birth) and male hoverflies are not included in the model from this stage onwards as their activities are assumed not to influence the distribution of larvae and thus the mortality of the aphids.

The transition from larvae to adult is modelled with the assumption that the larvae need to eat a minimum of 120 aphids in total to reach a weight at which they are able to pupate (28 mg) (Ankersmit et al. 1986). Thus, once this number of aphids has been consumed by an individual larva it pupates and becomes an adult (if male, it is then removed from the model).

Reproduction

In this model oviposition occurs once within a single 1 m^2 area (i.e. grid cell) per day. This occurs providing aphids are present and the location has no other larvae. It is assumed only one egg is laid per day within the cell, and the egg is assumed to become a larva the next day. This is probably an underestimate; however, it can easily be modified at a later stage. A suggested estimate may be up to 49 eggs within a 1 m^2 area per day, based upon Harmel et al. (2007), where a high oviposition rate of *E. balteatus* was observed when aphid-infested potato was studied (a mean of 48.9 eggs per laying and per female). This study also found that no eggs were produced by the hoverfly on healthy aphid-free plants.

Mortality

The scenarios shown here do not include adult hoverfly mortality. Experiments with mortality in the model showed that adult mortality has a high impact upon the population dynamics of the syrphids and should be included in further developments of the model.

Mortality of larvae occurs when no aphids are present to feed them (possible if aphids are consumed or are alate and fly away); otherwise there is no mortality of larvae.

Movement and Dispersal

Movement of syrphids and oviposition is key to this model. A number of rules govern the oviposition of larvae by female adult syrphids:

- Search for prey is not random (Kindlmann and Dixon 1993).
- Refrains from ovipositing in the presence of conspecific larvae (Hemptinne et al. 1993).
- Avoids laying eggs close to old aphid colonies, recognized by the presence of winged aphids (Hemptinne et al. 1993).

In this model, rules govern a non-random search for prey, where eggs are only laid where aphid colonies are present and oviposition does not occur where larvae are already present. The model does not include a rule to recognise old aphid colonies at present, but this information is available in the model and could be included at a later stage.

Basic Movement

A model of syrphid predator movement proposed by Kareiva and Odell (1987) is that predators move at constant speed but change direction of movement more often when satiated (area restricted search), and that increase in prey density increases the feeding rate and satiation of the predators (applied to *Uroleucon nigrotuberculatum* and *Coccinella septempunctata*). However, this may have restricted applicability to the early stages of aphid colony development (Kindlmann and Dixon 1993) and it has not been proved that this strategy is optimal (it was arbitrarily chosen).

This model will use a simplified movement rule based upon this principle – the adult female hoverflies move in a random direction, but move a greater distance if no aphids are present or the crop is early in season. It has been shown that crop growth stage and habitat type may influence syrphid movement patterns and oviposition (Powell et al. 2004), providing the foundations for this behavioural rule.

It is assumed that hoverflies move between 4 and 6 m a day (given that a mark-recapture study of Holloway and McCaffery (1990) found hoverflies moved between 20–30 m in a 5 day period). Thus, in the model, 'focused' movement in favourable habitat (margins or late season crop) or around aphid colonies is set between 0 and 4 m, and in unfavourable habitat (early season crop), movement is set at 4–6 m per day.

Foraging Optimisation

It has been suggested that the model of Kareiva and Odell (1987) can be improved by adding terms to describe foraging optimisation (Kindlmann and Dixon 1993). This will enable the model to function at later stages of aphid colony development. The ability of the predator to assess the present and future quality of an aphid colony for their larvae should be included in the model. The effect of more than one aphid colony present in a landscape should also be considered – the presence of other colonies is likely to reduce the optimal number of eggs laid by the predator in a particular aphid colony (Kindlmann and Dixon 1993).

This is applied in the model through a simple behavioural rule: if there are aphids present within a given 1 m^2 location but other larvae are also present, the hoverfly does not oviposit but moves on a short distance.

Parasitation/Predation

A very simple model of aphid consumption was constructed based on the research of Ankersmit et al. (1986):

$$MORT = (0.3119e^{0.0337(A\times 24)} \times D + (2.512e^{0.0253(A\times 24)}) \qquad (14.1)$$

where *MORT* is the predation rate per day; *A* is the age of the Syrphid larvae in days; and *D* is the density of aphids per cm^2 (which is scaled down from 1 m^2 in the model). More recent, complex models exist, e.g. the use of a Holling type-III function by Tenhumberg (1995). However, the nature of the model presented here at this stage does not require this level of complexity.

Glossary

Please note this glossary is largely taken from Parry (2009).

Beowulf cluster A scalable performance computer cluster (distributed system) based on commodity hardware, on a private system network, with open source software (Linux) infrastructure (see http://www.beowulf.org/)

Block Mapping A method of partitioning an array of elements between cores of a distributed system, where the array elements are partitioned as evenly as possible into blocks of consecutive elements and assigned to processors. The size of the blocks approximates to the number of array elements divided by the number of processors.

Central Processing Unit (CPU) May be referred to as a 'core' or 'node' in parallel computing: computer hardware that executes (processes) a sequence of stored instructions (a program).

Cyclic Mapping A method of partitioning an array of elements between cores of a distributed system, where the array elements are partitioned by cycling through each core and assigning individual elements of the array to each core in turn.

Grid Computer 'Grids' are comprised of a large number of disparate computers (often desktop PCs) that are treated as a virtual cluster when linked to one another via a distributed communication infrastructure (such as the internet or an intranet). Grids facilitate sharing of computing, application, data and storage resources. Grid computing crosses geographic and institutional boundaries, lacks central control, and is dynamic as cores are added or removed in an uncoordinated manner. BOINC computing is a form of distributed computing where idle time on CPUs may be used to process information (http://boinc.berkeley.edu/)

Graphics Processing Unit (GPU) Computer hardware designed to efficiently perform computer graphics calculations, particularly for 3-dimensional objects. It operates in a similar manner to a vector computer, but is now widely available as an alternative to the standard CPU found in desktop computers.

Message passing (MP) Message passing (MP) is the principle way by which parallel clusters of machines are programmed. It is a widely-used, powerful and general method of enabling distribution and creating efficient programs (Pacheco 1997). Key advantages of using MP architectures are an ability to scale to many processors, flexibility, 'future-proofing' of programs and portability (Openshaw and Turton 2000).

Message passing interface (MPI) A computing standard that is used for programming parallel systems. It is implemented as a library of code that may be used to enable message passing in a parallel computing system. Such libraries have largely been developed in C and FORTRAN, but are also used with other languages such as Java (MPJ-Express http://mpj-express.org/). It enables developers of parallel software to write parallel programs that are both portable and efficient.

Multiple Instruction Multiple Data (MIMD) Parallelisation where different algorithms are applied to different data items on different processors.

Parallel computer architecture A parallel computer architecture consists of a number of identical units that contain CPUs (Central Processing Units) and function as ordinary serial computers. These units, called cores, are connected to one another. They may transfer information and data between one another (e.g. via MPI) and simultaneously perform calculations on different data.

Single Instruction Multiple Data (SIMD) SIMD techniques exploit data level parallelism: when a large mass of data of a uniform type needs the same instruction performed on it. An example is a vector or array processor and also a GPU. An application that may take advantage of SIMD is one where the same value is being added (or subtracted) to a large number of data points.

Stream Processing Stream Processing is similar to a **SIMD** approach, where a mathematical operation is instructed to run on multiple data elements simultaneously.

Vector Computer/Vector Processor Vector computers contain a CPU designed to run mathematical operations on multiple data elements simultaneously (rather than sequentially). This form of processing is essentially a SIMD approach. The Cray Y-MP and the Convex C3880 are two examples of vector processors used for supercomputing in the 1980s and 1990s. Today, most recent commodity CPU designs include some vector processing instructions.

References

Abbott, C. A., Berry, M. W., Comiskey, E. J., Gross, L. J., & Luh, H.-K. (1997). Parallel individual-based modeling of Everglades deer ecology. *IEEE Computational Science and Engineering, 4*, 60–78.

Ankersmit, G. W., Dijkman, H., Keuning, N. J., Mertens, H., Sins, A., & Tacoma, H. M. (1986). *Episyrphus balteatus* as a predator of the aphid *Sitobion avenae* on winter wheat. *Entomologia Experimentalis et Applicata, 42*, 271–277.

Barlow, N. D., & Dixon, A. F. G. (1980). *Simulation of lime aphid population dynamics*. Wageningen: Centre for Agricultural Publishing and Documentation.

Barnes, D. J., & Hopkins, T. R. (2003). The impact of programming paradigms on the efficiency of an individual-based simulation model. *Simulation Modelling Practice and Theory, 11*, 557–569.

Bithell, M., & Macmillan, W. (2007). Escape from the cell: Spatial modelling with and without grids. *Ecological Modelling, 200*, 59–78.

Bokma, A., Slade, A., Kerridge, S., & Johnson, K. (1994). Engineering large-scale agent-based systems with consensus. *Robotics and Computer-Integrated Manufacturing, 11*, 81–91.

Bouzid, M., Chevrier, V., Vialle, S., & Charpillet, F. (2001). Parallel simulation of a stochastic agent/environment interaction model. *Integrated Computer-Aided Engineering, 8*, 189–203.

Castiglione, F., Bernaschi, M., & Succi, S. (1997). Simulating the immune response on a distributed parallel computer. *International Journal of Modern Physics C, 8*, 527–545.

Chave, J. (1999). Study of structural, successional and spatial patterns in tropical rain forests using TROLL, a spatially explicit forest model. *Ecological Modelling, 124*, 233–254.

Cornwell, C. F., Wille, L. T., Wu, Y. G., & Sklar, F. H. (2001). Parallelization of an ecological landscape model by functional decomposition. *Ecological Modelling, 144*, 13–20.

Da-Jun, T., Tang, F., Lee, T. A., Sarda, D., Krishnan, A., & Goryachev, A. (2004). Parallel computing platform for the agent-based modeling of multicellular biological systems. Parallel and distributed computing: Applications and technologies. *Lecture Notes in Computer Science, 3320*, 5–8.

Dibble, C., Wendel, S., & Carle, K. (2007). Simulating pandemic influenza risks of US cities. In *Proceedings of the 2007 winter simulation conference*, Vols 1–5 (pp. 1527–1529). New York: IEEE Press.

Dupuis, A., & Chopard, B. (2001). Parallel simulation of traffic in Geneva using cellular automata. In E. Kühn (Ed.), *Virtual shared memory for distributed architecture*. Commack: Nova Science Publishers, Inc.

Foster, I. (1995). *Designing and building parallel programs*. Reading: Addison-Wesley.

Gasser, L., Kakugawa, K., Chee, B., & Esteva, M. (2005). Smooth scaling ahead: Progressive MAS simulation from single PCs to Grids. Multi-agent and multi-agent-based simulation. Joint Workshop MABS 2004, 19 July 2004. New York: Springer.

Gilbert, N. (2007). *Agent-based models*. London, UK: Sage.

Grimm, V., & Railsback, S. F. (2012). Designing, formulating and communicating agent-based models. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems*. Dordrecht: Springer. pp. 361–377.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., et al. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling, 198*, 115–126.

Haefner, J. W. (1992). Parallel computers and individual-based models: An overview. In D. L. DeAngelis & L. J. Gross (Eds.), *Individual-based models and approaches in ecology: Populations, communities and ecosystems* (pp. 126–164). New York: Routledge, Chapman and Hall.

Harmel, N., Almohamad, R., Fauconnier, M.-L., Jardin, P. D., Verheggen, F., Marlier, M., et al. (2007). Role of terpenes from aphid-infested potato on searching and oviposition behaviour of *Episyrphus balteatus*. *Insect Science, 14*, 57–63.

Hellweger, F. L. (2008). Spatially explicit individual-based modeling using a fixed super-individual density. *Computers and Geosciences, 34*, 144–152.

Hemptinne, J.-L., Dixon, A. F. G., Doucet, J.-L., & Petersen, J.-E. (1993). Optimal foraging by hoverflies (Diptera, Syrphidae) and ladybirds (Coleoptera: Coccinellidae): Mechanisms. *European Journal of Entomology, 90*, 451–455.

Holloway, G. J., & McCaffery, A. R. (1990). Habitat utilisation and dispersion in *Eristalis pertinax* (Diptera: Syrphidae). *Entomologist, 109*, 116–124.

Host, G. E., Stech, H. W., Lenz, K. E., Roskoski, K., & Mather, R. (2008). Forest patch modeling: Using high performance computing to simulate aboveground interactions among individual trees. *Functional Plant Biology, 35*, 976–987.

Immanuel, A., Berry, M. W., Gross, L. J., Palmer, M., & Wang, D. (2005). A parallel implementation of ALFISH: Simulating hydrological compartmentalization effects on fish dynamics in the Florida Everglades. *Simulation Modelling Practice and Theory, 13*, 55–76.

Ishida, T., Gasser, L., & Nakashima, H. (2005). Massively multi-agent systems I. First international workshop, in MMAS 2004. Heidelberg: Springer-Verlag.

Jamali, N., Scerri, P., & Suguwara, T. (Eds.) (2008). Massively multi-agent technology: AAMAS workshops, MMAS 2006, LSMAS 2006, and CCMMS 2007 Hakodate, May 9, 2006 Honolulu, May 15, 2007, Selected and Revised Papers, LNAI 5043, Heidelberg: Springer-Verlag.

Kadau, K., Germann, T. C., & Lomdahl, P. S. (2006). Molecular dynamics comes of age: 320 billion atom simulation on BlueGene/L. *International Journal of Modern Physics C, 17*, 1755.

Kareiva, P., & Odell, G. (1987). Swarms of predators exhibit “preytaxis” if individual predators use area-restricted search. *The American Naturalist, 130*, 233–270.

Khronos (2010). OpenCL implementations, tutorials and sample code. Beaverton. http://www.khronos.org/developers/resources/opencl

Kindlmann, P., & Dixon, A. F. G. (1993). Optimal foraging in ladybird beetles (Coleoptera: Coccinellidae) and its consequences for their use in biological control. *European Journal of Entomology, 90*, 443–450.

Kirk, D. B., & Hwu, W. W. (2010). *Programming massively parallel processors: A hands-on approach*. Burlington: Morgan-Kaufmann.

Lomdahl, P. S., Beazley, D. M., Tamayo, P., & Gronbechjensen, N. (1993). Multimillion particle molecular-dynamics on the CM-5. *International Journal of Modern Physics C: Physics and Computers, 4*, 1075–1084.

Lorek, H., & Sonnenschein, M. (1995). Using parallel computers to simulate individual-oriented models in ecology: A case study. In *Proceedings, ESM ‘95 European Simulation Multiconference*, Prague, June 1995.

Lozano, M., Morillo, P., Lewis, D., Reiners, D., & Cruz-Neira, C. (2007). A distributed framework for scalable large-scale crowd simulation. In R. Shumaker (Ed.), *Virtual reality, HCII 2007. Lecture Notes in Computer Science*, *4563*, 111–121.

Lysenko, M., & D’Souza, R. M. (2008). A framework for megascale agent-based model simulations on graphics processing units. *Journal of Artificial Societies and Social Simulation*, *11*(4), 10. Available at: http://jasss.soc.surrey.ac.uk/11/4/10.html

Massaioli, F., Castiglione, F., & Bernaschi, M. (2005). OpenMP parallelization of agent-based models. *Parallel Computing, 31*, 1066–1081.

Mellott, L. E., Berry, M. W., Comiskey, E. J., & Gross, L. J. (1999). The design and implementation of an individual-based predator-prey model for a distributed computing environment. *Simulation Practice and Theory, 7*, 47–70.

Metz, J. A. J., & de Roos, A. M. (1992). The role of physiologically structured population models within a general individual based model perspective. In D. L. DeAngelis & L. J. Gross (Eds.), *Individual based models and approaches in ecology: Concepts and models* (pp. 88–111). New York: Routledge, Chapman and Hall.

Minson, R., & Theodoropoulos, G. K. (2008). Distributing RePast agent-based simulations with HLA. *Concurrency and Computation: Practice and Experience, 20*, 1225–1256.

Nagel, K., & Rickert, M. (2001). Parallel implementation of the TRANSIMS micro-simulation. *Parallel Computing, 27*, 1611–1639.

Nichols, J. A., Hallam, T. G., & Dimitrov, D. T. (2008). Parallel simulation of ecological structured communities: Computational needs, hardware capabilities, and nonlinear applications. *Nonlinear Analysis-Theory Methods & Applications, 69*, 832–842.

Openshaw, S., & Turton, I. (2000). *High performance computing and the art of parallel programming: An introduction for geographers, social scientists, and engineers*. London: Routledge.

Pacheco, P. S. (1997). *Parallel programming with MPI*. San Francisco: Morgan Kauffman Publishers.

Parry, H. R. (2006). Effects of land management upon species population dynamics: A spatially explicit, individual-based model (Unpublished PhD thesis, University of Leeds, Leeds).

Parry, H. R. (2009). Agent based modeling, large scale simulations. In R. A. Meyers (Ed.), *Encyclopedia of complexity and systems science* (pp. 148–160). New York: Springer.

Parry, H. R., & Evans, A. J. (2008). A comparative analysis of parallel processing and super-individual methods for improving the computational performance of a large individual-based model. *Ecological Modelling, 214*, 141–152.

Parry, H. R., Evans, A. J., & Heppenstall, A. J. (2006a). Millions of agents: Parallel simulations with the Repast agent-based toolkit. In Trappl, R. (Ed.), Cybernetics and Systems 2006, *Proceedings of the 18th European Meeting on Cybernetics and Systems Research*. Vienna: Austrian Society for Cybernetic Studies.

Parry, H. R., Evans, A. J., & Morgan, D. (2006). Aphid population response to agricultural landscape change: A spatially explicit, individual-based model. *Ecological Modelling, 199*, 451–463.

Popov, K., Vlassov, V., Rafea, M., Holmgren, F., Brand, P., & Haridi, S. (2003). Parallel agent-based simulation on a cluster of workstations. *EURO-PAR 2003 Parallel Processing*, *2790*, 470–480.

Powell, W., A'Hara, S., Harling, R., Holland, J. M., Northing, P., Thomas, C. F. G., & Walters, K. F. A. (2004). 3D Farming: Making biodiversity work for the farmer. Report to Defra LK0915.

Railsback, S. F., Lytinen, S. L., & Grimm, V. (2005). StupidModel and extensions: A template and teaching tool for agent-based modeling platforms. Available at: http://condor.depaul.edu/~slytinen/abm/StupidModelFormulation.pdf

Ramachandramurthi, S., Hallam, T. G., & Nichols, J. A. (1997). Parallel simulation of individual-based, physiologically structured population models. *Mathematical and Computer Modelling, 25*, 55–70.

Rao, D. M., Chernyakhovsky, A., & Rao, V. (2009). Modelling and analysis of global epidemiology of avian influenza. *Environmental Modelling and Software, 24*, 124–134.

Richmond, P., Coakley, S., & Romano, D. (2009a). A high performance agent-based modelling framework on graphics card hardware with CUDA. In K. Decker, J. Sichman, C. Sierra, and C. Castelfranchi (Eds.), *Proceeding of 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, 10–15 May 2009. Budapest

Richmond, P., Coakley, S., & Romano, D. (2009b). Cellular level agent-based modelling on the graphics processing unit. In *International Workshop on High Performance Computational Systems Biology HIBI'09*, 14–16 Oct 2009, Trento.

Scheffer, M., Baveco, J. M., DeAngelis, D. L., Rose, K. A., & van Nes, E. H. (1995). Super-Individuals: A simple solution for modelling large populations on an individual basis. *Ecological Modelling, 80*, 161–170.

Schuler, A. J. (2005). Diversity matters: Dynamic simulation of distributed bacterial states in suspended growth biological wastewater treatment systems. *Biotechnology and Engineering, 91*, 62–74.

Springel, V. (2005). The cosmological simulation code GADGET-2. *Monthly Notices of the Royal Astronomical Society, 364*, 1105–1134.

Stage, A. R., Crookston, N. L., & Monserud, R. A. (1993). An aggregation algorithm for increasing the efficiency of population models. *Ecological Modelling, 68*, 257–271.

Standish, R. K., & Madina, D. (2008). Classdesc and graphcode: Support for scientific programming in C++, arXiv:cs.CE/0610120. Available from http://arxiv.org/abs/cs.CE/0610120

Stone, J. E., Phillips, J. C., Freddolino, P. L., Hardy, D. J., Trabuco, L. G., & Schulten, K. (2007). Accelerating molecular modelling applications with graphics processors. *Journal of Computational Chemistry, 28*, 2618–2640.

Takeuchi, I. (2005). A massively multi-agent simulation system for disaster mitigation. In *Massively Multi-Agent Systems I*: *First International Workshop MMAS 2004*, Kyoto Dec 2004. Heidelberg: Springer-Verlag.

Tenhumberg, B. (1995). Estimating predatory efficiency of *Episyrphus balteatus* (Diptera: Syrphidae) in cereal fields. *Environmental Entomology, 24*, 687–691.

Tenhumberg, B. (2004). Predicting predation efficiency of biocontrol agents: Linking behavior of individuals and population dynamic. In C. Pahl-Wostl, S. Schmidt, T. Jakeman (Eds.), *iEMSs 2004 International Congress: Complexity and Integrated Resources Management*. Osnabrueck: International Environmental Modelling and Software Society.

Timm, I. J., & Pawlaszczyk, D. (2005). Large scale multiagent simulation on the grid. In *Proceedings of the Workshop on Agent-based Grid Economics (AGE 2005) at the IEEE International Symposium on Cluster Computing and the Grid (CCGRID)*. Cardiff: Cardiff University

Wang, D., Gross, L., Carr, E., & Berry, M. (2004). Design and implementation of a parallel fish model for South Florida. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS '04)*, 5–8 Jan 2004, Big Island: IEEE Computer Society

Wang, D., Carr, E., Gross, L. J., & Berry, M. W. (2005). Toward ecosystem modeling on computing grids. *Computing in Science and Engineering, 7*, 44–52.

Wang, D., Berry, M. W., Carr, E. A., & Gross, L. J. (2006). A parallel fish landscape model for ecosystem modeling. *Simulation, 82*, 451–465.

Wang, D., Berry, M. W., & Gross, L. J. (2006). On parallelization of a spatially-explicit structured ecological model for integrated ecosystem simulation. *International Journal of High Performance Computing Applications, 20*, 571–581.

Wendel, S., & Dibble, C. (2007). Dynamic agent compression. *Journal of Artificial Societies and Social Simulation*, *10*(2), 9. Available at: http://jasss.soc.surrey.ac.uk/10/2/9.html

Wilkinson, B., & Allen, M. (2004). *Parallel programming: Techniques and applications using networked workstations and parallel computers* (2nd ed.). New Jersey: Pearson Prentice Hall.

Woods, J. D. (2005). The Lagrangian Ensemble metamodel for simulating plankton ecosystems. *Progress in Oceanography, 67*, 84–159.

Woods, J., & Barkmann, W. (1994). Simulating plankton ecosystems by the Lagrangian ensemble method. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 343*, 27–31.

Wu, Y. G., Sklar, F. H., Gopu, K., & Rutchey, K. (1996). Fire simulations in the Everglades Landscape using parallel programming. *Ecological Modelling, 93*, 113–124.

Chapter 15
Uncertainty and Error

Andrew Evans

Abstract Errors in input data, parameterisation, and model form cause errors and uncertainty in model outputs. This is particularly problematic in non-linear systems where small changes propagate through models to create large output differences. This chapter reviews the issues involved in understanding error, covering a broad range of methodologies and viewpoints from across the spatial modelling sciences.

15.1 Introduction to Error and Its Terminology

There are known knowns. These are things we know that we know. There are known unknowns. That is to say, there are things that we know we don't know. But there are also unknown unknowns. There are things we don't know we don't know.

Donald Rumsfeld: February 12, 2002

The quote above outlines, as best as it can, an important truth in modelling the real world: that the ramifications of ignorance can be tempered by meta-information on the level of that ignorance. Whatever the appropriateness of Donald's statement at the time (on which, see Žižek 2004), the Rumsfeld 'Ladder of Ignorance' nevertheless summarises nicely that it is one thing not to know something, and it is quite another to be able to quantify that ignorance and to summarise it.[1] While there are things we know with perfect accuracy in modelling the real world, in general these are few and far between. It is much more the case that we know that there is some *error* in our understanding, and this leads to *assumptions* in our models and *uncertainty* about our model results that need to be communicated to users of the results. If we are

[1] Rumsfeld largely repeated the terminology of risk assessment in engineering, see, for example, Suter et al. (1987).

A. Evans (✉)
Centre for Applied Spatial Analysis and Policy (ASAP), University of Leeds, Leeds, UK
e-mail: a.j.evans@leeds.ac.uk

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_15,

lucky, we can quantify this error and/or the resultant uncertainty. If we are very unlucky we either can't do this, or we don't know about the error in the first place: we have an unknown unknown, a situation to be avoided at all costs.[2]

Generally in agent-based modelling we have a difficult job, as we tend to deal with concealed and non-linear systems which may be influenced by multiple variables. Some of these variables we may not recognise as important (an error of understanding, that is, an *epistemic error*; see Faber et al. 1992). Indeed, we are often uncertain as to how closely the broader model form which these variables slot into replicates the system in the real world. It is often the case that we have data and need to infer at least part of the model from it. It may be that other models would do the same quality job against the data we have, and a better job against new data: the so-called model *equifinality* problem (Beven and Binley 1992). Even if we have the right form and variables, we may have multiple options for the weights of particular variables within the model (the *inverse problem*). These difficulties have led many to suggest that such models should be regarded as grand "thought experiments", not so much designed to predict accurately as to allow us to reflect on the systems we are studying and our understanding of them (Di Paolo, Bullock and Noble 2000).

Traditionally, however, modellers tend to feign confidence in their model forms and concentrate on error issues associated with another feature of dealing with real-world multivariate systems: that some of the variables we don't want to use will cause *noise* in the real data records of those we do. Noise is essentially variation in our variables of interest around the values we expect to represent their 'true' or 'important' value, but it is difficult to define objectively. At best 'true' in this context means, tautologically, uninfluenced by the variables causing the noise, while 'important' means, equally tautologically, the signal that we need to understand the system given the variables we've chosen to include. Noise produces a *sampling error*, as we hope sampling for our model inputs will generally give us the 'true' or 'important' behaviour of a sampled system, but what we get instead is varied by outside influences, including the mechanics of the measurement process. If we use the data as the foundation of a model prediction, such an error will plainly cause problems.

Noise is frequently treated as an *aleatory* [i.e. random] *error* (which may be regarded as, a type of *ontological* (Walker et al. 2003) or *phenomenological* (Faber et al. 1992) *uncertainty*), added to an underlying signal. The apparently random nature of noise is both problematic and of use. More often than not our definition of something as 'random' is an admission of our ignorance of the influences on a system, or our inability to model them directly and deterministically. However, even though the acknowledgement of noise represents something of an admission of failure, if we know something of the form of the errors involved we can build a description of them into our model. If our model is also a perfect representation of

[2]Here we will largely deal with ignorance from the viewpoint of uncertainty. For more detailed discussions of wider types of ignorance in modelling see: Faber et al. (1992), Walker et al. (2003), and Brown (2004).

the bit of the system we are interested in, this gives us the so-called *Perfect Model Scenario*. As noise-based errors can usually be treated as random, one simple way we can include such errors is by developing *stochastic models*, which include some randomisation of the key variables within strictly controlled ranges or distributions. This is usually achieved through *Monte Carlo testing*: the distribution of each input variable and/or parameter in the model is sampled randomly, but with an emphasis on more probable values appearing proportionally more often (so-called *Monte Carlo sampling*); the model is then run and the process repeated multiple times. Such stochastic models will give a distribution of results if run enough times, and this is often treated probabilistically (for a clear agent-based example centred on generating uncertainty statistics, see Bobashev and Morris 2010). However, for social modellers at least, the top-down analysis of final aggregate results isn't facilitated by the fact that, by-and-large, we lack the very large samples over time other modelling disciplines have and therefore struggle to understand whether the final probabilistic results match the real world well or poorly. This data scarcity sometimes perversely encourages social modellers to abandon randomisation and make one-number 'dart-board' predictions that attempt to hit the few real-world points we have as closely as possible. The alternative to top-down probabilistic assessments of results are bottom-up attempts to delimit the effects of different sources of error as they *propagate* through the system. Unfortunately these are far from simple in non-linear systems. There is a large body of literature on understanding the propagation of error from model inputs, through the model, to outputs/predictions, and for linear/linearisable mathematical models there are well-trodden solutions. However, these solutions usually rely on us being able to characterise the distribution of the noise or other error involved. For social modellers the lack of data highlighted above often makes this problematic. Moreover, many of the techniques assume the distribution is Gaussian/normal. For non-linear systems like ours this need not be the case – indeed, noise may additionally have a changing character (*heteroscedasticity*), and the system may have inputs that vary in importance (i.e. be *non-stationary*) – all of which render many of the traditional methods for dealing with the propagation of errors problematic.

In some cases, then, we may be limited to following some traditional non-linear systems analysis (e.g. Smith et al. 2010b) in bounding worst-case scenarios. This is the position that non-linear uncertainty analysts have endeavoured to move away from for years, not least because identifying and quantifying a "worst" case is usually difficult (Suter et al. 1987). Moreover, it may be that the error propagating through our systems renders even that approach problematic. Non-linearities tend to accentuate small initial data fluctuations (Lorenz 1963) until the small differences between our noisy model input data and the 'true' signal we were hoping for at the start have exploded to cause wild behaviour in our final model results. In such situations, the resultant uncertainty range swamps the range of predictions we suspect the system might have given if presented with the 'true' data. Such errors bloom equally where the *digital precision* with which we can deal with numbers in computer systems fails us and we get initially small changes to our figures through, for example, truncation. If our model is predictive over time, such exploding differences will only increase as we move further from the starting conditions. As such, a nuanced approach to error at different stages of the model process seems critical. However, the peculiarly untraditional architecture of agent-based systems, and their complicated interactions and iterations, do create difficulties in applying the techniques developed for segregating and quantifying errors in more traditional non-linear systems modelling. For example, one standard technique used to more easily understand how errors are propagated is to linearize non-linear models at particular points of equilibrium (for example, with a Taylor expansion), under the assumption that these equilibria are the key modes of operation of the system. The view is taken that the loss of accuracy at these points due to (often low-order) linearization is a worthwhile sacrifice to make to understand the propagation. In agent-based systems, however, the large number of interactions between elements with mixed-method rulesets make any such approach difficult, on top of which most agent-based modellers see non-linear dynamics and a lack of equilibrium in a far more positive light than those working in alternative modelling paradigms.

Along with errors of the types above, uncertainty is also produced by *biases* in our system or inputs: systematic shifts in our model or results away from the 'true' picture. Although traditional modellers make a clear distinction between bias and error, for most models of any complexity the distinction is not always so clear – a missing variable from a model, for example, may be an error of understanding and a systematic bias, but may display as a set of variations that appear to be noise; each problem is related but often handled separately before the overarching issue is appreciated.

This chapter will outline some of the errors and uncertainties associated with modelling the real world, and introduce some of the techniques to deal with such issues. It is worth noting that the chapter only really deals with error from the point of view of uncertainty (the assessment of error in the calibration, verification, sensitivity testing, and validation of models is dealt with more fully in Ngo and See 2012). This chapter is broadly divided into the sources of error and uncertainty, following through the modelling process from inputs to outputs, and ends on an optimistic note with a discussion of why we stand some chance of dealing with this difficulty. It is probably beholden of the author to note that the size and scope of the uncertainty literature is significant beyond the limitations of a single chapter in a book, so this review, by necessity, is more selective in some areas than others.

15.2 Uncertainty Associated with Input Data

Most agent models are based, in some manner, on the real world. Even the most abstract models contain rulesets built on qualitative or quantitative data collection. Real world data can be directly used as an input during formation of a model's structure, the calibration of parameters, or for driving the model. This section looks at the errors that result from the recording process for this data; having insufficient data; missing data within a dataset of otherwise continuous values; and errors that result from the pre-processing of data, such as classification binning. Generally it has to be noted that the complexities of dealing with sources of data mean that we often assume little error in our input data, any prediction error being usually attributed to our parameterisation. This is far from ideal.

15.2.1 Data Measurement and Transcription Errors

Plainly, however data is measured there will generally be errors associated with the process, including transcription errors.

Input error is most successfully quantified for instrumental noise, where the instrument can be checked against multiple readings of the same physical property. In this situation, errors are represented by metrics of *accuracy* (closeness of the sample or a derived statistic to the real value) and *precision* (the tightness or lack of variance of a sample repeated under the same set of conditions). Provided there is no consistent bias in the sample, levels of accuracy will largely be determined by the measurement precision. Standard measures of variance will provide a representation of the error associated with a lack of precision, and, as most instrument errors are Gaussian, the usual figures reported are the standard deviation of the sampling distribution (the *standard error*) and the sample mean, in the form mean ±SD (see Nagele 2001). Such reported figures may be useful in the remaining modelling process provided the error distribution is Gaussian or the figures adapted to a reported alternative distribution. The JCGM/ISO GUM methodology (http://www.bipm.org/en/publications/guides/gum.html) is the standard in this area, and utilises a probabilistic treatment of the belief one might hold in a measurement and standard propagation of error techniques (see Sect. 15.5.1, below).

For spatially located data, particularly data that arrives without a clear error distribution associated with it, more care has to be taken that the data error is not heteroscedastic. That said, information about the spatial field, for example that all points within an area should be the same, or there is positive spatial autocorrelation (i.e. points should be more similar to nearer neighbours), can allow estimates of the distribution of errors to be made. Heuvelink (1998) gives details (see also, for spatio-temporal autocorrelation, Powers 2005). Spatial/locational uncertainties in spatial data are covered by an extensive literature, but Zhang and Goodchild (2002) provide a very comprehensive overview of standard techniques in raster, vector, and spatial object uncertainty modelling. Research issues in semantic uncertainty associated with objective and subjective spatial data are reviewed in Fisher et al. (2004) and Evans and Waters (2007), respectively.

Measurement errors that are non-instrumental, for example errors encouraged by qualitative survey design, are complicated issues of psychology and semiotics. They are one of the most important areas of concern for agent-based modelers wishing to deal with qualitative rulesets and decision making. Good survey design can go a long way – a good starting point on minimizing errors in quantitative judgments is Poulton (1989), while Groves et al. (2009) concentrate on minimizing errors in surveys more generally. In addition, the use of fuzzy sets defined from surveys as model inputs can at least acknowledge and embrace the problem (Zadeh 1975, 1976; Verkuilen 2005; see Evans and Waters 2007, for a spatial example).

If we are lucky, such issues are simple and systematic biases we can recognize and may, infact, be of interest: for example, a bias from mis-understanding the intended levels of a Likert scale survey, or a genuine attempt at fraud. Issues of genuine fraud during data collection might be revealed by comparison with normal (or other) distributions, or through comparison of chosen digits in the data with the Benford distribution (Kiesling undated; Cho and Gaines 2007; Mebane and Kalinin 2009), but more usually they require detailed stakeholder knowledge and trust to be developed during the modelling project to solve them.

Transcription errors should become increasingly rare as more data is collected electronically at source. Most will be treated as noise, unless we are lucky enough to have a consistent bias, though some will be recognisable as outliers. The standard source on recognising and dealing with outliers is Barnett and Lewis (1994). For an updated treatment in multivariant space, see Cerioli and Farcomeni (2011), while López (1997) and Rogers et al. (2009) give good starting points for recognising geographical/spatio-temporal outliers.

15.2.2 Appropriate Sample Size

The inherent complexities of most of the systems agent-based modellers deal with mean than there is a complicated, multivariant, and non-linear relationship between variables of interest and system predictions. To capture the complete set of potential combinations of variables would involve considerable sampling efforts, in systems that are often hard, if not impossible, to sample well. In addition, we have the problem of noise distorting our samples. To understand noise we need repeated measurements of the same quantity/system state, with enough samples taken to define the distribution of the measurements under the influence of the noise sources. Once we have this distribution we may use it probabilistically (see above), or we may try to estimate what the noiseless data would look like. In general, to get as close to the noiseless value as possible, we have to pick a representative statistic to estimate that is as noiseless as possible: for example, if the noise is Gaussian, the arithmetic mean of the population. Where we want continuous data we may smooth out the noise. Keesman and van Straten (1989) summarise some of the opportunities for data smoothing, while Beck (1987) summaries some of the issues. However, it isn't especially clear whether treating data to remove noise is always appropriate. We generally try to minimise the effects of noise on inputs, especially with systems that explode errors non-linearly, as it is usually regarded as a detrimental influence from things we'd like to exclude from our models. However, this needs to be determined on a case-by-case basis; if the real systems suffer from such noise, are we correct to exclude it by, for example, statistical pre-processing? A better approach (Sect. 15.6) may be to build systems that show the same resilience to noise that we see in real world systems.

If we are going to try to remove noise, and we've identified our statistic of interest, we need to sample sufficiently highly that we can estimate that value in the population accurately from the samples. But how do we decide how large a sample is 'sufficient? Traditionally this has come down to trading off expensive increases in sample size against inherent risk that with small samples your value of interest may be unrepresentative. For situations where the value of interest has a well characterised, independent, and constant variation, we can directly calculate the size of sample needed for us to be able to make the estimate of the true value at some level of precision within some pre-defined levels of confidence. For example, with random independent samples, the standard error of the sample mean is the population standard deviation, divided by the square root of the sample size. By adjusting the sample size, we can reduce the error by a known degree. It is usual to trade off sample error and size for a specific confidence, such that if multiple samples were taken the number for which the true value of the statistic would fall within the range of the sample-based estimate ± the sample error would be, say, 95%.

However, this process is not so simple when the data/noise is non-normal and not independent, as it frequently is in non-instrumental noise cases. For basic non-normal distributions, appropriate sample sizes can be estimated for a given confidence using Kolmogorov-Rényi statistics (Spear 1970). However, for time series and spatial data, this process becomes more complicated. Spatial and temporal autocorrelation (where nearby points have related values) can have a significant effect on the apparent sample size of sampled datasets by introducing sample redundancy (Getis 2007). These issues become particularly important when datasets for validating models against are drawn from the same area by sample splitting (Araújo et al. 2005). Significance testing based on autocorrelated data should take autocorrelation into account (though rarely does). A summary of some of the general methods for dealing with spatial autocorrelation can be found in Legendre (1993) and Getis (2007). Kelejian and Prucha (2010) outline something of the size of the problem facing spatial modellers in their discussion of the relationships between sample size, spatial correlation, and missing data, in regression models with spatial lags.

Where non-spatial data suffers from heteroscedasticity or non-independence of noise, it can be treated (see Gallagher and Doherty 2007, for details) which helps with some issues. Spatial heteroscedasticity can further complicate the picture though (see Lauridsen and Kosfeld 2007); for example, positive autocorrelation in errors can falsely reduce error levels (Araújo et al. 2005; Getis 2007). With more complicated non-linear systems, we often have to take a slightly wider viewpoint and concentrate instead on how input variation affects the modelling we are trying to do. When we are trying to model a non-linear system, and the function that we are trying to estimate parameters for is known, it is possible to sample repeatedly to simultaneously build up a picture of the error and the resultant sample size needed. Methodologies can be found in the comparison provided by Malinarič and Ďuríšek (2004). If we know something of the variation in the model error at key points, we can sample these more frequently (O'Neill et al. 1980; Beck 1987). However, with complex systems and limited sampling budgets it is sometimes more practical to use more distribution-free methods, for example 'sampling to redundancy methods', like species area curves, where sample novelty across multiple samples is plotted against sample numbers or size to determine when sampling is sufficient to capture all new elements in a population (Lyman and Ames 2007). While such methods give a poor statistical confidence, they do at least ensure a sample across the potential range of values has been taken. A final issue is that in many of the systems we study the relationships are non-Markovian, that is their future may be influenced by the specific pathway the system has taken historically, rather than just the current instantaneous, autocorrelated, state. This introduces considerable complications into the process of determining appropriate sample sizes.

Adequately sampling the set of potential combinations of variables and predictands is difficult in complex and non-linear systems with non-normal and interdependent variables, and it is doubly so if we also wish to understand the errors in the data. As such, we are generally thrown back on validating models at output, rather than trying to statistically validate the representativeness of the inputs.

15.2.3 Missing Data

Missing data is usually a result of disrupted sampling, or the repurposing and combination of previously collected datasets. For traditional models missing data can often be problematic, especially where data is iteratively re-fed into the model. Artificial intelligence systems based around weight-adjusted learning (like artificial neural networks) and/or case-by-case decision making (like agent-based models) tend to respond better to missing data in both calibration and prediction than traditional models. Nevertheless, sometimes it is necessary to deal directly with missing data for computational or predictive purposes.

Where data is reasonably well distributed it is sometimes possible to interpolate new data into the gaps using a function based on the data we have. For simple datasets with well-known statistical properties, the techniques used for developing the functions (such as linear least-squares regression) have well-known error assessments that utilise all the data brought in to calculate the function. However, for complex non-linear datasets – especially spatial datasets – where the error and function are not easily characterised, it is more common to assess the error by rotational cross-validation (repeated removal of known data points, construction of the interpolation, and comparison of the interpolation against the removed points). This has the strengthening advantage of maintaining some distinction between the datasets used to construct and validate the function. A short but useful review of the relevant literature on missing spatial data can be found in Kelejian and Prucha (2010) and interpolation in Isaaks and Srivastava (1990). When a distribution of new data points is needed, rather than values on a continuous surface, popular techniques revolve around resampling what is already present to generate larger datasets. For example, in bootstrapping, systems are trained on data derived by sampling a distribution multiple times to generate a training set, unselected data giving a validation dataset. Such techniques are common when datasets are too small to use as-is. In addition to generating new data with its own or inherited error, where data limitations are known resampling can be used to constrain errors, particularly where based on Bayesian or error-led assessments (Luoto et al. 2010).

Where data is poorly distributed, missing data can lead to biases. Such biases can be quite subtle, particularly when dealing with spatial autocorrelation. Where a surface is needed, it is sometimes possible to adjust the importance of samples to account for an inappropriate sample distribution. For example, spatially clustered data can be declustered to reduce the importance of over-sampled areas by weighting each value by a function of the distance to its neighbours (for techniques, see Dubois and Saisana 2002). When a distribution is required, resampling can remove some kinds of biases (for example, autocorrelation in sequential, or spatial, samples) if the sampling is carefully randomised (Luoto et al. 2010). However, ultimately biases caused by missing data usually necessitate additional data collection exercises to resolve the problem.

In the absence of good data, models often rely on strongly believed deterministic relationships or qualitative theory, where they might be better off including Bayesian entities in the relationships so that they can be updated as information comes in, and uncertainty can be properly quantified (Young et al. 1996). Bayesian approaches are, of course, only really worthwhile where we know more data may be forthcoming. This is not always the case in the kinds of systems agent-based modellers deal with, at least currently.

15.2.4 Classification Discretisation Error

Almost all data is an aggregation or interpretation of facts about the world. Direct measurements of unique physical properties are very rare, even in such apparently concrete subjects as physics. There will, therefore, always be some loss of information in data recording and use. Even in the event that our instruments are recording at an accuracy/precision we are happy with, we generally add an additional uncertainty, or find one introduced in post-production, through data classification into bins.

Binning data into classifications can be problematic, especially where classification schemes are multivariate and prototypical (that is, very broadly, objects are classed by, and into, examples). Real-world membership of a set is usually fuzzy, and fuzzy sets are generally a more realistic way of dealing with the world. Where crisp sets are needed, entropy statistics can be used to represent multivariate classification uncertainty, and their relative simplicity provides a useful option for spatial mapping (van der Wel et al. 1996). The more common uncertainty, however, usually concerns the granularity of the bins and where the original data point fell within the bin. Data that has already been binned appropriately is not usually problematic where we have control over it (the uncertainty is easily quantified, and can be included in with other assessments of precision). It is only where we wish to use the data for other classification systems or as a continuous dataset, that binning causes issues. For numerical data, there is the possibility of shifting the data back into a continuous sequence by stochastically distributing the data within each class to match an overall distribution fitted to the totality of the classified data. However, once such a distribution has been identified, sampling directly from the distribution becomes simpler.

Redistributing the original sample is only really worthwhile if the classified data is n-tuples, carrying ancillary data with the data that was originally binned. One common use of such a redistribution is within spatial microsimulation (Ballas et al. 2005), in which a population of individuals, which have been lumped together at some geographical scale (say a national bin) are redistributed to smaller areas (say electoral districts bins) such that their distribution matches some statistic (say, employment) in that area. If people can broadly be divided into socio-economic types, with correlated traits, we might expect ancillary traits (say, newspaper readership) to be recreated in the smaller areas (with an error associated with the strength of correlation between the two traits). Such models are increasingly used as the starting conditions for agent-based models where individual-level census data is unavailable, though assessing the accuracy of the recreation of ancillary variables is not easy without detailed new sampling, because we're usually trying to recreate distributions which are essentially unknown. Generally even where we are just trying to recreate the location of individuals with a set of traits which we have constraining distributions for, the geographical location is rarely accurate; commonly individuals are assigned to the smaller geographical boundary set itself or randomly allocated a home within the area. More sophisticated pycnophylactic (Tobler 1979), or other types of reallocation, are rarely completed, meaning there is also a considerable distribution error within each area.

A further major error during classification is caused by conversion between classification schemes, for example the placing of classified and geographically binned census data into new classes and geographical boundaries (Martin et al. 2002). Usually error can only be avoided by aggregating up bins or spatial boundaries to some common aggregate level (for example, Martin 2003).

15.3 Uncertainty Associated with Model Choice

As well as errors and uncertainty associated with data, we recognise that there are also epistemic uncertainties: those associated with our knowledge of the system. Essentially we may regard ourselves as being on a fruitless quest: languages (computer code included) are not the physical real world. Not only does this mean that we are unlikely to ever get a perfect computational representation of the real world (what Faber et al. 1992, call *hermeneutic ignorance*), but it also means we're unlikely to ever understand it properly, as we simply don't have the tools to do so (Faber et al.'s *axiomatic ignorance*), and those we do have may be fatally flawed (Faber et al.'s *logical ignorance*, following Gödel's incompleteness work). Not only this, but we essentially have to limit our modelling attempts in a way that the interconnected real systems are not limited (the *closure problem*: Lane 2001). Nevertheless, as languages, and mathematics in particular, have shown, we can get a useful approximation that carries us forward. This is especially true for real systems that are mediated though language. In this section we examine some of the epistemic uncertainties we will have to deal with to do so. We shall assume a simple model where input variables are utilised via some kind of weighting against each other, or mathematical relationship, or ruleset, and the component *parameters* of these forms control the conversion of the variable inputs to one or more model outputs. The parameters are *calibrated*, that is fixed based on the real world to give as realistic output as possible. The parameterised section of the model may include actions by one or more agents.

15.3.1 Error in Choosing Variables

Simultaneous with finding data for our models is the process of deciding which data we are going to use, and which we are going to exclude. The tendency to load a model with variables is a particular problem with those branches of agent-based modelling where the model is developed to accurately replicate reality. An increased number of variables may lead to a more realistic model, but it also leads to increased levels of error through the need to calibrate more parameters (the so-called *Information Paradox*, O'Neill 1973; Rowe 1977). Moreover, added detail often adds little to a model, and a shift from parsimony can obscure simpler models that perform just as adequately. Generally measures of model quality trade off accuracy of representation against model complexity (see Spiegelhalter et al. 2002, for a discussion of classical and Bayesian methods for achieving this tradeoff). This said, however, the option for adding additional variables is sometimes worth investing effort in early in the modelling process. Flexible code that allows for the addition and removal of variables through a well-structured object hierarchy and generic programming (parameterized types: Gamma et al. 1994) will pay considerable dividends on its investment in the longer term.

While we would hope that the choosing of variables was part of a linear progression, from thinking about the system we are interested in, to deciding how to model it, to picking data, it rarely works out so simply. Investigating our data often suggests we may have to settle for different, less than ideal, proxies for the data we would like to have, or, indeed, different data altogether. It may also be that we are using too small or too large a number of variables to represent the system (O'Neill 1973's *aggregation error*). To some extent having too many variables should reveal itself through covariance, but having too few variables, or the wrong type of variables, will result in errors or biases. In addition, there are often problems of scale: we may misunderstand the boundaries between objects in the real world (Suter et al. 1987), or, more simply, have the wrong time or spatial scale for the model.

There are broadly two sets of techniques for choosing/excluding variables. We can either examine the real system statistically, independent of the model, to see which variables might be appropriate, or we can run the model and use its ability to predict the real system to determine how well we've identified the variables needed.

The former methodologies have the advantage that we are dealing directly with the system. However, for agent-based models they have the issue that they tend to assume aggregate statistical tests on lumped data can identify variables acting at the individual level. With model-testing, we often assume our model form is appropriate, and any differences between the model outputs and the real world are due to poorly chosen variables/parameters, which is plainly untrue. However, the advantage with this approach is that testing is achieved at the same scale as the final model.

In both cases, the choice of variables is often (though not always) compared with a single dependent predictand, with the strength of the relationship being used to exclude variables. It should be noted that is not necessarily ideal. Utilisation of a single output statistic (or, indeed, multiple statistics) is always going to be problematic, as it will fail to calibrate the system to the nuances of the detailed individual characteristics of the system (Wagener et al. 2003) even if the model is at the individual level. Optimisation against a single output may only be sufficient to identify between three and five parameters with any accuracy. It may be necessary to consider multiple outputs to gain any further distinction (Wagener et al. 2003). Moreover, following Benjamini and Hochberg (1995), there is an argument that more attention should be given to the false-positive (Type I) errors when variables are kept, to ensure than random variation doesn’t allow in variables that could be trimmed out (Green and Babyak 1997). The probability of Type I and II errors in multi-model assessment can usefully be balanced with reference to the costs to policy makers that result from the different errors (Hartley et al. 2006).

In the first category of techniques, examining the real system, the simplest method is just to examine the size of the variables. For linear models, variables can be removed on the basis that smaller variables are less likely to have an effect than larger ones, and small co-varying variables, particularly those on the same time-cycles, can be removed or aggregated (O’Neill and Rust 1979). However, this is less possible for non-linear models, where small variations in variables can have large effects. Looking in more detail at the relationships rather than the size, Stepwise Linear Regression has been used since the 1940s to exclude insignificant variables (Glahn and Lowry 1972). Although the core technique is broadly distribution-insensitive, it does assume variables are uncorrelated and related to the final dependent variable linearly. Stepwise variables really need to be on a common range to avoid size effects. While there are issues with this (see King 1986), a range transformation can aid when working with some non-linearities.

Where there is co-linearity between variables it may be that an underlying variable or process may be responsible. While we may be able to tease apart the relationships with an instrumental variable approach, the usual method for proceeding in such cases is to use Principle Components Analysis (PCA) to combine variables into independent components representing the latent variables. This can both indicate variables that are essential/non-essential and provide combined-variable components that represent the missing ‘true’ variable influencing the system. PCA analysis of model parameter sets following calibration runs can additionally reveal potential points of investigation for new processes not directly captured by the model (Keesman and van Straten 1990). Plainly, we may also find ourselves in the situation of having ‘known unknowns' – knowing a variable is missing, but being unable to discover what it is. Provided we know something of the part played by the variable we may still be able to represent such unknowns as latent variables within an agent based model, as they are within Bayesian (Kavetski et al. 2006), Hierarchical Bayesian (Clark 2005), or Structural Equation Modelling. In these techniques the explicit representation of uncertainty usefully shifts the models away from only assessing uncertainty at the level of inputs (through Monte Carlo sampling) and outputs (Clark and Gelfand 2006). However, embedding Bayesian techniques themselves, for example, is not always simple in agent-based models, not least because Bayesian assessments of any detail often rely on an assumption of independent Gaussian output noise (see, for example, Kavetski et al. 2006).

In general, for non-linear systems that are sensitive to small variable changes, it is usually the case that attempts to identify variables statistically from the original data are of limited success. For such systems we really need to consider all possible variable combinations and their effects on model runs, though generally a subset of the combinotronic space is used. There is a large literature on variable selection that utilises models. George (2000) provides an overview of the key issues. Statistical representations of the model may suggest the number of parameters that can reasonably be extracted from the data (e.g. Young et al. 1996), but more usually selection proceeds by running the model with a set of variables and assessing how well it runs, either through significance testing (for example, in Structural Equation Modelling: Green and Babyak 1997) or, more commonly, by ranking the errors associated with different selections.

The spread of values of parameters that match model inputs to model results can tell us if the associated variables are important to the sensitivity of the model. If we are confident in our model structure, parameters which vary a great deal between calibrations while still producing viable results may not be especially *important* to the detailed behaviour of a system (Spear 1970; though see below) and might be discarded. The Generalised/Regional Sensitivity Analysis (GSA/RSA) Hornberger-Spear-Young Algorithm utilises this *rejection sampling* and Monte Carlo testing of inputs and parameters to determine which variables a model should contain (Hornberger and Spear 1981; see Beck 1987; Young et al. 1996 for summaries of developments). Although such techniques tend to be tied to statistical models, the general principles are applicable in agent-based systems. A popular alternative to GSA, sometimes merged with it in hybrid methodologies, is to allow the weighting of variables to be dynamically set during single model runs, and to prune weightings associated with the variables dynamically as model calibration moves towards highlighting some variables over others (essentially the non-linear equivalent of the above pruning of small-sized variables). This pruning can, for example, be done with a Bayesian approach (George 2000). Of course, the danger with this is that parameters extracted from the real-world system may not be stable, and the relationships as represented may vary (Matott et al. 2009). In one hybrid example, Wagener et al. (2003) suggest that by splitting up the parameters' range and different modelling time-windows it should be possible to identify which parameters are important at specific model periods. This also allows an assessment of the sensitivity of specific model components formed by combining parameters.

Looking at parameter variation has another useful by-product: variation over time may also tell us whether variables are *missing*. When we think that variables should be related in a stable fashion, variation may result from the current parameters adapting to make up for missing parameters (Beck 1987). Moreover, Beck (1987) and (for an agent-based system) Heppenstall et al. (2007) have suggested that for recursive estimation and Genetic Algorithm based parameter calibration respectively, trajectories through parameter space may reveal underlying processes in the real data. Beck notes that calibration can often clash with model forms, suggesting adjustment is necessary.

15.3.2 Model Representation – Is This the Right Functional Form?

Even if we can correctly identify the variables involved in our model, we still have the potential for *model error*, that is, error in our final outputs resulting from a structural problem with our model. We need to tackle the *identifiability problem*, for both variables and the relationships between them captured in the model. In general this is not an area of error much investigated by agent-based modellers. This perhaps reflects our general feeling that we are better placed than most modellers to claim our models match reality and are directly representative of true objects and relationships in the world. Even if we believe our agents represent active agencies in the real world, we can be much less certain that we have no *functional error* (van der Sluijs et al. 2003), that is, that we are using the correct relationships between them.

In general agent rulesets will be built up from other studies that generate/test hypotheses about relationships in the real world, and give them a significance value that, broadly, represents the likelihood that the relationships are not falsely identified as real. Plainly there are potential errors here associated with identifying the incorrect hypothesized relationships, and most statistical tests include terms to allow for sample size and degrees of freedom, and will have a particular power representing the likelihood of false positives and false negatives. The question then, really, is how these sub-models/rulesets are combined when no, or relatively little, information on the combination process exists. Frequently this combination in agent systems is achieved through choosing weighted elements based on a ranking process, or combining them arithmetically, but there are many alternatives (see, for examples, Wooldridge 2009). This problem arises beyond areas of, for example, decision making – we may lack a coherent understanding of even relatively deterministic elements of the model.

On the simplest level, we can examine the performance of a single model run under different starting conditions and parameterisations to gain an idea of the range of probabilistic outcomes. Differences between the space of model responses and the real data may allow us to explore model deficiencies and even go some way to separating out model structural error from input uncertainties (Keesman and van

Straten 1989). Alternatively, we can build our models by evolving them to have the right components, through Genetic Programming, with sub-models as genes (see Poli et al. 2008 as a starting point).

However, multiple model testing is now becoming the preferred option in many modelling fields. Indeed, if one looks at subject areas where models are entrenched in the testing of hypotheses, multiple model testing is replacing single model vs null hypothesis testing as the standard methodology (Johnson and Omland 2004), with the likelihood of gaining a correct hypothesis considerably enhanced by multiple hypothesis testing as significances can be ranked and filtered, and likelihoods enhanced through Bayesian techniques (see Farcomeni 2008).

A general methodology for multiple model testing of parameters was developed by Hornburger, Spear, and Young (see Sect. 15.3.1, above). We shall come back to examine this in detail when we look at calibration, however, the basic idea is that multiple models with different parameters are run and only those models that can hit a given set of targets are kept (so called *rejection sampling*). This algorithm was developed into the GLUE (Generalised Likelihood Uncertainty Estimation) procedure by Bevan and Binley (1992). This utilises multiple model runs which may vary in form or parameterisation, and assigns a likelihood to each. Results can then be ranked by likelihood and/or summary statistics generated by weighted combinations of the predictions. Poor models sets can be removed when new data from the real world is available to validate against. O'Neill et al. (1980) have suggested that by filtering out model runs by validation criteria at different stages of the model evolution (e.g. days 10, 20, and 30 of a model run) it is possible to constrain the error of the final models that survive. The days for this filtering are best taken when the inter-model variation is high (O'Neill et al. 1980). By adjusting the parameters on the basis of their co-variance it is possible to reduce their error further (O'Neill et al. 1980). Gupta et al. (1998) extend these broad techniques to multi-objective (~output) models and review alternative developments.

The potential for combining model results under GLUE marks it out as an early basic example of a broader set of methodologies for *Multimodel Ensemble Forecasting.* With ensemble modelling the issue of which model to run is avoided, to an extent, by running multiple models and then selecting the best or combining their results. In ensembles, one can either run very different models, or the same model can be run multiple times with a variety of initial states drawn from the potential distribution of real conditions and their potential errors. Once ensemble models have run, they can be combined to give an overall prediction including an uncertainty measure dictated by not only within-forecast variation, but between forecasts as well, for example, using Bayesian Model Averaging (Raftery et al. 2005). In general, combining multiple predictions will improve forecast reliability in the same way that generating the mean of noisy data is usually a better estimate of the true value than picking a single sample (Leith 1974). The combination of predictions means that forecast *sharpness* (closeness of forecasts) can be assessed as an additional uncertainty measure (Gneiting and Raftery 2005). A good review of multi-model selection criteria and combination techniques can be found in Johnson and Omland (2004). Generally multiple-model ensemble methods are most frequently used in climate/weather studies and hydrology.

They are rarer elsewhere, where single models with randomisation of key components and a probabilistic assessment are more likely (Brown 2010; for a review in meteorology, see Gneiting and Raftery 2005). This reflects the considerable costs involved in multiple model development and the limited number of researchers working in very specific fields, particularly in the social sciences.

15.3.3 Picking Scale

One of the problems with data-driven identification of models/variables is that the system explicitly represents the spatial and temporal scales at which the data is sampled, rather than that most appropriate for the system (Young et al. 1996). To an extent this is mitigated in agent-based systems which have the potential for modelling different components at the appropriate spatio-temporal scales with less of the cross-scale errors that creep into other kinds of models. Multi-scale modelling and validation where there was any doubt would be an ideal solution, but data and computational effort are strongly limiting factors in this. To an extent the issue can be investigated by using cross-scale validation techniques (Costanza 1989) both during calibration and to examine key scales at which the model best represents the system (Malleson 2010).

15.3.4 Model Fitting – Picking Parameters

For any given sub-component of a model there is usually a need to estimate parameters from the real world as represented in training datasets. Such parameters are almost certainly going to be a 'fudge' on real-world processes, and therefore be associated with errors of verisimilitude, and there will be additional errors associated with accurately estimating them: inversion errors of picking the correct weights from the vast number that may model training datasets, and accuracy errors associated with picking them well. A large number of parameters can potentially lead to cryptic equifinality, with erroneous models matching training data simply by providing so many tunable parameters that they can match any function. Having a wide variety of parameters may also enhance overfitting unless care is taken to prevent it, that is, model weights may be adjusted to the point that they very accurately model training datasets, but don't have the flexibility to capture the alternative behaviours of the real system. There is a perverse relationship between the number of parameters and overfitting because poor models with lower levels of parameters won't overfit, whereas those with the right number are more likely to do so. The usual solution to overfitting is to reserve some data as a test set not involved in training, but this is often difficult to justify where data is thin on the ground or critical and unique, as it tends to be in social sciences, and, as mentioned above, spatial and temporal autocorrelation can cause problems in determining the appropriate size of dataset necessary to do a good calibration job.

More generally the errors associated with parameters are adjusted largely by minimising the error of the output of the model, either assuming that the error is entirely due to the parameters being mis-calibrated, or trying to segregate the errors from different sources, i.e. inputs, parameter calibration, and the model form. A key element of this may be *sensitivity testing*: perturbing inputs and/or parameters to see what the result is on the final model output. This allows an assessment of the importance of the input/parameter on the model behaviour, and, if the perturbation is drawn from an input error distribution using Monte Carlo sampling, an idea of how much those errors change the range of results (*uncertainty testing*). The standard text on sensitivity testing is Saltelli et al. (2000), but a good introductory review is Hamby (1994).

Once errors are assessed they can be used to adjust the parameters to improve the match, either statically, at the end of the model run, or dynamically as the model is running. The estimate/adjustment of unknown weights associated with variables can be achieved in a variety of ways:

1. through expert/stakeholder advice,
2. real-world experimentation in aggregate,
3. or automatic fitting to known input–output data.

15.3.4.1 Expert Advice

There are a wide range of methodologies for involving experts and stakeholders in model design and assessment. At the simplest, this involves expert *face validation* of parameters determined automatically, that is getting experts to agree the model looks ok. Seminal work on the process and problems of eliciting uncertainty assessments from experts was presented by Spetzler and von Holstein (1975). More recently a sophisticated expert-analysis process, which includes quantitative sensitivity testing, was developed by Funtowicz and Ravetz (1990). Their NUSAP (Numeral Unit Spread Assessment Pedigree) methodology builds up a 'pedigree' for a model based on evidence including expert opinion on proxy use, the empirical basis of parameters, theoretical understanding, methodological rigour, and model validation (Van der Sluij et al. 2002; http://www.nusap.net). Alternatively expert advice can be incorporated at one remove from the assessment process, by getting experts to design the metrics for uncertainty assessment, rather than completing the assessment themselves (Bevan and Binley 1992).

At the other end of the scale, experts can directly choose parameters. Because of the complication of most models and the lack of absolute verisimilitude, it is rare for experts to choose the values that parameters are fixed at. It is more usual for expert advice to be used in initialising weights that are then adjusted through calibration against the real world. For example, expert advice can be: incorporated into the development of priors in Bayesian treatments of parameters/parameter uncertainty (for a summary, see Clark 2005; for a clear discussion on options for very non-informed priors, see Kavetski et al. 2006); used to constrain the ranges parameters are sampled from (Lutz et al. 1996); or alternatively incorporated though the development of inputs or parameters as fuzzy sets (Janssen et al. 2010). The balance between automatic calibration and expert input can vary considerably, with attempts made to integrate expert calibration into an otherwise automatic procedure (Gupta 1999) and to replicate the actions of experts automatically (Boyle et al. 2000).

15.3.4.2 Real-World Experimentation in Aggregate

Sadly governments seem strangely unwilling to give agent-based modellers the complete control over national policies they need and deserve. Real-world experimentation in aggregate is more common in the physical sciences, where ethical issues play out less. For social science modellers, such parameters are usually taken from the quantitative literature outlining statistical treatments of society, but these more rarely generate laws and sets of parameters that can be built directly into larger-scale models in the same way. Large scale experiments to derive rulesets are rare, even in these days of internet data collection and citizen scientists.

15.3.4.3 Fitting to Known Input–Output Data

Most commonly models follow a process of *data assimilation*, in which forecasts (or, more rarely, backcasts) are generated and compared with real-world data, with the model being adjusted automatically on the basis of the difference. With agent-based models this adjustment is commonly a static process – the model runs to some completion and then the adjustment takes place. This is because agent-based systems are generally initiated and allowed to run on their internal dynamics without the injection of external driving data as the model progresses. However, sequential/dynamic data assimilation (that is, adjustment as the model runs) is common in other fields and likely to become an increasingly important element of agent-based modelling as it attempts to take on predicting large scale and dynamic socio-economic systems (as we shall see, machine learning does represent a middle-way taken by many agent-based systems).

The calibration process has to find optimal parameter weights in a variable space of potential solutions. For simple mathematical functions with a limited number of variables, the technique used is usually to assume the function includes one or more error terms, and then to fit the function to the data by minimising the error term. The classic example of this is linear least-squares fitting, which seeks to place a line representing data through scattered data points by minimising the residual error between the line and the points along its length. Such techniques make a number of assumptions, not least that the errors are random and limited to specific variables. For example, the standard least-squares method assumes there is only an error on the independent variable, not the dependent variable that is being predicted. This is rarely the case where two datasets are being used to derive model rules.

Unfortunately for most agent-based models the non-linearities and considerable interactions involved render mathematical treatments impossible for almost all components. The solution spaces involved are complicated and too extensive to try all parameter combinations. In the absence of expert advice and experimental results, we are usually left with imputing the parameters from data. The worse-case scenario is where we have clear input data, but only a very qualitative understanding of what potential outputs might look like (for example, in predicting urban form). Choosing parameters by manually manipulating their values to see what gives appropriate-looking results is generally to be avoided. The inversion problem plays out particularly badly against researchers with limited time on their hands and it is likely that local or sub-optima will be chosen. Nevertheless, this technique is frequent in agent-based modelling, as the computational resources needed for model runs are high (removing automated checking as an option), and the variables are often interdependent in non-linear manners (rendering mathematical optimisations inappropriate/impossible). Experts should always be involved in the face validation process where it can't be avoided to limit the potential errors.

Where the computational demands are less restrictive, but still prevent a full characterisation of the solution space, we have the option of adjusting the parameters through either a greedy algorithm (adjusting the weights by some rule and keeping those changes that improve the final fit) or some mathematical equivalent (distributing the error to individual components and adjusting them to reduce the local error). As part of this data-led process we usually have to identify some optimisation function to minimise (usually the error between reality and the model output, but not always), and heuristics to control the selection of adjustments.

Standard treatments in non-agent-based models are, at their simplest, recursive greedy treatments with parameters updated on the basis of new data (Gupta et al. 1998, review standard methods for multi-input/multi-output calibration). Many modelling techniques rely on transfer functions to convert between input sources and output objectives (one can visualise a matrix that stores the functions that convert between the two). Given output errors, it is possible, if we know the form of the relationship between input parameters and outputs, to estimate the error in the functions' parameters (Beck 1987), in a manner similar to back-propagation in neural-networks (though with a more flexible set of functional relationships). A great many techniques rely on linearising these functions through Taylor expansions for key conditions or dynamically, as a precursor to allocating error to the parameters. As one can imagine, the mathematics of updating the associated parameters becomes quite complicated. Many models rely on Bayesian methodologies to cope with the updating process, though this is still far from simple. Furthermore as many inputs and parameters are non-normal and cross-correlated, inputs and parameters are often sampled using Monte Carlo techniques when looking at the error in the model due to noise and calibration issues. Generally the sensitivity of non-linear models to small changes in parameters means that multiple parameter sets need to be tested uniquely (Smith et al. 2010a). Multiple runs of the same model utilising different starting conditions and parameter sets allows for the quantification of the error and its effects and the use of this information in the updating process. The full adjustment process is therefore of considerable complexity. While most such techniques are heavily embedded in statistical modelling, it is nevertheless worth considering the application of their core ideas to agent-based modelling.

The standard technique used is the *Extended Kalman Filter*. The idea behind an Extended Kalman Filter is essentially that we know the output (real predicted values plus model-caused error) is a function of the model components, i.e. the model inputs and parameters, along with errors associated with both. Knowing the output, real values, input values, parameters, and the error associated with the input measurements, we can estimate the remaining missing element, the parameter errors, and adjust them on this basis. The parameter error is only an estimate, and (with an adaptive Kalman filter: see Evensen 1992 or Young 2002 for an introduction) will change with each new input/output pair, but if we know the parameter error, even roughly, we can adjust the parameters to remove that error. When, as usually is the case, there are multiple outputs from the model and multiple parameters, the adjustment is in the form of a Kalman gains matrix, which is used to adjust the parameters' actions in the next iteration along with the error value. The process generally moves recursively. The uncertainty is usually represented through Bayesian-like probabilities (as the error cannot be assumed Gaussian, these are usually dealt with through Monte Carlo methods: see Young 2002 for an introduction), and the adjustment takes place preferentially when we know more about the real world than we do about the model (i.e., there's no adjustment if we're more sure about the model than the current real-world values). Beck (1987) gives a summary of both this technique, and recursive estimation techniques in general, along with a summary of the issues with Extended Kalman Filters, chief of which, from our point of view, is the usual assumption of Gaussian input noise throughout. To gain a best estimate of parameters where there is error, assumptions must be made about the variables the error relates to directly and the error distribution (Smith et al. 2010a), but this is frequently not well characterised for social-science models.

Generally when multiple model runs are used with an algorithm from the Hornburger/Spear/Young family the spread of results gives a minimal estimate of the parameter uncertainty (Gupta et al. 1998). However, with multiple models there is the potential for intelligently utilising cross-model comparison to further limit the parameter uncertainty. The Ensemble Kalman filter, after Evensen (1994), can be utilised on single-model multiple-run ensembles to reduce the combinatorial load needed to characterise the parameter change. It uses Monte Carlo sampling (commonly Markov Chain Monte Carlo) to take an initially naive distribution for each parameter and update it using a Bayesian treatment of new data to gain a better parameter distribution. An alternative methodology by Toth and Kalnay (1993) utilises the differences between perturbed and unperturbed ensemble models to adjust the unperturbed models, removing potential errors caused by specific system instabilities. Of promise is also the SIMEX methodology (Cook and Stefanski 1994) in which a system that has well-understood input errors has increments of those errors added to the inputs across multiple model runs, and the output error assessed. As the output error increases tell us about the relationship between the stepped input and output errors, the remaining output error due to poor parameterisation can be identified as the equivalent of the intercept on a graph of input vs. output errors (in a perfectly modelled system). By building a relationship between the final and input errors, it is therefore possible to estimate the parameter error, and, thereafter, to correct the parameters. Chowdhurya and Sharma (2007) review the literature on this technique, the adjustments necessary under a variety of conditions, and compare it with methodologies like GLUE. More generally, ensemble re/starting conditions can be subjected to a variety of algorithms, including evolutionary algorithms, to constraint the errors and lower computational effort (see NRC 2006, for a review).

In general, however, agent-based social and ecological modellers don't tend to follow the techniques generated in other fields of similar complexity. Instead, they are turning to artificial intelligence (AI) mechanisms to calibrate their models. In part this is because most agent models are extremely computationally expensive to run, but the subject area doesn't have the computational, personnel, or data resourcing seen in, for example, climate modelling. AI represents a sound and relatively fast method of calibration. It is usual for most models to be a mix of parameters fixed on the basis of the literature, parameters fixed by an AI method like a Genetic Algorithm, and parameters that vary stochastically across a distribution, picked with Monte Carlo sampling. Such multi-method models are difficult to assess for parameter quality except by validation of their outputs, though there is no reason some of the algorithms above could not be applied to elements of the models.

In addition to fixed parameters, most agent-based techniques include some form of machine learning, essentially doing the same job as dynamic data assimilation for a limited sub-set of parameters; parameters are derived by experiencing the system and optimising behaviour based on one or more objective functions. These objective functions are generally more internalised than simply the error between model outputs and the real world. In many senses agent-based modellers would rather see a model that learns well, than one that minimises an output error, but which has unrealistic internal dynamics. The problem is, of course, that such learning is hard to assess for reliability, except to the degree to which the overall model works.

Where information comes from experimentation or the literature, rather than model testing, confidence intervals are usually used to represent input and parameter uncertainty because inter-relationships are rarely known (Young et al. 1996). Confidence intervals for model parameters are more difficult to calculate properly when there is covariance between inputs/parameters, when the solution surface is complicated, and where input errors are poorly understood (Gallagher and Doherty 2007; who give some indications of ways forward). Under these conditions, and for relatively simple parameterisations, uncertainty associated with inputs can be represented through a sensitivity coefficient matrix – more detail on these will be given below, but essentially they are the covariance matrix showing how the output/s change as each input varies. In terms of parameter uncertainty, for large numbers of more complicated tests statistical significances can be generated to reduce parameter errors, with significances adjusted to pare down the potential for Type I and II errors, which would be high using traditional one-test *p*-values (Farcomeni 2008). The variation during parameterisation can also be used to give uncertainty statistics (see Matott et al. 2009, for a review). Equally, some calibration tests, notably those based on Fuzzy or Baysian and/or Monte Carlo techniques can give uncertainty estimates (Keesman and van Straten 1990; Kennedy and O'Hagan 2001; Clancy et al. 2010) and assign likelihoods to parameter sets (Mitchell et al 2009). However, while such tools exist, it is nevertheless more common in model calibration to simply take the best result without considering the potential identification error.

So far we've dealt with uncertainty as something that is outside of the model and to be assessed for reduction. A more realistic way of dealing with it may be to build it into the model, so that the model reacts as more information comes online (through Baysian probabilities or more general Dempster–Shafer methodologies, or by including a more explicit error distribution in the model) or, furthermore, to assume that such an uncertainty is inherent in reality through the use of Fuzzy Sets and Logic (see Hassan et al. 2010, on agent-based systems; also Zadeh 2005 which goes further in handling modelling uncertainty explicitly using methods including Fuzzy Logic).

More generally, however, there is a fundamental question to be asked about many of these calibration techniques, including those used currently by many agent-based modellers. Many traditional model calibration/inversion techniques fail to cope with agent-based systems simply on the basis that they adjust parameter weightings to an average across a system, which isn't what an individual agent would respond to. Ideally each agent needs calibrating separately, rather than picking up average behaviours. However, if traditional calibration is to be utilised, the space to explore for individual calibration is considerable and the number of parameters fitting the system very large. In this sense, giving each agent some degree of machine learning may be the closest we can get to appropriate parameterisation in agent-based systems.

15.4 Model Mechanics – Errors Generated by Running the Model

In general, agent-based modellers assume models run well, not least because processor time renders multiple-platform runs difficult. Our confidence in this matter may be misplaced. *Model-fix Errors* come in when elements are added to the model that are not in the real system, either for simplification or because an element of the real system is not understood (van der Sluijs et al. 2003). These errors can be distinguished from *Process Error*, in which a complex element of the real system is simplified to make calculation tractable (van der Sluijs et al. 2003). On top of such accepted errors, it may be that our software is not well formed, either because of software bugs, or because the digital precision needed is not sufficient.

15.4.1 Model Bugs

It is an unpleasant truth that many of our models probably contain coding errors. Les Hatton produced a devastating report on the quality of coding in industrial programs influenced by academia (Hatton 1997). He noted that on average the C programs he looked at contained 8 lines hiding serious faults per 1,000 lines of code. Programs written in the academic favourite Fortran were generally so over-parameterised and poorly written that the average rose to 12 lines in 1,000. Moreover, the situation with Fortran wasn't helped by the fact that software written in Fortran contained 2.5 times as many lines as the equivalent C software. When a single algorithm for seismic processing was tested with the same data across multiple programs and platforms, Hatton found that the results were only comparable to within one significant figure. Before anyone gets too smug about this never happening in their code, let us not forget that these programs, albeit starting out as academic software, were finalised by software houses who work to specific quality standards and testing regimes. Some recent changes in programming will have reduced these issues: the removal of some error prone areas of code, such as pointers, from languages like Java will have helped considerably, as will the rise of Programming by Contract, Unit Testing, and the inclusion of Assertions. However, it remains true that most academic code, particularly that written in older versions of languages like Fortran, is likely to be replete with issues. Galán et al. (2009) offer practical advice, for agent-based modellers specifically, on model verification and code-checking.

15.4.2 Uncertainty Due to Representation

Computers can only usually hold memory-limited binary representations of numbers. As such, some numbers are, by necessity, stored as approximations. Such digital imprecision can, if unchecked and/or propagated, result in catastrophic macro-scale errors (see, for an example, Hayes 2003). Good, programmer-centred, discussions on mitigating this issue can be found in Warren (2002) or Hyde (2004), while Izquierdo and Polhill (2006) and Polhill et al. (2006) provide sound practical advice and concentrate specifically on the propagation of these errors in agent-based modelling. Ultimately, however, the issue is constraining. The unification of most platforms around IEEE 754 as the standard for floating-point arithmetic has helped coders at least tackle the issue consistently (though utilising IEEE 754 standard routines in some languages is still far from direct – yes, Java, I'm talking about you). Nevertheless, one still has to take care with the transfer of code involving other data types between platforms (for example, the maximum integer size can change considerably). In general it is good practice to assess model error due to differences in processor, complier, and memory architecture, by transferring models to different platforms. However, the implementation of such transfers is limited by the lack of common code representation schemes of sufficient detail and the coding time needed. Common runtime environments such as the Java and .Net virtual machines mitigate the effort required to some extent, but don't stress-test code to a great enough degree as some issues that usually play out more apparently on different platforms are ameliorated at the virtual machine level. For problems of representation specifically, efforts to work using *Interval Computation* (essentially arithmetically treating the potential upper and lower bounds of representations as they interact) seem promising, if at an early stage for complex models. A good introduction can be found in Hayes (2003), while further material can be found at http://www.cs.utep.edu/interval-comp/.

Related issues for spatio-temporal modellers include the granularity within which space and time are represented, which controls the *resolution* of the data – the size of the smallest useful object. These issues stretch from the appropriateness of different styles and sizes of neighbourhood in Cellular Automata (see Moreno et al. 2008), through to the synchronous or asynchronous updating of agent states (see Schönfisch and Roos 1999). This is a vast area of potential error; however, in general, the most recognised response is to build up models across a variety of tested landscapes, starting with abstract plains, and to test models on multiple systems as above.

15.5 Output Uncertainties

More often than not, the problems involved in quantifying input and parameter uncertainties mean that agent-based modellers deal with uncertainty at the point of model output. While outputs can be assessed for overall uncertainty, it is also at output that we most often consider the representation of uncertainty to stakeholders, and the recording of uncertainty in metadata.

15.5.1 Assessing Overall Uncertainty

In general a large number of agent-based studies either make no direct comparison with the real world (in the sense that they are abstract behavioural models), or treat the error between predictions and reality as the single expression of model uncertainty. If this error is low, the assumption is that inputs are realistic and parameters well estimated. While there is some truth to this, such characterisations give us little idea of how a model will respond to change, or where the model or data needs investment. If, instead, we can examine the contribution of specific input, parameter, and model-form errors to the final prediction we stand a better chance of commenting on, and tackling, these issues. Of course, if a model isn't sensitive to errors, it matters less if they are present; but if a model changes in a strongly non-linear fashion under error, then that has important ramifications for its predictive power.

Traditionally the contribution of errors in mathematical models is examined by tracking the noise from the inputs and using the difference between model outputs (including the noise) and the real world (the, so-called, *prediction error*) to estimate the errors due to parameters. Generally a traditional error propagation/sensitivity analysis utilises the following formula (commonly after Ku 1966), which gives the standard deviation of the results Y of a function, based on the standard deviations

(s)/ variances (s^2) of the input variables ($X, Z\ldots$), and the relationship between each variable and Y:

$$s_y = \sqrt{\left(\frac{\partial Y}{\partial X}\right)^2 s_x^2 + \left(\frac{\partial Y}{\partial Z}\right)^2 s_z^2 + \cdots + \left(\frac{\partial Y}{\partial X}\right)\left(\frac{\partial Y}{\partial Z}\right) s_{xz}^2 + \cdots}$$

where $s_{xz}^{\ 2}$ is the estimated covariance between the variables X and Z, and $\frac{\partial Y}{\partial X}$ is the partial derivative of the function Y with respect to X, known in this context as the *sensitivity coefficient* (see NIST/SEMATECH 2010 for a summary). Where more than one output value is predicted, the equation needs expanding to Jacobean matrices (that relate each variable to each output via a partial derivative). Even when the relationship between Y and each variable is poorly characterised, as long as the variables can be shown to be independent (i.e. with no covariance), input variation can be empirically correlated with outputs individually to give the sensitivity coefficients (O'Neill et al. 1980). For independent variables and relatively simple relationships this leads to reasonably simple predictors for error which can be used within models and which can give rankable information on the importance of variables to the model sensitivity and confidence intervals (Walker 1982).

However, there are considerable issues in applying this methodology in the kinds of systems agent-based modellers deal with, and the kinds of models they generate. Variables are rarely completely independent in non-linear systems, and in such cases a more sophisticated development based around variance-covariance matrices is necessary (O'Neill et al. 1980; van Straten 1985; for developed details, see Beck 1987). In combination the error terms can gain strange distributions if the same variables link together multiple mathematical representations within a model (Tang and Wang 2001 – see references therein). If these relationships vary with time, the matrices may need updating with new input data iteratively (see parameter estimates, Sect. 15.3.4). In addition, spatial systems have their own problems, both with spatial autocorrelation of errors, and with large combinotronic spaces when multiple spatial locations contribute to a final prediction at one or more points. Heuvelink (1998) details the use of this technique when mathematically modelling simple spatial systems with well-known input errors, however, there are considerable issues with more complex non-linear spatial models.

In general, non-linear relationships are usually linearized in such treatments through Taylor expansions. This may be limited to points of assumed equilibrium, that is, where it is assumed that if there is no change in inputs there is no change in outputs (Young et al. 1996); but, as mentioned above, this is not always appropriate in the kinds of systems agent-based modellers tackle. Alternatively the linearization may be around dynamic model points, but such schemes do not cope well with the kinds of relationships modelled by agent-based systems, which tend not to be continuously differentiable, if, indeed, they can be represented mathematically at all, and where function-changing relationships between two variables and a third can make partial differentials difficult to work with consistently. Either way, for more complicated functions under large variances the first-order linear approximations generally used introduce their own errors. Replacements for this technique, which take more account of the non-linear nature of most systems, still tend to rely on an overly mathematical treatment in which noise is regarded as an additional component to a signal (see, for example, Smith et al. 2010a, and for a review Matott et al. 2009).

Given these problems, for models of any level of complication it is usual to resort to Monte Carlo Sensitivity (MCS) testing, in which the model is run multiple times with input data perturbed in some fashion. In uncertainty testing the perturbations are usually drawn from the error distribution of the appropriate inputs, and the parameter distributions are usually also sampled to provide the parameters for each run. Although sensitivity testing can proceed by targeted 'manual' manipulation of the inputs, automated Monte Carlo sample selection based on input/parameter distributions is needed for full output distribution uncertainty testing. There are some broad variations on the scheme: Bayesian models, for example, generally explore uncertainty by sampling their parameter distributions, and then adding white noise to the inputs, while GLUE simply varies the parameter values (Kavetski et al 2006). Either way, multiple model runs using such carefully selected inputs and/or parameters allow for an assessment of the variation in the model outputs on the basis of their errors, and statistical summaries can be generated, along with confidence statistics. In this way, the technique avoids the middle stage of traditional error assessments: the stage of directly calculating the error propagation. A good introduction to Monte Carlo techniques in a spatial context can be found in Walker et al. (2003), along with references to work on sensitivity testing, while a clear detailing of the technique from the point of view of tracking input errors can be found in JCGM (2008a). A more generic study of uncertainty testing, concentrating on statistical summaries, can be found in Bobashev and Morris (2010).

Because the run-time of models can be long, Monte Carlo simulation of thousands of runs length may be inappropriate, even with parallel processing. Some spatial analysts have claimed that much of a distribution can be determined with a small number of runs (up to a hundred: Openshaw 1989; Bobashev and Morris 2010), but this is of considerable contention (Heuvelink 1998). Given this, more restricted tests have been devised which control the sampling of inputs to ensure a small but representative sample of their distributions is taken into account. One could, for example, sample the parameter space regularly: a so-called *Grid Sample*. While this has the advantage that it is easy to see the sensitivity of one parameter against others (Urban and Fricker 2010), this still generates large numbers of runs. More notable is the Latin Hypercube sampling technique (see McKay et al. 1979), in which each input is divided up into n number of sections, each with an equal probability of occurring (i.e. for a normal distribution the sections are larger at the distribution limbs, where probabilities are generally lower). Each section is then Monte Carlo sampled once and only once. Each value from the series of sections for one variable is then combined with an equivalent value from each of the other variables, generating n sets of input values. Essentially this ensures the full range of sample distributions will be sampled, but only generates n tests. The combination of samples from each distribution can be random, or the combination can be chosen to enhance or dampen correlations between the values (for a discussion, see Wyss and

Jorgensen 1998; Urban and Fricker 2010). An excellent summary and pointers to the literature on sensitivity testing can be found in Wyss and Jorgensen (1998) which is associated with software for generating both Monte Carlo and Latin Hypercube input sample datasets. Of course, the use of this technique in testing output variation based on input error assumes the modeller has some idea of the distribution of errors in their inputs, which may not always be true.

Increasingly, in climate modelling, researchers are avoiding the use of full models in parameter sweeps. Instead, they train an emulator (for example, an Artificial Neural Network) on the outputs of a sub-set of parameter sweeps, and then use the emulator to predict the results of a more comprehensive sweep through parameter space with appropriate significance values for the predictions. Of particular interest is Gaussian Process emulators, which use the equivalent of Bayesian kriging to estimate the form of a solution space, in the same way that kriging can be used to estimate missing data in geographical space (Urban and Fricker 2010; see also Young et al. 1996, for a statistical approach that provides a statistical linearization of complex deterministic models). As with kriging, it may be appropriate to feed in training samples in areas of particular variation, worrying less about other areas (Urban and Fricker 2010).

In addition to quantifying uncertainty and error using the outputs, it is also possible to process outputs to reduce the uncertainty by redefining the objective function we are aiming at. For example, where thresholds are involved, uncertainty in models can be reduced by predicting event occurrences rather than continuous probabilities; indeed, generally the prediction of statistical aggregations of outputs, or aggregations related to model outputs can reduce the uncertainty if the relationships are more robust to variance (Glahn and Lowry 1972). More generically, *Forecast Post-Processing* can include interpolation and adjustment for biases and local conditions (NRC 2006). If the outputs are to be used in sequential/dynamic data assimilation (i.e. as the model runs and real data comes in) they will plainly have an effect on the non-linear behavior of the model, and filtering results to remove small-scale instabilities can stop non-linearities getting out of hand (Evensen 1992).

Finally, it is worth noting that in agent-based systems there is interesting work to be done at the meta-assessment level. One direction here is the push towards more objective and automatic hands-off model assessment by allowing meta-agents to assess the models (Li et al. 2007). A second area of interesting potential is the broadening of our criteria of assessment. It is worth noting, with Mearns (2010), that even with the best models, metrics of uncertainty may well increase in some modelling efforts before they decrease. The improvement of models is not always about improving very specific error metrics; structural change may bring greater verisimilitude and future error constraints, without these resulting immediately. We have to be wary of measuring success on the basis of error metrics. Indeed, it may be that with our software, like any engineered solution, we might actually wish to trade error off against alternative values, such as model versatility, adaptability, evolvability or interoperability, and there are a number of techniques from engineering that allow us to manually examine these trade-offs on a cost basis (see Hasings and McManus 2004, for an introduction). This may be an interesting area for meta-agents to additionally explore.

15.5.2 Representing Uncertainty

In general model error is calculated as the total absolute difference between the real and predicted values, normalised in a variety of ways. The bias of a model can be examined by using the total non-absolute difference, as this will highlight consistently great or lesser predictions (if our inputs are reliable, a consistent bias is usually indicative of an issue with the model form). Despite these relatively simple outlines, where the error between a model and reality is given, the statistic used in detail needs considerable and careful thought, especially where the error is heteroscedastic. Gupta et al. (1998) summarise some of the measures commonly used for aspatial value predictions, especially multi-objective predictions that need combining, while Knudsen and Fotheringham (1986) discuss errors in the context of spatial predictions.

Uncertainty itself is usually reported as an estimated statistic (like the mean of model runs) and an uncertainty or set of confidence intervals. For Gaussian sample data, for example, this is usually the sample mean ± standard error. As Smith et al. (2010a) point out, this type of representation is appropriate for linear systems where behaviour varies predictably and slowly with a shift from the mean, but means considerably less in sensitive non-linear systems. In addition error measures like the standard deviation of a sampling distribution drawn from a Gaussian population are well understood for standard statistical estimators like sample means, and the biases between them and population figures are well characterised. The biases in the statistics can therefore be taken into account by readers or augmented when reporting results. For complex and novel model errors, however, this is less easy, and generally it is simpler to quote the distribution-free summaries of model runs. For example, model 95% output ranges are quoted more often than formal 95% confidence intervals (for reasonably clear details of generating confidence intervals from Monte Carlo runs, see Lodwick 1989; Heuvelink 1998; or Bobashev and Morris 2010). However, almost all simple metrics can hide considerable useful information; for example, with Bayesian predictions summary statistics usually hide the fact that forecasts are influenced by the prior belief used to initialise the system.

The relationships between model inputs and outputs can be represented, as discussed above, by sensitivity coefficients. Where the relationship is linear, standard regression between the inputs and outputs, along with a correlation coefficient, is useful, but this becomes more complicated with non-linear non-normal data. For non-linear but independent variables there are less powerful representations of the relationships between inputs and outputs that allow the contribution of the inputs to be quantified, such as the Importance Index, and Relative Deviation Ratio. For more co-linear variables, there is the partial correlation coefficient. A wide range of such basic sensitivity statistics are reviewed in Hamby (1994).

For spatial modellers, it is key to understand the distribution of uncertainty in space and time. Uncertainty can, therefore, usefully be displayed on maps. For example, based on output confidence limits, maps displaying all possible results within 95% confidence limits can be displayed (e.g. Hartley et al. 2006).

For ensemble predictions, Bayesian Model Averaging (Raftery et al. 2005) will produce uncertainty maps that take into account both intra and inter-model variation. Laffan (1999), Reinke and Hunter (2002), Drecki (2002) and Kardos et al. (2003) explore some of the theoretical issues and solutions associated with communicating uncertainty using 2D maps. Uncertainty representation in 3D spatial datasets is explored by Viard et al. (2011).

However, it is plainly important that we consider not only the display of uncertainty to other scientists, but also to policy makers and the public at large. This is equally plainly problematic, and an area in which contentions about the relationship between science, decision-making, the public, and trust are extremely likely to arise (see Brown 2010 for a review). Scientific uncertainty can be converted into policy reticence, even when the science points strongly to action. Equally, however, the exposition of uncertainty can lead to increasingly targeted investment in areas with high uncertainty (Suter et al. 1987). Agent-based systems, with their individual-level processes, may be well placed to bring policy-centred discussions of uncertainty back to a more detailed level of treatment (Zellner 2008), arguably lost for non-linear systems since the move to Monte Carlo assessments.

Shackley and Wynne (1996) discuss some of the mechanisms by which scientists mitigate the effects of uncertainty, while Walker et al. (2003) subdivide some of the uncertainties in ways more pertinent to the interaction of modellers and policy-makers (for example, they identify *scenario uncertainty*, in which it is not clear what scenario is going to occur). Specifically spatial uncertainties and decision making are examined from a policy-makers' viewpoint by Cornélis and Bruet (2002). Morss et al. (2008) give a useful template study for determining how the public understand uncertainty and want it displayed, while a detailed discussion of stakeholder engagement (a very large area) is provided by Dewulf et al. (2005). A formal approach to uncertainty in decision making may be formulated by embedding uncertainty representation within the demands of Quality Assurance (QA) guidelines (see, for examples, Refsgaard et al. 2005; van der Sluijs et al. 2003; JCGM 2008b), potentially including schemes designed under the ISO 9000 standards family. Such guidelines can also include detailed frameworks for decision-making under uncertainty (for an example of a formal quantitative decision-making framework centred on risk and uncertainty see Marin et al. 2003). On the flip-side, Brown (2004, 2010) and Couclelis (2003) provide usefully discussions on the place of uncertainty in science as a social process, and uncertainty's place in scientific self-reflection and knowledge production, while some of the more cognitive uncertainties associated with science-led decision making are described in van der Sluijs et al. (2003).

While we have dealt here with uncertainty associated with the advancement of knowledge, there is one further uncertainty or error that doesn't impact the quality of knowledge advancement, but is nevertheless important for scientists and society because it reduces the speed of progress: the uncertainty that scientific work is novel. Smithson (1989, 3) identified the difficulty of constructing models in a world in which scientific pursuits are becoming increasingly swamped by knowledge and separated into different areas. Recent developments have suggested that scientists are already starting to "re-invent the wheel" (Monaco and Anderson 1994). With over 1,350,000 scientific papers a year published (Björk et al. 2009), this novelty error or (at best) uncertainty can only increase, and represents a very real threat to modelling, if not a significant barrier science needs to avoid as it becomes a mature human endeavour.

15.5.3 Metadata Systems

A structured meta-framework for uncertainty may be built into the model itself, as it is in Bayesian treatments (see Clark 2005) or, for example, through Zadeh's (2005) GTU (Generalized Theory of Uncertainty). However, there are increasing efforts to develop separate metadata systems that focus on uncertainty assessments (Dassonville et al. 2002; Gan and Shi 2002). Such efforts are key to the chained interoperability of models, and the transmission of uncertainty with results. In particular, eXtensible Markup Language schemata that encapsulate uncertainty promise to take uncertainty recording and manipulation from the current level of the dataset down to the specific datum, storing detailed uncertainty information with each data point. A notable example for spatial modellers is UncertML (Williams et al. 2008; http://www.uncertml.org/) which has the potential to be used with the Geographical Markup Language (Cornford 2011), along with the web-based framework supplied to aid in its more general use, UncertWeb (http://www.uncertweb.org/).

15.6 Conclusions

> *To be able to predict only that all things are more or less equally probable is not a useful basis for decision making*
>
> M.B.Beck (1987)

All the above may seem terribly depressing. We work with non-linear, non-normal, high-combinatronic-space, models, highly demanding of computing power and memory storage. Why, then, do we believe we can do any better at modelling the world than astrology or the I-Ching? Are we not generating just the same kinds of largely random outputs and imputing meanings to them beyond rational boundaries? I don't think so, and part of the reason for this confidence comes down to the way the world works. By and large, at the scale at which we deal with it, the world is not completely random; rivers do not leap 50 m into the air and turn into a shower of goldfish; economic systems do not contrive to feed everyone bullion from ATMs. Systems are, generally, very stable compared with the wide range of potential states they could be in. Self-regulatory elements in the systems act to dampen the effect of noise and constrain the propagation of errors. However, my confidence rests in modelling that concentrates to a far greater degree than we currently do in identifying elements that act to dampen systems and drive them towards attractors; elements like social negotiation, averaging, buffering, and thresholding. If we can centre our investigations of the real world on these, and then represent them in our models, we stand a much greater chance of building reasonable models of our apparently highly unpredictable systems.

Finally, it is also worth highlighting an alternative viewpoint put forward by Beck (1987), who suggests that rather than asking what the future will be, given parameters now, we instead ask what parameters would be necessary now, to create a reasonable future. It is generally true that as agent-based modellers we usually model current systems to predict what they might be like in the future, with very little reflection on our duty as academics to imagine a better world, and critique the fundamental components of the systems we are modelling. It is easy to point out when current policies will be disastrous, and even to see how small tweaks may mitigate this, but it is much harder for us to consider wholesale changes that might make the world a notably better, if stranger, place.

15.7 Further Reading

For a good overview of the subject area, which weights scientific methodology and stakeholder engagement, see Refsgaard et al. (2007). NRC (2006: Climate models) reviews uncertainty assessment and control methods, with good sections on uncertainty and decision making, while a formal strategy for conveying uncertainty to policymakers can be found in van der Sluijs et al. (2003). Funtowicz and Ravetz (1993) provide a solid attempt to place uncertainty in the context of both critical theory and the democratization of science.

The Royal Society special issue “Ensembles and probabilities: a new era in the prediction of climate change” (Collins and Knight 2007) gives an insight into the state of the art in much of the field of complex systems modelling outside of agent-based modelling, including the use of emulators, while Brown (2010: Physical models) and Matott et al. (2009: Environmental models) provide good technical overviews of these areas.

Matott et al. (2009) additionally give a very complete review of uncertainty software, broken down into data analysis, identifiability analysis, parameter estimation, uncertainty analysis, sensitivity analysis, multimodel analysis, and Bayesian networks. This is supplemented by an ongoing website at: http://www.epa.gov/athens/research/modeling/modelevaluation/index.html

A good review of Monte Carlo techniques, with a meta-review of other sensitivity and uncertainty testing techniques, is Helton et al. (2006), and Hamby (1994) gives a good review of sensitivity statistics. Finally, Bobashev and Morris (2010) provide a very clear walkthrough of one such sensitivity/uncertainty analysis for an agent-based system.

References

Araújo, M., Pearson, R. G., Thuiller, W., & Erhard, M. (2005). Validation of species-climate impact models under climate change. *Global Change Biology, 11*, 1504–1513.

Ballas, D., Rossiter, D., Thomas, B., Clarke, G. P., & Dorling, D. (2005). *Geography matters: Simulating the local impacts of national social policies*. York: Joseph Rowntree Foundation.

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data* (3rd ed.). Chichester: Wiley. pp.604.

Beck, M. B. (1987). Water quality modeling: A review of the analysis of uncertainty. *Water Resource Research, 23*(8), 1393–1442.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B, 57*, 289–2300.

Beven, K. J., & Binley, A. M. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes, 6*, 279–298.

Björk, B-C., Roos, A., & Lauri, M. (2009). Scientific journal publishing: Yearly volume and open access availability. *Information Research 14*(1): Paper 391. http://InformationR.net/ir/14-1/paper391.html. Accessed 7 Feb 2011

Bobashev, G. V., & Morris, R. J. (2010). Uncertainty and inference in agent-based models. In *Proceedings of the 2010 Second International Conference on Advances in System Simulation*. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5601895. Accessed 31 Mar 2011.

Boyle, D. P., Gupta, H. V., & Sorooshian, S. (2000). Towards improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. *Water Resources Research, 36*, 3663–3674.

Brown, J. D. (2004). Knowledge, uncertainty and physical geography: Towards the development of methodologies for questioning belief. *Transactions of the Institute of British Geographers NS, 29*, 367–381.

Brown, J. D. (2010). Prospects for the open treatment of uncertainty in environmental research. *Progress in Physical Geography, 34*(1), 75–100.

Cerioli, A., & Farcomeni, A. (2011). Error rates for multivariate outlier detection. *Computational Statistics & Data Analysis, 55*(1), 544–553.

Cho, W. K. T., & Gaines, B. J. (2007). Breaking the (Benford) Law: Statistical fraud detection in Campaign Finance. *The American Statistician, 61*(3), 218–223.

Chowdhury, S., & Sharma, A. (2007). Mitigating parameter bias in hydrological modelling due to uncertainty in covariates. *Journal of Hydrology, 340*, 197–204.

Clancy, D., Tanner, J. E., McWilliam, S., & Spencer, M. (2010). Quantifying parameter uncertainty in a coral reef model using Metropolis-Coupled Markov Chain Monte Carlo. *Ecological Modelling, 221,* 1337–1347.

Clark, J. S. (2005). Why environmental scientists are becoming Bayesians. *Ecological Letters, 8*, 2–14.

Clark, J. S., & Gelfand, A. E. (2006). A future for models and data in environmental science. *Trends in Ecology and Evolution, 21*(7), 375–380.

Collins, M., & Knight, S. (2007). Theme issue 'Ensembles and probabilities: A new era in the prediction of climate change'. *Philosophical Transactions of the Royal Society A, 365*, 1957–2191.

Cook, J. R., & Stefanski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association, 89*(428), 1314–1328.

Cornélis, B., & Brunet, S. (2002). A policy-maker point of view on uncertainties in spatial decisions. In W. Shi, P. F. Fisher, & M. F. Goodchild (Eds.), *Spatial data quality* (pp. 168–185). London: Taylor and Frances.

Cornford, D. (2011). Uncertainty and the OGC: An Aston University take. Presented 03/03/2011 at OGC TC Meeting, Universities DWG Session, Bonn. http://www.uncertweb.org/documents. Accessed 1 Mar 2011.

Costanza, R. (1989). Model goodness of fit: A multiple resolution procedure. *Ecological Modelling, 47*, 199–215.

Couclelis, H. (2003). The certainty of uncertainty: GIS and the limits of geographic knowledge. *Transactions in GIS, 7*(2), 165–175.

Dassonville, L., Vauglin, F., Jakobsson, A., & Luzet, C. (2002). Quality management, data quality and users, metadata for geographical information. In W. Shi, P. F. Fisher, & M. F. Goodchild (Eds.), *Spatial data quality* (pp. 202–215). London: Taylor and Frances.

Dewulf, A., Craps, M., Bouwen, R., Taillieu, T., & Pahl-Wostl, C. (2005). Integrated management of natural resources: Dealing with ambiguous issues, multiple actors and diverging frames. *Water Science & Technology, 52*, 115–124.

Di Paolo, E. A, Noble, J, & Bullock, S (2000) Simulation models as opaque thought experiments. In *Seventh International Conference on Artificial Life* (pp. 497–506) Cambridge: MIT Press,. http://eprints.ecs.soton.ac.uk/11455/

Drecki, I. (2002). Visualisation of uncertainty in geographical data. In W. Shi, P. F. Fisher, & M. F. Goodchild (Eds.), *Spatial data quality* (pp. 140–159). London: Taylor and Frances.

Dubois, G., & Saisana, M. (2002). Optimal spatial declustering weights — Comparison of methods. In *Proceedings of the Annual Conference of the International Association for Mathematical Geology*, 15–20 Sept Berlin, 479–484. http://composite-indicators.jrc.ec.europa.eu/Document/Optimizing Spatial Declustering Weights – Comparison of Methods.pdf. Accessed 31 Mar 2011.

Evans, A. J., & Waters, T. (2007). Mapping vernacular geography: Web-based GIS tools for capturing "fuzzy" or "vague" entities. *International Journal of Technology, Policy and Management, 7*(2), 134–150.

Evensen, G. (1992). Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *Journal of Geophysical Research, 97*(C11), 17905–17924.

Evensen, G. (1994). Sequential data assimilation with nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research, 99*(C5), 143–162.

Faber, M., Manstetten, R., & Proops, J. (1992). Humankind and the environment: An anatomy of surprise and ignorance. *Environmental Values, 1*, 217–241.

Farcomeni, A. (2008). A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion. *Statistical Methods in Medical Research, 17*, 347–388.

Fisher, P., Wood, J., & Cheng, T. (2004). Where is Helvellyn? Fuzziness of multi-scale landscape morphometry. *Transactions of the Institute of British Geographers, 29*(1), 106–128.

Funtowicz, S. O., & Ravetz, J. R. (1990). *Uncertainty and quality in science for policy*. Dordrecht: Kluwer Academic Publishers.

Funtowicz, S. O., & Ravetz, J. R. (1993). Science for the post-normal age. *Futures, 25*, 739–755.

Galán, J. M., Izquierdo, L. R., Izquierdo, S. S., Santos, J. I., del Olmo, R., López-Paredes, A., & Edmonds, B. (2009). Errors and artefacts in agent-based modelling. *Journal of Artificial Societies and Social Simulation 12*(1): 1. http://jasss.soc.surrey.ac.uk/12/1/1.html. Accessed 25 Mar 2011.

Gallagher, M., & Doherty, J. (2007). Parameter estimation and uncertainty analysis for a watershed model. *Environmental Modelling and Software, 22*, 1000–1020.

Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1994). *Design patterns: Elements of reusable object-oriented software*. Reading, Mass: Addison Wesley. pp.416.

Gan, E., & Shi, W. (2002). Error metadata management systems. In W. Shi, P. F. Fisher, & M. F. Goodchild (Eds.), *Spatial data quality* (pp. 251–266). London: Taylor and Frances.

George, E. I. (2000). The variable selection problem. *Journal of the American Statistical Association, 95*(452), 1304–1308.

Getis, A. (2007). Reflections on spatial autocorrelation. *Reg Sci Urban Econ, 37*, 491–496.

Glahn, H., & Lowry, D. (1972). The use of Model Output Statistics (MOS) in objective weather forecasting. *Journal of Applied Meteorology, 11*, 1203–1211.

Gneiting, T., & Raftery, A. E. (2005). Weather forecasting with ensemble methods. *Science, 310*(5746), 248–249.

Green, S. B., & Babyak, M. A. (1997). Control of type I errors with multiple tests constraints in structural equation modeling. *Multivariable Behavioural Research, 32*, 39–51.

Groves, R. M., Fowler, F. J., Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology*. Hoboken: Wiley-Blackwell. pp.488.

Gupta, H. V. (1999). Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *Journal of Hydrologic Engineering, 4*(2), 135–143.

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research, 34*(4), 751–763.

Hamby, D. M. (1994). A review of techniques for parameter sensitivity analysis of environmental models. *Environmental Monitoring and Assessment, 32*, 135–154.

Hartley, S., Harris, R., & Lester, P. J. (2006). Quantifying uncertainty in the potential distribution of an invasive species: Climate and the Argentine ant. *Ecology Letters, 9*, 1068–1079.

Hasings, D., & McManus, H. (2004). A framework for understanding uncertainty and its mitigation and exploitation in complex systems. In *2004 Engineering Systems Symposium*, 29–31 Mar 2004, MIT. http://esd.mit.edu/symposium/pdfs/papers/hastings.pdf. Accessed 28 February 2011.

Hassan, S., Garmendia, L., & Pavon, J. (2010). Introducing uncertainty into social simulation: Using fuzzy logic for agent-based modelling. *International Journal of Reasoning-based Intelligent Systems, 2*(2), 118–124.

Hatton, L. (1997). The T experiments: Errors in scientific software. *Computational Science & Engineering, IEEE, 4*(2), 27–38.

Hayes, B. (2003). A Lucid interval. *American Scientist, 91*(6), 484–488.

Helton, J. C., Johnson, J. D., Salaberry, C. J., & Storlie, C. B. (2006). Survey of sampling based methods for uncertainty and sensitivity analysis. *Reliability Engineering and System Safety, 91*, 1175–1209.

Heppenstall, A. J., Evans, A. J., & Birkin, M. H. (2007). Genetic algorithm optimisation of a multi-agent system for simulating a retail market. *Environment and Planning B, 34*(6), 1051–1070.

Heuvelink, G. B. M. (1998). *Error propagation in environmental modelling with GIS*. London: Taylor and Francis. pp.127.

Hornberger, G., & Spear, R. (1981). An approach to the preliminary analysis of environmental systems. *Journal of Environmental Management, 7*, 7–18.

Hyde, R. (2004). *Write great code 1: Understanding the machine*. San Francisco: No Starch Press. pp. 440.

Isaaks, E. H., & Srivastava, R. M. (1990). *Applied geostatistics*. Oxford: Oxford University Press. pp.592.

Izquierdo, L. R., & Polhill, J. G. (2006). Is your model susceptible to floating-point errors? *Journal of Artificial Societies and Social Simulation 9*(4): 4. http://jasss.soc.surrey.ac.uk/9/4/4.html. Accessed 25 Mar 2011.

Janssen, J. A. E. B., Krol, M. S., Schielen, R. M. J., Hoekstra, A. Y., & de Kok, J.-L. (2010). Assessment of uncertainties in expert knowledge, illustrated in fuzzy rule-based models. *Ecological Modelling, 221*, 1245–1251.

JCGM (2008a) Evaluation of measurement data – Supplement 1 to the "Guide to the expression of uncertainty in measurement" – Propagation of distributions using a Monte Carlo method, Joint Committee for Guides in Metrology 100. http://www.bipm.org/utils/common/documents/jcgm/JCGM_101_2008_E.pdf. Accessed 15 Feb 2011.

JCGM (2008b) Evaluation of measurement data – Guide to the expression of uncertainty in measurement, Joint Committee for Guides in Metrology 100. http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf. Accessed 15 Feb 2011.

Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution, 19*, 101–108.

Kardos, J. D., Moore, A., & Benwell, G. L. (2003). Visualising uncertainty in spatially-referenced attribute data using hierarchical spatial data structures. In *Proceedings of the 7th International Conference on GeoComputation*. University of Southampton, 8–10 Sept 2003. http://www.geocomputation.org/2003/. Accessed 25 Mar 2011

Kavetski, D., Kuczera, G., & Franks, S. W. (2006). Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. *Water Resources Research, 42*, W03407.

Keesman, K., & van Straten, G. (1989). Identification and prediction propagation of uncertainty in models with bounded noise. *International Journal of Control, 49*, 2259–2269.
Keesman, K., & van Straten, G. (1990). Set-membership approach to identification and prediction of lake eutrophication. *Water Resources Research, 26*, 2643–2652.
Kelejian, H., & Prucha, I. (2010). Spatial models with spatially lagged dependent variables and incomplete data. *Journal of Geographical Systems, 12*(3), 241–257.
Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society B, 63*(3), 425–464.
Kiesling, J. B. (undated) Charting Electoral Fraud: Turnout Distribution Analysis as a Tool for Election Assessment. *Diplomacy Lessons*. http://www.bradykiesling.com/election_fraud_analysis.htm. Accessed 10 Jan 2010.
King, G. (1986). How not to lie with statistics: Avoiding common mistakes in quantitative political science. *American Journal of Political Science, 30*, 666–687.
Knudsen, C. D., & Fotheringham, A. S. (1986). Matrix comparison, goodness-of-fit and spatial interaction modeling. *International Regional Science Review, 10*(2), 127–147.
Ku, H. (1966). Notes on the use of propagation of error formulas. *Journal of Research of National Bureau of Standards C. Engineering and Instrumentation, 70*(4), 263–273.
Laffan, S. W. (1999). Spatially assessing model error using geographically weighted regression. In *Proceedings of the 4th International Conference on GeoComputation.* Fredericksburg: Mary Washington College, 5–28 July 1999. http://www.geocomputation.org/1999/. Accessed 25 Mar 2011.
Lane, S. N. (2001). Constructive comments on D Massey 'Space-time, "science" and the relationship between physical geography and human geography'. *Transactions of the Institute of British Geographers, 6*, 243–256.
Lauridsen, J., & Kosfeld, R. (2007). Spatial cointegration and heteroscedasticity. *Journal of Geographical Systems, 9*(3), 253–265.
Legendre, P. (1993). Spatial Autocorrelation: Trouble or New Paradigm? *Ecology, 74*, 1659–1673.
Leith, C. E. (1974). Theoretical skill of Monte Carlo forecasts. *Monthly Weather Review, 102*(6), 409–418.
Li, Y., Brimicombe, A. J., & Li, C. (2007). Agent-based services for validating multi-agent models. In *Proceedings of the 9th International Conference on GeoComputation*, Maynooth: Eire.
Lodwick, W. A. (1989). Developing confidence limits on errors of suitability analyses in geographical information systems. In M. Goodchild & S. Gopal (Eds.), *Accuracy of spatial databases* (pp. 69–80). London: Taylor and Francis.
López, C. (1997). Quality of geographic data – Detection of outliers and imputation of missing values (Unpublished PhD thesis, Universidad ORT, Uruguay). http://www.thedigitalmap.com/~carlos/papers/PhDthesis/phdthesis.pdf. Accessed 1 Mar 2011.
Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences, 20*, 130–141.
Luoto, M., Marmion, M., & Hjort, J. (2010). Assessing spatial uncertainty in predictive geomorphological mapping: A multi-modelling approach. *Computers & Geosciences, 36*, 355–361.
Lutz, W., Sanderson, W., & Scherbov, S. (1996). Probabilistic population projections based on expert opinion. In W. Lutz (Ed.), *The future population of the world. What can we assume today?* (pp. 397–428). London: Earthscan Publications. Chapter 16.
Lyman, R. L., & Ames, K. M. (2007). On the use of species-area curves to detect the effects of sample size. *Journal of Archaeological Science, 34*, 1985–1990.
Malinarič, S., & Ďuríšek, P. (2004). Sensitivity coefficients analysis. In *Proceedings of Thermophysics 2004*. http://www.tpl.fpv.ukf.sk/engl_vers/thermophys/2004/Mal-Dur.pdf. Accessed 14 Feb 2011.
Malleson, N. (2010). Agent-based modelling of Burglary (Unpublished PhD thesis, School of Geography, University of Leeds, Leeds). http://www.geog.leeds.ac.uk/fileadmin/downloads/school/people/postgrads/n.malleson/thesis-final.pdf. Accessed 1 Mar 2011.
Marin, C. M., Guvanasen, V., & Saleem, Z. A. (2003). The 3MRA risk assessment framework – A flexible approach for performing multimedia, multipathway, and multireceptor risk assessments

under uncertainty. *Human and Ecological Risk Assessment: An International Journal, 9*(7), 1655–1677.

Martin, D. (2003). Extending the automated zoning procedure to reconcile incompatible zoning systems. *International Journal of Geographical Information Science, 17*, 181–196.

Martin, D., Dorling, D., & Mitchell, R. (2002). *Linking censuses through time: problems and solutions Area, 34*, 82–91.

Matott, L. S., Babendreier, J. E., & Parucker, S. T. (2009). Evaluating uncertainty in integrated environmental models: A review of concepts and tools. *Water Resources Research 45*: WO6421. http://www.agu.org/journals/wr/wr0906/2008WR007301/2008WR007301.pdf. Accessed 2 Feb 2011.

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics, 21*(2), 239–245.

Mearns, L. O. (2010). The drama of uncertainty. *Climatic Change, 100*, 77–85.

Mebane, W. R Jr., & Kalinin, K. (2009). Comparative election fraud detection. In *Annual Meeting of the American Political Science Association*, Toronto 3–6 Sept 2009. http://www-personal.umich.edu/~wmebane/apsa09.pdf. Accessed 10 Jan 2010.

Mitchell, S., Beven, K., & Freer, J. (2009). Multiple sources of predictive uncertainty in modeled estimates of net ecosystem CO_2 exchange. *Ecological Modelling, 220*(23), 3259–3270.

Monaco, J. H., & Anderson, R. L. (1994). Tai's formula is the trapezoidal rule. *Diabetes Care, 17*(10), 1224–1225.

Moreno, N., Wang, F., & Marceau, D. J. (2008). An object-based land-use cellular automata model to overcome cell size and neighborhood sensitivity. In *Proceedings OF GEOBIA 2008 – Pixels, Objects, Intelligence GEOgraphic Object Based Image Analysis for the 21st Century*. Calgary, 5–8 Aug. http://www.isprs.org/proceedings/XXXVIII/4-C1/Sessions/Session6/6753_Marceau_Proc_pap.pdf. Accessed 25 Mar 2011.

Morss, R. E., Demuth, J. L., & Lazo, J. K. (2008). Communicating uncertainty in weather forecasts: a survey of the US public. *Weather and Forecasting, 23,* 974–91.

Nagele, P. (2001). Misuse of standard error of the mean (SEM) when reporting variability of a sample. *A critical evaluation of four anaesthesia journals. British Journal of Anaesthesia, 90*(4), 514–516.

National Research Council of the National Academies (NRC). (2006). *Completing the forecast: Characterizing and communicating uncertainty for better decisions using weather and climate forecasts*. Washington, DC: National Academies Press.

Ngo, T. A., & See, L. M. (2012). Calibration and validation of agent-based models of land cover change. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 181–196). Dordrecht: Springer.

NIST/SEMATECH (2010). *e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/. Accessed 10 Jan 2011.

O'Neill, R. V. (1973). Error analysis of ecological models. In D. J. Nelson (Ed.), *Radionuclides in ecosystems. CONF-710501* (pp. 898–908). Springfield: National Technical Information Service.

O'Neill, R. V., & Rust, B. (1979). Aggregation error in ecological models. *Ecological Modelling, 7*(2), 91–105.

O'Neill, R. V., Gardner, R. H., & Mankin, J. B. (1980). Analysis of parameter error in a nonlinear model. *Ecological Modelling, 8*, 297–311.

Openshaw, S. (1989). Learning to live with errors in spatial databases. In M. Goodchild & S. Gopal (Eds.), *Accuracy of spatial databases* (pp. 263–276). London: Taylor and Francis.

Polhill, J. G., Izquierdo, L. R., & Gotts, N. M. (2006). What every agent-based modeller should know about floating point arithmetic. *Environmental Modelling & Software, 21*(3), 283–309.

Poli, R, Langdon, W. B., and McPhee, N. F., with Koza, J. R. (2008). *A Field Guide to Genetic Programming*. http://www.gp-field-guide.org.uk/. Accessed 28 February 2011.

Poulton, E. C. (1989). *Bias in quantifying judgments*. Hove: Erlbaum. pp.328.

Powers, A. C. (2005). Simulating patterns of uncertainty in postclassification change detection. In *Proceedings of the 8th International Conference on GeoComputation*. University of Michigan, 31 July–3 Aug 2005. http://www.geocomputation.org/2005/. Accessed 25 Mar 2011.

Raftery, A. E., Gneiting, T., Balabdaoui, F., & Polakowski, M. (2005). Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review, 133*, 1155–1174.

Refsgaard, J. C., Henriksen, H. J., Harrar, W. G., Scholten, H., & Kassahun, A. (2005). Quality assurance in model based water management – Review of existing practice and outline of new approaches. *Environmental Modelling & Software, 20*(10), 1201–1215.

Refsgaard, J. C., van der Sluijs, J. P., Etejberg, A. L., & Vanrolleghem, P. A. (2007). Uncertainty in the environmental modeling process – A framework and guidance. *Environmental Modelling and Software, 22*, 1543–1556.

Reinke, K., & Hunter, G. J. (2002). A theory for communicating uncertainty in spatial databases. In W. Shi, P. F. Fisher, & M. F. Goodchild (Eds.), *Spatial data quality* (pp. 76–101). London: Taylor and Frances.

Rogers, J. P., Barbara, D., & Domeniconi, C. (2009). Detecting spatio-temporal outliers with kernels and statistical testing. In *17th International Conference on Geoinformatics*. http://dx.doi.org/10.1109/GEOINFORMATICS.2009.5293481. Accessed 1 Mar 2011.

Rowe, W. D. (1977). *An anatomy of risk*. New York: John Wiley and Sons.

Saltelli, A., Chan, K., & Scott, E. M. (2000). *Sensitivity analysis*. Chichester: Wiley. pp.475.

Schönfisch, B., & de Roos, A. (1999). *Synchronous and asynchronous updating in cellular automata Biosystems, 51*(3), 123–143.

Shackley, S., & Wynne, B. (1996). Representing uncertainty in global climate change science and policy: Boundary ordering devices and authority. *Science Technology Human Values, 21*, 275–302.

Smith, L. A., Cuéllar, M. C., Du, H., & Judd, K. (2010a). Exploiting dynamical coherence: A geometric approach to parameter estimation in nonlinear models. *Physics Letters A, 374*, 2618–2623.

Smith, A. H. C., Ponci, F., & Monti, A. (2010b). Bounding the dynamic behavior of an uncertain system via polynomial chaos-based simulation. *Simulation, 86*, 31–40.

Smithson, M. (1989). *Ignorance and uncertainty: Emerging paradigms*. New York: Springer.

Spear, R. C. (1970). The application of Kolmogorov-Renyi statistics to problems of parameter uncertainty in systems design. *International Journal of Control, 11*, 771–778.

Spetzler, C. S., & von Holstein, S. (1975). Probability encoding in decision analysis. *Management Science, 22*(3), 340–358.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit (with Discussion). *Journal of the Royal Statistical Society B, 64*, 583–639.

Suter, G. W., Barnthouse, L. W., & O'Neill, R. V. (1987). Treatment of risk in environmental impact assessment. *Environmental Management, 11*, 295–303.

Tang, S., & Wang, Y. (2001). A parameter estimation program for error-in-variable model. *Ecological Modelling, 156,* 225–236.

Tobler, W. R. (1979). Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association, 74*, 519–530.

Toth, Z., & Kalnay, E. (1993). Ensemble forecasting at NMC: The generation of perturbations. *Bulletin of American Meteorological Society, 74*, 2317–2330.

Urban, N. M., & Fricker, T. E. (2010). A comparison of Latin hypercube and grid ensemble designs for the multivariate emulation of an Earth system model. *Computers & Geosciences, 36*, 746–755.

Van der Sluijs, J. P., Potting, J., Risbey, J., van Vuuren, D., de Vries, B., Beusen, A., Heuberger, P., Quintana, S. C., Funtowicz, S., Kloprogge, P., Nuijten, D., Petersen, A. C., & Ravetz, J. (2002). *Uncertainty assessment of the IMAGE = TIMER B1 CO2 emissions sScenario, using the NUSAP method.* (Report No. 410 200 104). Bilthoven: Dutch National Research Program on Climate Change.

Van der Sluijs, J. P., Risbey, J., Kloprogge,P., Ravetz, J. R., Funtowicz, S. O., Quintana, S. C., Pereira, A. G., De Marchi, B., Petersen, A. C., Janssen, P. H. M., Hoppe, R., & Huijs, S. W. F. (2003). RIVM/MNP guidance for uncertainty assessment and communication. http://www.nusap.net/downloads/detailedguidance.pdf. Accessed 14 Feb 2011.

van der Wel, F. J. M., van der Gaag, L. C., & Gorte, B. G. H. (1996). Visual exploration of uncertainty in remote sensing classifications. In *Proceedings of the 1st International Conference on GeoComputation*. Leeds: University of Leeds, 17–19 Sept 1996. http://www.geocomputation.org/1996/. Accessed 25 Mar 2011.

van Straten, G. (1985). Analytical methods for parameter-space delimitation and application to shallow-lake phytoplankton-dynamics modeling. *Applied Mathematics and Computation, 17*, 459–482.

Verkuilen, J. (2005). Assigning membership in a fuzzy set analysis. *Sociological Methods & Research, 33*(4), 462–496.

Viard, T., Caumon, G., & Lévy, B. (2011). Adjacent versus coincident representations of geospatial uncertainty: Which promote better decisions? *Computers & Geosciences, 37*(4), 511–520.

Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S., & Gupta, H. V. (2003). Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. *Hydrological Processes, 17*(2), 455–476.

Walker, W. W. (1982). A sensitivity and error analysis framework for lake eutrophication modeling. *Water Resources Bulletin, 18*(1), 53–60.

Walker, W. E., Harremoes, P., Rotmans, J., van der Sluijs, J. P., van Asselt, M. B. A., Janssen, P., et al. (2003). Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment, 4*(1), 5–18.

Warren, H. S., Jr. (2002). *Hacker's delight*. Boston: Addison Wesley. pp. 306.

Williams, M., Cornford, D., Bastin, L., & Ingram, B. (2008). UncertML: an XML schema for exchanging uncertainty. In *Proceedings of GISRUK 2008*. http://www.uncertml.org/publications.php. Accessed 1 Mar 2011.

Wooldridge, M. (2009). *An introduction to multiagent systems* (2nd ed.). Chichester: Wiley. pp. 484.

Wyss, G. D., & Jorgensen, K. H. (1998). *A user's guide to LHS: Sandia's Latin Hypercube Sampling Software*. SAND98-0210 Albuquerque: Sandia National Laboratories.

Young, P. C. (2002). Advances in real-time flood forecasting. *Philosophical Transactions of the Royal Society of London A, 360*, 1433–1450.

Young, P. C., Parkinson, S., & Lees, M. J. (1996). Simplicity out of complexity in environmental modelling: Occam's razor revisited. *Journal of Applied Statistics, 23*, 165–210.

Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning–1. *Information Sciences, 8*, 199–249.

Zadeh, L. A. (1976). A fuzzy-algorithmic approach to the definition of complex or imprecise concepts. *International Journal of Man-Machine Studies, 8*(3), 249–291.

Zadeh, L. A. (2005). Toward a generalized theory of uncertainty (GTU) – An outline. *Information Sciences, 172*, 1–40.

Zellner, M. L. (2008). Embracing complexity and uncertainty: The potential of agent-based modeling for environmental planning and policy. *Planning Theory & Practice, 9*(4), 437–457.

Zhang, J., & Goodchild, M. (2002). *Uncertainty in geographical information* (p. 266). London: Taylor and Francis.

Žižek, S. (2004). What Rumsfeld doesn't know that he knows about Abu Ghraib. *In These Times,* 21 May 2004. http://www.lacan.com/zizekrumsfeld.htm. Accessed 10 Jan 2010.

Chapter 16
Agent-Based Extensions to a Spatial Microsimulation Model of Demographic Change

Belinda M. Wu and Mark H. Birkin

Abstract New technologies and techniques now enable us to construct complex social models with more sophistication. In this paper we introduce an individual-based model, which combines the strengths of both microsimulation models and agent-based model approaches to project the UK population 30 years into the future. The hybrid modelling approach has been adopted to add flexibility and practicality in order to capture individual characteristics, especially in terms of individual movements, interactions and behaviours in the absence of suitable microdata. Such characteristics during the life courses of individuals are modelled through an event-driven model that simulates discrete processes that represent important demographic transitions.

16.1 Introduction

People and society, as well as their past, present and future, have always been of fundamental interest to both academic social scientists and to planners and policy-makers in both government and commercial organisations. New research methods enabled by the capabilities of modern computers may radically transform human ability to reason systematically about complex social systems. This has become increasingly important as our world today confronts rapid and potentially profound transitions driven by social, economic, environmental, and technological changes.

Indeed, "one of demography's main contributions to societal planning is to provide projections of the future population" (Rees 2009). Despite the simplification and uncertainty of the modelled population, population projections still play an

B.M. Wu (✉) • M.H. Birkin
School of Geography, University of Leeds, Leeds, UK
e-mail: b.wu@leeds.ac.uk; m.h.birkin@leeds.ac.uk

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_16,

indispensible role in our society today, as population evolution affects all levels of planning and policy making. MoSeS (Modelling and Simulation for e-Social Science) is a research node of the UK National Centre for e-Social Science.[1] MoSeS aims to develop a national demographic model and simulation of the UK population specified at the level of individuals and households. Our interest is not only in the construction of more sophisticated and effective models, but also how simulations might be transferred into a real-world environment. The approach builds up from cities and regions, with an aim to create simulation models of interactions between individuals, groups, or neighbourhoods within large metropolitan areas. Such simulations can form the basis of a wide range of applications in both e-Research and public policy analysis.

This paper describes the hybrid approach used in the MoSeS Dynamic Model, which combines the strengths of both MSM and ABM to enable the modelling of complex social systems. The results generated by the model and our experiments with different sub-populations will also be discussed in this chapter.

16.2 Background

Modelling population and their societies have always been challenging, due to the complex nature of such systems. They normally consist of a large number of individual components that interact in a strongly nonlinear fashion. Population models concern collections of discrete entities with stochastic behaviours, as well as complex population dynamics caused by temporal population changes. Such changes are bounded and result from something other than a monotonic approach to a stable equilibrium. The dynamics of these systems are typically irregular and may include oscillations and chaos that are produced by some combination of exogenous and endogenous stochastic and/or deterministic factors. Such a complex system requires the capability to handle real data and phenomena that are not necessarily tractable mathematically. Therefore modelling such a system often requires heavy use of simulations and other computational tools to reproduce the picture of the dynamics and behaviours within such systems (Turchin 2003).

There are two main types of population simulation models: microsimulation models (MSMs) and macrosimulation models (see Birkin and Wu 2012, for further details). In MSMs each individual is described with its particular attributes and behaviour, while a macrosimulation model is based on stochastic differential equations, where the population is described in aggregated terms by the number of individuals in different states (Gustafsson and Sternad 2007). Traditional macroscopic population models project populations by age and sex only. Aggregated probabilities are applied to grouped populations by age and sex in the fundamental demographic transitions such as mortality, fertility and migration to project the components of

[1] http://www.ncess.ac.uk/

changes in future. Due to their aggregate nature, it is difficult to capture the individual characteristics in such models. MSMs, on the other hand, can model the impact on individual decision units from the changes in strategic planning or government policies. With the increasing complexity of such social and economic programmes, MSMs, with a used and tested record, have become an essential part of the decision making process, allowing identification of the magnitude of gains and losses from policy changes in such areas as taxation, social security, pensions and social services (Bourguignon and Spadaro 2006; Fredriksen et al. 2007). As MSMs can model the population and their past, present and future at the individual level (Wolf 2001), its usage has quickly spanned the whole spectrum of the social sciences to model complex social systems where individual characteristics are important.

However, MSMs require realistic micro-data and can be more difficult to apply in situations where appropriate data are not available. For example, in many applications of MSMs to problems involving taxation, housing or the public finances, real individual records can be manipulated under alternative scenarios to evaluate the impact of policy alternatives. In a population MSM, the individual attributes and transition processes are typically driven by probabilities from aggregate data or samples, and this can make it less flexible in modelling interactions and behaviours of various population groups that occur in the real world. Agent-based models (ABMs) can strengthen our understanding of the interactions between people and their behaviour by modelling demographic processes through interactions with other agents and/or the environment that they live in and model actions according to their unique built-in rules of behaviour. The flexibility of being able to construct heterogeneous agents and their rules makes it particularly helpful when there is a knowledge gap or data are unavailable (Axtell 2000; Epstein 1999). Crooks and Heppenstall (2012) present a useful overview of ABMs.

The MoSeS hybrid model attempts to combine the strength of a dynamic spatial MSM and an ABM to simulate discrete demographic processes at the level of an electoral ward to project the individuals into the future from the year 2001 to 2031.

16.3 Methodology

16.3.1 Microsimulation Models (MSMs)

MSMs can model the impact on individual decision units from the changes in strategic planning or government policies. Referring to the probabilistic generation of artificial data on an individual basis of the kind that would be observed in real life, MSM techniques have several advantages over the macrosimulation approach (van Imhoff and Post 1998). It allows more factors to be taken into account in demographic processes, and it is possible to model the interactions between individuals and to incorporate continuous covariates that are not possible in the macrosimulation approach (Siegel 2002). It can also produce data with much richer outputs and yield probabilistic results allowing confidence intervals to be created for the resulting projections. MSMs are able to generate individual event histories that are fully consistent with a set of transition intensities (probabilities) and produce estimates of the full distribution of an outcome, in addition to the expected value that is produced analytically by most models. They are also particularly useful where the projections are produced to answer a “what if” question such as the effect of a policy on a demographic characteristic (Wittenberg et al. 1998).

However, social simulation models are normally targeted to analyze complex social outcomes, i.e. macro outcomes that strongly depend on systemic processes of interactions between individuals that are co-located within a given environment. Microsimulation and macrosimulation are also recognized in previous research as alternative methods for making similar statements about the future. Despite its power, the MSM approach is also complex with onerous data requirements. It relies on quality microdata to produce realistic results, especially for important demographic transitions. When suitable data are not available, it will struggle. MSMs are also more prone to error than macrosimulation due to the sources of randomness such as the extra sources of error from the rich attributes of the individuals, a degree of inherent randomness in the Monte Carlo simulation process and the randomness in the initial population samples (van Imhoff and Post 1998; Wilson and Rees 2005). Some researchers continue to pursue macro-micro linkages, either in linking general equilibrium macro-economic models to MSMs (Aaberge et al. 2007) or allowing behavioural responses in MSMs. Among them, the MicMac project aims to offer a bridge between aggregate projections of cohorts (Mac) and individual cohort members (Mic) (Willekens 2005). Mac and Mic both use the same set of aggregated transition rates, but extra attributes in addition to age and sex have been introduced into Mic to address demographic events and other life transitions at the individual level (Gampe et al. 2009). Mac focuses on transitions among functional states by age and sex to produce cohort biographies, while Mic addresses demographic events and other life transitions at the individual level using a multistate model to produce individual biographies (Willekens 2005).

However, the difficulty of incorporating behavioural responses into a pure MSM becomes clear in practice. The fundamental dilemma of the behavioural vs. probabilistic model has been experienced (O’Donoghue 2001).

16.3.2 Agent-Based Models (ABMs)

More recently, the ABM approach has been used in various areas of social science, as agents in an ABM seem to be able to naturally mimic human population. A definition is provided by Axtell (2000): “An ABM consists of individual agents with states and rules of behaviour. Running such a model is simply creating a population of such agents and letting agents interact, and monitoring what happens”. Typically, individual agents can move around and interact with each other and the environment that they live in according to their built-in rules. They can also store information gathered through such movements and interactions and make decisions or act upon it.

One of the most useful features of ABMs, as suggested by Epstein (1999), is that it demonstrates that a sufficient microspecification can generate a macrostructure of interest. Through the repetition of application interaction rules for individual agents at the microscopic level, a target macrostructure can be effectively attained. As ABM tries to bridge the micro and macro by identifying microspecifications that are sufficient to generate the macroscopic picture in a robust and replicable manner, it also allows us to study the micro-to-macro mapping.

Furthermore, each individual/society is multi-dimensional and such dimensions can often not be considered separately in the real world, unlike the artificial division of different disciplines of academic social science. The ABM approach reflects the complexity of social systems and provides powerful new forms of hybrid theoretical computational work, especially in studies of non-equilibrium systems. In fact, ABM "invites the interpretation of society as a distributed computational device, and in turn the interpretation of social dynamics as a type of computation. This interpretation raises important foundational issues in social science – some related to intractability, and some to undecidability proper". For instance, "it is not obvious how we can formulate (or solve) the equations to represent large populations of discrete heterogeneous agents coevolving on a separate space, with which they interact" (Epstein 1999).

In demography, there is a poor level of precision in theoretical constructions. Quite often it not only lacks sufficient theory for applications of statistical models and data collection, there is also insufficient accounting for observability of important quantities used in the theory. Therefore ABM is very helpful for problems where "writing down equations is not a useful activity" (Billari et al. 2003). Despite such useful features that can benefit demographic studies, especially for those who are interested in understanding demographic behaviour, limited use of ABMs has been made in the area of demography. However, there is a stream of researchers who believe that the use of ABM in demography is of interest to all scientists interested in studying demographic behaviour, as well as to computer scientists and modellers who are looking for a promising field of application. Different to the approach based on statistical analysis of behavioural data that aims to understand why specific rules are applied by humans, ABMs pre-suppose (realistic) rules of behaviour and try to challenge the validity of these rules by showing whether they can or cannot explain macroscopic regularities. They argue that in order to study human populations, agent-based approaches are particularly useful from various theoretical perspectives, and as a promising stream of research, agent based approaches can improve our understanding of demographic behaviour (Billari et al. 2003).

16.3.3 The MoSeS Hybrid Approach

A hybrid modelling approach that combines the strength of both MSMs and ABMs has been adopted to capture the individual characteristics in the simulation model. Microsimulation in the hybrid model provides the capability to work with large scale data through the list processing power and the theoretical basis of dealing with real microdata and empirical data, the macro consequences of a certain behaviour from in depth studies, as well as the analysis of the impact of policies through a predefined group of transition rates, which drives the simulation.

However using the pure MSM, it is difficult to model the movements, interactions and behaviours of individuals or sub-populations due to its statistical nature and central controlled structure, as well as a lack of appropriate data on important transitions. Therefore a hybrid approach combining MSM and ABM techniques is tested to provide the flexibility and practical solution for modelling the subtlety of population changes and the heterogeneous behaviour of the individuals among a large population with rich details. This is less well studied and lacks an appropriate theoretical basis in previous MSMs. The use of ABMs also provides us with the ability to introduce heterogeneous agents in the model whose distinctive behaviours are not necessarily mathematically tractable (Billari et al. 2003).

ABMs mimic the human population naturally in the way that individuals can move around and interact with each other and/or the environment where they live. They can also have some intelligence according to their built-in rule bases where memories/changes can be stored for future decisions/actions. Within an ABM, individuals in the hybrid model are not socially isolated. On the contrary, they demonstrate the autonomy and interdependency at the same time. Thus this hybrid approach allows us to better understand social/demographic processes such as migration and marriage, where the movements, interactions and behaviours are playing an important role. The feature discussed above also provides a way to experimentally test hypotheses on decision processes and behaviours at an individual level.

16.4 Creating Agents from a Microsimulated Population

16.4.1 MoSeS Model Description

MoSeS is a dynamic spatial MSM that simulates discrete demographic processes at a ward level and projects the individuals into the future from the year 2001 to 2031. At the core of the MSM is the population model consisting of six modules, which model the following demographic processes: ageing, mortality, fertility, health change, marriage and migration. It models the demographic lifecycle at an annual interval. As a dynamic MSM, this model uses dynamic ageing and the simulation is driven by probabilities applied to each individual in each demographic transition at each simulation step. These transition probabilities are underpinned by a variety of data sources, most notably the ‘British Household Panel Survey’ (BHPS),[2] a longitudinal survey of individuals and households which is enormously rich in its

[2] http://www.esds.ac.uk/longitudinal/access/bhps/L33196.asp

demographic, social, economic and behavioural content. As a spatial MSM, all probabilities used in the MSM are not only age and gender specific, but also spatially disaggregated at the ward level.[3] Such probabilities have been calculated to reflect the individual and local difference, using a combination of both individual-level sample records and aggregate small area data from the UK census. The probabilities also reflect other vital factors regarding specific demographic processes. The transition of population status, movements and interactions of individuals are the focus in the model.

MoSeS models the individuals within the households, where they interact with the rest of the world through interactions with other people and the environment that they live in. Although the studied population is modelled as individuals, there is an interdependency between the household, individual and environment. The attributes of individuals, households and their environment can change due to various interactions between the individuals and: (a) other individuals; (b) households; or (c) the area that they live in. For example, during the process of marriage, the formation of a new household will result in changes in at least one individual's location; for households, this will result in changes in both existing and new households; for the areas that they used to/are going to live in, it will see changes in both local housing and the local population. Similar changes will be experienced in migration processes, too. Due to this interdependency, the operation of these demographic processes of individuals also leads to the formation and dissolution of households during the simulation.

MoSeS uses an ABM for some demographic processes to allow more flexibility and practicality where individual movements, interactions and behaviours are playing an important role. The MoSeS ABM naturally complements its spatial dynamic MSM components, as MoSeS agents reflect four of the most important features highlighted by Epstein (1999): heterogeneity, autonomy, explicit inclusion of space and bounded rationality:

Firstly, MoSeS agents have a rich portfolio of attributes from various data sources of human population samples and such characteristics change over time. Secondly, there is no central control over behaviours of individual MoSeS agents, and each agent acts autonomously according to their own rules. Thirdly, all events happen in an explicit space in MoSeS and local interactions are modelled typically through agents' interactions with others or their local spatial environments. Finally, MoSeS agents only make decisions/take actions according to simple rules that are based on local information. They do not have global information or infinite computational power. Such features of the MoSeS agents are "especially powerful in representing spatially distributed systems of heterogeneous autonomous actors with bounded information and computing capacity who interact locally" (Epstein 1999). Such a description of the system reflects the characteristics of a dynamic spatial population model extremely well.

[3] A ward is an administrative unit varying in size from 2,000 to 10,000 households. They are typically more populous in urban areas than in the countryside.

Fig. 16.1 Map of the wards of Leeds (Source: Generated by the authors using 2001 Census statistics and ward boundaries. Crown Copyright 2003)

Agents have been created from the microsimulated population to model the heterogeneous migration behaviours and the impact of migration histories in two experiments. Student migration and the impact of personal migration history on mortality will be described in the following sections. We use Leeds as an example (Fig. 16.1), which is a city in northern England with a population of approximately 760,000, but the MSM is generalizable between local areas across the UK.

16.4.2 Student Migration

Migration is a complex demographic process where individual movements, interactions and behaviours are of obvious importance (Champion et al. 2002). Using an ABM, individual activities and diversity of migration decisions leading to observed, complex migration patterns can be simulated in detail. Some attempts have produced fruitful outcomes (Espindola et al. 2006; Makowsky et al. 2006). In this chapter we investigate the usefulness of a hybrid modelling approach in a series of experiments where the MSM is combined with an ABM. We experimented on three important properties of agents: their ability to interact with their environment; to interact with others and to carry their own personal history around; and to use the history as a reference for current/future decisions/actions.

We chose student migrants as the sub-population in our first experiment due to their distinctive migration patterns, as well as the fact that in a large UK city such as Leeds, the frequent local migrations of approximately 30,000 university students make an important impact on local population structures. Previous studies have recognized the difficulty of modelling student migration in small areas (Champion et al. 2002). Students do not only move frequently and they also exhibit a distinctive pattern from the rest of the migrants, i.e. they tend to reside only in certain areas, mostly around universities during the period of their study. Most then leave while other new students move in. Due to the replenishment of the student population each year, the population in such wards stays younger than in other wards. As an ABM is very flexible in terms of constructing heterogeneous agents with different rules, we experiment with an ABM for the student migration process using hypothetical rules in a similar way to Schelling's model on individual decisions to move and the composition of neighbourhoods (Schelling 1971).

In the MoSeS ABM, we identify four types of agents/students: first year undergraduates, year 2 and 3 undergraduates, masters students and doctoral students. Based on the above assumptions, we then apply some general rules to the students in terms of the areas they stay in and the duration of their stay. Specific rules for individual groups vary according to their specific characteristics. For example, a year one undergraduate student will choose to stay in available accommodation on campus and then move into off-campus accommodation where they can stay in the city for two more years. They may then study a masters degree for one more year or leave. A masters student can then stay in the area for one year and leave or continue with doctoral study for three more years. The typical interaction between the agents/students themselves would be to look for fellow students to live within a certain area, and the interaction with the environment would be checking if there is a vacancy in that area. An agent/student keeps searching in areas where their fellow students live until they find a vacancy. The agent without success at the end of the search stays in the current area.

The simulation results of the student population using the pure MSM and using the ABM approach have been compared to the observed distribution of the student population. Through the implementation of simple rules at the individual level, the heterogeneity of student migration is captured in the model. Instead of students almost evenly scattering around the whole city in the MSM, the hybrid model provides a much better reflection of the observed student population concentration around the city centre, close to a university. More importantly, as new students come into the area each year, most existing students leave upon completion of their study in the hybrid model. Students are no longer ageing with the rest of the population in the suburban areas as in the pure MSM (Fig. 16.2). The number of students in wards is indicated by the shade of colour, the darker colour indicates more students in areas. Further details and discussion of these model results can be found elsewhere (Wu et al. 2008).

16.4.3 Mortality

Simulating geographically identified populations can demonstrate the local (environmental, economic, etc.) impact on the individuals to a degree. However, sometimes not only the current locations, but where individuals came from or used to live also

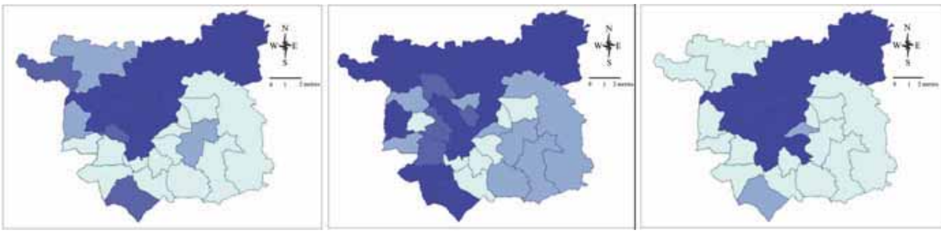


Fig. 16.2 Student migration in small areas 2001 and 2031: observed, pure MSM and hybrid results. Darker areas indicated higher migration (Source: Generated by the authors using 2001 Census statistics and ward boundaries. Crown copyright 2003)

contributes to individual heterogeneities. For instance, if a person has worked as a miner all his life, his/her mortality/morbidity rates should not suddenly change a great deal just because he/she retired to a pleasant residential area. An ABM can complement the MSM by retrieving personal histories with great ease. In this experiment, we explored three scenarios of mortality projections based on the impact of: the current residence location, the first residence location in the system/birth places and the mortality dependent on personal migration histories. In the first scenario, all individuals are simulated in the MSM. Their survivals are determined against an age, sex and location specific mortality probability generated on the basis of local information about the current location of residence. An ABM approach is used in the second and third scenarios, where agents carry their own histories along with them and have to check on such histories to determine their mortality probabilities. In the second scenario, the survivals are determined on the basis of the mortality rates of individual first residence location/birth places. In the third scenario, we tried to model the impact of personal migration history on mortality. Mortality is projected on the basis of the mortality rates of the area where the individual stays the longest to date.

In the experiments, the whole population of Leeds has been simulated under the three assumptions for 30 years separately and the results of the year 2031 are then compared spatially to assess the difference in the mortality distribution within the city. Although the distribution pattern of mortality is similar on the whole, the experiments still reveal some interesting variations in small areas. Unsurprisingly, the mortality projection based on origins in the second scenario is more different from the projection based on current locations in the first scenario, while the projection based on individual migration history in the third scenario demonstrates more similarity with the first scenario. However, interestingly, we can see from the map that there tends to be a higher mortality in the more established suburban wards in the northern area of the city in the first residence based projection compared to the current residence. This reflects the fact that new migrants may take time to absorb the benefits of favourable demographic conditions in the robustly healthy and affluent northern suburban areas. In comparison, the majority residence based projection indicates a reduction of mortality compared with the current residence based projection in the northern suburban areas, but indicates an increase of mortality in the traditionally less affluent areas in the south and eastern areas around the city centre. Such findings demonstrate that personal history could have an important impact on mortality (see Fig. 16.3).

Fig. 16.3 Mortality projections in small areas 2001 and 2031: by current, first and majority residence. Darker areas indicated higher mortality (Source: Generated by the authors using 2001 Census statistics and ward boundaries. Crown copyright 2003)

Empirical research on the relationship between limiting long-term illness and migration established that the illness status of migrants is mid-way between that of their origin and destination locations (Norman et al. 2004). If this finding also applies to mortality, then a combination of all three scenarios may be needed to represent the mortality chances of migrants properly. We will continue to improve the mortality projection in the light of such evidence. Needless to say, the migration histories of international migrants may also have a considerable impact on their health and mortality compared to the local people. Although the mortality experiments discussed here are purely based on hypothesis (as a result of lack of data), it demonstrates that there are many more aspects of the population MSM that can be strengthened through the use of personal history. Important elements of the model such as marriage behaviour, fertility patterns and change in health status might all benefit in a similar way (e.g. the recent trends of adopting the Total Fertility Rate (TFR) approach, marriage behaviour dependency on previous marital status and health on previous health history). Such explorations are not only just interesting experiments, but can potentially play a vital role in facilitating decision making where the impact of personal history should be taken into account.

16.5 Conclusions

In this paper, we introduced the hybrid approach used in an ambitious social model (MoSeS) which combines the strength of both a MSM and ABM to model heterogeneity in a complex social system. MSM in the hybrid model provides the capability to work with large scale data through the list processing power and the theoretical basis of dealing with real datasets, the macro consequences of certain behaviour from in-depth studies, as well as the analysis of the impact of policies through a predefined group of transition rates, which drives the simulation. With the four important features of heterogeneity, autonomy, explicit inclusion of space and bounded rationality, the MoSeS agents in the hybrid model naturally mimic human populations and complement the spatial dynamic MSM components. The hybrid model provides a powerful way to model the local interactions in spatially distributed systems of heterogeneous autonomous actors with bounded information and computing capacity.

In the two experiments we described above, we deliberately chose some straightforward processes and applied ABM with considerable simplification. However, the modelling of interaction, behaviour and personal history impact can be significantly more complex with more intelligent agents. For example, people can have much more complex interactions with others and their environment to make a decision or take an action, assisted by information from their personal histories or requirements during the household formation process. The hybrid approach can also present opportunities to recent demographic modelling trends such as using total fertility rate (TFR) instead of age-specific fertility rates (ASFRs), as "maternal" agents can easily track the history throughout their childbearing lives. The examples discussed in this chapter are simply used to demonstrate the potentials of the hybrid model.

As a demographic planning tool, MoSeS can monitor the evolution of population structures and various demographic changes on a fine geographical scale. This provides vital information for demographic planning/policy making (especially location-based policies). MoSeS can also benefit other public policy making or public service planning. For instance, the ageing trends in certain suburban areas may promote changes in the health service and public transportation provision in order to enable easy access to such services for the old population in the area. The rich attributes captured in the system are also very useful in various policy analyses or research purposes.

The MoSeS hybrid model has provided a framework to enable the effective modelling of heterogeneous decision making units on a large scale, as well as adding the flexibility to introduce different modelling techniques to strengthen various aspects of the model. The model itself provides a useful tool in assisting decision making, exploring various "what if" situations and testing different hypotheses. As we have discussed above, the hybrid modelling approach demonstrates great potential in demographic modelling and we will continue trying to improve various aspect of this model using this approach to provide a better groundwork for more wide-ranging social science studies.

Acknowledgments This research has been funded by the Economic and Social Research Council through the National Centre for e-Social Science (MoSeS RES-149-25-0034 and GENeSIS RES-149-25-1078).

The British Household Panel Study data were originally collected by the Economic and Social Research Council Research Centre on Micro-social Change at the University of Essex, now incorporated within the Institute for Social and Economic Research. Neither the original collectors of the data nor the UK Data Archive bear any responsibility for the analyses or interpretations presented here.

2001 Census: Special Licence Household Sample of Anonymised Records (SL-HSAR) were deposited by the University of Manchester, Cathie Marsh Centre for Census and Survey Research. Although all efforts are made to ensure the quality of the materials, neither the original data creators, depositors or copyright holders, the funders of the Data Collections, nor the UK Data Archive bear any responsibility for the accuracy or comprehensiveness of these materials.

References

Aaberge, R., Colombino, U., Holmøy, E., Strøm, B., & Wennemo, T. (2007). Population ageing and fiscal sustainability: Integrating detailed labour supply models with CGE models. In A. Harding & A. Gupta (Eds.), *Modelling our future: Population ageing, social security and taxation*. Amsterdam: North-Holland.

Axtell, R. (2000). Why agents? On the varied motivations for agent computing in the social sciences. In *Proceedings of the workshop on agent simulation: Applications, models and tools*. Argonne: Argonne National Laboratory.

Billari, F., Ongaro, F., & Prskawetz, A. (2003). *Agent-based computational demography: Using simulation to improve our understanding of demographic behavior*. London/Heidelberg: Springer/Physica.

Birkin, M. A., & Wu, B. (2012). A review of microsimulation and hybrid agent-based approaches. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 51–68). Dordrecht: Springer.

Bourguignon, F., & Spadaro, A. (2006). Microsimulation as a tool for evaluating redistribution policies. Working paper 2006–20, Society for the Study of Economic Inequality, (available from www.ecineq.org).

Champion, T., Fotheringham, S., Rees, P., Bramley, G., et al. (2002). *Development of a migration model*. London: Office of the Deputy Prime Minister and the Stationery Office.

Crooks, A. T., & Heppenstall, A. J. (2012). Introduction to agent-based modelling. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 85–105). Dordrecht: Springer.

Epstein, J. M. (1999). Agent-based computational models and generative social science. *Complexity, 4*(5), 41–60.

Espindola, A. L., Silveira, J. J., & Penna, T. J. P. (2006). A Harris-Todaro agent-based model to rural–urban migration. *Brazilian Journal of Physics, 36*(3A), 603–609.

Fredriksen, D., Heide, K. M., Holmøy, E., & Solli, I. F. (2007). Macroeconomic effects of proposal pension reforms in Norway. In A. Harding & A. Gupta (Eds.), *Modelling our future: Population ageing, social security and taxation*. Amsterdam: North-Holland.

Gampe, J., Zinn, S., Willekens, F., Gagg, N., van der, de Beer, J., Himmelspach, J., & Uhrmacher, A. (2009). The microsimulation tool of the MicMac-Project. In *Proceedings of the 2nd general conference of the international microsimulation association*, Ottawa, 8–10 June 2009.

Gustafsson, L., & Sternad, M. (2007). Bringing consistency to simulation of population models: Poisson simulation as a bridge between micro and macro simulation. *Mathematical Biosciences, 209*(2), 361–385.

Makowsky, M., Tavares, J., Makany, T., & Meier, P. (2006). An agent-based model of crisis-driven migration. In *Proceedings of the complex systems summer school 2006*. Santa Fe: Santa Fe Institute.

Norman, P., Boyle, P., & Rees, P. (2004). Selective migration, health and deprivation: A longitudinal analysis. *Social Science and Medicine, 60*(12), 2755–2771.

O'Donoghue, C. (2001). Dynamic microsimulation: A methodological survey. *Brazilian Electronic Journal of Economics, 4*(2).

Rees, P. (2009). Population: Demography. In N. Thrift & R. Kitchen (Eds.), *International encyclopedia of human geography*. Oxford: Elsevier. Publication in summer 2009.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology, 1*, 143–186.

Siegel, J. (2002). *Applied demography*. London: Academic.

Turchin, P. (2003). *Complex population dynamics: A theoretical/empirical synthesis*. Princeton, NJ: Princeton University Press.

van Imhoff, E., & Post, W. (1998). Microsimulation methods for population projection. *Population: An English Selection, 10*, 97–138.

Willekens, F. (2005). Biographic forecasting: Bridging the micro–macro gap in population forecasting. *New Zealand Population Review, 31*(1), 77–124.

Wilson, T., & Rees, P. (2005). Recent developments in population projection methodology: A review. *Population, Space and Place, 11*, 337–360.

Wittenberg, R., Pickard, L., Comas-Herrera, A., Davies, B., & Darton, R. (1998). *Demand for long-term care: Projections of long-term care finance for elderly people*. London: PRRRU.

Wolf, D. A. (2001). The role of microsimulation in longitudinal data analysis. *Canadian Studies in Population, 28*, 165–179.

Wu, B. M., Birkin, M. H., & Rees, P. H. (2008). A spatial microsimulation model with student agents. *Computers, Environment and Urban Systems, 32*, 440–453.

Chapter 17
Designing, Formulating, and Communicating Agent-Based Models

Volker Grimm and Steven F. Railsback

Abstract Designing, formulating, and communicating agent-based models (ABMs) poses unique challenges, especially how to choose the right level of complexity and how to describe – and, even, think about – a model in a way that captures both its essential characteristics and its complete detail. Two techniques for dealing with such challenges have become established among agent-based modellers. The "ODD" (Overview, Design concepts, Details) protocol is a standard for describing ABMs in publications, but also provides design patterns for the model developer. ODD starts with an overview of what the model is and does, and then describes how the model implements ten "design concepts" that capture essential and unique characteristics of ABMs. Last come all the details needed to completely replicate the model. "Pattern-oriented modelling" (POM) is a set of strategies for using patterns observed in the systems to ensure that an ABM captures the right "essence" of the system. POM starts with identifying multiple patterns of behaviour in the real system and its agents that seem to capture the essential internal mechanisms for the problem being modelled. These patterns are then used to decide what kinds of entities, state variables, and processes need to be in the model; compare and test alternative "theory" for key agent behaviours; and filter potential parameter values to limit uncertainty. ODD and POM are important steps toward the acceptance of agent-based approaches as established, credible ways to do science.

V. Grimm (✉)
Department of Ecological Modelling, Helmholtz Center for Environmental Research - UFZ, Leipzig, Germany
e-mail: volker.grimm@ufz.de

S.F. Railsback
Department of Mathematics, Humboldt State University, Arcata CA, USA
e-mail: steven.railsback@humboldt.edu

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_17,

17.1 Introduction

Agent-based models (ABMs) can be hard to communicate. They usually include different types of agents and spatial units, each distinguished by a suite of state variables, attributes, behaviours, and environmental processes. Model results can depend strongly on how model entities are initialised, how processes and events are scheduled, the data used to drive the simulated environment, and the details of how submodels represent model processes. As a consequence, many, if not most, descriptions of ABMs are lengthy but nevertheless incomplete.

Communication is, however, part and parcel of using ABMs as a scientific tool. Models are the "Materials and Methods" used by modellers to obtain scientific results. Thus, as with any other method used in science, the very feature that renders an ABM "scientific" is replicability: that the same results can be obtained by peers if they use exactly the same materials and methods. But most published ABMs would be impossible, or at least challenging, to replicate. This situation cannot be tolerated because it undermines the credibility of ABMs as a scientific method. Moreover, because ABMs are increasingly built to support real-world decision making, model structure and assumptions need to be transparent to decision makers; otherwise the models are likely to be (justifiably) ignored, or used inappropriately.

Incomplete and inefficient communication is linked to a second challenge of agent-based modelling: the lack of a framework for designing ABMs. Current practice is that most ABMs are developed from scratch and that the choice of model structure and process representation is more or less *ad hoc*. Model design and formulation often reflect the disciplinary background, experience, and personal preferences of the modeller more than general principles of model design that would lead to similar model designs for similar problems.

Lack of efficiency and coherence in design and communication is to be expected for an emerging scientific approach. But agent-based modelling is no longer in its infancy: hundreds of ABMs have proven the potential of this approach, both in theory and application, in a wide array of disciplines; many research projects and programs dealing with complex systems include ABMs in their portfolio of approaches; and the issues of standardisation, communication, and transparency have increasingly been addressed over the last 10 years.

There are several examples of the drive for a general framework for ABMs. Railsback (2001) listed design concepts that are important for designing ABMs, in particular when adaptive behaviour is included. Grimm and Railsback (2005) devoted their book on individual-based modelling in ecology to making this approach more "coherent and efficient". A series of workshops addressed the issue of model replication (e.g., Hales et al. 2003). Standards for model development (Richiardi et al. 2006) and communication were proposed (Grimm et al. 2006, 2010), and a community has been established with the mission to "foster the agent-based modelling development, communication, and dissemination for research, practice and education" (Janssen et al. 2008): the Open Agent Based Modelling Consortium (www.openabm.org).

Standardisation is an indicator for a maturing approach: the approach has been used enough to understand its strengths and weaknesses, what decisions have to be made when using it, and how standards can make its use more efficient and coherent. Standards must of course avoid restricting the approach's flexibility and usefulness.

Here we present the ODD protocol, a standard for communicating and formulating ABMs; and pattern-oriented modelling (POM), a general strategy for tying model structure to multiple observed patterns to make model design and analysis less *ad hoc*. Both approaches were formulated in ecology but are relevant in any discipline using ABMs. Since both ODD and POM have been described in detail elsewhere, we focus here on what these approaches are, how they can be used, and what benefits they provide to the agent-based modeller.

17.2 The ODD Protocol

The ODD protocol was developed as a standard format for describing individual- and agent-based models (Grimm et al. 2006; Grimm and Railsback 2005). The acronym stands for the three blocks by which ODD's elements are grouped: Overview, Design concepts, Details. ODD is designed to be used for *all* ABMs, independent of their domain, purpose, complexity, and computer implementation. The main idea of ODD is to present ABMs in a hierarchical way: an overview of model structure and processes first, with details on processes last. Between overview and details is a checklist that explains how important general concepts for designing ABMs have been taken into account. This checklist ensures that modellers make important design decisions consciously and that readers understand how and why certain design decisions were made.

Using ODD implies referring to this standard explicitly by stating: "The model description follows the ODD (Overview, Design concepts, Details) protocol for describing individual- and agent-based models (Grimm et al. 2006, 2010)" and then presenting the seven elements of ODD (Table 17.1) using exactly the given sequence and element labels. By using ODD always in this way, writers do not waste time figuring out how to organise their model description and readers avoid frustration by knowing exactly where to find information about a model. Moreover, understanding of a model is greatly facilitated by first providing an overview of its structure and processes. For most readers, this information is sufficient to grasp the overall rationale of the model and to understand its results. Readers hooked by the overview can go into more detail and see how certain processes have been represented. Finally, an ODD description should be complete enough to allow the model's replication.

For complex models, the Details and possibly also the Design concepts parts may be too long, so they can be moved to an electronic appendix, or the full ODD model description could be published in a separate report.

Table 17.1 The seven elements of the ODD protocol (See also the ODD webpages at www.ufz.de/oesatools/odd)

ODD	ODD element	Questions to be answered
Overview	Purpose	What is the purpose of the model?
	Entities, state variables, and scales	What kind of entities are in the model? By what state variables, or attributes, are these entities characterised? What are the temporal and spatial resolutions and extents of the model?
	Process overview and scheduling	Which entities do what, in what order? When are state variables updated? How is time modelled — as discrete steps or as a continuum over which both continuous processes and discrete events can occur?
Design concepts	Design concepts	There are ten design concepts (see text). How have these concepts been taken into account in the model's design?
Details	Initialisation	What is the initial state of the model, i.e. at time $t=0$?
	Input data	What input does the model use from external sources such as data files or other models to represent processes that change over time?
	Submodels	What, in detail, are the submodels that represent the processes listed in "Process overview and scheduling"? What are the model parameters, their dimensions, and reference values? How were submodels designed or chosen, tested, and parameterised?

17.2.1 Design Concepts

The design concepts of ODD were first summarised by Railsback (2001), then discussed in detail in Grimm and Railsback (2005), and most recently updated by Grimm et al. (2010). The ten concepts are not actually needed to replicate a model, but they are important for communicating the essence of how (and why) an ABM has been designed—because traditional formats such as equations and diagrams cannot capture the essence of most ABMs. For example, the first design concept is "emergence". Its discussion should explain whether key model results are imposed, or whether instead they emerge from agent behaviours and other processes – information that is the key to understanding the rationale of the model. Here "imposed" means that model rules and equations force the model to behave in a predictable way. For example, if a model is designed so that an event like a traffic jam occurs with a constant probability of 5%, then it is no surprise that on average the model produces traffic jams 5% of the time. If, however, traffic jams emerge from the behaviour and decisions of the agents and the roads they follow, there is no simple way to predict how often traffic jams occur: we have to run the model and see what emerges.

In many descriptions of ABMs, it is not entirely clear what the authors wanted to emerge and what they imposed. Similar "ad hoceries" often occur with the other design concepts. The following description of the concepts is adopted from Grimm et al. (2010):

- *Emergence*. What key outputs of the model are modelled as emerging from the adaptive behaviour of its agents? Are there other outputs that are more tightly imposed by model rules and hence less dependent on what individuals decide to do?
- *Adaptation*. What rules do agents have for changing behaviour in response to changes in themselves or their environment? Do these traits explicitly aim at increasing some measure of individual objectives or success? Or do they instead cause individuals to reproduce observed behaviours that are implicitly assumed to convey success?
- *Objectives*. If adaptive behaviour is represented as explicitly seeking some objective, what is the objective and how is it measured? Examples of "objectives" are "fitness" for organisms, "utility" for economic reward in social models, or simply "success".
- *Learning.* Do individuals change their adaptive behaviour over time as a consequence of their experience? How?
- *Prediction*. To make decisions, model agents often need to predict future consequences of their alternatives. What internal models are used by the agents to estimate future conditions or consequences of their decisions? What "tacit" or hidden predictions are implied in these internal models?
- *Sensing*. What information (state variables of other model entities and themselves) can agents sense and consider in their adaptive decisions? Are the mechanisms by which agents obtain information modelled explicitly, or are agents simply assumed to "know" these variables?
- *Interaction*. What kinds of interactions among agents are in the model? Are there direct interactions, or are the interactions indirect, e.g. via competition for a mediating resource? How do agents interact with their environment?
- *Stochasticity*. What processes are modelled by assuming that they are random or partly random? Why is stochasticity used – to represent variability in a simple way, or to cause events or behaviours to occur with a specified frequency?
- *Collectives*. Are there aggregations of agents that affect, and are affected by, the agents? Examples include social groups, fish schools and bird flocks, human networks and organisations, or cells constituting an organ. Are collectives represented as emergent properties of the agents or as a separate kind of entity with its own state variables and traits?
- *Observation*. What data and patterns must be observed from the ABM for testing, understanding, and analyzing it, and how are they collected?

ODD model descriptions should describe how these design concepts were taken into account. It is possible to leave out some of the concepts if they are not included in the ABMs design at all (e.g., if adaptive behaviour is simple, then Objectives, Learning, and Prediction may be irrelevant). Four concepts are relevant to almost every ABM: emergence, interaction, stochasticity, and observation.

17.2.2 Examples

As a simple example, we apply the ODD protocol to the "Segregation" model published as part of NetLogo (Wilensky 1997), which was inspired by the work of T. Schelling (e.g., Schelling 1971).

Purpose. The model addresses segregation of households in cities: why members of two different groups tend to separate into different neighbourhoods. The model explores the relationship between how tolerant individuals are of the opposite group and how segregated neighbourhoods are, when individuals express intolerance by moving.

Entities, state variables, and scales. The model entities include mobile agents that represent households, and square grid cells that represent houses. Households are characterised by their location (which house they occupy) and their colour, which is either blue or red. Households also have a state variable "happy?", a boolean variable set to false if the household is intolerant of its neighbours. The grid cells make up a square 51×51 cells in extent, with no depiction of roads or other spaces between them. The space is "toroidal": if an agent moves off one edge of the space, it reappears on the opposite edge; and the neighbours of a household on one edge of the space include those on the opposite edge.

The length of a time step is unspecified but represents the time in which a household would decide whether to move. The number of time steps in a model run is an emergent outcome: the model runs until no households are intolerant of their neighbours and want to move (see submodel *move*).

Process overview and scheduling. The following actions are executed, in this order, once per time step.

- If no households are intolerant of their neighbours ("happy?" is true for all households), then the model stops.
- The households who are intolerant of their neighbours ("happy?" is false) execute the submodel *move*. The order in which these households execute is randomly shuffled each time step.
- All households update their "happy?" variable (see submodel *update*).
- Outputs for system-level results are updated.

Design concepts. The key outcomes of the model are segregation patterns, especially how strongly segregated the entire system is; these outcomes *emerge* from how tolerant households are to unlike neighbours. The households' *adaptive behaviour* is to move when their *objective* – to live in a neighbourhood with the fraction of unlike neighbours below their tolerance threshold – is not met. The behaviour does not involve learning, or prediction other than the implicit prediction that moving will lead to a neighbourhood where the tolerance objective is met. Households *sense* the colour of other households on the eight surrounding grid cells. *Stochasticity* is used in only two ways: to initialise the model so that it starts unsegregated; and to determine the new location of households when they move, because modelling the details of the movement is beyond the scope of this model. *Observations* include a visual display of which colour household is on each grid cell, and two numerical results: the mean (over all households) percent of neighbours of similar colour, and the percent of households with "happy?" false.

Initialisation. A user-chosen number (typically, 2000) of households are initialised. They are each placed on a random empty grid cell and given a colour randomly, with equal probability of red and blue.

Input data. The model does not use input from external models or data files.

Submodels. The submodel *move* is performed by individual households if they are not tolerant of their neighbours. The household chooses a direction randomly from a uniform real distribution between 0° and 360°, then moves forward a distance drawn randomly from a uniform real distribution of 0–10 grid cell widths. If there is already a household on the grid cell at this new location, the household repeats the move. If the new grid cell is empty, the household moves to its centre.

The submodel *update* is conducted by all households, to determine whether they tolerate their neighbourhood. The tolerance of households is determined by a parameter "%-similar-wanted", the value of which ranges from 0 to 100 and applies to all households. A household's "neighbours" are all households on the eight surrounding patches. The household's variable "happy?" is set to false unless the number of neighbours with the household's colour is greater than or equal to "%-similar-wanted" divided by 100 and multiplied by the number of neighbours.

Using ODD for such simple models may look overdone. However, the ODD model description has a clear hierarchical structure, different elements of the model are easy to find, and all information required for implementation is provided. Moreover, the design concepts provide some explanation of why the model was designed as it was.

An important benefit of ODD is that any ABM can be described in exactly the same format as the segregation model. ODD thus provides a unifying format. Consequently, using ODD to rewrite model descriptions made it much easier to compare three different models of land use/land cover changes (Polhill et al. 2008a): SLUDGE (Parker 1999; Parker and Meretsky 2004), SOME (Brown et al. 2004, 2005), and FEARLUS (Polhill et al. 2001; Gotts et al. 2003) with the ELMM extension (Polhill et al. 2005, 2008b).

Just identifying the entities and state variables of the three models made it easy to see the conceptual similarities of these models, but also their differences in complexity, resolution, and scope (Table 17.2). Obviously, SLUDGE and SOME are conceptually very simple and similar, whereas FEARLUS + ELMM can represent quite complex landscapes and scenarios. FEARLUS + ELMM could, however, probably be mapped to the two simpler models by choosing simplified parameterisations and initialisations. Model-to-model comparison, and thus transfer of questions, model designs, and results, is greatly facilitated by using ODD. This applies also to comparing process overviews and schedules, and design concepts (Polhill et al. 2008a).

Table 17.2 Comparison of the entities (*bold*) and their state variables (*itemized lists*) of three land use model (see text for references). For a full ODD description, in addition the precise number and type of state variables needs to be specified, for example for "biophysical characteristics" or "aesthetic quality" (See also Polhill et al. 2008a)

SLUDGE		SOME		FEARLUS + ELMM	
Land-scape cell	– Productivity – Externalities – Output price for land use 0 – Coordinates	**Land-scape cell**	– Aestethic quality – Coordinates	**Land parcel**	– Biophysical characteristics – Land use (of the past) – Yield (of most recent land use) – Owning land manager
Land owner	– Land use – Coordinates	**Resident**	– Relative importance of three different aspects of landscape cell – Utility – Coordinates	**Land manager**	– List of owned land parcels – Land manager's subpopulation – Land managers market – Account – Aspiration threshold – Memory size – Land offer threshold – Bidding strategy – Selection strategy
Market location	– Coordinates	**Service center**	– Coordinates	**Subpopulation**	– Land manager list – Incomer offer price distribution – Imitative probability distribution – Aspiration threshold distribution – Land offer threshold distribution – Bidding strategy distribution
Land-scape	– Dimensions			**Land market**	– Auction type – Parcels for sale list – Bid list
				Environment	– Spatial topology – External conditions – Array of probabilities of changes in external conditions – Break-even threshold – Possible land uses

17.2.3 ODD as Design Patterns for Formulating ABMs

Since 2006, ODD has been used for over 70 individual- and agent-based models. ODD has mainly been used by ecologists, but recently applications are appearing in other domains as well, e.g., microbiology (Hellweger et al. 2008), socio-ecology (Guzy et al. 2008), biomedical research (Galvão and Miranda 2009), and anthropology (Premo and Hublin 2009). After reviewing most of these applications of ODD and summarising the questions and comments from many ODD users, Grimm et al. (2010) revised ODD slightly but completely rewrote its description and explanation to make its use easier.

Reviewing these applications of ODD revealed unexpected benefits, including that ODD provides a general format for thinking about and designing ABMs. ODD makes us ask questions about model structure and design in a hierarchical way: structure and overview first, processes and details later. This facilitates translating the often confusing and vague conceptual models which are the starting point of any modelling process (Grimm and Railsback 2005) into model formulations. Scientists with no background in agent-based modelling found ODD helpful for understanding what would be involved to implement an ABM of their system and problem. Likewise, communication between students who develop ABMs and their supervisors is made much more efficient by using ODD. ODD also facilitates communication between different domains, which usually have their own styles of modelling and communicating models.

Last but not least, ODD corresponds to "design pattern" in software engineering (Gamma et al. 1994), which describe "recurring solutions to common problems in software design" (Wikipedia 2009). In agent-based modelling, ODD is a design pattern for formulating new models, and for re-formulating existing models.

17.3 Patterns and Stylized Facts

While ODD has proven very important and useful, it provides only a general structure for formulating ABMs. It cannot, by itself, prevent "ad hocery" in the choice of entities and state variables. ABMs can still be too simple or too complex to be useful for inferences about how their real counterparts are working. We need a strategy for increasing the chance that our models explain the internal organisation of real complex systems by guiding our choices of what to include in a model and in what detail.

Such a strategy emerged in ecology over the last 15 years: pattern-oriented modelling (POM). It is based on the notion that complex systems provide indicators of their internal organisation: patterns, signals, or regularities that cannot be explained by random processes and therefore call for an explanation (Heine et al. 2007). For example, if vegetation in semi-arid regions shows a characteristic banded structure (Fig. 17.1), we can take this pattern as an indicator of how vegetation is affected by,

Fig. 17.1 Banded vegetation pattern in semi-arid regions produced by a simple grid-based model (Thiery et al. 1995; model reimplementation courtesy of U. Berger). The 100×100 cells each have the size of a tree. *Dark cells* are covered by woody vegetation, *white cells* are not. Implicitly, the landscape is assumed to have a gentle slope so that runoff from rain flows from *top* to *bottom*

and affects, the distribution of water. Our task thus is to "decode" the patterns observed in the real systems. We do this inversely by trying to find models that produce the same patterns.

This approach is certainly not new but just the basic idea of science, or inference (Platt 1964). Patterns were key to revealing the internal organisation of atoms and the universe (spectral patterns), identifying quasars (periodic signals), and proposing asteroids as the cause of mass extinctions (unusual Iridium concentrations in geologic strata). However, one important and often neglected point in the science of agent-based complex systems is that a single pattern is rarely enough to decode the internal organisation. Many different models are, for example, able to reproduce cycles in the population dynamics of small mammals (Czárán 1998). How can we select the right model or falsify models that produce cycles for the wrong reasons?

The basic and simple idea of POM is to use multiple patterns, observed at different scales and hierarchical levels, as multiple criteria for selecting among alternative model structures, submodel formulations, and parameter sets. One single pattern might be relatively "weak" and contain little information so that still, say, half of the models considered would reproduce it. But by adding a second pattern, we possibly reduce the degrees of freedom in model structure. Thus, every single pattern serves as a filter. Again, this method of inference is old and regularly used by detectives who use motives, alibis, witness statements, and evidence to filter possible suspects. Or, in the famous story of the discovery of the structure of DNA (Watson 1968), X-ray diffraction patterns indicated a spiral structure, but many such structures were compatible with this patterns. Only after two additional patterns, Chargaff's rule and the geometry of purine and pyridimine bases, were taken into account was the real structure identified.

In social sciences and in particular in economics, the notion of "stylised facts" (after Kaldor 1961) corresponds to what we here mean by patterns. In the modelling literature of these domains, stylised facts often seem to be disregarded as too vague to use for rigorous inferences. The point of POM is, however, to increase inferential power by *combining* multiple vague, weak, or qualitative patterns or stylised facts (see also the discussion of 'middle range models' by Gilbert (2008, pp. 41–44)).

17.4 Pattern-Oriented Modelling

In POM, multiple patterns are used for three elements: designing model structure (what entities, state variables, and processes to include); selecting among alternative submodels; and determining parameter values.

17.4.1 Patterns for Model Structure

The key question of POM's first element is: what do experts who know the system well consider to be characteristic features, or essentials, of the system? For example, in growing cohorts of trees, competition among neighbouring trees increases as the trees grow. Initial differences in size and the trees' spatial distribution cause some trees to be more suppressed by competition so they die, which is called "self-thinning" in ecology. The resources used by the trees that die (space, light, nutrients) become available to survivors, who can grow further until competition is again strong enough that more trees die.

A characteristic pattern of this process is that the relationship between average tree biomass and tree density follows a power law: a straight line on a log-log graph. Moreover, the slope of this line is similar for virtually all plant species and environments. Until about 2000 it was believed that the slope has to be, for theoretical reasons, –3/2 but a new theory predicts a slope of −4/3 (Enquist et al. 1998).

The consequence of this pattern for modelling self-thinning was that models focussed on tree density and average biomass as state variables. Different models, however, can each produce a power law with the right slope. So, what other patterns characterise self-thinning? This question is not fully answered yet, but it seems that the spatial distribution of trees is quite regular during the entire process, which indicates that neighbour competition, depending on local tree density, is important.

To potentially explain this regular spatial distribution, a model should be individual-based so that space and local interactions can be represented. Trees might be represented just by their biomass, which corresponds to a certain "zone of influence" over which they interact with neighbours (Berger and Hildenbrandt 2000). No further patterns seem to be known yet that would suggest including further state variables of the trees, e.g. height, crown shape, etc. Interestingly, power laws are also found in the size distribution of cities, where probably similar mechanisms of neighbour competition are the underlying mechanism (Batty 2005).

To use POM to design a model structure, we identify multiple patterns, typically two to five, which seem to characterise the system and problem being modelled. Making it possible for the model to reproduce each pattern requires adding some state variables and processes. These make the model more complex but also rich enough in structure to be tested against not only one, but multiple patterns. The models are also "mechanistically rich" (DeAngelis and Mooij 2003) so that they can be checked for patterns that were not used for model formulation and calibration.

An example of a mechanistically rich model is the natural beech forest ABM of Rademacher et al. (2004), who found characteristic patterns in the age structure and spatial distribution of canopy trees in model results. These patterns were not at all considered during model formulation and verification, which were driven by other characteristic patterns, but agreed very well with observations (Rademacher et al. 2001). This discovery indicated that the model captures essentials of the system's internal organisation instead of just being fit to observations. One generic feature of structurally realistic models is that they reproduce observed patterns at different hierarchical levels simultaneously. They capture essential behaviours of both the system and its agents.

Quite a few ABMs have been oriented towards reproducing sets of observed patterns, but usually this is not stated explicitly. POM means to use multiple patterns systematically and explicitly.

17.4.2 Patterns for Selecting Submodels

ABMs usually comprise a set of submodels that describe the model entities' behaviour. The segregation model described above, for example, has two submodels, *move* and *update*. Formulating submodels includes many degrees of freedom. The decision of whether and where to move could, instead of being very simple and partly random, have been a complex algorithm using knowledge about distant locations and prediction of future changes in neighbourhoods. Submodels, like entire models, must compromise between being simple and unrealistic vs. more realistic and harder to understand. How much detail, and what detail, must we have to solve a particular problem?

The idea of POM for selecting submodels is to use the entire ABM as a virtual laboratory in which we separately test alternative submodels: which submodel, or theory (Grimm and Railsback 2005), is able to reproduce multiple observed patterns simultaneously? If more than one submodel can do so, which is the most simple one? For example, Railsback and Harvey (2002) distilled from the literature six broad patterns in the habitat selection behaviour of trout; their ABM was designed so these patterns *could* emerge but were not *forced* to emerge. Three alternative habitat selection submodels were tested, and only one could reproduce all six patterns.

Contrasting alternative submodels or theories by their ability to reproduce sets of patterns corresponds, in principle, to more formal model selection algorithms from information theory, which are used to select among alternative simple descriptive models (Piou et al. 2009).

One important advantage of contrasting alternative submodels is also that it helps communicate that "the model can be wrong" (J. Goss-Custard personal communication) and thus counter the widespread notion that complex ABMs cannot be wrong because you can fit them to any result you want. The segregation model, for example, does *not* reproduce segregation if the decision to move is entirely random. Likewise, Janssen et al. (2009) found that two 'naïve' models of behaviour in a common dilemma game, random walk and greedy agents, did not reproduce all four patterns identified in controlled laboratory experiments.

Of course, this kind of high-level, pattern-oriented submodel selection cannot be applied to all submodels of an ABM but only to those that represent key behaviours; examples from ecology are habitat selection (Railsback and Harvey 2002) and foraging strategy (Goss-Custard et al. 2006).

17.4.3 Patterns for Parameterisation

Pattern-oriented models are usually of moderate complexity with typically 10–20 parameters. Ideally, all parameter values could be determined directly from information about the agents being modelled, but this is virtually never the case. Estimating parameter values from expert knowledge is often sufficient to get the qualitative behaviour of the model right, because experts often know much more than can be expressed by hard numbers. One important benefit of POM is that this qualitative knowledge can be included via empirical if-then rules and tested against multiple patterns.

Nevertheless, if models are to be used to make inferences about the real world and support decision-making, "guestimated" parameterisations can be too uncertain. For example, Wiegand et al. (2004) developed a spatially explicit ABM of brown bears re-invading the Alps from Slovenia. Uncertainty in the predicted population growth rate in a certain area was very high (Fig. 17.2). Wiegand et al. therefore used multiple patterns in spatial distributions and census time series to reduce uncertainty in parameter values.

They created a large number of parameter sets by sampling values of the uncertain parameters from their assumed ranges. Then, each observed pattern was used as a filter: parameter sets leading to model output which did not reproduce the pattern, according to quantitative criteria, were discarded. Some patterns could be reproduced by many parameter sets, others contained more information and reduced

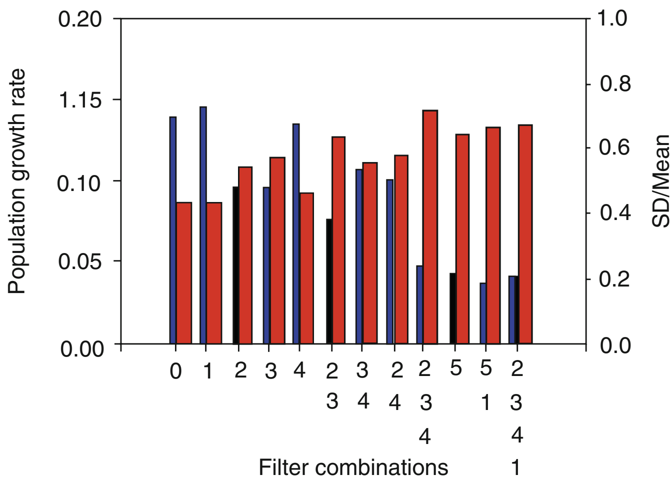


Fig. 17.2 Population growth rate (*red bars*) and its coefficient of variation (*black bars*) of a brown bear population as predicted by the model of Wiegand et al. (2004). Each *pair of bars* represents one set of model parameterisations, as (*left to right*) more patterns are used as "filters": parameterisations that do not reproduce the pattern are no longer used. Uncertainty decreases greatly as more patterns are used

the set of possible parameterisation more. In combination, however, only a few (typically 10–30) parameter sets fulfilled all patterns. Using only these "filtered" parameterisations strongly reduced uncertainty in the model output (Fig. 17.2).

Again, this technique of "Monte Carlo filtering" is not new and is used in other domains under the name "inverse modelling", but Wiegand et al. (2004) first demonstrated its power for parameterising ABMs and linked it to the other two elements of POM. Inverse parameterisation, like the entire POM strategy, is independent of its origin in ecology (Grimm et al. 2005) and applicable to ABMs in general (e.g. Janssen et al. 2009).

17.5 Discussion

ABMs are fundamentally different from the traditional models of many fields, especially from models based on system-level equations. We now have enough experience to know that the processes of designing, building, and doing science with ABMs are different from traditional approaches. Traditional system-level models are designed and described using established "languages" such as differential equations and flow diagrams, and have their complexity determined mainly by mathematical tractability. With ABMs, we need new standards and strategies. This chapter describes two such standards and strategies that have already proven very useful.

The ODD standard addresses the need for a way to describe both the essential characteristics and the full details of an ABM. The value and success of ODD is illustrated by how widely and rapidly it has been adopted. Benefits of ODD include allowing readers to quickly get an overview of a model's most important characteristics, providing sufficient detail for replication, and providing the modeller with a framework for thinking about and designing an ABM from the start.

The POM strategy addresses three needs in the process of designing and using an ABM. First, it helps with the fundamental model design problem: determining what entities, variables, and processes must be in a model, so it is complex enough but not unnecessarily complex. Second, it provides a way to develop and test theory: rules for how individual agents behave that are shown, via pattern-oriented analysis, to reproduce essential dynamics of the system. Finally, POM can be used to filter potential parameter values in an efficient and rigorous way.

Both ODD and POM seem "messier" than the standard techniques of traditional modelling, but that is simply a reflection that ABMs are messier than simpler models: if we want to model complex systems, we must use more complex models. But the messiness of ABMs is exactly why we need techniques like ODD and POM: to help us cope with the complexity. We can view these techniques as evidence of the establishment and acceptance of ABMs as a way of doing science.

References

Batty, M. (2005). *Cities and complexity: Understanding cities through cellular automata, agent-based models, and fractals*. Boston: The MIT Press.

Berger, U., & Hildenbrandt, H. (2000). A new approach to spatially explicit modelling of forest dynamics: Spacing, ageing and neighbourhood competition of mangrove trees. *Ecological Modelling, 132*, 287–302.

Brown, D. G., Page, S. E., Riolo, R., & Rand, W. (2004). Agent-based and analytical modeling to evaluate the effectiveness of greenbelts. *Environmental Modelling and Software, 19*, 1097–1109.

Brown, D. G., Page, S. E., Riolo, R., Zellner, M., & Rand, W. (2005). Path dependence and the validation of agent-based spatial models of land use. *International Journal of Geographic Information Systems, 19*, 153–174.

Czárán, T. (1998). *Spatiotemporal models of population and community dynamics*. London: Chapman & Hall.

DeAngelis, D. L., & Mooij, W. M. (2003). In praise of mechanistically-rich models. In C. D. Canham, J. J. Cole, & W. K. Lauenroth (Eds.), *Models in ecosystem science* (pp. 63–82). Princeton: Princeton University Press.

Enquist, B. J., Brown, J. H., & West, G. B. (1998). Allometric scaling of plant energetics and population density. *Nature, 395*, 163–165.

Galvão, V., & Miranda, J. G. V. (2009). Modeling the Chagas' disease after stem cell transplantation. *Physica A, 388*, 1747–1754.

Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1994). *Design patterns: Elements of reusable object-oriented software*. Reading: Addison-Wesley.

Gilbert, N. (2008). *Agent-based models. Quantitative applications in the social sciences*. Thousand Oaks: Sage.

Goss-Custard, J., Burton, N. H. K., Clark, N. A., Ferns, P. N., McGrorty, S., Reading, C. J., Rehfisch, M. M., Stillman, R. A., Townend, I., West, A. D., & Worrall, D. H. (2006). Test of a behavior-based individual-based model: Response of shorebird mortality to habitat loss. *Ecological Applications, 16*, 2215–2222.

Gotts, N. M., Polhill, J. G., & Law, A. N. R. (2003). Aspiration levels in a land use simulation. *Cybernetics and Systems, 34*, 663–683.
Grimm, V., & Railsback, S. F. (2005). *Individual-based modeling and ecology*. Princeton: Princeton University Press.
Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., Thulke, H.-H., Weiner, J., Wiegand, T., & DeAngelis, D. L. (2005). Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science, 310*, 987–991.
Grimm, V., Berger, U., Bastiansen, F., et al. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling, 198*, 115–126.
Grimm, V., Berger, U., DeAngelis, D. L., Polhill, G., Giske, J., & Railsback, S. F. (2010). The ODD protocol for describing individual-based and agent-based models: A first update. *Ecological Modelling, 221*(23), 2760–2768. doi:10.1016/j.ecolmodel.2010.08.019.
Guzy, M.R., Smith, C.L., Bolte, J. P., Hulse, D. W., & Gregory, S. V. (2008). Policy research using agent-based modeling to assess future impacts of urban expansion into farmlands and forests. *Ecology and Society*, 13, 37. [online] URL: http://www.ecologyandsociety.org/vol13/iss1/art37/
Hales, D., Rouchier, J., Edmonds, B. (2003). Model-to-model analysis. *Journal of Artificial Societies and Social Simulation,* 6, URL: http://jasss.soc.surrey.ac.uk/6-4/5.html
Heine, B.-O., Meyer, M., & Strangfeld, O. (2007). Das Konzept der stilisierten Fakten zur Messung und Bewertung wissenschaftlichen Fortschritts. *Die Betriebswirtschaft, 67*, 583–608.
Hellweger, F. L., Kravchuk, E. S., Novotny, V., & Gladyshev, M. I. (2008). Agent-based modeling of the complex life cycle of a cyanobacterium (*Anabaena*) in a shallow reservoir. *Limnology and Oceanography, 53*, 1227–1241.
Janssen, M. A., Alessa, L. N., Barton, M., Bergin, S., & Lee, A. (2008). Towards a community framework for agent-based modelling. *Journal of Artificial Societies and Social Simulation, 11*(2), 6. http://jasss.soc.surrey.ac.uk/11/2/6.html.
Janssen, M. A., Radtke, N. P., & Lee, A. (2009). Pattern-oriented modeling of common dilemma experiments. *Adaptive Behavior, 17*(6), 508–523. doi:10.1177/1059712309342488.
Kaldor, N. (1961). Capital accumulation and economic growth. In F. A. Lutz & D. C. Hague (Eds.), *The theory of capital* (pp. 177–222). London: Macmillan.
Parker, D. C. (1999). Landscape outcomes in a model of edge-effect externalities: A computational economics approach. *SFI Publication 99-07-051*. Santa Fe: Santa Fe Institute. http://www.santafe.edu/sfi/publications/Working-Papers/99-07-051.pdf
Parker, D. C., & Meretsky, V. (2004). Measuring pattern outcomes in an agent-based model of edge-effect externalities using spatial metrics. *Agriculture, Ecosystems and Environment, 101*, 233–250.
Piou, C., Berger, U., & Grimm, V. (2009). Proposing an information criterion for individual-based models developed in a pattern-oriented framework. *Ecological Modelling, 220*, 1957–1967.
Platt, J. R. (1964). Strong inference. *Science, 146*, 347–353.
Polhill, J. G., Gotts, N. M., & Law, A. N. R. (2001). Imitative versus nonimitative strategies in a land use simulation. *Cybernetics and Systems, 32*, 285–307.
Polhill, J. G., Parker, D. C., & Gotts, N. M. (2005). Introducing land markets to and agent based models of land use change. In K. G. Troitzsch (Ed.), *Representing social reality: Pre-proceedings of the third conference of the European Social Simulation Association* (pp. 150–157). Koblenz: Dietmar Fölbach.
Polhill, J. G., Brown, D. G., & Grimm, V. (2008a). Using the ODD protocol for describing three agent-based social simulation models of land use change. *Journal of Artificial Societies and Social Simulation*, *11* URL: http://jasss.soc.surrey.ac.uk/11-2/3.html
Polhill, J. G., Parker, D. C., & Gotts, N. M. (2008b). Effects of land markets on competition between innovators and imitators in land use: Results from FEARLUS-ELMM. In C. Hernandez, K. Troitzsch, & B. Edmonds (Eds.), *Social simulation technologies: Advances and new discoveries* (pp. 81–97). New York: IGI Global.
Premo, L. S., & Hublin, J.-J. (2009). Culture, population structure, and low genetic diversity in Pleistocene hominins. *PNAS, 106*, 33–37.

Rademacher, C., Neuert, C., Grundmann, V., et al. (2001). Was charakterisiert Buchenurwälder? Untersuchungen der Altersstruktur des Kronendachs und der räumlichen Verteilung der Baumriesen mit Hilfe des Simulationsmodells BEFORE. *Forstwiss Centralbl, 120*, 288–302.

Rademacher, C., Neuert, C., Grundmann, V., et al. (2004). Reconstructing spatiotemporal dynamics of Central European natural beech forests: The rule-based forest model BEFORE. *Forest Ecology and Management, 194*, 349–368.

Railsback, S. F. (2001). Concepts from complex adaptive systems as a framework for individual-based modelling. *Ecological Modelling, 139*, 47–62.

Railsback, S. F., & Harvey, B. C. (2002). Analysis of habitat-selection rules using an individual-based model. *Ecology, 83*, 1817–1830.

Richiardi, M., Leombruni, R., Saam, N., & Sonnessa, M. (2006). A common protocol for agent-based social simulation. *Journal of Artificial Societies and Social Simulation*, *9*, URL: http://jasss.soc.surrey.ac.uk/9/1/15.html

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology, 1*, 143–186.

Thiery, J. M., D'Herbes, J. M., & Valentin, C. (1995). A model simulating the genesis of banded vegetation patterns in Niger. *Journal of Ecology, 83*, 497–507.

Watson, J. (1968). *The double helix: A personal account of the discovery of the structure of DNA*. New York: Atheneum.

Wiegand, T., Revilla, E., & Knauer, F. (2004). Dealing with uncertainty in spatially explicit population models. *Biodiversity and Conservation, 13*, 53–78.

Wikipedia (2009). Software design patterns. July 24, 2009. URL: http://en.wikipedia.org/wiki/Software_design_patterns

Wilensky, U. (1997). NetLogo segregation model. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston. http://ccl.northwestern.edu/netlogo/models/Segregation

Chapter 18
Agent Tools, Techniques and Methods for Macro and Microscopic Simulation

Ateen Patel and Andrew Hudson-Smith

Abstract Many situations exist that require virtual crowds to be modelled via computer simulations on varying scales. Such simulations often have conflicting goals; the need for large and complex worlds with rich behaviours in agents, but at the same time, the need for fast performance provided by simpler agents with reasonable crowd authoring. Our goal in this chapter is to establish the tools and techniques required for simulating large-scale virtual crowds. We identify both macroscopic and microscopic simulation methods and detail application areas where there is the need for navigation and behaviour of agents around the simulation environment, to correspond to realism. Hence, we actively identify different classes of applications that form balances between the conflicting goals that exist in simulating virtual populations.

18.1 Introduction

Different applications have specific needs, and therefore, dedicated techniques to solve navigation and behavioural problems. Applications for these macroscopic and microscopic simulations can be described in three broad classes: safety and urban planning, entertainment, including the video games and the movie industry, and lastly virtual reality.

Safety and urban planning models require simulations that correspond to realism related to calibrated environments. The navigation of the crowds and their behaviour has to correspond to reality. The general objective of these types of simulations are to model the flow of pedestrians around a building, e.g. when trying to exit a building or fire, to test the suitability of the design of a building, or as an aid to large

A. Patel (✉) • A. Hudson-Smith
Centre for Advanced Spatial Analysis (CASA), University College London, Westminster, UK
e-mail: ateen.patel@ucl.ac.uk; asmith@geog.ucl.ac.uk

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_18,

scale planning, such as shopping centres, stadiums or town planning. Therefore, the simulation has to correspond to real world conditions; however interactivity and real time simulation is not a requirement. The data are generally analysed during a post processing phase.

Entertainment applications, such as video games or movies, require the behaviour and navigation of agents to appear real, but exact realism is not necessary. Interactivity is a crucial requirement for video games. Examples include recent games that provide an open world, such as Crysis 2 for the Xbox 360,[1] where players inhabit a virtual city. The realism here is enhanced by simulations of computer controlled agents with subsets reacting to real time human input.

While safety and urban simulations require exact simulations, and entertainment applications require interactivity, virtual reality simulations require both interactivity and believability. Virtual reality models can also enhance urban planning models by adding a level of immersion that gives the realism required to aid town planning. Here, the immersion of a user into the virtual world requires pedestrians moving around the scene to look and behave in a natural and real manner. A number of resources are used in a virtual reality application, from the rendering of the scene in the virtual world, to the pedestrians walking around the city during the crowd simulation. In order to solve the navigation problem, the global movement of a pedestrian around the virtual world is generally planned (see Sect. 18.2). We will explore the meaning of planned movement in the following sections. In addition to global planned behaviour, local behaviour of pedestrians interacting at street level must be taken into account. The local behaviour is commonly reactive, taking into account the interactions of pedestrians with the environment and other pedestrians (see Sect. 18.3).

The section that follows will survey some of the relevant models for pedestrian simulations to give a broad idea of techniques used and the tools that exist as well as alternatives to these tools. Further surveys can be found in Pelechano et al. (2008) and Thalmann and Musse (2007).

18.2 Macroscopic Agent Behaviour

In order to simulate crowds, individuals need to behave realistically at a local level as well as at a global level when navigating the environment. Therefore, coordinating the movements of these individual agents plays an important role. When simulating large and complex environments, local behavioural techniques such as the social forces model (Helbing and Molnár 1995; Helbing et al. 2000) and the boids model (Reynolds 1987) can get caught up in local minima. Several techniques have been proposed in order to deal with the crowd navigation problem at a global scale. The techniques can be divided into three major classes: navigation graphs (Pettre

[1] http://www.crytek.com

et al. 2005; Pelechano and Badler 2006; Lerner et al. 2006), potential fields (Loscos et al. 2003; Chenney 2004; Treuille et al. 2006), and roadmaps (Bayazit et al. 2002; Sung et al. 2005; Sud et al. 2007). This section will describe the techniques for each class of navigation.

18.2.1 Navigation Graphs

Graphs were first introduced by Teller (1992), with the aim to aid architectural walk-throughs in real time. Graphs are made up of a series of nodes and portals. The agents need to navigate through these nodes and portals defined by the graph to reach a destination. Nodes may represent rooms or corridors indoors and pavements outdoors, and portals may represent doors internally and crossings outdoors.

In Pettre et al. (2005), navigation graphs use a cell decomposition technique represented as a data structure that captures the topological structure of the navigable space in the environment. The navigable space is divided into a set of intersecting circles. Edges are defined as the line segments that join the intersecting circles, which divide the adjacent navigable areas. Navigation graphs only have to be computed once for each environment. Navigation for agents from a source to a destination is performed by scattering the number of people navigating from a source A to a destination B by $x\%$ using Dijkstra's graph search algorithm. Alternative paths for pedestrians to follow are then computed depending on the scattering parameter x. This scattering provides for variety, saving computation time during simulation. Collision avoidance is, therefore, disabled for agents at far distances, as collisions are not detectable by a viewer.

In Pelechano and Badler (2006), Helbing and Molnár's (1995) social forces model is used for the local behaviour, and combined with a wayfinding approach to build up a high level global view of the environment. This global view is constructed via agents exploring the environment using current knowledge. Wayfinding is computed by using four components: building a mental model of the environment, its current position within the mental model, processes that help it learn features such as doors and walls within the environment, and finally, the navigation process that allows it to move through the environment. The mental map is made up of a cell and portal graph with nodes added as the agent navigates the environment. The algorithm calculates the shortest path to the exit based on the agent's knowledge of the environment. When agents interact, they can share information to the adjacent cells only, to share knowledge such as whether it is passable or there is danger through the portal (e.g. a door).

Agents have different levels of knowledge to conform to real world situations within three classes: agents who are leaders and trained, i.e. they have complete knowledge of the environment; those that do not have complete environmental knowledge and make their own decisions in stressful environments such as a fire evacuation; and agents who are not leaders and untrained, i.e. they do not have complete environmental knowledge and may be incapable of making their own decisions in a highly stressful environment. High level wayfinding then takes place. Leaders share their mental map with other agents. Agents then check their shortest known path based on their own and shared knowledge, reacting differently to a hazard based on the agent class.

In Lerner et al. (2006), outdoor environments are represented using cells and portals. The cells and portals are created by a two-pass algorithm; the first pass creates the initial partition, which is refined by a second pass. It proves more efficient than traditional Binary Space Partitioning (BSP), and creates an efficient automatic cell and portal partitioning; as such it can deal with arbitrary orientation of walls. Half edges, i.e. a single sided oriented edge obtained from a 3D model, are passed to the algorithm in 2D, which will create the partitioning. Pedestrian paths are represented as cells, whereas the intersections between pedestrian paths and crossings are represented by portals. The average number of visible cells and portals in a rendered scene is significantly lower than a BSP tree, which allows for more agents to be accommodated in real time.

18.2.2 Potential Fields

In methods that involve potential fields, the environment is modelled as a 2D discretised grid. Each cell in the grid has a certain potential. The goal has attractive potentials while the obstacles have repulsive potentials. A gradient can therefore be applied that guides the agents from the source to the destination. These methods help reduce the complexity at a macroscopic level. In Thompson and Marchant (1994), Simulex used potential fields for the agents to navigate around the scene.

Loscos et al. (2003) also uses a potential field method whereby the space is subdivided into a 2D grid to indicate the areas that are accessible such as crossings and pathways, and collision detection with buildings is achieved with a collision map (Tecchia and Chrysanthou 2000). The agents are not entirely represented as flows, but have individual-based behaviour. A graph of goals is around the urban environment, covering each pavement area. Individual behaviour is achieved by creating a trajectory, where the agents start at one of the goals in the graph, and are assigned another goal as a destination. The flows are achieved by using local information rather than global, making use of the graph of goals. When an agent reaches a goal, it stores a direction vector that fades over time. When another agent reaches that goal, it checks for a previous direction vector, and if found, makes a decision based on that vector. These methods have the potential of getting stuck in local minima. Figure 18.1 shows a representation of the agents moving around in the environment.

Another similar technique to potential fields is flow tiles (Chenney 2004). Each flow tile contains a precomputed velocity field. These flow tiles can be combined to form a continuous flow for various applications, where the edges and corners maintain continuity. For a crowd simulation, the streets can be represented with flow tiles (Fig. 18.2). Internal boundary conditions on the tiles can denote buildings preventing the agents bumping or walking through buildings.

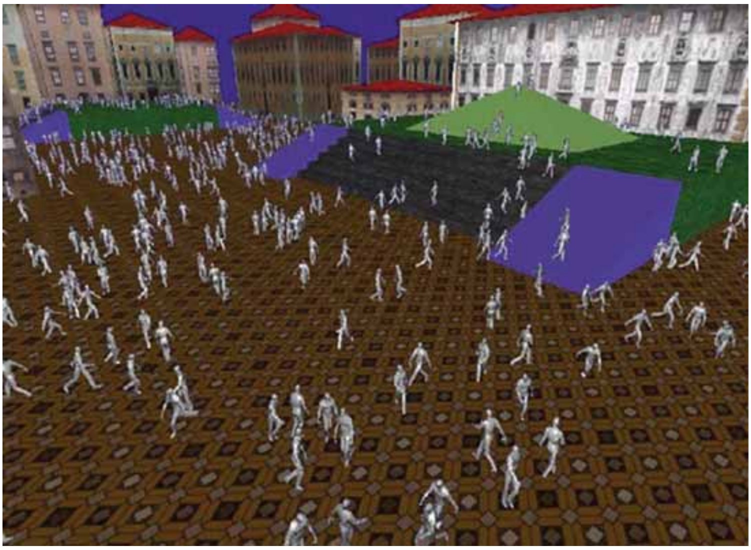


Fig. 18.1 Behaviours achieved by agents based on underlying grids containing behavioural maps (Taken from Tecchia and Chyrsanthou (2000))

Fig. 18.2 Flow tiles on the streets drive the flow of agents through the city (Taken from Chenney (2004))

In Treuille et al. (2006), the model uses a continuum-based approach. The environment is modelled as a 2D grid including the mapping of uneven terrain. Large numbers of agents can be accommodated as the computations are based spatially rather than per-each individual agent. The motion of crowds is controlled by a dynamic potential field, which allows it to avoid moving obstacles without the need

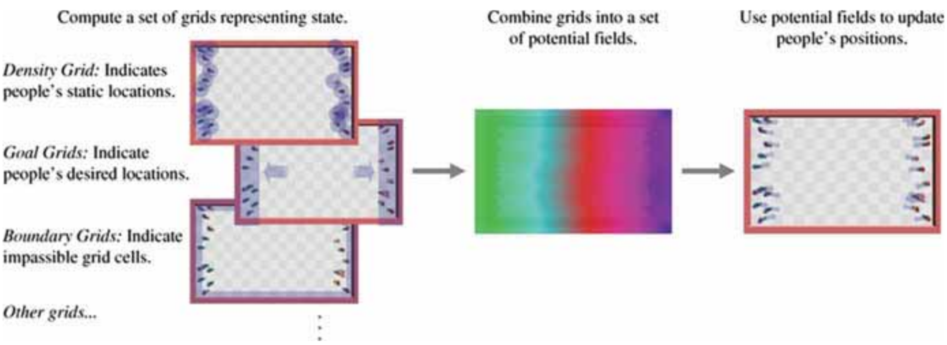


Fig. 18.3 An underlying dynamic potential field representing density, goals and boundaries. Other conditions control the positions of the agents (Taken from Treuille et al. (2006))

for explicit collision avoidance. The potential field is composed of superimposing grids made up of different types of information, such as the density grid (the location of people), the goal grid (their goals), boundary grids (impassible grid cells), etc. (Fig. 18.3). The environment uses a cost function that determines the speed of the agents. The cost function takes in factors such as the slope of the terrain, the density of the agents, obstacles, etc. This cost function helps compute the optimal path for the agents to travel through by using the gradient of the computed potential field. Most crowd simulations combine global path planning and local behaviour planning methods with the potential of conflicts between the two methods. The continuum dynamics method integrates local path planning with the global navigation for the agents. A dynamic potential field integrates agents as moving obstacles, and therefore, a dynamic global path can be planned that will avoid static obstacles as well as other agents. As the path is dynamic and is updated based on the current environment, agents do not get stuck in local minima.

18.2.3 Road Maps

Road maps use robotics as a basis, as well as algorithms for route planning of agents around an urban environment developed from the motion planning solutions of robots. The real time global navigation of only a small number of agents will run efficiently when computing road maps based solely on the robot motion planning solution. When the number of agents increases, global collision-free routes have to be computed for each independent agent, which can result in exponential computation time, therefore becoming a bottleneck. Dynamic objects can cause an even greater problem for motion planning with the potential of agents getting stuck in local minima. Various solutions have been developed from road maps that will allow large numbers of agents to navigate around scenes that contain static as well as dynamic objects.

Bayazit et al. (2002) incorporates Reynold's seminal work on flocking (Reynolds 1987) with probabilistic roadmaps (PRM) to allow global navigation of agents.

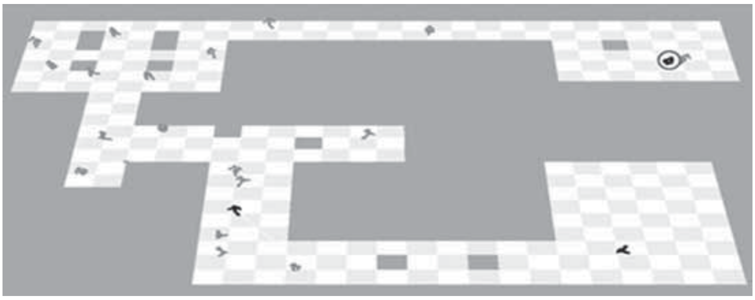


Fig. 18.4 The black character uses roadmaps to navigate through the rooms avoiding obstacles to lift a box (*circled*) in the *upper right* room (Taken from Sung et al. (2005))

Flocking behaviour uses local behaviour planning, where each flock member's behaviour is determined by its local environment. Integrating a roadmap based path planning method with the flocking behaviour allows complex group behaviours among the flock such as homing towards a goal, exploring an environment and shepherding. The roadmaps that are computed have maximum clearance from obstacles, allowing the flock to pass through a collision-free path. Individual flock members may have sub-goals within the group, and on reaching the sub-goal, continues along the global path of the group.

Lien et al. (2005) extend the road map integration of Bayazit et al. (2002), specifically the shepherding behaviour of the flock when following a global path. Shepherding behaviour is a type of group behaviour where one or more group members shepherds or controls the motion of another group, i.e., the flock. Lien et al. (2005) introduce multiple shepherds that control the flock individually without communicating with each other, and still manage to efficiently control a flock. The environment, once divided into a roadmap, contains milestones that are the nodes of the roadmap. The shepherds attempt to steer the flocks towards the milestones in order to help the flock reach the goal. The roadmap here is dynamic.

Kamphuis and Overmars (2004) created roadmaps for groups to follow. A path is created for a single agent known as the backbone path. The clear area around the path is termed the corridor, in which other agents in the group must stay. The corridor's width and the extent to which other agents can move away from the main group is limited by using a social forces model. This ensures the group stays together, and allows agents in the group to catch up or even pass through obstacles from the same side.

Sung et al. (2005) used the concept of roadmaps to create a global motion planning solution. A probabilistic roadmap (PRM) was computed for the pedestrians to navigate through the complex environments to avoid collisions (Fig. 18.4). The PRM does not create an accurate navigation path, but provides an initial guess for the search that approximately satisfies the constraints. It then uses Dijkstra's search to further refine the search to allow for smooth adjustments to the position, and direction of the pedestrian, providing an accurate motion that exactly satisfies the constraints. All the possible sets of actions are represented in a motion graph, and the search algorithms just described are used to generate motion. The crowd motion is authored by taking into account constraints such as duration, position, orientation and body pose. This allows two characters to meet at particular places face-to-face at specified times, or even to go to certain events together.

Lau and Kuffner (2005) introduced a behavioural motion planning approach using a finite-state machine (FSM), which determines the motion of the pedestrians by avoiding collisions with other pedestrians as well as obstacles. This motion is determined during runtime via a planning algorithm that uses the FSM to create a global search to create the trajectory that the pedestrian has to follow. This technique was further extended by Lau and Kuffner (2006) to accommodate a larger number of pedestrians in the simulation. It also allowed for the motion planning technique to re-plan in real time, allowing for a more dynamic environment. The environment here is modelled as a 2D grid where the cells are determined to either contain an obstacle or not. A search tree is precomputed to represent all the behavioural states that are reachable based on the current state of the pedestrian. A search algorithm also computes a global path to the goal of each pedestrian by avoiding obstacles mapped onto the environment grid map. A path finding algorithm using reverse path lookup is then used to calculate the shortest path to the destination for the pedestrian to follow. Using the behavioural states that were determined using the precomputed search tree, the path calculated is able to handle the presence of certain obstacles and characters by making the pedestrians jump or pass under them. A sequence of behaviours is then converted to actual believable motion using a motion synthesis system blending frames at transition points.

Sud et al. (2007) introduced a technique called Adaptive Elastic Roadmaps (AeRO). The roadmap is based on a Voronoi diagram. It uses an approach that allows agents to have distinct goals and individual-based behaviour without getting stuck in local minima similar to Treuille et al. (2006). The road map algorithm adapts to dynamic environments and computes a dynamic global navigation path for the agents by taking into account moving obstacles and inter-agent collision avoidance. The global road map updates incrementally, which reduces the computation time and offsets the requirement of continuous updating by introducing a concept called *link bands*. Link bands are introduced at a local level to augment the global navigation. The link bands deform and reform the road map at a local level, and therefore, guide agents through a collision free path avoiding other agents and moving obstacles.

18.3 Microscopic Agent Behaviour

There are varying techniques used to solve the crowd motion planning problem at a local scale. The technique used depends on the application that needs to be solved, and the specific requirements based on the crowd density and the scale of the area in which it needs to be simulated. This section will focus on the crowd behaviour at a local scale, and will divide local navigation techniques based on the application, as the application's requirements define the technique.

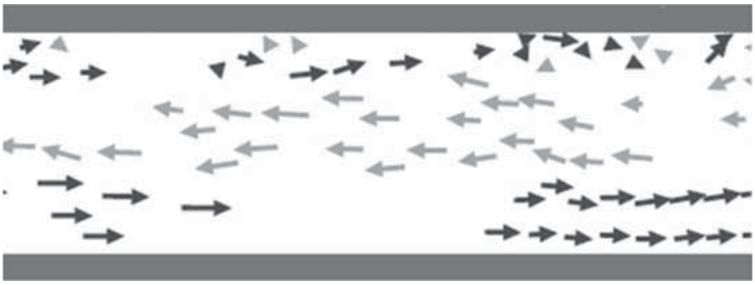


Fig. 18.5 The flow of agents leads to lane formation in this example. The direction and *arrow* lengths represent the walking direction and the speed of the agents (Taken from Helbing et al. (2001))

18.3.1 Safety, Urban Modelling and Architectural Applications

The analysis of video films of individual and groups of pedestrians has given a good insight into the way pedestrians behave, and it could give us an idea of how they influence crowds – see also Johnasson and Kretz (2012). There are certain observations made by Helbing et al. (2001) and Loscos et al. (2003). With regards to groups of pedestrians, they will walk at the same speed, follow the same goals, and will wait for each other if one goes missing. In order to simulate these pedestrians, the main goal of the simulator is to provide results that match real world data, e.g. evacuation times, walking speed, influence among individuals, individual personal space. Therefore, it is important to calibrate the models with real world data. There are three common approaches that are used to simulate pedestrians for safety applications. These can be classed as fluid, cellular automata (CA) and agent-based models (ABMs). Details of CA and ABMs can be found in Iltanen (2012) and Crooks and Heppenstall (2012), respectively.

Flow-based simulations are used to model pedestrians at a macroscopic scale, where an individual does not have a powerful influence on the behaviour of the crowd. These pedestrians are, therefore, not individual entities but are part of a network of flows. These flows are based on the Navier-Stokes equations, where Henderson's (1971) early work showed that it was possible to model pedestrians based on these equations. The Navier-Stokes equations describe the motion of fluid substances based on Newton's second law. These equations describe compressible and incompressible fluids, but to apply it to pedestrians, only the incompressible part of the equations need to be used. Helbing et al. (2001) and Hughes (2003) are examples of recent work that has used fluids to simulate crowds. Figure 18.5 illustrates pedestrian flow represented by arrows from Helbing et al. (2001).

CA approaches include the environment that is divided into a group of cells. Each cell has a finite amount of states. The pedestrians occupy the cells, and they move from cell to cell based on parameters such as crowd density and walking speed. ABMs are quite popular among safety sciences, social sciences and the urban modelling area. An agent as defined by Jennings et al. (1998) is "a computer

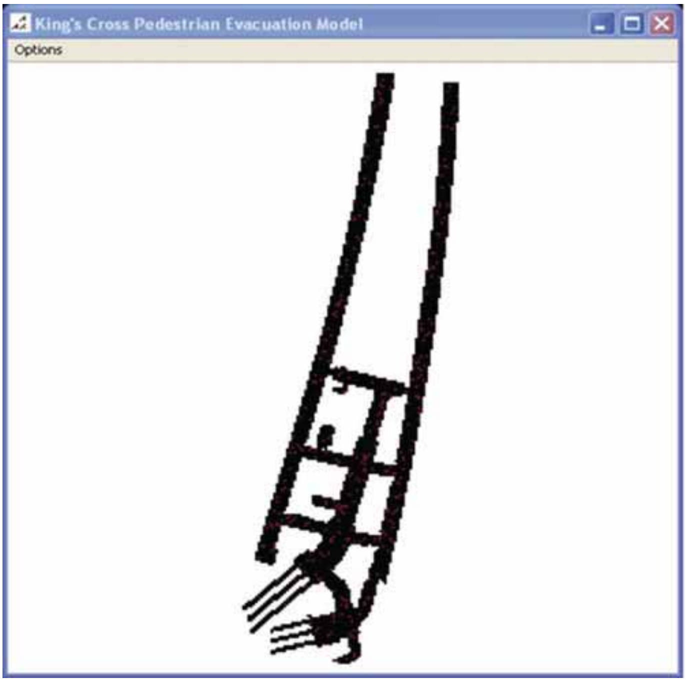


Fig. 18.6 Kings Cross St. Pancras underground station evacuation model (Taken from Castle (2006))

system, situated in some environment, that is capable of flexible autonomous action in order to meet its design objective". These agents can also represent pedestrians at an individual level, unlike flows. This allows us to set micro-specifications of a crowd system, which generate emergent behaviours at a macro-scale. Castle's (2006) work is one example of an ABM for evaluating evacuation scenarios at King's Cross St Pancras underground station in London (Fig. 18.6). Batty et al.'s (2003) simulation of the Notting Hill carnival is another example of the use of an ABM in this area.

Helbing and Molnár (1995) introduce a 'social forces model' for pedestrians. They describe a model where a sum of the forces exists that influences the locomotion of pedestrians. There are three forces. The first is acceleration since the velocity of the pedestrian can vary over time. The second is repulsion, which is the repulsive force experienced from other pedestrians and obstacles, and the third is attraction, which is the attractive force experienced by other people (e.g. friends, family) or other objects (e.g. popular attractions, window displays). These forces are combined with a term to account for behavioural fluctuations to form the equation for the 'social forces model'. This model is one of the most popular approaches to pedestrian modelling in safety applications, and has been reused and refined in many other approaches such as Helbing et al. (2001) to simulate escape panic, and others used for modeling evacuation scenarios (Braun et al. 2003; Cordeiro et al. 2005).

18.3.2 Entertainment Applications

The aim of crowd simulation in entertainment applications is interactivity, especially in video games. Interactivity is also needed in the movie industry, but applied in a different context to video games. Whereas video games require real-time interaction with crowds for the end user, the movie industry handles all the crowd simulation offline in order to create impressive scenes containing huge crowds. The user interacts with the simulator running several trials and set-ups to create the final scene. This also gives the user global and local control, allowing precise control over the crowds' movements. Although the simulation is computed offline, it has to be fast enough for the several trials that will take place.

Work by Reynolds (1987) first introduced artificial life to the computer animation/graphics area. He demonstrated how the behaviour of bird flocks could emerge from an extension of a particle system, where the birds act as the particles. This was a computer model of coordinated animal motion such as bird flocks or fish schools, which accounted for collision avoidance and goal seeking (Fig. 18.7). This behaviour can be applied equally to crowds, although commercially, this model has had various applications. One such application is the bat swarms and penguin flocks in the 1992 movie 'Batman Returns'. More recent work by Reynolds (1999) extended the behaviours that were simulated, and introduced steering behaviours in addition to the collision avoidance and goal seeking introduced in Reynolds (1987). The steering behaviours involved seeking, fleeing, pursuit, evading, path following, wall following, etc. The crowd simulation has been further extended, and moved onto the Sony PlayStation 3 platform (Reynolds 2006). The general navigation and behaviour remains the same, but accommodates more agents in real time (Fig. 18.8).

Particle systems have been used in other ways to simulate crowds (Fig. 18.8). Bouvier et al. (1997) uses a particle system combined with Newton's law of motion to model crowds. This again uses a microscopic approach, which allows for the navigation to be planned at a local level. The local planning is based on taking into account reactive behaviour, where individuals avoid each other based on density as well as the avoidance of obstacles. The model by Bouvier et al. (1997) also uses higher level decision making to create the simulation, taking into account the parameters such as destination, density, the duration of stay for visiting certain places and

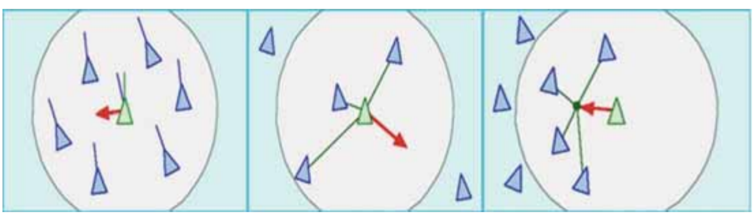


Fig. 18.7 Animated flock showing behaviours of separation (steering to keep a minimum distance among flockmates), alignment (steering to keep aligned with the direction of local flockmates) and cohesion (steering to move closer to flockmates) (Taken from Reynolds (1999))

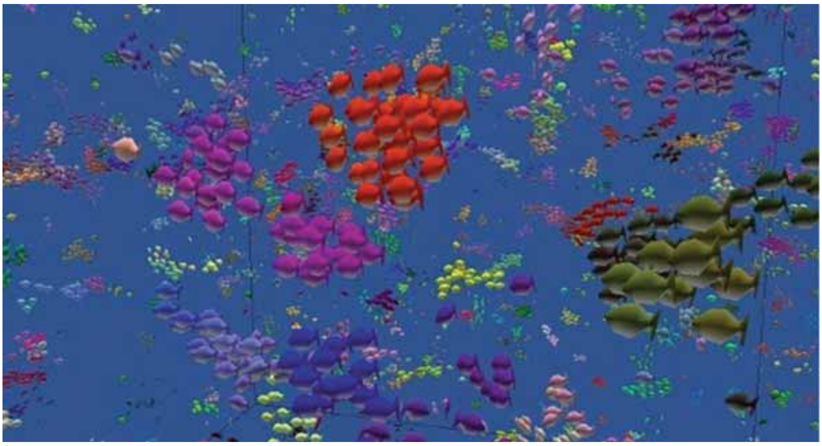


Fig. 18.8 Flock of 10,000 fish rendered on a Sony PlayStation 3 (Taken from Reynolds (2006))

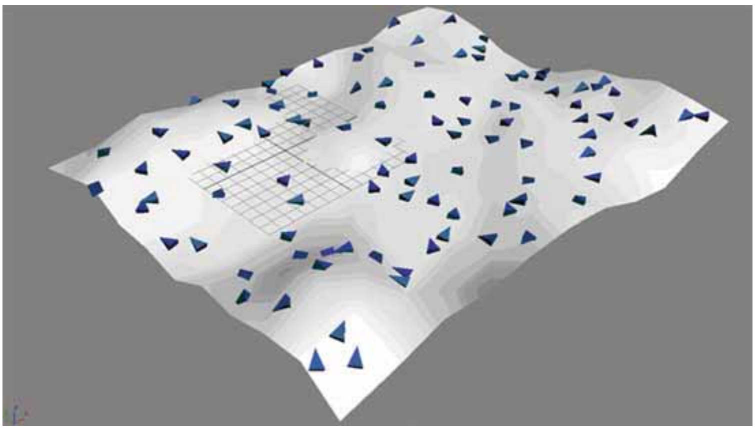


Fig. 18.9 3ds Max crowd and delegate system: follow surface and wander behavior

other global events such as taking into account the presence of obstacles. This higher level is based on Newton's law.

Models created for the entertainment industry are often created in packages outside of the custom created research toolkits. While not being a specialised modelling system per se, 3ds Max from Autodesk is an industry leader in game production linking into the wider Autodesk suites. Built into the package is a 'Crowd and Delegate' system allowing groups of 3D objects and characters to be animated according to a series of computed behaviours. Using simple rules such as 'avoid' 'follow', 'seek' or custom written in the form of scripts, one can create crowds with highly complex behaviours. Figure 18.9 illustrates the most basic level of model behaviour in 3ds Max using the 'follow surface' and 'wander behaviour'. Each delegate is represented by the triangles forming part of a team assigned with a wander rule for a set period of time while following the 3D object surface. Once set up, key frames are assigned according to a predefined time sequence, and the simulation is pre-computed allowing for later export or rendering. Using such a method, it can be quickly extended to cityscapes, fictional landscapes or any realm.

Models created in such a software not only utilize advances in graphic card technology, but also advances in physics-based engines such as Havok Physics, which allows the user to easily add additional elements such as mass and gravity to influence the behavior of the agents.

The various built-in components of 3ds Max enables high quality graphic outputs as well as real-time previews and outputs to game engines such as Crysis. This allows researchers to achieve 'semi movie-like' results. Indeed 3ds Max is arguably the most powerful mass market simulation engine although it does need viewing in context. The package itself is essentially a 'black box' compared to more traditional approaches with routes and behaviours pre-computed with minimal feedback on the underlying dynamics. We would argue, however, that it is within the entertainment industry with increasing consumer demand for realistic reaction from agents within gaming environments where future innovation will emerge.

18.3.3 Virtual Reality Applications

The techniques used for virtual reality applications combine the techniques used in the previous two sections. This is an area where we see more of a convergence between the two broad areas of crowd simulation, i.e. realism of crowd behaviour, and high quality visualisation. This convergence helps to immerse the user into the virtual crowd.

Tecchia et al. (2001) introduced a layered approach to simulate agent behaviour. The approach is based on using a 2D grid with four layers. Two layers are used for collision detection with the environment, and between other agents. The other two layers are used for other complex behaviours, one layer which involves behaviours such as waiting, turning a certain direction or computing a new direction based on the environment. The other layer deals with highly complex local behaviour such as pushing a button to call a lift (Fig. 18.10). This layered approach allows for highly realistic behaviour to be simulated at a real-time rate, with thousands of agents.

Musse and Thalmann (2001) introduced scalability in simulating virtual crowds. The crowd simulation used an agent-based approach. The hierarchy was comprised of virtual crowds, groups and individuals where the groups are the more intelligent and complex structures. The individual follows the specification of the group. The group can be controlled at different levels of the autonomy, which refers to the independence of the agents without the need for user intervention including the amount of information needed for simulating the crowds. There are three different levels of control of the groups. The first is through the use of scripted behaviour for the crowds to follow, the second is by defining behavioural rules to user events and reactions to create more complex behaviours, and the third is by guiding crowds to follow orders by the user during run-time. Individual agents, therefore, group together based on having common sociological factors defined for the group.

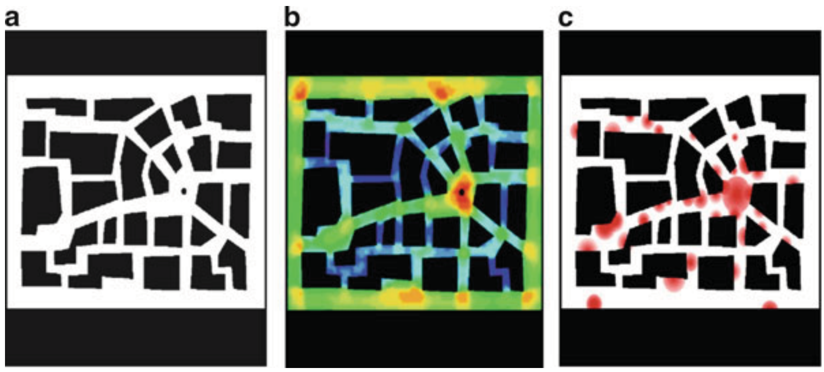


Fig. 18.10 (**a**) An example of a collision map. The regions where agents can move are encoded in *white* and inaccessible regions in *black*. Examples of behaviour maps: (**b**) visibility map (**c**) attraction map (Taken from Tecchia et al. (2001))

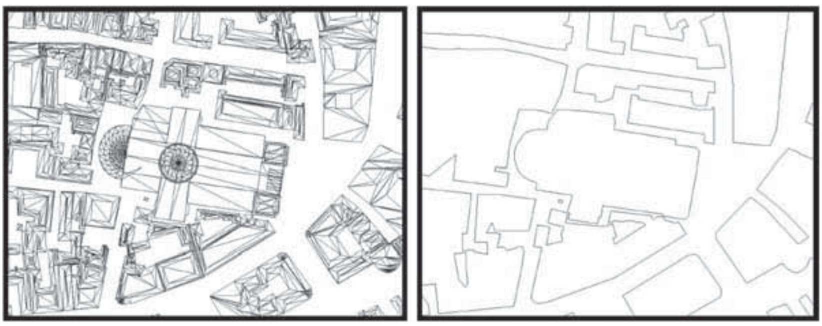


Fig. 18.11 A 2D map extracted from a 3D database (Taken from Lamarche and Donikian (2004))

Lamarche and Donikian (2004) introduce a different type of hierarchical approach using the geometry of the virtual environment. The topological structure of the geometry enables the capability for global path planning in real-time for each agent while taking into account visibility and reactive behaviours. To create the topological structure, the 3D geometry in the database is converted into a 2D map (Fig. 18.11). This is done by using two parallel planes separated by the height of a humanoid, and cutting the database geometry starting at the floor. This area cut is the navigation area, which is projected onto an XY plane to create the 2D map. A cell decomposition technique is then used on the 2D map, and the resulting triangulation is then optimised to create convex cells of the environment, which is then captured in a graph.

The cell decomposition enables the ability to generate road maps for path planning. The topological abstraction of the geometry reduces the path planning graph size, enabling real time path planning computations. While agents are navigating the path, a reactive technique is used to take into account the presence of other agents in each cell. An iterative optimisation approach is used to reach the goal and

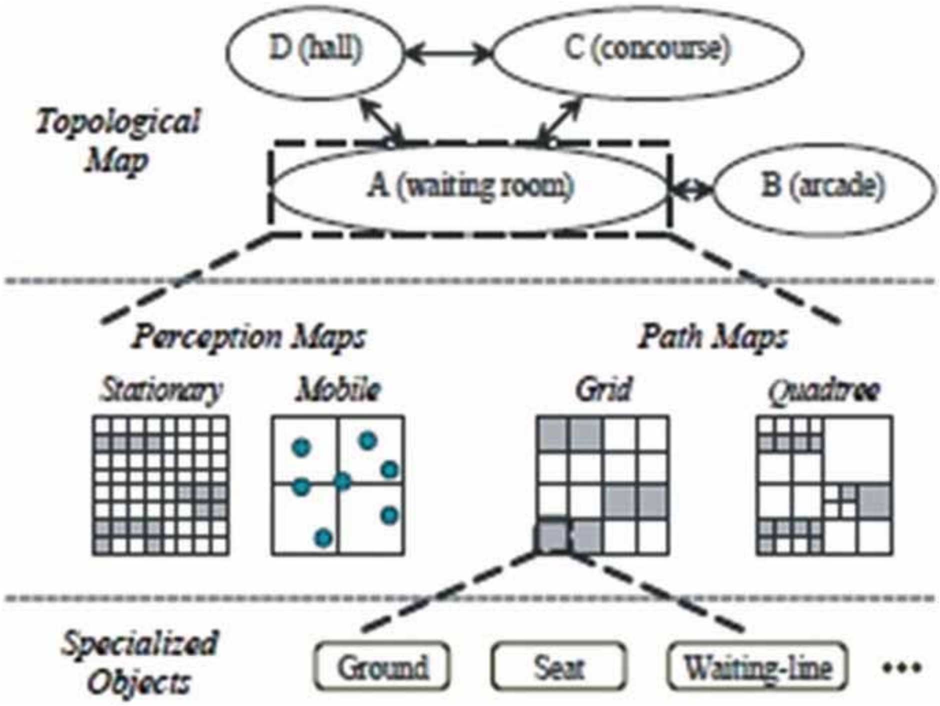


Fig. 18.12 Hierarchical collection of maps representing the environment (Taken from Shao and Terzopoulos (2005))

avoid collisions. The way the trajectory of the agent is computed depends on the pedestrian viewing angle and the next cell. A free personal space is kept around each agent, which is a minimum distance between other agents, and obstacles. By predicting the positions of the other agents, the model can predict if a collision may occur, and in turn modify its trajectory.

Shao and Terzopoulos (2005) use a conceptually similar hierarchical structure for the virtual environment to Lamarche and Donikian (2004), but with different methods. Their method uses a hierarchical collection of maps (Fig. 18.12). At the highest level, it includes a topological map that represents the structure of the different parts of the virtual world. This topological structure results in a graph structure that represents the environment as a set of rooms or corridors that are interconnected. At the middle level, perception maps provide information related to perceptual queries by storing lists of stationary objects as well as agents in the vicinity. This level also includes grid maps and quadtree maps for the geometry of the environment. The lowest level consists of path maps.

These maps enable path planning online for navigating through the environment. The level also includes special representations such as seats and waiting lines. Pedestrians in this model use a reactive and deliberative approach. The reactive behaviours with which the pedestrians are equipped are primitive and based on Tu

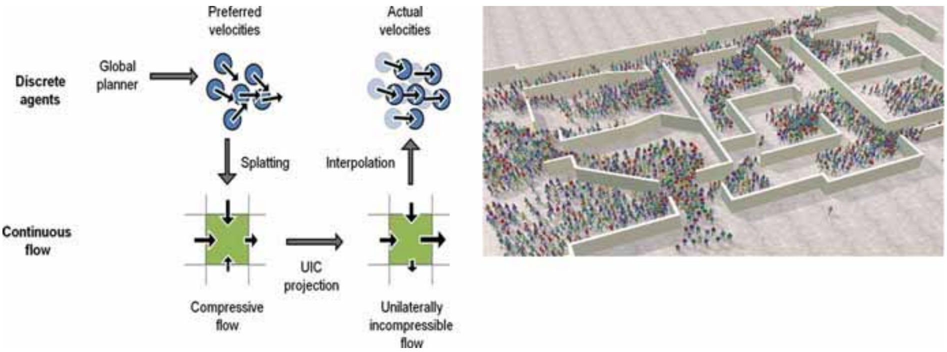


Fig. 18.13 Overview of the algorithm (*left*) with a screenshot of the simulation (*right*) (Taken from Narain et al. (2009))

and Terzopoulos (1994) using a bottom-up strategy. The behaviours include standing still and moving forward as well as collision avoidance. These local behaviours in turn support more complex motivational behaviours. The motivational and navigational behaviour enable the pedestrians to reach their goal. The topological map introduced here allows for the achievement of this task. The graph structure is used to find the global path for the pedestrian, which is then refined using the grid maps.

In Narain et al. (2009), a dual representation approach is used, where agents are represented as both discrete, and as part of a continuous system which is described by density and flow velocity. Therefore, the discrete representation of agents is combined with the representation of the crowd as a continuum fluid. Local collision avoidance is mapped into the continuous domain. A Unilateral Incompressibility Constraint (UIC) is introduced in this mapping to obtain a variational constraint on the crowd flow depending on the region of the environment. The constraint accelerates the collision avoidance step, as it increases the density constraint to maximum where obstacles exist.

Figure 18.13 illustrates the UIC algorithm as well as a rendered scene of an evacuation. This approach can be used as a local planner along with a global planning approach such as a road-map based algorithm (Sud et al. 2007) or continuous based optimisation (Treuille et al. 2006) on a coarse grid. The global planning approach needs to be computed to determine a preferred velocity for each agent. The global path is computed by avoiding large static obstacles in the environment, and ignoring the presence of agents. The UIC solver introduced is then used for collision avoidance among agents, and therefore, computes the corrected velocity field for the crowds, taking agents into account. Using the corrected crowd velocity field, the agent flow is constrained in high density regions, but the agents flow will be over-constrained in low density region. In order to overcome this problem, the velocity of the agent is computed by interpolating between the continuum velocity and the agents own preferred velocity. A final step is then added to create a personal space around each agent by introducing a minimum distance based on the method used in (Treuille et al. 2006).

18.4 Tools

18.4.1 Safety and Urban Planning

A number of commercial tools exist for creating pedestrian simulators for safety applications, particularly for evacuation scenarios. We will cover the most popular tools, and we will try to describe the technique used behind each tool where possible. Due to the commercial nature of some of the tools, the information in the literature is gathered from publicly available data.

Legion™ is a popular pedestrian simulator used in the commercial sector, which resulted from extensive research at the Maia Institute in Monaco. It uses ABMs for all of its simulations, with the environment layouts based on computer aided design (CAD). Therefore, agents are treated as individuals. Legion™ is based on work carried out by Still (2000), trying to simulate and analyse crowd dynamics in environments such as stadiums. Environments in the simulation are known as *iSpace*, where agents and the environment communicate with each other. Agents inform the environment with their observations such as obstacles, while it asks the environment for the direction to take based on its objective. In this way, the agents can communicate using the environment. The path planning algorithm has not been disclosed, but the agent uses a least effort algorithm to reach its goal, based on satisfying a number of constraints, such as speed distribution and collision (Thalmann et al. 2005). The least effort means the maximum speed, but a minimum cost, time and congestion, and therefore, optimality may not be guaranteed. The path costs are based on their length, travel time and effort. A number of parameters in Legion™ for the agents give the simulations a factor of believability, e.g. they may move around randomly in a particular space, and come together to congregate before moving apart again. Other random events can be entered into the simulation such as size, speed and age.

As mentioned in the previous sections, flow-based models are based on a macro-scale, where the individual is not significant, and pedestrians are part of a network of flows. EVACNET4 (Kisko et al. 1998) is an example of a flow-based model used to simulate evacuation scenarios. The environment of the simulator is represented as a network of nodes. This network is called the EVACNET network model, and consists of a set of nodes and arcs. The nodes are the building components or the bounded parts of the environment, such as rooms, corridors, lobbies and halls. The arcs are the passageways that connect these parts. Each arc has to be set up with a flow capacity and a traversal time by the user. The user also has to set the initial placement of occupants, i.e. each node has an initial state, such as the number of people, and the destination (fire exits). The maximum number of people for each node must also be defined. The program then searches for a minimum evacuation time, and the network flow algorithm finds a solution for the evacuation. Note that the model does not have the environment described in explicit geometry, and it is rather an implicit model using the nodes. There are other examples of flow-based models such as EESCAPE and FIREWIND.

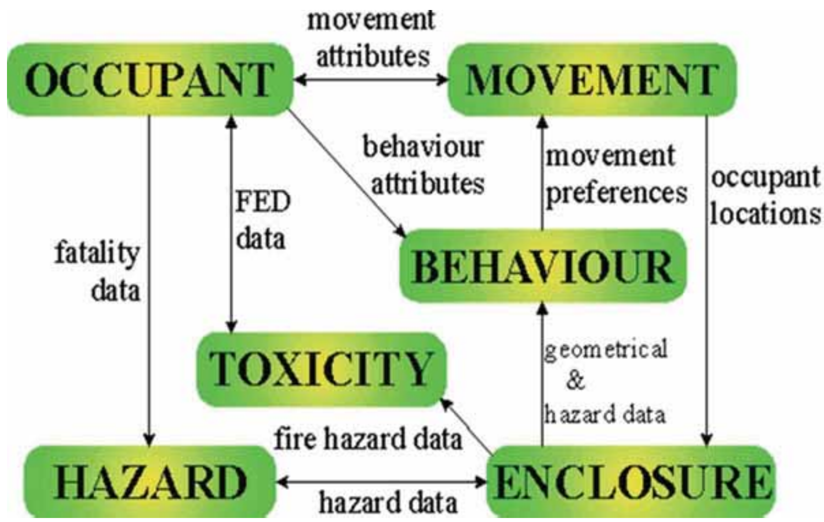


Fig. 18.14 The submodels representing the algorithm of the EXODUS model

Simulex (Thompson and Marchant 1994) is another agent-based simulator. Like Legion™, Simulex uses CAD files, but uses these files to define 2D grids which represent floor plans. It uses distance maps on top of the 2D grid, which is a discrete vector field. The aim of Simulex is to model evacuation scenarios, and individual agents have various attributes such as walking speed, acceleration, position, and angle of orientation. Agents use the gradient of the vector field to reach their goal, while the attributes help solve the motion planning problem.

EXODUS is a simulation tool created by the Fire Safety Engineering Group of the University of Greenwich.[2] This tool is also an agent-based simulator. It is used for both evacuation scenarios and pedestrian dynamics, and was first introduced in Galea et al. (1993). It is made up of five interactive submodels, namely, occupant, movement, behaviour, toxicity and hazard (Fig. 18.14). The navigation of the occupants uses a rule-based system and is adaptive. CAD files are used again to define 2D grids for the floor plans as well as the general geometric structure. Each grid is made up of nodes and arcs, but unlike EVACNET4, these nodes represent a small region of space on the grid, while the arcs represent the distance between nodes. Individual agents travel using the arcs from node to node to reach their goal. This model incorporates many psychological factors for the simulation.

In addition to creating the first prototype for Legion™, Still (2000) developed Myriad II along with colleagues at Crowd Dynamics Ltd.[3] Myriad II is unique compared to other tools seen thus far, where it integrates three different techniques to create the modelling tool (Fig. 18.15). It comprises network analysis, spatial analysis, and agent-based analysis. This integration enables it to create the best possible model based on the situation in the environment, e.g. a network model representing

[2] http://fseg.gre.ac.uk/exodus/

[3] http://www.crowddynamics.co.uk/

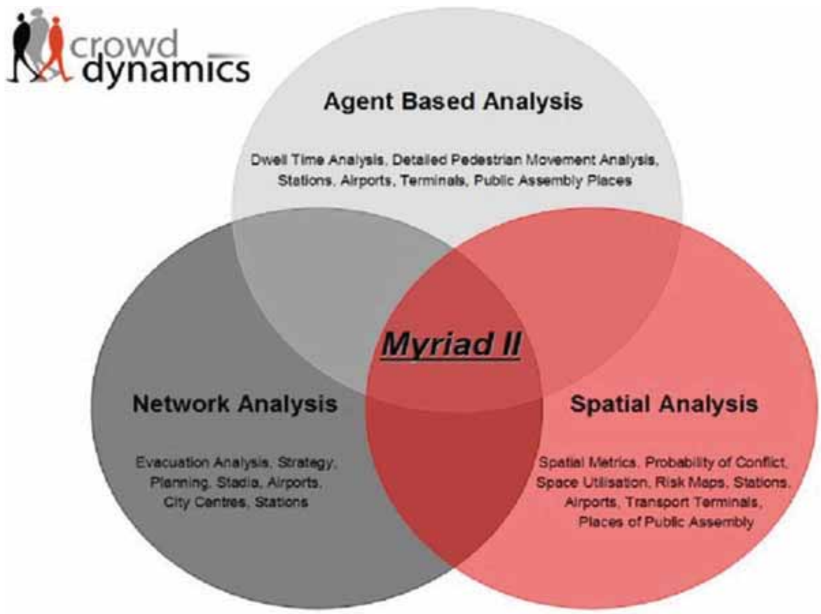


Fig. 18.15 The overlapping of different methods in Myriad II

simple roads can be integrated into an ABM representing complex interactions in parts of the environment such as a shopping centre. Data can then be passed between the two models of the environment. Four attributes define the crowd behaviour, namely, objective, motility (walking speed), constraint (eg. crowd density) and assimilation (acceleration) (Still 2000).

Mass Motion (Challenger et al. 2009) is a 3D agent-based simulation tool developed by Erin Morrow at Arup.[4] The ABM's simulation core runs on a 64 bit multi-core architecture, and it may be the only currently available commercial simulation software that is multi-threaded. Each individual agent has a position, orientation and velocity. Each agent has a goal, and has to reach it in the minimum amount of time by using both local and global navigation. Figure 18.16 shows a crowd simulation using the Mass Motion software.

Each individual agent is aware of the environment using a 2D projection of all static obstacles within a defined volume in order to map the obstacles (Fig. 18.17). A modified version of Dijkstra's algorithm is used to define the complete paths between origin and destination within the map. Each agent also accounts for the visibility of other agents within their immediate neighbourhood.

Figure 18.17 shows the structure of a simulation environment within Mass Motion via a sparse node network and its implicit relationships between floors via links. The distance for each agent to the exit is precomputed using the modified Dijkstra's algorithm and is stored at link nodes. This helps in considering the best possible route to the destination (exit) via a perceived cost, which is randomised per

[4] http://www.arup.com/

Fig. 18.16 Screenshot of Mass Motion simulation software (Transbay terminal in San Franciso) (Image courtesy of Erin Morrow, Arup)

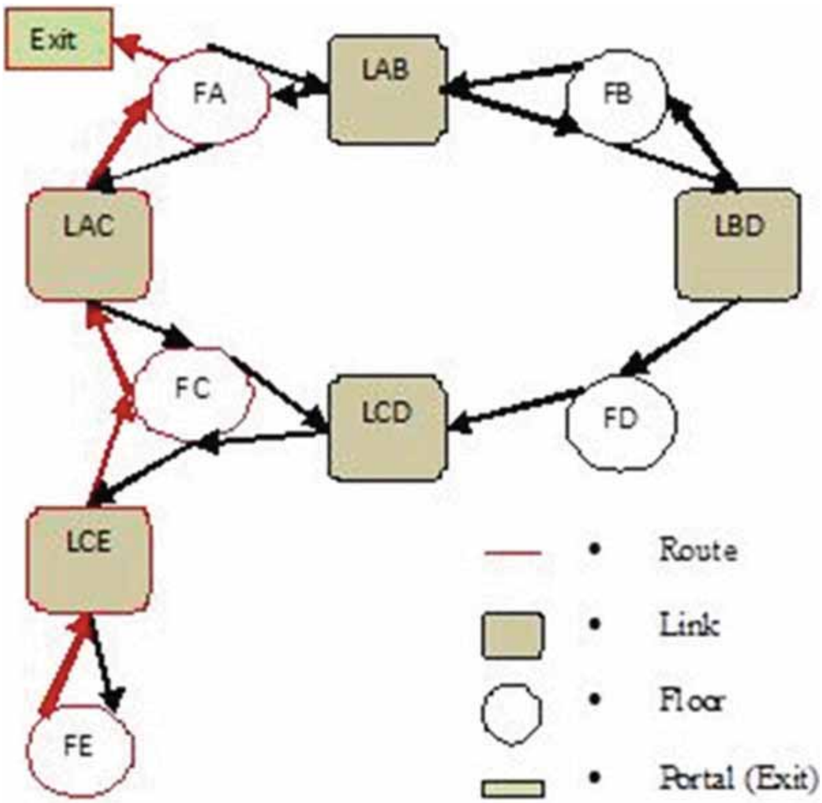


Fig. 18.17 Sparse node network structure of a simulation environment within Mass Motion (Image courtesy of Erin Morrow, Arup)

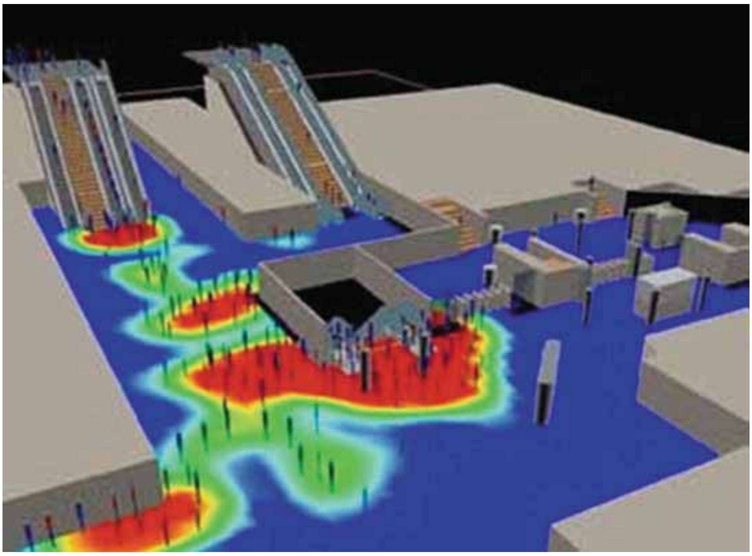


Fig. 18.18 Screenshot of the STEPS software

agent using randomised weights for the cost components of the routes. A simplified algorithm for the total route cost is given as:

$$cost = W_D * \left(\frac{D_G}{V} \right) + W_Q * Q + W_L * L \tag{18.1}$$

where W_D is the distance weight; D_G is the total distance from agent position to goal; V is the agent's desired velocity; W_Q is the queue weight, Q is the expected time in the queue before reaching the link entrance; W_L is the link traversal weight; and L is the link type cost (level, ramp, stair, etc.). The variables W_D, V, W_Q and W_L are random agent properties. The structure shown in Fig. 18.17 has the advantage of allowing the environmental geometry to be easily replaced, where the sparse node network will then update itself based on the new geometric relationships.

STEPS[5] is another agent-based simulation tool developed by Mott MacDonald and uses coarse grid geometry. Agents do not have a global view of the environment but move around towards a final exit. STEPS takes in CAD files for modelling geometry. It uses a CA where each individual occupies one cell at a given time. Agents know the shortest route but not the density or obstacles that they may encounter. Each agent then moves in the desired direction if the next cell is empty. The routes that agents take can generally be defined by a crude form of global view, mainly for normal operation, i.e. not during evacuation scenarios. Costs can be accounted for by the route taken. Each agent also has its own characteristics and familiarity behaviour. Figure 18.18 shows a screenshot of the software.

[5] http://www.mottmac.com/skillsandservices/software/stepssoftware/

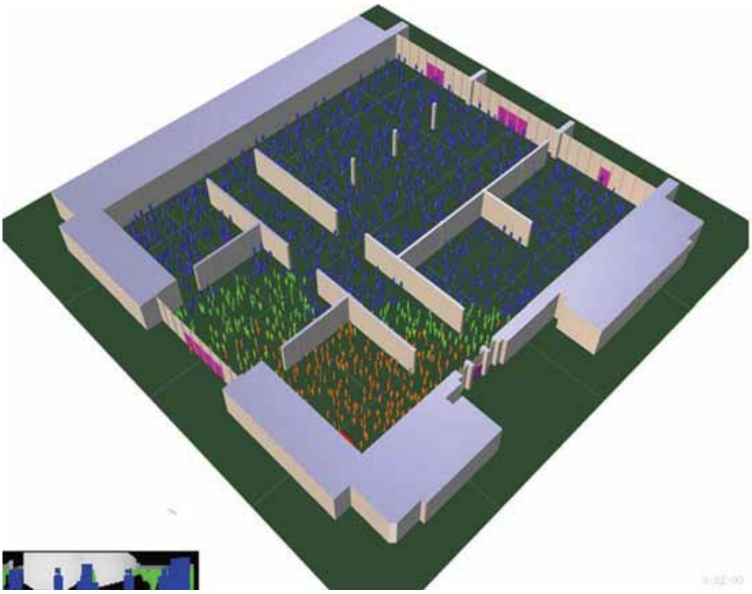


Fig. 18.19 Screenshot of the MASSIVE software

18.4.2 Entertainment

As the entertainment field is based on interactivity and real time simulation where computer games are concerned, a number of tools exist to facilitate the creation of interactive environments for a cinematographer or the user.

MASSIVE (Multiple Agent Simulation System in Virtual Environment) is one of the most advanced crowd animation and simulation systems for off-line productions (Fig. 18.19). It was first developed by Stephen Regelous in order to address the animation problems for the Lord of the Rings[6] trilogy, and is now used in many visual effects systems. The software does not contain any explicit Artificial Intelligence (AI), the agents can be assigned AI, and it uses a simple way to deploy intelligent agents (Thalmann et al. 2005). It does not use global path planning, and the user can author the entire crowd motion although autonomous agents can be introduced. It uses a hierarchical system where behavioural systems vary (Pettre et al. 2006), and at the lowest level, techniques such as potential fields can control the flow of the crowds (Thalmann and Musse 2007). MASSIVE uses a bottom-up approach, which leads to emergent behaviour based on reactive motion planning, with high user control.

The user specifies inputs such as vision and internal states. Through the use of fuzzy logic, agents then respond to the environment after taking into account the static and dynamic obstacles (such as other agents). In this way agents can endlessly walk

[6] http://www.lordoftherings.net/

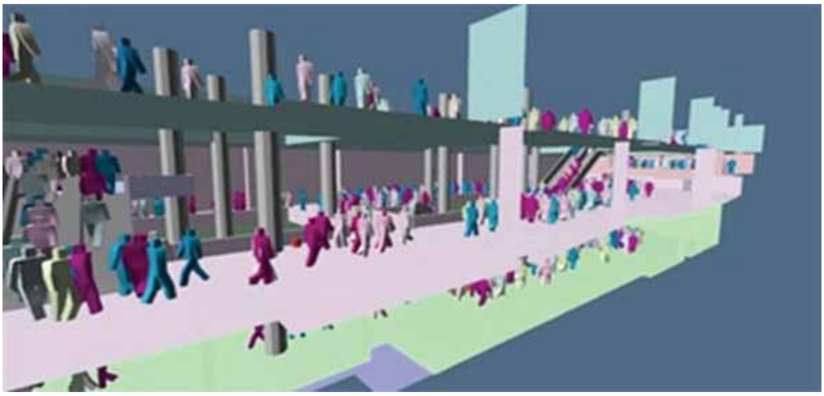


Fig. 18.20 Screenshot of the Spirops Crowd software

through the environment without the need for elaborate path planning. The realisation of long term goals of the agents in this case is not important, as cinematic shots average at around 5 s, and therefore, the main aim is high quality visualisation during the flow. The design decision in this case allows for a very large number of agents (hundreds of thousands) to be accommodated during the simulation. MASSIVE Insight[7] is a spin-off from MASSIVE, and is at a beta stage that is aimed for safety, architecture and urban applications. It uses the original bottom-up approach of MASSIVE and a global planning algorithm will be incorporated in the system.

There are a number of middleware tools available to author crowds for video games. Spirops[8] is a middleware tool for game development, which focusses on AI problems in the game development cycle (see Fig. 18.20). This allows the in-game characters to behave realistically. Spirops crowd is a component of Spirops focused on crowd simulation. It creates paths for the pedestrians to follow and avoid collisions. It was first developed by Axel Buendia based on his research thesis. Due to its commercial nature, not many details are available with regards to the path planning of pedestrians. However, it currently uses a hard linked behaviour to plan the global path although a future release will include dynamic planning. A screenshot is provided in Fig. 18.20.

PathEngine[9] is another toolkit to provide realistic movements of agents in virtual worlds such as games. The toolkit provides collision avoidance and paths for agents to follow. However, path finding is based on visibility of agents, and it does not provide a global path planning solution.

OpenSteer[10] has been developed by Reynolds based on his boids work (Reynolds 1987, 1999). It is an open source library written in C++ and uses OpenGL, which is available to build the steering behaviours of autonomous agents. It is a cross

[7] http://www.massivesoftware.com/real-world-simulation/

[8] http://www.spirops.com/

[9] http://www.pathengine.com/

[10] http://opensteer.sourceforge.net/doc.html.

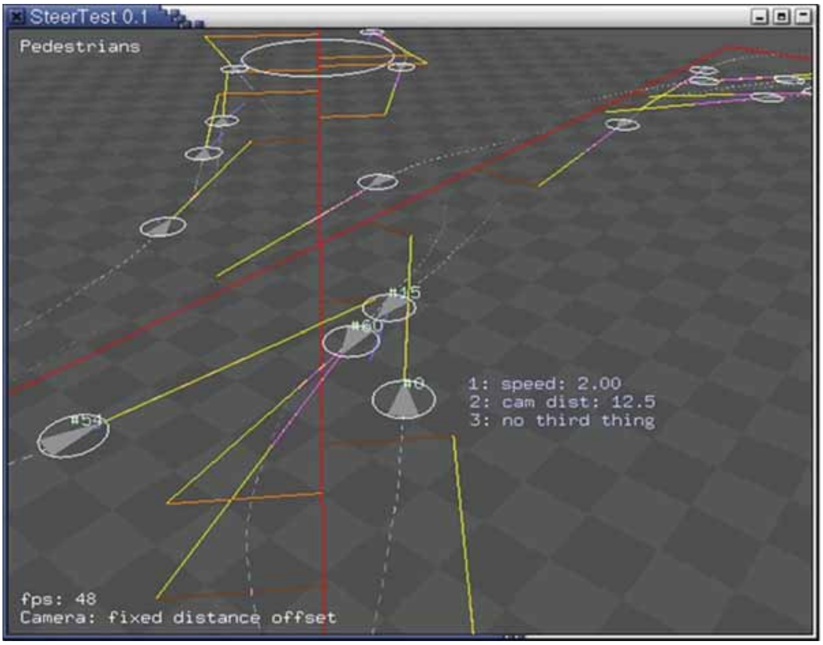


Fig. 18.21 OpenSteer demo window

platform toolkit available as a demo which demonstrates a number of steering behaviours supplied as sample plug-ins (Fig. 18.21). Bespoke plug-ins can be written by using the steering library provided, which defines the behaviours for the autonomous agents such as wander, goal seeking, flee, path following as well as collision avoidance. The individual steering behaviours as described in Reynolds (1999) combine to form a more complex behaviour. There are two ways it happens: either by switching between individual behaviours as the environment changes, or by combining the individual behaviours together, therefore, working in parallel with each other.

18.4.3 Virtual Reality

As this section deals with the convergence between two fields, not many tools exist. Research in academia has focused on the realistic behaviour of individuals combined with rendering huge numbers (hundreds of thousands) on desktop computers. Crowd simulation systems in this area have mainly been developed in the academic field. A few tools mainly relevant to this section will be described in greater detail.

ViCrowd is an academic tool developed by Musse (2000) and detailed in Musse and Thalmann (2001). It includes the three-tier hierarchy for the different degrees of autonomy, and uses Reynold's flocking model for behavioural rules. A screenshot

Fig. 18.22 Screenshot of ViCrowd (Taken from Musse and Thalmann (2001))

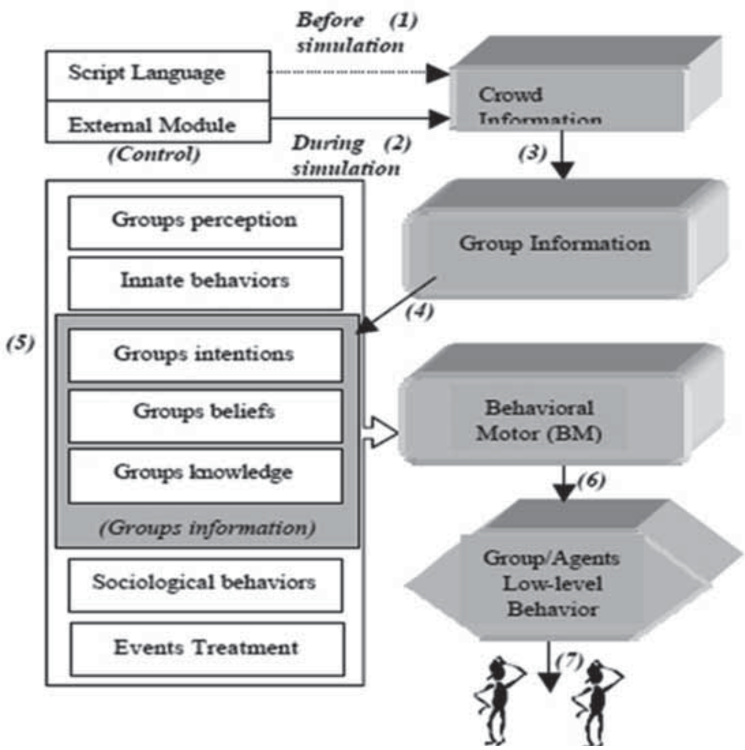


Fig. 18.23 Diagram representing the behavioural algorithm (Taken from Musse and Thalmann (2001))

of ViCrowd can be seen in Fig. 18.22. The architecture for the tool is shown in Fig. 18.23. A behaviour motor exists to process various low level behaviours such as perception and other sociological behaviours. The architecture of ViCrowd is described below, and the steps are indicated in Fig. 18.23.

18.4.3.1 Procedure

Step 1: Scripted behaviours are specified before the simulation takes place
Step 2: The user may also provide behaviours for the crowd to follow during the simulation.
Step 3: The behaviours indicated in the last two steps, as well as information emerging using behavioural rules, are then distributed among the groups.
Step 4: The group information that is passed describes the groups' goals, intentions and knowledge.
Step 5: This information is then combined with other information such as the events and other sociological behaviours. These sociological behaviours include beliefs describing the internal state of the groups, and knowledge representing the environmental information.
Step 6: The behavioural motor represents the process where the low level behaviour of the groups and individuals is generated.
Step 7: The low level behaviour of the groups is then created such as goals, walking speed and actions.

The behaviour motor's process uses a number of priority rules for the behaviours. It first analyses the group information where the low level behaviours for each group are created. If sociological factors exist, the low level behaviours can be changed again by getting the agents to change groups. If certain events occur, the reaction defined by the user or scripted behaviour can also change the group's low level behaviour. This low level behaviour is then passed to the individual agents of the group, where perception and innate abilities are used to satisfy the group's intentions. This procedure described allows for complex behaviours of crowds to be simulated in real time.

Crowdbrush is another tool that comes out of the same research lab as ViCrowd. This tool was designed and developed by Ulicny et al. (2004) to make the authoring of complex crowd scenes with different scenarios simpler. It introduces the capability to create, modify and control the crowd member of a scene in real time with visual feedback. A brush metaphor allows for crowd authoring. The main focus of the tool is to author crowd scenes, which is beyond the scope of this section. Crowdbrush does allow the creation of pedestrian paths, but is done manually, and no global motion planning algorithms are involved. Low level behaviours are defined using a rule-based behaviour engine (Ulicny and Thalmann 2002). A simple reactive rule system is defined in order to achieve fast real-time simulation of thousands of agents. The agents use simple displacement of humans when reacting to any type of internal or external event defined by the behavioural rules. A simple collision avoidance system also exists based on Helbing and Molnár's (1995) social forces model. Collision queries are minimised by using a bin-lattice space subdivision (Reynolds 2000). The behaviour can be applied both directly and indirectly in real-time using the same brush metaphor used for authoring crowd scenes by sending events to activate behavioural rules or tagging agents with a tagging brush that selectively triggers different behavioural rules for different agents. This allows for various actions happening in different parts of the crowd scene.

18.5 Summary

As demonstrated through the three application areas within this chapter, there are multiple micro and macro simulation techniques dependent on the environment to be modelled and the industry involved. In-house or third party toolkits are common place, each with their own pros and cons and methods to achieve the desired simulation. Packages such as 3ds Max are available free of charge to the global academic community, yet their use is limited outside of the entertainment field. As the simulation industry widens in scope and moves towards more real time inputs and outputs rather than pre-computed ones, we expect this to change. Simulation is increasingly moving towards a real-time input/output environment as processing speeds and data collection techniques advance.

References

Batty, M., Desyllas, J., & Duxbury, E. (2003). The discrete dynamics of small-scale spatial events: Agent-based models of mobility in carnivals and street parades. *International Journal of GIS, 17*(7), 673–697.

Bayazit, O. B., Lien, J.-M., & Amato, N. M. (2002). Roadmap-based flocking for complex environments. In *Proceedings of the 10th Pacific conference on computer graphics and applications* (p. 104). New York: IEEE Computer Society.

Bouvier, E., Cohen, E., & Najman, L. (1997). From crowd simulation to airbag deployment: Particle systems, a new paradigm of simulation. *Journal of Electronic Imaging, 6*(1), 94–107.

Braun, A., Musse, S. R., de Oliviera, L. P. L., & Bodmann, B. E. J. (2003). Modeling individual behaviors in crowd simulation. *Proceedings of the 16th International Conference on Computer Animation and Social Agents*. Los Alamitos: IEEE Computer Society. http://doi.ieeecomputersociety.org/10.1109/CASA.2003.1199317

Castle, C. J. E. (2006). *Developing a prototype agent-based pedestrian evacuation model to explore the evacuation of King's Cross St Pancras underground station*. London: Centre for Advanced Spatial Analysis (UCL). http://eprints.ucl.ac.uk/3327/. Sept 2006.

Challenger, W., Clegg W. C., & Robinson A. M. (2009). *Understanding crowd behaviours: Guidance and lessons identified*. Technical Report prepared for UK Cabinet Office, Emergency Planning College, University of Leeds, 2009.

Chenney, S. (2004). Flow tiles. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on computer animation* (pp. 233–242). Grenoble: Eurographics Association.

Cordeiro, O. C., Braun, A., Silveria, C. B., Musse, S. R., & Cavalheiro, G. G. (2005). Concurrency on social forces simulation model. *First International Workshop on Crowd Simulation*. EPFL, Lausanne.

Crooks, A. T., & Heppenstall, A. J. (2012). Introduction to agent-based modelling. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 85–105). Dordrecht: Springer.

Galea, E. R., Galparsoro, J. M. P., & Pearce, J. (1993). A brief description of the EXODUS evacuation model. *Proceedings of the International Conference on Fire Safety*, *San Francisco*, Vol 18 (pp. 149–162).

Helbing, D., Farkas, I. J., & Vicsek, T. (2000). Simulating dynamical features of escape panic, Nature, 407, pp. 487–490.

Helbing, D., & Molnár, P. (1995). Social force model for pedestrian dynamics. *Physical Review E, 51*(5), 4282–4286.

Helbing, D., Molnár, P., Farkas, I. J., & Bolay, K. (2001). Self-organizing pedestrian movement. *Environment and Planning B: Planning and Design, 28*(3), 361–383.

Henderson, L. F. (1971). The statistics of crowd fluids. *Nature, 229*(5284), 381–383.

Hughes, R. L. (2003). The flow of human crowds. *Annual Review of Fluid Mechanics, 35*(1), 169–182.

Iltanen, S. (2012). Cellular automata in urban spatial modelling. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 69–84). Dordrecht: Springer.

Jennings, N. R., Sycara, K., & Wooldridge, M. (1998). A roadmap of agent research and development. http://eprints.ecs.soton.ac.uk/2112/

Johnasson, A., & Kretz, T. (2012). Applied pedestrian modeling. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 451–462). Dordrecht: Springer.

Kamphuis, A., & Overmars, M. H. (2004). Finding paths for coherent groups using clearance. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on computer animation* (pp. 19–28). Grenoble: Eurographics Association.

Kisko, T. M., Francis, R. L., & Nobel, C. R. (1998). *Evacnet4 user's guide*. University of Florida.

Lamarche, F., & Donikian, S. (2004). Crowd of virtual humans: A new approach for real time navigation in complex and structured environments. *Computer Graphics Forum, 23*(3), 509–518.

Lau, M., & Kuffner, J. J. (2005). Behavior planning for character animation. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on computer animation* (pp. 271–280). Los Angeles: ACM.

Lau, M., & Kuffner, J. J. (2006). Precomputed search trees: Planning for interactive goal-driven animation. In *Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on computer animation* (pp. 299–308). Vienna: Eurographics Association.

Lerner, A., Chrysanthou, Y., & Cohen-Or, D. (2006). Efficient cells-and-portals partitioning: Research articles. *Computer Animation and Virtual Worlds, 17*(1), 21–40.

Lien, J- M., Rodriguez, S., Malric, J., & Amato, N. M. (2005). Shepherding behaviors with multiple shepherds. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation* (pp. 3402–3407).

Loscos, C., Marchal, D., & Meyer, A. (2003). Intuitive crowd behaviour in dense urban environments using local laws. In *Theory and practice of computer graphics (TPCG'03)* (pp. 122–129). Los Alamitos: IEEE Computer Society.

Musse, S. R. (2000). *Human crowd modelling with various levels of behaviour control*. Unpublished PhD thesis, EPFL.

Musse, S. R., & Thalmann, D. (2001). Hierarchical model for real time simulation of virtual human crowds. *IEEE Transactions Visualization and Computer Graphics, 7*(2), 152–164.

Narain, R., Golas, A., Curtis, S., & Lin, M. C. (2009). Aggregate dynamics for dense crowd simulation. In *Proceedings of the 2009 ACM SIGGRAPH/Eurographics symposium on computer animation*. New York: ACM.

Pelechano, N., Allbeck, J. M., & Badler, N. I. (2008). Virtual crowds: Methods, simulation and control. *Synthesis Lectures on Computer Graphics and Animation, 3*(1), 1–176.

Pelechano, N., & Badler, N. I. (2006). Modeling crowd and trained leader behavior during building evacuation. *IEEE Computer Graphics and Applications, 26*(6), 80–86.

Pettre, J., Ciechomski, P. H., Maim, J., Yersin, B., Laumond, J.-P., & Thalmann, D. (2006). Realtime navigating crowds: Scalable simulation and rendering. *Computer Animation and Virtual Worlds, 17*(3–4), 445–455.

Pettre, J., Laumond, J., & Thalmann, D. (2005). A navigation graph for real-time crowd animation on multilayered and uneven terrain. Presented at the *First International Workshop on Crowd Simulation (V-CROWDS'05),* Lausanne.

Reynolds, C. (1987). Flocks, herds and schools: A distributed behavioral model. *Computer Graphics, 21*(4), 25–34.

Reynolds, C. (1999). Steering behaviours for autonomous characters. *Proceedings of the 1999 Game Developers Conference, San Jose, California* (pp. 763–782). URL: http://www.red.com/cwr/steer/

Reynolds, C. (2000). Interaction with groups of autonomous characters. *Proceedings of the 2000 Game Developers Conference,* San Francisco.

Reynolds, C. (2006). Big fast crowds on PS3. In *Proceedings of the 2006 ACM SIGGRAPH symposium on videogames* (pp. 113–121). Boston: ACM.

Shao, W., & Terzopoulos, D. (2005). Autonomous pedestrians. In *Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on computer animation* (pp. 19–28). Los Angeles: ACM.

Still, G. (2000). *Crowd dynamics*. Unpublished PhD thesis, Warwick University, Coventry.

Sud, A., Gayle, R., Andersen, E., Guy, S., Lin, M., & Manocha, D. (2007). Real-time navigation of independent agents using adaptive roadmaps. In *Proceedings of the 2007 ACM symposium on virtual reality software and technology* (pp. 99–106). Newport Beach: ACM.

Sung, M., Kovar, L., & Gleicher, M. (2005). Fast and accurate goal-directed motion synthesis for crowds. In *Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on computer animation* (pp. 291–300). Los Angeles: ACM.

Tecchia, F., & Chrysanthou, Y. (2000). Real-time rendering of densely populated urban environments. In *Proceedings of the eurographics workshop on rendering techniques 2000* (pp. 83–88). New York: Springer.

Tecchia, F., Loscos, C., Conroy-Dalton, R., & Chrysanthou, Y. L. (2001). Agent behaviour simulator (ABS): A platform for urban behaviour development. In *Proceedings of the first international game technology conference and idea expo (GTEC'01)*. Hong Kong: ACM.

Teller, S. J. (1992). *Visibility computations in densely occluded polyhedral environments*. Unpublished PhD thesis, University of California, Berkeley.

Thalmann, D., Kermel, L., Opdyke, W., & Regelous, S. (2005). Crowd and group animation. In *ACM SIGGRAPH 2005 courses* (p. 1). Los Angeles: ACM.

Thalmann, D., & Musse, S. R. (2007). *Crowd simulation*. London: Springer (Online service, and Adriana Braun).

Thompson, P. A., & Marchant, E. W. (1994). *Simulex; developing new computer modelling techniques for evaluation*. Unit of Fire Safety Engineering, University of Edinburgh.

Treuille, A., Cooper, S., & Popovic, Z. (2006). Continuum crowds. *ACM Transaction on Graph, 25*(3), 1160–1168.

Tu, X., & Terzopoulos, D. (1994). Artificial fishes: Physics, locomotion, perception, behavior. In *Proceedings of the 21st annual conference on computer graphics and interactive techniques* (pp. 43–50). New York: ACM.

Ulicny, B., Ciechomski, P., & Thalmann, D. (2004). Crowdbrush: Interactive authoring of real-time crowd scenes. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on computer animation* (pp. 243–252). Grenoble: Eurographics Association.

Ulicny, B., & Thalmann, D. (2002). Towards interactive real-time crowd behavior simulation. *Computer Graphics Forum, 21*(4), 767–775.

Part IV
Fine-Scale, Micro Applications of Agent-Based Models

Chapter 19
Using Agent-Based Models to Simulate Crime

Nicolas Malleson

Abstract Due to the complexity of human behaviour and the intricacies of the urban environment, it is extremely difficult to understanding and model crime patterns. Nevertheless, a greater understanding of the processes and drivers behind crime is essential for researchers to be able to properly model crime and for policy-makers to be able to predict the potential effects of their interventions. Traditional mathematical models that use spatially aggregated data struggle to capture the low-level dynamics of the crime system – such as an individual person's behaviour – and hence fail to encapsulate the factors that characterise the system and lead to the emergence of city-wide crime rates.

This chapter will outline a realistic agent-based model that can be used to simulate, at the level of individual houses and offenders, occurrences of crime in a real city. In particular, the research focuses on the crime of residential burglary in the city of Leeds, UK. The model is able to predict which places might have a heightened burglary risk as a direct result of a real urban regeneration scheme in the local area.

19.1 Introduction

Understanding the processes and drivers behind crime is an important research area in criminology with major implications for both improving policies and developing effective crime prevention strategies (Brantingham and Brantingham 2004; Groff 2007). Advances in environmental criminology theory (e.g. Cohen and Felson 1979; Clarke and Cornish 1985; Brantingham and Brantingham 1993) have highlighted a shift in the field towards understanding the importance of the social and environmental contexts in which crimes occur, rather than focussing purely the behaviour of offenders.

N. Malleson (✉)
School of Geography, University of Leeds, Leeds, UK
e-mail: n.malleson06@leeds.ac.uk

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_19, © Springer Science+Business Media B.V. 2012

Furthermore, the complexity of the crime system – which consists of the dynamic interactions between the individuals involved in each crime event as well as their interactions with others and with their environment – means that individual-level approaches are the most suitable modelling methodologies for simulating the crime system.

This chapter will discuss how agent-based models (ABM's), coupled with realistic geographic environments, can be used to simulate crime. In particular, it will focus on the crime of residential burglary and outline a current agent-based simulation model that can be used to make predictions about future burglary rates in the real world. The model described is based on the city of Leeds, UK.

The chapter is organised as follows. The next section will outline the important drivers of the crime system that must be included in a model followed by a discussion on how crime has been modelled previously. The remainder of the chapter will then discuss a model that can be used to simulate residential burglary and will demonstrate how it can be used to simulate the effects that urban-regeneration can have on burglary.

19.2 Background: Environmental Criminology

Crime is a highly complex phenomenon. An individual crime event is the result of the convergence of a multitude of different factors including the motivations and behaviours of the offender, influences of the physical surroundings, community-wide effects such as community cohesion, the actions of the victim and the behaviour of other people such as the police or passers-by. Associated with this already complex framework are additional factors such as a diverse urban geography and obscure human psychology.

Criminology can help to understand patterns of crime. However, pre-1970 criminology research was largely dominated by studies into victims, the law and offenders (Andresen 2010) and thus omitted a vital element; the *place* in which the crime occurs. It was to this end that the field of "environmental criminology" arose as a discipline to study the *spatial* variations of crime and the underlying reasons for these variations (Johnson et al. 2002). The remainder of this section will discuss examples from environmental criminology research for a crime model. Although the focus is on the crime of residential burglary, many of the factors are relevant for most other types of inquisitive crime.

19.2.1 Physical Factors

Major advancements in criminological theory in the 1970s solidified the link between the physical form of an area and its affect on crime (Jeffery 1971; Newman 1972). With respect to burglary, the important physical factors that determine a house's vulnerability can be classified into three groups as identified by Cromwell et al. (1991).

The first group, *accessibility*, relates to how easy it is to actually enter a property. For example, detached houses and ground-floor flats have been found to be vulnerable because there are more potential entry points (Robinson and Robinson 1997; Felson 2002). The second category of physical factor that might influence burglary is *visibility* and refers to the extent to which a residence can be seen by neighbours and passers-by (Cromwell et al. 1991). Buildings that are less visible are generally easier for offenders to access without being seen by others. Visibility can be affected by objects such as large hedges or other buildings that can obscure the view of the property as well as factors like the distance between the house and its connecting road, levels of street lighting and the amount of passing traffic. Finally, *occupancy* represents whether the residents are at home or not.

19.2.2 The Social Environment

Although physical factors are clearly important determinants of burglary risk, the "environmental backcloth" (Brantingham and Brantingham 1993) extends well beyond these simple physical factors. It is also important to consider the *social* factors that surround a crime event. Unfortunately, whereas the relationship between physical factors and burglary risk is often fairly straightforward, that of the social environment and crime is not. For example, deprived communities often suffer disproportionately high crime rates (Baldwin and Bottoms 1976; Sampson et al. 1997) but the reverse has also been found (Wilkström 1991; Bowers and Hirschfield 1999).

Fortunately, the relationship between other variables is more straightforward. Students, for example, are often a highly victimised group (Tilley et al. 1999; Barberet et al. 2004) as student households are often seen as an easy targets (Deakin et al. 2007) and can contain an abundance of attractive goods. Other demographic factors that can increase burglary risk include the age of residents, the tenure type (e.g. publicly rented compared to privately owned) and the number children/young people in the area (Tilley et al. 1999).

Another factor that is not necessarily related to socioeconomic status, but can have a strong impact on crime rates, is community cohesion. It is hypothesised that if a community looses the ability to police itself then crime is the "natural response" by individuals. This process can occur when an area contains a transient population as people do not stay in area long enough for make friends and develop a feeling of "community" and ownership over the area. The importance of community cohesion is evidenced by the seminal theories it has provoked (e.g. Shaw and McKay 1942; Jeffery 1971; Newman 1972; Wilson and Kelling 1982) and by the large body of empirical research that supports it (Hope 1984; Brown and Bentley 1993; Wright and Decker 1996; Sampson et al. 1997; Kawachi et al. 1999).

In summary, this section has illustrated that the relationship between crime and the surrounding environment is complex. In order to model the system, it must be determined if a high crime rate is due to the types of housing in the area, the houses' physical properties, the number of and behaviour of potential burglars, the amount of community cohesion or for other reasons that have yet to be identified. However, using the appropriate methodology it is nevertheless possible to account for all these features in a crime model as the following section will discuss.

19.3 Modelling Crime

19.3.1 The Geography of Crime

Since the first pioneering work on the geography of crime in the nineteenth century (Quetelet 1831; Glyde 1856), crime research has moved to smaller and smaller units of analysis. However, with the exception of a small number of "crime at place" studies (e.g. Eck 1995; Weisburd et al. 2009), most research still uses aggregated data and there has been very little work into what the most appropriate unit of analysis should be (Weisburd et al. 2009). Modern environmental criminology theories (e.g. Cohen and Felson 1979; Brantingham and Brantingham 1981; Clarke and Cornish 1985) suggest that an individual crime depends on the behaviour of *individual* people or objects and should thus be analysed at the level of the individual (Weisburd et al. 2004). This is extremely relevant with the crime of burglary because burglars choose *individual* homes based on their *individual* characteristics (Rengert and Wasilchick 1985). Models that uses aggregate-level crime or demographic data will therefore suffer, to a greater or lesser extent, from the ecological fallacy (Robinson 1950). Indeed, recent crime research has shown that individual- or street-level events exhibit considerable spatial variation which would be hidden if analysed at even the smallest administrative boundaries (Bowers et al. 2003; Weisburd et al. 2004; Groff et al. 2009; Andresen and Malleson 2011).

That said, the majority of crime models to date employ regression techniques and look for relationships using aggregate data. For a review of commonly used approaches the reader is directed to Kongmuang (2006) but, in general, the central drawback is that statistical models fail to address the importance of the individual: individual people, incidents, locations and times.

Following this, ABM appears to be the most appropriate methodology for modelling crime and the following section will explore the use of ABM for crime analysis in more detail.

19.3.2 Agent-Based Crime Modelling

19.3.2.1 Advantages and Disadvantages

An obvious advantage with ABM is its ability to capture emergent phenomena. Environmental criminology research tells us that the geographical patterning of crime rates is an emergent phenomenon, resulting from the interactions between individual people and objects in space. Only "bottom-up" approaches truly capture this phenomenon.

Closely related to it ability to reproduce emergent phenomena is the ability of ABM to create a *natural description* of the system under observation (Bonabeau 2002). There are many systems, particularly in the social sciences, that cannot be sensibly modelled using mathematical equations (Axtell 2000; O'Sullivan 2004; Moss and Edmonds 2005). Because, with an agent-based model, rules are specified directly for each individual unit there is no need to try to coax a higher-level model into performing as if it were modelling individuals directly. Therefore, by using ABM the "natural variety" of cities becomes part of the model, rather than smoothed out by aggregate methods (Brantingham and Brantingham 2004).

Of course there are some disadvantages to using agent-based modelling for crime analysis. Crime systems are highly dependent on human characteristics such as seemingly irrational behaviour and complex psychology. However, formally defining these characteristics in a computer model is extremely difficult and can lead to reduced behavioural complexity (O'Sullivan and Haklay 2000). If the behavioural complexity of the agents is adequate, then computation power can become a problem as each decision made by each agent becomes more computationally expensive.

19.3.2.2 Incorporating Geography

To gain a better understanding of the spatial nature of crime, geographic information systems (GIS) are routinely used to analyse crime data sets (Hirschfield et al. 2001) and are becoming an increasingly important tool for crime analysts (Chainey and Smith 2006; Weir and Bangs 2007). They are also being used for another purpose; agent-based crime modelling.

In order to make predictive analyses (i.e. predicting future crime rates in a real city or neighbourhood) it is essential that the environment is a realistic representation of the physical area under study. Therefore the coupling of agent-based models with GIS is essential. This is not such a daunting task as it once was as many toolkits are now available to support researchers in this activity such as Repast Simphony (North et al. 2005a, b) and Agent Analyst (The Redlands Institute 2009).

However, a researcher must be aware that incorporating a GIS with an ABM can result in an *overly-complex* model that is as difficult to understand as the underlying system itself. Too much complexity can detract from our understanding of the dynamics that are at the heart of the system (Elffers and van Baal 2008). As Axelrod (1997) notes, if the goal of a simulation is to more fully understand the underlying dynamics then it is the fundamental model assumptions which are important, not the accuracy of the surrounding environment.

19.3.2.3 Existing Agent-Based Crime Models

Following the remarks made by eminent environmental criminologists (such as Brantingham and Brantingham 1993), researchers are starting to realise the benefits of ABM for studying crime. Initial models, (e.g. Gunderson and Brown 2000; Winoto 2003; Melo et al. 2005; Malleson et al. 2009, 2010) were relatively simple and did not necessarily incorporate realistic urban environments. They were typically used to explore theory or determine how changing variables such as offender motivation or police behaviour impacted on offending rates. More recently, advanced models have begun to emerge that can explore crime rates in real cities and can be used to make real-world predictions. For example: Dray et al. (2008) used ABM to explore drug market dynamics in Melbourne; Liu et al. (2005) present an agent-based/cellular-automata model of street robbery in the city of Cincinatti; Birks et al. (2008) and Hayslett-McCall et al. (2008) have independently developed agent-based burglary simulations; and Groff and Mazerolle (2008) have developed an urban simulation for street robbery with a realistic vector road network. It is not possible to discuss these models in more detail here. For more information about current agent-based crime modelling applications the reader is directed to the recent book entitled "*Artificial Crime Analysis Systems: Using Computer Simulations and Geographic Information Systems*" (Liu and Eck 2008) or a special issue of the Journal of Experimental Criminology entitled "*Simulated Experiments in Criminology and Criminal Justice*" (Groff and Mazerolle 2008).

19.4 A Simulation of Burglary

Having suggested that ABM is the most appropriate methodology for modelling crime, this section will strengthen the case for ABM by outlining, in detail, an advanced burglary simulation. Then Sect. 19.5 will show how the model can be used to predict crime patterns after an urban regeneration scheme. For more information about any aspects of the model, the interested reader is directed to Malleson (2010).

19.4.1 The Virtual Environment

The virtual environment is the space that the agents inhabit and, in a crime model, must incorporate many of the factors that form the "environmental backcloth" (Brantingham and Brantingham 1993). Along with a road and public transport networks that the agents can use to navigate the city, the environment must include individual buildings – to act as homes for the agents and as potential burglary targets – and community-wide factors such as deprivation and community cohesion.

19.4.1.1 The Community Layer

In Sect. 19.2 it was noted that people other than the offender can have an affect on crime by acting as victims or guardians. This is particularly relevant to burglary because an offender is unlikely to attempt to burgle if they are aware that the house is occupied or if they are being observed by passers-by. In an ABM, people are represented as agents. This approach demonstrated success when it was included in a burglary model that operated on an abstract environment (Malleson et al. 2010).

However, creating a simulation of every person in a *real city* is an immense undertaking. Instead, the behaviour of people other than offenders can be simulated through a *community* layer in the virtual environment. In this manner, factors that would otherwise originate directly from agent behaviour can be estimated for each community based on the socio-demographic information about that community. For example, houses in student communities are likely to be vacant at different times (e.g. in the evenings) than communities who predominantly house families with small children. Rather than simulating individual household behaviour, it is possible to *estimate* occupancy rates for the whole community based on demographic data.

UK data for the layer can be extracted from the 2001 UK census (Rees et al. 2002b) and also from deprivation data published by the UK government such as the Index of Multiple Deprivation (Noble et al. 2004).[1] These data can then be spatially referenced through the use of administrative boundary data available through the UKBORDERS service (EDiNA 2010). It was noted in Sect. 19.3 that the use of administratively-defined areal boundaries can pose serious problems to research because the boundaries are not designed to be homogeneous. To mediate these problems in this research, individual-level data will be used wherever possible (houses and roads, for example, are represented as individual geographic objects).

An obvious requirement of the community layer is a measure of *occupancy*. In this simulation, occupancy is calculated at different times of day based on the proportions of the following demographic variables: *students*; *working part time*; *economically inactive looking after family*; *unemployed*. These four variables were chosen because they are able to represent common employment patterns. Another important relationship noted in Sect. 19.2 was that *community cohesion* has a large influence on crime; residents in cohesive communities are more likely to be mindful of their own and their neighbours' property. For this model, community cohesion is calculated from three variables that have been identified in the literature (Shaw and McKay 1969; Sampson et al. 1997; Bernasco and Luykx 2003; Browning et al. 2004) as important: *concentrated disadvantage*; *residential stability*; *ethnic heterogeneity*. With the exception of concentrated disadvantage which is obtained directly from the Index of Multiple Deprivation, all other variables can be established from the UK census.

In a similar manner to community cohesion, research has shown that potential burglars feel more comfortable in areas that are similar to their own because they do not feel that they will "stand out" (Wright and Decker 1996). This concept can be formalised through the creation of a *sociotype* which is a vector containing values for all the available census and deprivation data for each area. Therefore, the similarity between a target community and a burglar's home community can be calculated as the Euclidean distance between the two sociotypes.

---

[1] Census data is published through CASWEB (Mimas 2010), For more information about the census see Rees et al. (2002a, 2002b)

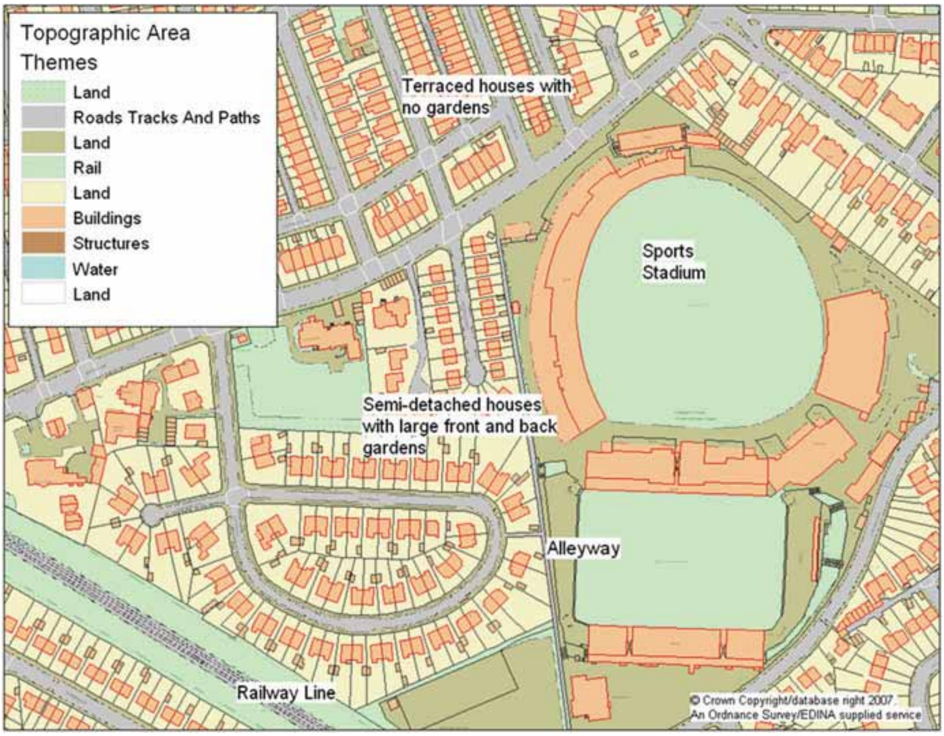


Fig. 19.1 An example of the OS MasterMap Topography layer which shows how different types of houses can be distinguished and the types of geographic objects that *could* be included in a crime model

The final community-level variable, *attractiveness*, incorporates a measure of the affluence of the target community and therefore the potential available returns from burglary. Ideally this would be calculated individually for each property but in the absence of individual-level affluence data a community-wide variable must be used, based on census data. Evidence suggests that the following census variables provide good affluence measures: *percentage of full time students*; *mean number of rooms per household*; *percentage of houses with more than two cars*; and *percentage of people with higher education qualifications* (Bernasco and Luykx 2003; Kongmuang 2006).

19.4.1.2 The Buildings Layer

For the burglary simulation discussed here, Ordnance Survey MasterMap data (Ordnance Survey 2009) was used to represent the virtual environment in a highly detailed way. The product contains a number of different "layers" which can, separately, be used to represent the network of roads as well as other features such as buildings, rivers, parks etc. Figure 19.1 illustrates the Topography layer which is used in the model to create residential houses. Some cleaning and filtering processes

Fig. 19.2 Number of adjacent neighbours, size of garden and the number of neighbours within 50 m. All normalised to the range 0–1

were required to extract *houses* from the set of all buildings (which includes structures such as cinemas, shopping centres, garages etc.) but otherwise the data is ready for input.

Along with the variables that represent household attractiveness and occupancy – which are modelled at the level of the community because insufficient individual-level data are available – Sect. 19.2 identified the following factors as important determinants of household burglary risk:

- **Accessibility** – how easy it is to gain entry to the house (e.g. the number of windows or doors);
- **Visibility** – the level of visibility of the house to neighbours and passers-by;
- **Security** – effective physical security e.g. dogs or burglar alarms;

Parameter values for *accessibility* and *visibility* can be calculated directly through an analysis of the geographic household boundary data. Visibility can be calculated by using a GIS to compute both the size of the garden that surrounds each property and the number of other properties within a given buffer distance. Using similar geographic methods, the accessibility of the house can be estimated by determining if the house is detached, semi-detached or terraced by counting the number of adjacent buildings to the house. Figure 19.2 presents values for these variables normalised into the range 0–1. Although the geographical techniques are coarse and there are some errors (for example some terraced houses towards the north of the map have a larger number of neighbours than should be expected) they are able to broadly distinguish between the different physical house attributes that will influence burglary.

With regards to household *security*, there is unfortunately limited national or local data that can be used to estimate individual household security precautions. Generally, therefore, this value is set to be the same for every house so does not influence household burglary risk.

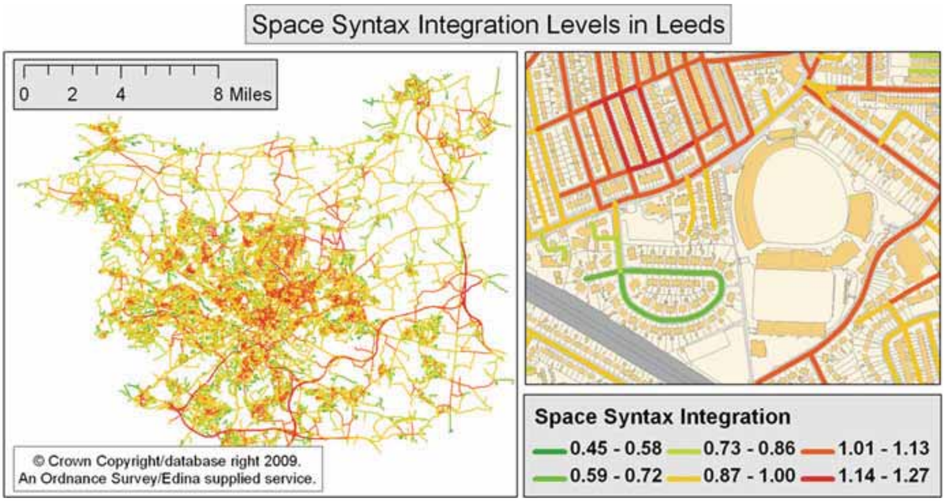


Fig. 19.3 Space syntax integration values for the entire city and a local area

19.4.1.3 The Transport Network

Transport networks are required in a geographic crime model because they restrict the agents' movements to certain paths and affect where and how the agents navigate the city. To include virtual roads, the Integrated Transport Network (ITN) MasterMap layer can be used. The ITN layer consists of line objects that represent all the different types of roads, including alleyways, motorways, pedestrianised areas etc. Using these data it is also possible to vary the speed that agents travel around the environment based on the transportation available to them.

Through an analysis of the roads data, it is possible to estimate the traffic volume on each road and this can affect the burglary risk associated with the houses on the road. Although most evidence suggests that houses which are situated on busy roads have a heightened burglary risk because they are more likely to be known by potential burglars (Brantingham and Brantingham 1993; Beavon et al. 1994), it is also possible that houses on busy roads are *less* of a risk at certain times of day because gaining undetected access can be more difficult.

Estimating traffic volume can be accomplished by using theories from the "space syntax" research area and analysing the *connectivity* of the road network.[2] Roads that are the most "integrated" (i.e. the most highly connected) have been found to correlate with large amounts of pedestrian and vehicle traffic and have been used in other crime studies (van Nes 2006). Figure 19.3 illustrates the integration values for all Leeds roads.

[2]For more information about space syntax techniques, refer to Hiller and Hanson (1984), Bafna (2003) or Park (2005).

19.4.2 The Burglar Agents

In the social sciences, agent-based models often use agents to represent people and this poses a substantial challenge: how should complex human psychology be included in a computer model? This section will address this issue and discuss how the burglar agents have been constructed for the burglary simulation.

19.4.2.1 Modelling Human Behaviour

Including human behavioural characteristics in agents – such as seemingly irrational behaviour and complex psychology (Bonabeau 2002) – can be a very difficult task to accomplish. However, agent cognitive architectures exist that can simplify the process of building a cognitively-realistic human agent. The most commonly used architecture is "Beliefs-Desires-Intentions" where *beliefs* represent the agent's internal knowledge of the world (i.e. its memory); *desires* represent all the goals which the agent is trying to achieve; and *intentions* represent the most important goals which the agent chooses to achieve first. Although the BDI architecture has been widely used (Rao and Georgeff 1995; Müller 1998; Taylor et al. 2004; Brantingham et al. 2005a, b), it has also suffered some criticism due mainly to its reliance on practical reasoning. No action is performed without some form of deliberation (Balzer 2000) but people rarely meet the requirements of rational choice models (Axelrod 1997).

A less widely used architecture is "PECS" (Schmidt 2000; Urban 2000) which stands for "Physical conditions, Emotional states, Cognitive capabilities and Social status". The authors of the architecture propose that it is possible to model the entire range of human behaviour by modelling those four factors. PECS is seen as an improvement over BDI because it does not assume rational decision making and is not restricted to the factors of beliefs, desires and intentions (Schmidt 2000). Instead, an agent has a number of competing *motives* (such as "clean the house", "eat food", "raise children", "sleep" etc.) of which the strongest ultimately drives the agent's current behaviour. Motives depend on the agent's internal state (an agent with a low energy level might feel hungry) as well as other external factors (an agent who smells cooking food might become hungry even if they do not have low energy levels). Personal preferences can also come into play, where some people feel a need more strongly than others even though their internal state variable levels are the same.

19.4.2.2 The Burglar Agents

The first decision to be made regarding the agents' behaviour is what internal state variables should be used as these, ultimately, dictate the range of possible motives

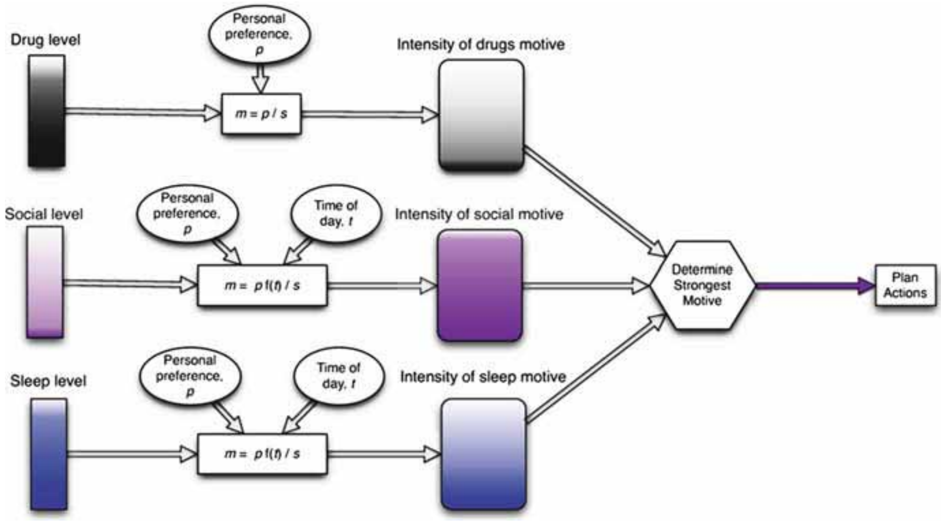


Fig. 19.4 How state variables, *s*, personal preferences, *p* and external factors (e.g. the time of day, *t*) are used in intensity functions to determine the strongest motive. In this example, the agent's *social* level is very low (the agent has not socialised in some time) and this is the strongest motive. The agent will make a plan that ultimately allows it to socialise (this could include burgling to make money first)

and behaviours. From the crime literature, it is apparent that a common motivation for burglary is the need to sustain a drug addiction or to maintain "high living" (i.e. socialising). Therefore, drug taking and socialising should be included as well as the ability to sleep when necessary.[3] With these behaviours in mind, the following state variables are sufficient:

- *Drugs* – the level of drugs in an agent's system. An agent's motivation to take drugs is based on the level of drugs in their system and a *personal preference* for drugs (i.e. how heavily they are addicted).
- *Sleep* – a measure of the amount of sleep an agent has had. The need for sleep is stronger at night than during the day.
- *Social* – a measure of how much the agent has socialised, felt more strongly during the day.

Levels of these internal state variables decrease over time and, as they decrease, the agents will be more strongly motivated to increase them. Figure 19.4 illustrates how state variable levels are combined with personal preferences and external factors (the time of day in this case) to determine the strongest motive which will drive an agent's behaviour. Although sleep can simply be sought at home, taking drugs and socialising require money which can only be gained through burglary.

Another important agent component is the *cognitive map*. As an agent moves around the environment, they remember all the houses and communities they have

[3]Legitimate employment (whether full-time or temporary) is also common and has been included in the model, but is not a feature that is used in the later case studies.

passed and also where they commit any burglaries. This allows two important characteristics of the burglary system to be included. Firstly, the agents' cognitive maps will be more detailed around their homes and the places they visit on a regular basis (e.g. drug dealers and social locations in this case). Secondly, it has been found that following a burglary, the victim and their neighbours have a substantially heightened burglary risk for a short time (Townsely et al. 2003; Johnson 2007) because the burglar is likely to re-visit the area.

19.4.2.3 The Process of Burglary

The process of actually committing a burglary in the model is broken into three distinct parts:

1. Deciding where to start looking for victims;
2. Searching for a victim;
3. Deciding upon a suitable target.

From the crime literature, some authors have suggested that burglars act as "optimal foragers" (Johnson and Bowers 2004; Bernasco and Nieuwbeerta 2005). Their decision regarding where to burgle is based on an analysis of potential rewards against risks. In this model the agents work in the same way and consider each area that they are aware of taking into account the distance to the area, its attractiveness, its similarity to the agent's home area and the number of previous successes they have had there. The area which is seen as the most appropriate to that burglar at that particular time is the one they travel to in order to start their search.

Research has shown that burglars do not search randomly for burglary targets, they exhibit identifiable search patterns (Johnson and Bowers 2004; Brantingham and Tita 2006). To reflect findings from the literature (e.g. Rengert 1996), in this model the agents perform a *bulls-eye* search; moving out from a starting location in increasingly large concentric circles (road network allowing). If an agent has not found a target within a certain amount of time, the burglary process is repeated; the agent chooses a new start location, travels there and begins the search again.

As the agents travels to their search location and performs their search, they inspect the houses they pass to determine if they are suitable for burglary. The assessment of suitability is based on the community cohesion and occupancy levels of the area, the traffic volume on the road and the accessibility, visibility and security levels of the individual house. The agent is also more likely to burgle if their motivation is high, i.e. as they become desperate to satisfy a need.

19.4.3 Model Implementation

For the simulation described here, the Repast Simphony tool was used (North et al. 2005a, b, c) which consists of a library of tools that can be used by computer programmers as well as a graphical-user-interface for non-programmers. Importantly, the software includes essential geographic functions that allow for the input/output of GIS data as well complex spatial queries. The simulation is written using the Java programming language and, due to the considerable computational complexity, was adapted to run on a high-performance computer grid provided by the National Grid Service (NGS: Geddes 2006).

19.4.4 Evaluating the Model – Verification, Calibration and Validation

Evaluating the predictive accuracy of ABMs (see Evans 2012) is a particularly problematic task although one that is extremely important. Not only are the models themselves usually highly complex, but there is often a lack accurate individual-level data against which the model can be evaluated. Following Castle and Crooks (2006), the process of evaluating this model was segregated into three distinct activities: verification, calibration and validation. Verification was accomplished by individually varying each model parameter and establishing its effect on the behaviour of the model. Calibration was manually undertaken based on knowledge of the dynamics of the model and model validity was achieved by testing the extent to which the model is able to represent the system it is attempting to simulate (Casti 1997).

19.5 Results of the Burglary Simulation

19.5.1 Scenario Context: EASEL

Parts of the south-east of Leeds, UK, contain some of the most deprived neighbourhoods in the country. To reduce deprivation in these areas, Leeds City Council has instigated an urban renewal scheme which is called EASEL (East and South East Leeds). By creating new houses, transport links, employment opportunities and green spaces, the council hopes to attract residents from outside the area (as well as many from within) to create more stable and less deprived neighbourhoods. Figure 19.5 illustrates where the EASEL boundary lies within Leeds as a whole and also shows how deprived the area is. Only the EASEL area (plus a 1 km buffer) will actually be simulated, i.e. agents within the model cannot move outside of this boundary.

At present, work has begun in two of the EASEL areas referred to here as sites *A* and *B*. The scenario is discussed here is “optimistic”; it assumes that the council’s plans succeed and the new communities are both cohesive and the new houses are well designed (secure from burglary). The scenario contains 273 individual offender agents (established through analysis of crime data).

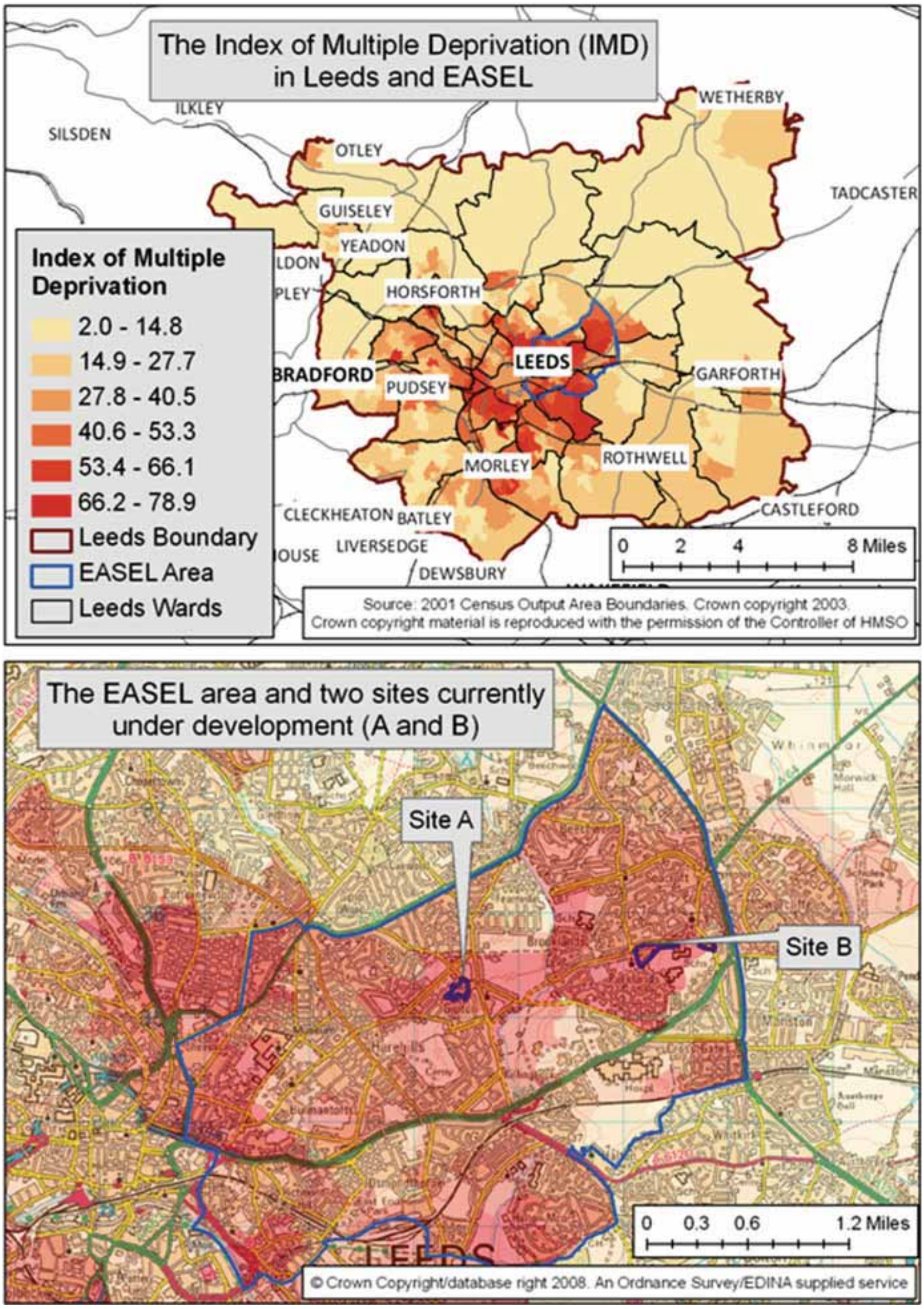


Fig. 19.5 The Index of Multiple Deprivation in Leeds and the EASEL area

19.6 Results

The model was first run *without* any of the proposed EASEL changes to create a benchmark. To ensure that the results were consistent, the simulation was run 50 separate times and the results from all simulations were combined. Having created a benchmark, the levels of security and community cohesion in the affected sites (A and B) were increased to reflect the planned EASEL regeneration changes and the simulation was executed again (50 times).

Figure 19.6 presents the difference in simulated crime rates before and after the proposed EASEL changes. Observing the entire EASEL area (upper-right map) it becomes apparent that, on the whole, the results of the two simulations are very similar. This is to be expected as the simulated environmental changes only cover very small areas. When observing the regeneration areas A and B in more detail, however, it appears that crime rates *within* the areas have fallen. This is not unexpected because the increased security and community cohesion make the houses in the area less attractive burglary targets. However, the orange and red areas surrounding the regeneration zones indicate that there are some houses which show a substantially higher risk of burglary than others. In other words, it appears that crimes are being *displaced* into the surrounding areas. The effect is highly localised which is unusual because it might be expected that burglaries

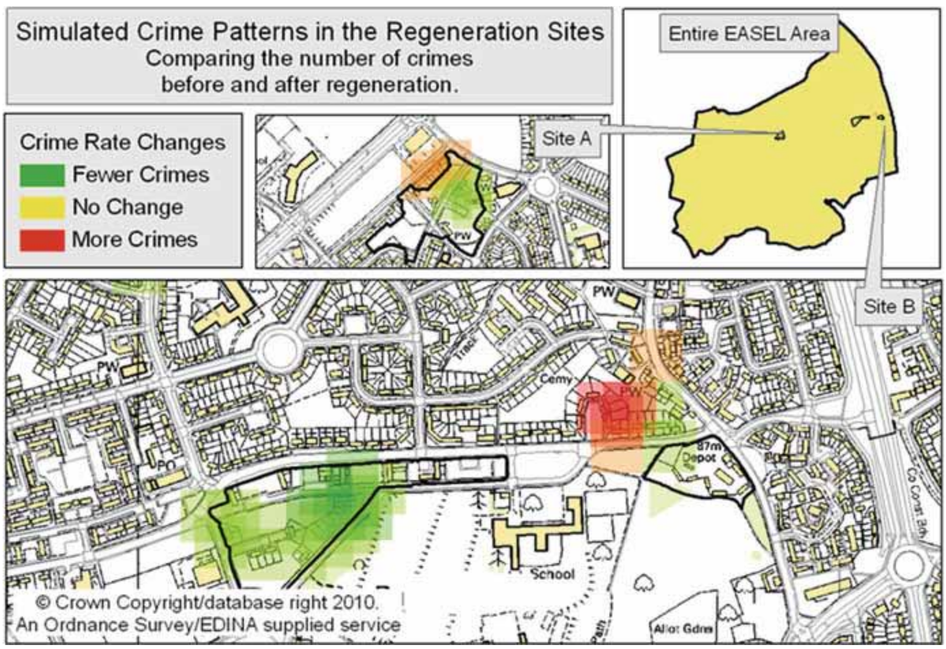


Fig. 19.6 Comparing simulated crime rates before and after regeneration of sites A and B

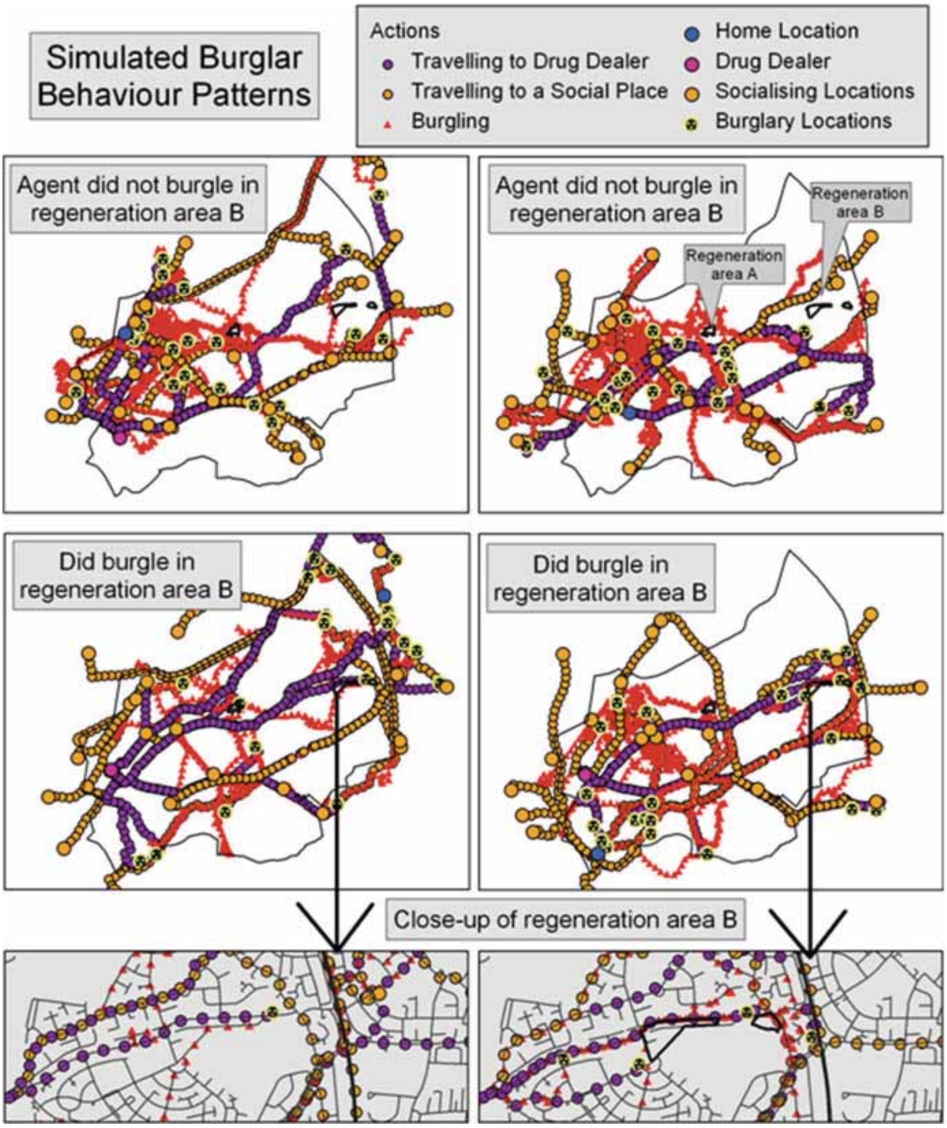


Fig. 19.7 Examples of simulated offender movement patterns in the post-regeneration simulation. Illustrative of the difference between the agents who did and did not burgle in development site B

would be more evenly distributed in the surrounding area (for example see Malleson et al. 2009).

The most substantial burglary increases are evident in a small number of houses to the north of the development site B. To explain why these houses in particular suffer a higher crime rate, Fig. 19.7 plots the movements of four agents; two who did not commit crimes in the highly burgled area and two that did. By observing the agents' travel patterns throughout the simulation it is obvious that even the agents who did commit crimes in the highly burgled area still left large parts of site B unexplored. The houses that suffered particularly high burglary rates are situated on a main road that runs along the northern boundary of the development area; a road that was regularly used by burglars. This explains part of their burglary risk; agents did not have to explore the area at length to become aware of them. Also, the houses themselves are slightly more visible and accessible than their non-regenerated neighbours which adds to their risk.

A close inspection of Fig. 19.7 indicates that the agents passed the houses whilst looking for a burglary target, not during legitimate travels on some other business (such as travelling to a social location). Figure 19.8 illustrates this in more detail. Therefore one can conclude, from this evidence, that the EASEL changes attracted the agents to the area specifically for burglary purposes and the location of some houses on the main road coupled with slightly more physical vulnerability (accessibility and visibility) increased their risk disproportionately to that of their neighbours. Although one might assume that the houses surrounding a regeneration area might experience increased burglary rates (indeed this can be explained by criminology theory), only an individual level model could not have predicted which *individual houses* might be susceptible to burglary above others. Only when crime theories were implemented in a model that is able to account for the low-level dynamics of the burglary system can specific real-world predictions such as this be made.

In conclusion, it is apparent that the effects of having a slightly higher burglary risk, coupled with their location on a main road, mean that on average particular houses received more burglaries after local regeneration. But only after an examination of the routine activities of the burglar agents as well as an inspection of the individual household characteristics does this become apparent. This result demonstrates the power of agent-based geographic models; here we are able to pinpoint which *individual houses* might suffer a high burglary risk as a direct but unintended consequence of urban regeneration. This also leads to a specific policy implication: the houses identified surrounding site B (as well as some in the site A) should be target hardened.

19.7 Conclusions

This chapter has discussed the use of ABM for analysing and predicting occurrences of crime. In particular, a model that has been used to simulate occurrences of residential burglary was outlined in detail. A brief review of crime research identified a number of key factors that should be included in a model. GIS data was used to create a realistic virtual environment that represents the study area in a high level of detail, including the individual roads that people use to travel around a city and the buildings that they pass on the way. Furthermore, through an analysis of the data it was possible to create estimates of the physical burglary risks associated with every individual house. Agents in the model (the "burglars") were equipped with an

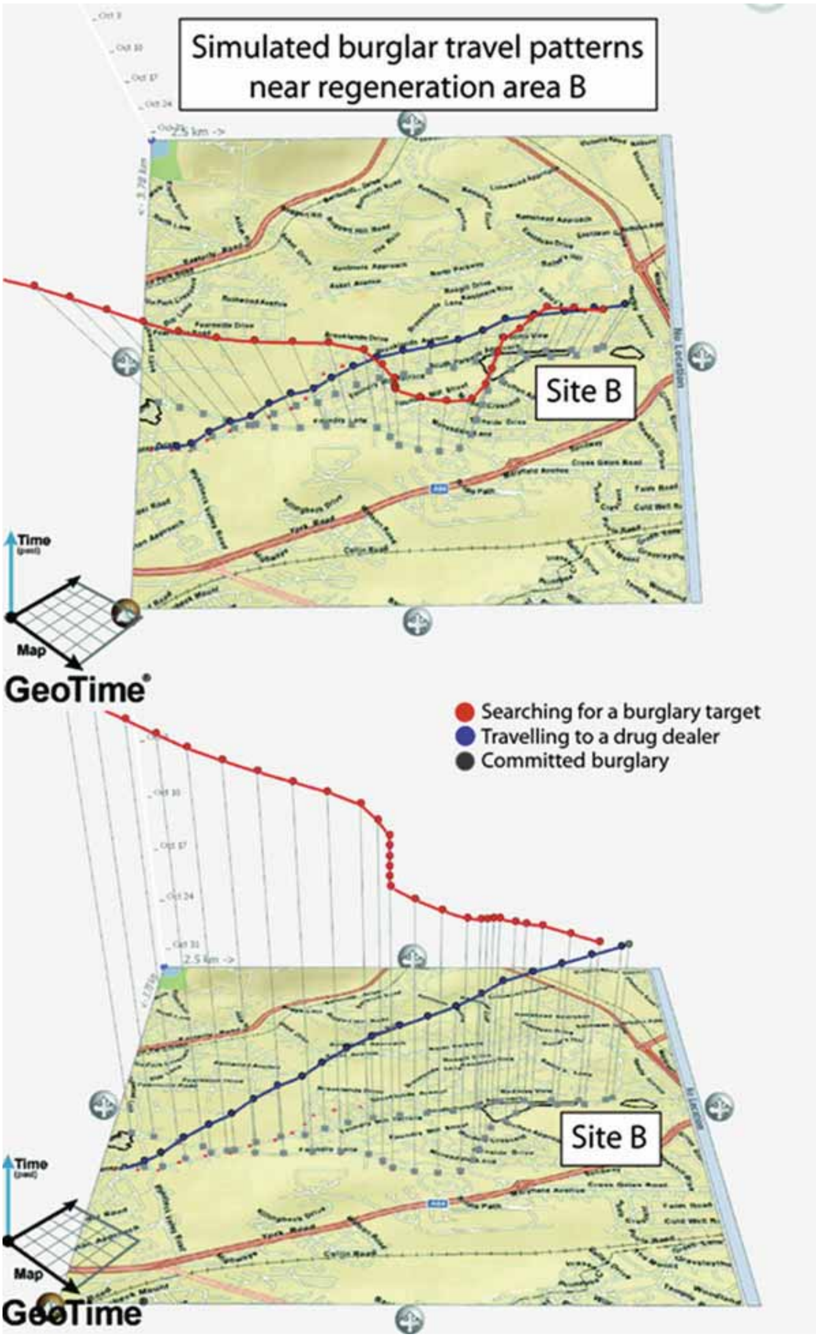


Fig. 19.8 Visualising the journey to and from a burglary close to regeneration area B. The agent travels to the area specifically for burglary. For clarity, both images illustrate the same journey but from different angles. GeoTime software used courtesy of Oculus Info Inc. All GeoTime rights reserved

advanced cognitive framework (PECS) and were able to make a comprehensive decision about what action they should take at any given model iteration. As important as the houses and the burglars, "communities" were incorporated into the model through the use of census and deprivation data.

The result is a comprehensive model that can directly account for the interactions and dynamics that drive the underlying system and can be used to make predictive analyses at a high resolution. As an example of the types of experiments that are possible with such a model, it was shown that a small number of houses might be at a higher risk of burglary after a regeneration scheme due to their spatial location and the resulting behaviour of the burglar agents. Although it inevitably has some drawbacks, the agent-based approach is the most appropriate technique for modelling such a system; one that is characterised by individual interactions and contains intelligent organisms that exhibit complex behaviour.

References

Andresen, M. A. (2010). The place of environmental criminology within criminological thought. In M. A. Andresen, P. J. Brantingham, & J. Kinney (Eds.), *Classics in environmental criminology*. Boca Raton: CRC Press.

Andresen, M. A., & Malleson, N. (2011). Testing the stability of crime patterns: Implications for theory and policy. *Journal of Research in Crime and Delinquency, 48*(1), 58–82. doi:10.1177/0022427810384136.

Axelrod, R. (1997). Advancing the art of simulation in the social sciences. In R. Conte, R. Hegselmann, & P. Terna (Eds.), *Simulating social phenomena* (pp. 21–40). Berlin: Springer.

Axtell, R. (2000). *Why agents? On the varied motivations for agent computing in the social science*. Center on Social and Economic Dynamics Working Paper No. 17. Available at http://www.brookings.edu/es/dynamics/papers/agents/agents.htm. Accessed Jan 2007.

Bafna, S. (2003). Space syntax: A brief introduction to its logic and analytical technique. *Environment and Behaviour, 35*(17), 17–29.

Baldwin, J., & Bottoms, A. E. (1976). *The urban criminal: A study in sheffield*. London: Tavistock Publications.

Balzer, W. (2000). SMASS: A sequential multi-agent system for social simulation. In R. Suleiman, K. G. Troitzsch, & N. Gilbert (Eds.), *Tools and techniques for social science simulation*, (Chap. 5, pp. 65–82). Heidelberg: Physica-Verlag.

Barberet, R., Fisher B. S., & Taylor H. (2004). *University student safety in the East Midlands* (Home Office Online Report 61/04). London: Home Office.

Beavon, D. J. K., Brantingham, P. L., & Brantingham, P. J. (1994). The influence of street networks on the patterning of property offenses. In R. V. Clarke (Ed.), *Crime prevention studies* (Vol. 2). New York: Criminal Justice Press.

Bernasco, W., & Luykx, F. (2003). Effects of attractiveness, opportunity and accessibility to burglars on residential burglary rates of urban neighborhoods. *Criminology, 41*(3), 981–1002.

Bernasco, W., & Nieuwbeerta, P. (2005). How do residential burglars select target areas? *British Journal of Criminology, 45*(3), 296–315.

Birks, D. J., Donkin, S., & Wellsmith, M. (2008). Synthesis over analysis: Towards an ontology for volume crime simulation. In L. Liu & J. Eck (Eds.), *Artificial crime analysis systems: Using computer simulations and geographic information systems* (pp. 160–192). Hershey: IGI Global. Information Science Reference.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America, 99*, 7280–7287.

Bowers, K., & Hirschfield, A. (1999). Exploring the link between crime and disadvantage in north-west England: An analysis using geographical information systems. *International Journal of Geographical Information Science, 13*(2), 159–184.

Bowers, K., Johnson S., & Hirschfield A. (2003). *Pushing back the boundaries: New techniques for assessing the impact of burglary schemes* (Home Office Online Report 24/03). London: Home Office.

Brantingham, P., & Brantingham, P. (1981). Notes of the geometry of crime. In P. Brantingham & P. Brantingham (Eds.), *Environmental criminology* (pp. 27–54). Prospect Heights: Waveland Press.

Brantingham, P. L., & Brantingham, P. J. (1993). Environment, routine, and situation: Toward a pattern theory of crime. In R. Clarke & M. Felson (Eds.), *Routine activity and rational choice* (Advances in criminological theory, Vol. 5). New Brunswick: Transaction Publishers.

Brantingham, P. L., & Brantingham, P. J. (2004). Computer simulation as a tool for environmental criminologists. *Security Journal, 17*(1), 21–30.

Brantingham, P., Glasser, U., Kinney B., Singh K., & Vajihollahi M. (2005a, October). A computational model for simulating spatial aspects of crime in urban environments. In *2005 IEEE international conference on systems, man and cybernetics,* Hawaii (Vol. 4, pp. 3667–3674). Available online at http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1571717. doi:10.1109/ICSMC.2005.1571717.

Brantingham, P., Glasser, U., Kinney B., Singh K., & Vajihollahi M. (2005b, March). Modeling urban crime patterns: Viewing multi-agent systems as abstract state machines. In *Proceedings of the 12th international workshop on abstract state machines*, Paris (pp. 101–117).

Brantingham, P. J., & Tita G. (2006). *Butterflies, bees and burglars: The behavioral ecology of criminal search strategies*. Presentation to the American Society of Criminology (ASC) 31st October – 4th November 2006.

Brown, B. B., & Bentley, D. L. (1993). Residential burglars judge risk: The role of territoriality. *Journal of Environmental Psychology, 13*, 51–61.

Browning, C. R., Feinberg, S. L., & Dietz, R. D. (2004). The paradox of social organization: Networks, collective efficacy, and violent crime in urban neighborhoods. *Social Forces, 83*(2), 503–534.

Casti, J. (1997). *Would-be-worlds: How simulation is changing the frontiers of science*. New York: Wiley.

Castle, C. J. E., & A. T. Crooks (2006). *Principles and concepts of agent-based modelling for developing geospatial simulations* (UCL working papers series, paper 110). London: Centre For Advanced Spatial Analysis, University College London. Available online at http://eprints.ucl.ac.uk/archive/00003342/01/3342.pdf

Chainey, S., & Smith C. (2006). *Review of GIS-based information sharing systems* (Home Office Online Report 02/06). Available on-line at www.homeoffice.gov.uk/rds/pdfs06/rdsolr0206.pdf. Accessed Mar 2007.

Clarke, R. V., & Cornish, D. B. (1985). Modeling offenders' decisions: A framework for research and policy. *Crime and Justice, 6*, 147–185.

Cohen, L., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review, 44*, 588–608.

Cromwell, P. F., Olson, J. N., & Avary, D. W. (1991). *Breaking and entering: An ethnographic analysis of burglary* (Studies in crime, law and justice, Vol. 8). Newbury Park: Sage Publications.

Deakin, J., Smithson, H., Spencer, J., & Medina-Ariza, J. (2007). Taxing on the streets: Understanding the methods and process of street robbery. *Crime Prevention and Community Safety, 9*, 52–67.

Dray, A., Mazerolle, L., Perez, P., & Ritter, A. (2008). Drug law enforcement in an agent-based model: Simulating the disruption to street-level drug markets. In L. Liu & J. Eck (Eds.),

Artificial crime analysis systems: Using computer simulations and geographic information systems. Hershey: IGI Global.
Eck, J. E. (1995). Crime places in crime theory. In J. E. Eck & D. Weisburd (Eds.), *Crime prevention studies* (Vol. 4). New York: Criminal Justice Press.
EDiNA. (2010). UKBORDERS. Online at http://edina.ac.uk/ukborders/
Elffers, H., & P. van Baal. (2008). Realistic spatial backcloth is not that important in agent based simulation: An illustration from simulating perceptual deterrence. In L. Liu & J. Eck (Eds.), *Artificial crime analysis systems: Using computer simulations and geographic information systems* (Chap. 2, pp. 19–34). Hershey: IGI Global.
Evans, A. (2012). Uncertainty and error. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 309–346). Dordrecht: Springer.
Felson, M. (2002). *Crime and everyday life* (3rd ed.). Thousand Oaks: Sage Publications.
Geddes, N. (2006). The national grid service of the uk. In *International conference on e-science and grid computing, Los Alamitos, CA, USA.*
Glyde, J. (1856). Localities of crime in suffolk. *Journal of the Statistical Society of London, 19*(2), 102–106.
Groff, E. (2007). Simulation for theory testing and experimentation: An example using routine activity theory and street robbery. *Journal of Quantitative Criminology, 23*, 75–103.
Groff, E., & Mazerolle, L. (2008). Simulated experiments and their potential role in criminology and criminal justice. *Journal of Experimental Criminology, 4*(3), 187–193.
Groff, E., Weisburd D., & Morris N. A. (2009). Where the action is at places: Examining spatio-temporal patterns of juvenile crime at places using trajectory analysis and GIS. In D. Weisburd, W. Bernasco, & G. Bruinsma (Eds.), *Putting crime in its place. Units of analysis in geographic criminology* (Chap. 3, pp. 61–86). New York: Springer.
Gunderson, L., & Brown D. (2000). Using a multi-agent model to predict both physical and cyber-criminal activity. In *2000 IEEE international conference on systems, man, and cybernetics* (Vol. 4, pp. 2338–2343). Nashville: IEEE.
Hayslett-McCall, K. L., Qiu, F., Curtin, K. M., Chastain, B., Schubert, J., & Carver, V. (2008). The simulation of the journey to residential burglary. In *Artificial crime analysis systems: Using computer simulations and geographic information systems* (Chap. 14). Hershey: IGI Global.
Hiller, B., & Hanson, J. (1984). *The social logic of space*. Cambridge: Cambridge University Press.
Hirschfield, A., Yarwood, D., & Bowers, K. (2001). Crime pattern analysis, spatial targeting and GIS: The development of new approaches for use in evaluating community safety initiatives. In G. Clarke & M. Madden (Eds.), *Regional science in business* (Advances in spatial science, Chap. 17). Berlin/Heidelberg: Springer.
Hope, T. (1984). Building design and burglary. In R. Clarke & T. Hope (Eds.), *Coping with Burglary* (International series in social welfare). Boston: Kluwer Academic Publishers.
Jeffery, C. R. (1971). *Crime prevention through environmental design*. Beverly Hills: Sage Publications.
Johnson, D. (2007, May). *Predictive analysis: Utilising the near repeat phenomena in Bournemouth*. Paper presented at the fifth national crime mapping conference, London.
Johnson, S., & Bowers, K. (2004). The stability of space-time clusters of burglary. *British Journal of Criminology, 44*, 55–65.
Johnson, S., Bowers, K., & Hirschfield, A. (2002). Introduction to the environmental criminology and crime analysis (ECCA) special edition. *Crime Prevention and Community Safety: An International Journal, 4*(1), 7–10.
Kawachi, I., Kennedy, B. P., & Wilkinson, R. G. (1999). Crime: Social disorganization and relative deprivation. *Social Science and Medicine, 48*, 719–731.
Kongmuang, C. (2006). *Modelling crime: A spatial microsimulation approach*. Ph. D. thesis, School of Geography, University of Leeds, Leeds.
Liu, L., & Eck, J. (Eds.). (2008). *Artificial crime analysis systems: Using computer simulations and geographic information systems*. Hershey: IGI Global.

Liu, L., Wang, X., Eck, J., & Liang, J. (2005). Simulating crime events and crime patterns in a RA/CA models. In F. Wang (Ed.), *Geographic information systems and crime analysis* (pp. 197–213). Reading: Idea Publishing.

Malleson, N. (2010). *Agent-based modelling of burglary*. Ph. D. thesis, School of Geography, University of Leeds, Leeds.

Malleson, N., Evans, A., & Jenkins, T. (2009). An agent-based model of burglary. *Environment and Planning B: Planning and Design, 36*, 1103–1123.

Malleson, N., Heppenstall, A., & See, L. (2010). Crime reduction through simulation: An agent-based model of burglary. *Computers, environment and urban systems, 31*(3), 236–250.

Melo, A., Belchior, M., & Furtado, V. (2005). Analyzing police patrol routes by simulating the physical reorganization of agents. In J. S. Sichman & L. Antunes (Eds.), *MABS* (Lecture notes in computer science, Vol. 3891). New York: Springer.

Mimas (2010). CasWeb: Web interface to census aggregate outputs and digital boundary data. Online at http://casweb.mimas.ac.uk/

Moss, S., & Edmonds, B. (2005). Towards good social science. *Journal of Artificial Societies and Social Simulation, 8*(4).

Müller, J. P. (1998). Architectures and applications of intelligent agents: A survey. *The Knowledge Engineering Review, 13*(4), 252–280.

Newman, O. (1972). *Defensible space*. New York: Macmillan.

Noble, M., Wright, G., Dibben, C., Smith, G., McLennan, D., Anttila, C., Barnes, H., Mokhtar, C., Noble, S., Avenell, D., Gardner, J., Covizzi, I., & Lloyd, M. (2004). *The English indices of deprivation 2004 (revised)*. London: Office of the Deputy Prime Minister.

North, M., Howe, T., Collier, N., & Vos, R. (2005a, October). The repast simphony development environment. In *Agent 2005 conference on generative social processes, models, and mechanisms*, Argonne National Laboratory, Argonne.

North, M., Howe, T., Collier, N., & Vos, R. (2005b, October). The repast simphony runtime system. In *Agent 2005 conference on generative social processes, models, and mechanisms*, Argonne National Laboratory, Argonne.

North, M., Howe, T., Collier, N., & Vos, R. (2005c, October). The repast simphony runtime system. In *Agent 2005 conference on generative social processes, models, and mechanisms*, Argonne National Laboratory, Argonne.

O'Sullivan, D. (2004, March 5–8). Complexity science and human geography. *Transactions of the Institute of British Geographers*, *29*, 282–295. Royal Geographical Society (with the Institute of British Geographers).

O'Sullivan, D., & Haklay, M. (2000). Agent-based models and individualism: Is the world agent-based? *Environment and Planning A, 32*(8), 1409–1425.

Ordnance Survey. (2009). Welcome to OS MasterMap. Available online http://www.ordnancesurvey.co.uk/oswebsite/products/osmastermap. Accessed Dec 2009.

Park, H. T. (2005). Before integration: A critical review of integration measure in space syntax. In *Proceedings of the 5th international space syntax symposium, 13–17 June, Delft*. Available on-line at http://www.spacesyntax.tudelft.nl/longpapers2.html. Accessed Dec 2009.

Quetelet, L. A. J. (1831). *Research on the propensity for crime at different ages*. Cincinnati: Anderson Publishing.

Rao, A. S., & Georgeff, M. P. (1995, June). BDI agents: From theory to practice. In V. Lesser & L. Gasser (Eds.), *Proceedings of the first international conference on multi-agent systems (ICMAS-95), San Francisco, USA*. Cambridge: MIT Press.

Rees, P., Martin, D., & Williamson P. (2002a). Census data resources in the united kingdom. In P. Rees, D. Martin, & P. Williamson (Eds.), *The census data system* (Chap. 1, pp. 1–24). Chichester: Wiley.

Rees, P., Martin, D., & Williamson, P. (Eds.). (2002b). *The census data system*. Chichester: Wiley.

Rengert, G. (1996). *The geography of illegal drugs*. Boulder: Westview Press.

Rengert, G., & Wasilchick, J. (1985). *Suburban burglary: A time and a place for everything*. Springfield: Charles Thomas Publishers.

Robinson, W. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*, 351–357.

Robinson, M. B., & Robinson, C. E. (1997). Environmental characteristics associated with residential burglaries of student apartment complexes. *Environment and Behaviour, 29*, 657–675.

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science, 277*, 918–924.

Schmidt, B. (2000). *The modelling of human behaviour*. Erlangen: SCS Publications.

Shaw, C. R., & McKay, H. D. (1942). *Juvenile delinquency and urban areas*. Chicago: The University of Chicago Press.

Shaw, C. R., & McKay, H. D. (1969). *Juvenile delinquency and urban areas*. Chicago: The University of Chicago Press.

Taylor, G., Frederiksen, R., Vane, R., & Waltz, E. (2004). Agent-based simulation of geo-political conflict. In *16th conference on innovative applications of artificial intelligence*. San Jose: AAAI Press.

The Redlands Institute. (2009). Agent-based modelling extension for ArcGIS users. http://www.spatial.redlands.edu/agentanalyst/

Tilley, N., Pease, K., Hough, M., & Brown, R. (1999). *Burglary prevention: Early lessons from the crime reduction programme* (Policing and reducing crime unit crime reduction research series paper 1). London: Home Office.

Townsely, M., Homel, R., & Chaseling, J. (2003). Infectious burglaries: A test of the near repeat hypothesis. *British Journal of Criminology, 43*, 615–633.

Urban, C. (2000). PECS: A reference model for the simulation of multi-agent systems. In R. Suleiman, K. G. Troitzsch, & N. Gilbert (Eds.), *Tools and techniques for social science simulation* (Chap. 6, pp. 83–114). Heidelberg: Physica-Verlag.

van Nes, A. (2006). The burglar as a space explorer in his own neighborhood. In U. Mander, C. Brebbia, & E. Tiezzi (Eds.), *The sustainable city IV. Urban regeneration and sustainability*. Wessex: WIT Press.

Weir, R., & Bangs, M. (2007, January). *The use of geographic information systems by crime analysts in England and Wales* (Home Office Online Report 03/07). Available online through RDS http://www.homeoffice.gov.uk/rds

Weisburd, D., Bernasco, W., & Bruinsma, G. (Eds.). (2009). *Putting crime in its place. Units of analysis in geographic criminology*. New York: Springer.

Weisburd, D. V., Bushway, S., Lum, C., & Yang, S.-M. (2004). Trajectories of crime at places: A longitudinal study of street segments in the city of seattle. *Criminology, 42*(2), 283–321.

Wilkström, P. (1991). *Urban crime, criminals and victims: The Swedish experience in an Anglo-American comparative perspective*. New York: Springer.

Wilson, J. Q., & Kelling, G. L. (1982, March). Broken windows: The police and neighborhood safety. *The Atlantic Monthly, 249*(3), 29–38.

Winoto, P. (2003). A simulation of the market for offenses in multiagent systems: Is zero crime rates attainable? In J. S. Sichman, F. Bousquet, & P. Davidsson (Eds.), *MABS* (Lecture notes in computer science, Vol. 2581, pp. 181–193). New York: Springer.

Wright, R. T., & Decker, S. H. (1996). *Burglars on the job: Streetlife and residential break-ins*. Boston: Northeastern University Press.

Chapter 20
Urban Geosimulation

Paul Torrens

Abstract The field of geosimulation represents one of the most innovative attempts to revitalize the usefulness and application of spatial modeling. With its generative emphasis on micro-dynamics of complex systems and its flexible treatment of space, time, pattern, and process, it marks a significant departure from traditionally employed coarse, static approaches. The primacy of geography in geosimulation also represents a departure for spatial simulation from its reliance on modeling methods borrowed from economics and physics, which were often ported to spatial applications because of tractability, but without consideration of the suitability of the fit. Research in geosimulation, while still nascent in its development, has been particularly active in urban applications, where the technique has considerably expanded the range of questions and ideas that can be explored in simulation. This chapter reviews the origins of urban geosimulation, discusses the state-of-the-art relative to urban applications, and speculates about potential future avenues of inquiry in the field.

20.1 Introduction

Geosimulation represents an innovative approach to constructing spatial simulations, building on the successes of previous generations of spatial simulation within the relatively unique context of a conventional era of 'big data', rapid advances in computing hardware and software, the convergence of modeling and simulation technologies across applications, and the growing utility of Geographic Information Science (Torrens 2010). Geosimulation has been developed in several disciplines, although much of its usefulness has been proven for *urban applications*. In essence, the geosimulation approach is characterized by information processing, and in that

P. Torrens (✉)
Department of Geography, The University of Maryland, College Park, MD, USA
e-mail: torrens@geosimulation.com

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_20,

way, it is no different than most conventional computer simulation schemes. However, the novelty, in geosimulation, is in using *geography* to map information processing directly to individual system elements and the processes that determine their dynamics in a massively interactive systems context, such that complete, realistic models of complex phenomena can be built, generatively, from the spatial atoms that comprise them. The emphasis in establishing such mappings, is on how geography – geographic processes, patterns, and context – can enable and advance more useful information processing.

In this chapter, I will describe the development of geosimulation for urban applications: its origins in the early introduction of computing to urban modeling in the 1970s, an overview of the current state-of-the-art, and a discussion of potential future avenues for research and development in the field.

20.2 The Origins of Geosimulation

The *premise* for geosimulation has quite a distinguished history. It dates back to Alan Turing's ideas for the digital computer, which were pioneered in his efforts to design devices that could crack the German Enigma code during World War II. In his original paper, Turing (1936, 1938) introduced the idea for an automaton (a term which had historically been associated with anthropomorphized but mechanical *machines*) that, given enough storage, power, and the right rule-set could automatically and efficiently compute solutions to mathematical problems. His later development of the idea to ascribe machine intelligence to such devices (Turing 1950) established the origins for modern-day artificial intelligence. Turing's use of neighborhood filters for information processing in these ideas was of key relevance to geography. Turing originally suggested that information processing could be treated as a quadruple of interacting factors: a serialized set of cells as containers for data on a tape-like manifold; state information which described data in the context of its unique location in space and time along the tape; a tape-reading head that could shift cell-by-cell along the tape to interpret neighboring state information on adjacent cells; and a table of rules that determined how states and neighbors should be contextualized. This introduced some core geographical concepts – space-time, relationships between pattern and process, action-by-proximity, neighborhood filtering, and perhaps even the trained eye of the geographer – into the early evolution of information processing.

The significance was not lost on geographers and the idea of using automata to formally treat geography in computer models of spatial process surfaced as early as the field began. Indeed, Waldo Tobler's (1970) concise expression of one of the tenets of exploration in the geographical sciences, the idea that near things are related to each other, neatly encapsulates the core components of Turing's automata and the heuristic was at the foundation of one of the first examples of automata modeling (and geosimulation), Tobler's model of the urbanization of Detroit. Other early examples included the land-use transition models developed by Chapin and

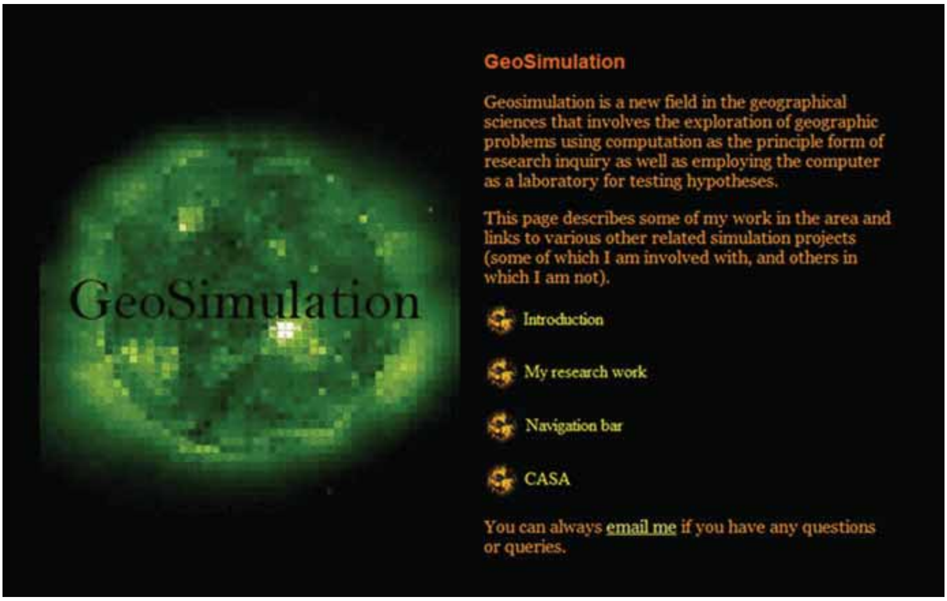


Fig. 20.1 The geosimulation.com website in 1999, complete with horrendous graphics, as befitting web design of the time

Weiss (1962), an urbanization model introduced by Nakajima (1977), and Peter Allen's work on modeling settlement hierarchy (Allen and Sanglier 1979).

The actual *term* geosimulation was introduced by Torrens in 1999 to describe efforts underway at the time in the Centre for Advanced Spatial Analysis (CASA) at University College London to build a next generation of spatial simulation methods, which were essentially building on the foundation introduced by Chapin, Weiss, and Tobler decades before. In a time before the popularity of Weblogs, Torrens launched a Website, http://www.geosimulation.com (later, .org) devoted to the topic (Fig. 20.1). Torrens outlined the idea for geosimulation in a talk at the 2000 Geocomputation meeting in Greenwich co-authored with David O'Sullivan (Torrens and O'Sullivan 2000). This was further developed in a 2004 special issue of the journal, *Computers, Environment and Urban Systems* by Torrens and Itzhak Benenson (Benenson and Torrens 2004b), who also co-authored a book on the topic, which was published in 2004 (Benenson and Torrens 2004a).

Many people at CASA were doing work in this area at the time, following the interests of Michael Batty and Yichun Xie in developing new forms of urban modeling around cellular automata (CA) (Batty and Xie 1994). Michael was building several extensions of the idea, for urbanization (Batty 1997a, b, 1999, 2001; Batty et al. 1999; Batty and Xie 1997) and movement of walkers within (Batty et al. 1998) and around (Batty et al. 2003) built spaces. Bin Jiang was collaborating with Michael on CA models for pedestrian simulation (Batty and Jiang 1999); David O'Sullivan was researching graph-based CA for gentrification modeling (O'Sullivan 2001); and Torsten Schelhorn, Muki Haklay, and David O'Sullivan were building the STREETS movement model for town centers (Schelhorn et al. 1999).

Several other groups were also developing the essential components of urban geosimulation in parallel at other sites. In North America, Keith Clarke's group at the University of California, Santa Barbara had long been developing an extensible urban growth model based on CA and his ideas for deltatrons (Clarke 1997; Clarke and Gaydos 1998; Clarke et al. 2007). This, perhaps, built on earlier work in applying CA to wildfire modeling (Clarke et al. 1994). Helen Couclelis, also at Santa Barbara, had experimented with CA modeling in the 1980s (Couclelis 1985) and her interest in the intersection between CA and GIS was revived around this time (Couclelis 1997; Takeyama and Couclelis 1997). Geographer, Michel Phipps (1989) at the University of Ottawa, developed the idea for the neighborhood coherence principle – initially used in biology, not urban analysis – using cellular automata. In the early-2000s, the land-use and land cover change modeling community in North America also began to pick-up CA modeling as a mechanism for developing what they referred to as spatially-explicit models, continuing the tradition started by Chapin and Weiss in the 1960s (Manson 2000; Brown et al. 2003; Evans and Kelley 2004; Lim et al. 2002). However, this work was mostly focused on *non-urban* areas, where ecologically significant canopies manifested as land cover.

In Europe, Roger White, at the Memorial University of Newfoundland was developing what would become the MURBANDY models (Engelen et al. 2002) with Guy Engelen and colleagues at the Research Institute for Knowledge Systems (RIKS) in the Netherlands (Engelen et al. 1995; White and Engelen 1994, 1997). From the outset, these were developed with the intention of becoming operational planning support systems (White and Engelen 1993). Denise Pumain and Lena Sanders at the Université Paris I were building the original SIMPOP model of demographic geography based around agent automata (Sanders et al. 1997). Itzhak Benenson and Portugali at Tel Aviv University were also working on urban segregation models based on the idea of agents in CA cells (Benenson 1998; Portugali 2000), echoing the idea of the "particle in a cell" (p. 99) introduced by Gipps and Marksjö (1985). Chris Webster and Fulong Wu were also building CA models of urban growth at Cardiff University, using fuzzy approaches and linguistic rule-sets (Webster and Wu 1998; Wu 1996). Ferdinando Semboloni at the University of Florence was developing 2.5 dimensional (land-use and height) urbanization models based on CA functionality (Semboloni 1997; 2000). Peter Mandl at Alpen-Adria Universitat in Austria was pursuing CA modeling research and adopted the term geosimulation for his work (Mandl 2000). Harry Timmermans and Jan Dijkstra at the Delft University of Technology in the Netherlands were also developing CA-based pedestrian models at the time (Dijkstra et al. 2000). Related work was ongoing in European physics and biophysics research, with some crossover in urban applications (Nagel and Schrekenberg 1995; Schweitzer 1997; Helbing and Molnár 1995; Ermentrout and Edelstein-Keshet 1993). Transport modelers in Europe had also begun to look at CA as a vehicle for simulating pedestrian traffic along streetscapes (Blue and Adler 2001), following early influential work by Gipps and Marksjö (1985).

In Asia, Anthony Gar-On Yeh and Xia Li at the University of Hong Kong were developing CA models of urbanization with GIS output functionality

(Li and Yeh 2000). Takashi Arai's group at the Tokyo University of Science was also developing well-calibrated CA models of urbanization on the basis of the White & Engelen model (Arai and Akiyama 2004).

In Australia, Robert Itami had long been developing agent automata models of hikers' movement along trails (Itami 1988), which although not urban was one of the first (as far as I know, *the* first) introductions of *agent* automata in geography. Martin Bell at the University of Adelaide developed a CA-like graphic model of urbanization that was coupled to geographic information systems (GIS) and that considered adjacency rules (Bell et al. 1999). Doug Ward, Stewart Phinn, and Alan Murray, then at the University of Queensland, also developed a CA-based urbanization model, which considered the role of road-building in fostering urban growth (Ward et al. 2000).

CA models rely on checking the information contained in automata through neighborhood filters and so geography featured implicitly in many of these models. Similarly, many of the models relied on GIS for data management and for visualizing model output. However, *geographical science*, which sits at the heart of the geosimulation approach, was not necessarily treated *explicitly* in the models. The contribution of geosimulation is mainly in reawakening interest in the developments introduced by early pioneers in the 1960s and 1970s, but also in infusing anew the idea of using geography to advance urban simulation amid more recent developments in computing technology. In this sense, geosimulation also draws upon the early work of Stan Openshaw in developing the field of geocomputation at the University of Leeds (Openshaw et al. 1987; Batty 1998) at the intersection of computing (rather than simply using computers) and geography (Longley et al. 1998).

Geography-specific automata modeling actually forms a smaller sub-set of the activity I have just described. Early work by Waldo Tobler really exemplifies a *dedicated* geographic consideration of the utility of employing automata for spatial modeling. His initial paper on the topic introduced variable neighborhood considerations as a vehicle for exploring the relationship between action and distance (Tobler 1979), perhaps following from his interest in automated cartography and projections (Tobler 1959). Later work by Couclelis extended geographic ideas, exploring the fundamental nature of information-gathering in geographic automata (Takeyama and Couclelis 1997). Similar ideas had been pursued by Phipps, in examining the utility of the neighborhood as a vehicle for spatial interaction (Phipps 1989). Clarke's careful exploration of sufficient geographic (and GIS) processes for the SLEUTH model (Clarke and Gaydos 1998) was also critical in laying the foundation for the development of dedicated geographic algorithms for urban automata: an area of research which still does not enjoy the attention that it deserves (Torrens and O'Sullivan 2000). Although not specifically urban, Robert Itami's work on ascribing spatial cognition as artificial intelligence for agent automata was pioneering in its early exploration of the role of spatial intelligence in allying automata models to human geography (Itami 2002, 1988). Recently, Bernard Moulin's group at Université Laval have developed a series of geosimulation applications, ranging from shopping behavior (Ali and Moulin 2005) and crowd modeling (Moulin et al. 2003) to disease propagation (Bouden et al. 2008).

Of course, much of the geography that finds its way into urban automata filters through GIS. Many of the urban automata modeling schemes built in the 1990s and early-2000s had components that *connected* to GIS. Usually, this was for simple data input and cartographic visualization of results. Many CA models, for example, would read-in polygonal, raster, or graph lattices as a cellular structure for automata. Similarly, the graphic user interface (GUI) components of GIS were often used to visualize model output cartographically, allowing for on-screen querying of results through brushing and other geovisualization procedures. For some time, there was debate about whether urban automata models should be run within standard GIS toolboxes (Wagner 1997; Batty et al. 1999; Park and Wagner 1997) and automata-based extensions for commercial GIS software were developed (Strout and Li 2006; Brown et al. 2005), as were GIS input-output functionality for popular open source (Dibble and Feldman 2004) or freeware automata model development packages (Blikstein et al. 2005). Similarly, there was debate about whether the two should be loose-coupled or tight-coupled (Brown et al. 2005; Clarke and Gaydos 1998; Torrens and Benenson 2005).

20.3 Geosimulation: A Primer

Geosimulation goes beyond issues of getting GIS data in and out of simulations. However, at its core, it deals with flexible handling of geographic information through process modeling (Torrens 2009) and matching those processes as realistically as possible to ideas, theory, hypotheses, or knowns of the system being considered. Geosimulation has several key components in interfacing geography with information processing generally and automata particularly.

First, traditional treatment of geographical units as average, spatially-modifiable geographical units, or (statistically) mean individuals (Openshaw 1983) in spatial modeling is expanded in geosimulation. This coarse approach is instead replaced with a regard for spatially non-modifiable entities, replete with individual descriptions and independent functionality. If spatial aggregates are indeed treated in simulation, they are handled generatively (Epstein 2006), as being built from the bottom up through assembly of individual entities and their connecting interactions for the purposes of producing aggregate behavior, phenomena, processes, or structures. This introduces a significant advantage as it allows for exploration of the genesis of spatial phenomena as the ‘atoms’ of the process. Additionally, it permits for the emergence of complexity from these assemblies across complicated mechanisms such as non-linearity, path-dependence, self-organization, feedback, scaling, bifurcation, fractality, and so on (O’Sullivan 2004).

Second, geosimulated entities are usually endowed with autonomy and independence in their behavior, even when collaborating or conflicting. This individuality is important as it shifts the attention in model-building and in exploring simulations to treatment of singular behavior in the context of larger systems (O’Sullivan and Haklay 2000). It also marks a departure from physics-based or economics-based modeling methods, from which spatial modeling has traditionally pilfered, in that the behavior of entities in simulation is not necessarily considered as being homogenous across the system, i.e., the spatial uniqueness of the behavior and the unique geography of its context matters, whether spatial, temporal, social, technical, environmental, built, economic, and so on. Moreover, these behaviors are not considered as being static within a simulation. Even if a transition rule is applied mechanically in the same way for each modeled entity, the unique experience of that entity will infuse the rule with unique information, producing variation in outcomes over space and time. This sort of sensitivity to micro-specification is one of the hallmarks of complexity studies (Arthur 1990). The computational flexibility of geosimulation also means that the approach is agnostic in its consideration of the sorts of behaviors, phenomena, agency, or processes that it can handle.

Third, geosimulations are usually designed as event-driven systems, as compared to the traditional approach of building time-driven (or even cross-sectional) models. Specifically, geosimulations generally treat interactions among modeled entities as events, with discrete bundles of change in space-time. These could be one-off events, or cyclical, seasonal, chain reactions, serials, and so on. They can also be considered synchronously or asynchronously among entities and spaces within the simulation. Treatment of timing in this manner has a number of advantages. It allows for representation of entities' internal 'clocks' (whether these are actual, mechanical within a simulation, or conceptual). This allows, for example, for the 'thought calculus' of a modeled entity to be worked through before it produces an interaction within the simulation, and for diversity in these calculi to be reconciled and scheduled parsimoniously across many interacting entities. When put together to form a system, update of modeled entities' clocks may be flexibly defined and the methodology can reconcile diverse temporal scales. Events can also be constructed heterogeneously per simulated entity with the result that the characteristic timing of a process, phenomenon, thought, collaboration, conflict, and so on can be represented in simulation. In essence, this allows for the treatment of entities at both their spatial and temporal atoms of behavior or process.

Fourth, geosimulation has a natural symbiosis with Geographic Information Science, GIS, spatial analysis, and related geospatial technologies. This connection to Geographic Information Science extends to spatial data models, including entity-relationship, object-oriented, raster, graph, hierarchical and so on. It also allies automata with spatial data access heuristics. This is perhaps not surprising, given the origins of geosimulation in *information processing* and the fundamental consideration of space, time, process, and neighborhood in relating information dynamics within the automata framework. It is, however, quite a significant development over traditional spatial modeling approaches, which quite often were designed for reading-in variables and parameters, but not for handling input data, output results, and the internal information processing dynamics of simulation with dedicated data models. Fundamentally, it increases the opportunities for information diffusion and interaction in models.

Fifth, with origins in the birth of digital computing, geosimulation is comfortably allied with computer science with the result that geosimulation models can be docked with other forms of computational modeling, including computer graphics and animation (Torrens 2007a), parallel and high-performance computing (Guan et al. 2006; Phipps and Langlois 1997), artificial neural networks (Li and Yeh 2002), Bayesian computing (Kocabas and Dragi evi 2006), swarm optimization (Liu et al. 2007), evolutionary computation (Manson 2005), and so on.

Sixth, because of its fundamental emphasis on dynamics and interaction, geosimulation is well-suited to representing complexity in simulation, and associated phenomena of feedback and path-dependence, non-linearity, emergence, fractality, allometry, bifurcation, autopoiesis, self-organization, and so on (see Batty (2005) for an overview).

20.4 Geographic Automata as a Vehicle for Geosimulation

The introduction of *geographic automata* has perhaps represented the most explicit conventional treatment of geosimulation. Development of the idea has come from a variety of sources, mostly organized around geographic CA, with extended (usually derived from GIS and spatial analysis) geographic functionality for relating cells to other cells through neighborhood filters. Often, these are developed to handle specific cellular geometries, such as layered rasters (Takeyama and Couclelis 1997), vectors (Moreno et al. 2008; Stevens and Dragi evi 2007), and graphs (networks) (Dibble and Feldman 2004; O'Sullivan 2001). Other approaches have used the geographic attributes of CA to accelerate computing in simulation (Guan et al. 2006; Liu et al. 2007).

The development if *geographically-enabled* CA has introduced fantastic geographic functionality to urban automata models, but in many ways they are extensions of existing CA approaches through spatial analysis. In the early-2000s, Torrens (2001) introduced a dedicated geographic automata system (GAS), designed to treat geography inherently in an automata framework. Starting with a basic, stripped-down automaton with processing capability (states, input, state transition), the approach infused geographic functionality into the basic working elements of the automaton. This included dedicated processing capabilities for space-time movement, malleable location conventions, dedicated neighborhood process rules that dictate how neighborhood filters should transform over space and time, and ontology of spatial primitives. In a paper with Itzhak Benenson (Torrens and Benenson 2005), Torrens demonstrated the concept with a working demonstration of the classic Schelling/Sakoda segregation model (Sakoda 1971; Schelling 1971), worked as a GAS, and a review of how all urban automata models at the time could be accommodated in the framework. The GAS framework goes beyond simply allying automata models with GIS, as it allows the model-designer to infuse core geographic principles into the essential functionality of the automata. These geographic primitives can then be used to build spatial entities or phenomena from the bottom-up. In essence, knowledge is created in model-building and simulation by experimenting with the geographical building-blocks of geographic complexity, from first principles.

Torrens has since published a series of demonstrations of the approach for urban geosimulation, including models of urbanization (Torrens 2006a), suburban sprawl (Torrens 2006b), residential location behavior (Torrens 2007b), gentrification dynamics (Torrens and Nara 2007), and behavioral geography (Torrens 2007a). Itzhak Benenson also developed the idea into a software package (Benenson et al. 2006).

The GAS framework has also been adopted for geosimulation in other fields. Shawn Laffan at the University of Queensland and Michael Ward at Texas A&M have developed a series of infection propagation models for veterinary studies using geographic automata (Doran and Laffan 2005; Ward et al. 2007; Laffan et al. 2007). Shen and colleagues (2009) have used geosimulation and geographic automata for land-use modeling. Hammam and colleagues (2007) developed an extended concept for geographic automata with geometry displacement. A series of related concepts for geographic automata have also been developed by Moreno and Marceau, with at least partial inspiration from the GAS approach (Moreno et al. 2008, 2009).

20.5 Epilog: The Future of Urban Geosimulation

The field of geosimulation is still quite nascent and developments are almost inextricably tied to the emergence of new forms of modeling and simulation in science generally. The emergence of new forms of dataware for modeling and simulation and the growth in computational social science around those developments could have a transformative impact on the future research trajectory for urban geosimulation. In particular, a set of promising avenues for future research are relevant.

The first is the development of semantic search on the Web (Berners-Lee et al. 2001), semantic computing (Egenhofer 2002), and the evolution of the "GeoWeb" (Elwood 2010; Haklay et al. 2008). The basic components of geosimulation are naturally amenable to ontological representation, which lends geosimulation interoperability with semantic computing. Coupled with the popularity of semantic approaches, there has been a recent swelling in the volume, availability, and semantic organization of geographic information on the Web. Already, applications that use geosimulation-like process functions are being used to extract and interpret space-time data on the Web or data generated using mobile devices tethered to the Web. These include so-called predestination models (Krumm and Horvitz 2007) that couple geosimulation-like modeling with location-based services to provide application to users of mobile devices based on their position in space and time and models of their (and others') past trajectories (Torrens 2010). Indeed, there exists great potential for the development of more sophisticated semantically-operable and Web-enabled geosimulation processing services, which can feed on a steady stream of newly-emerging geographic information (Goodchild 2007). The emergence of geoagents as Web-scraping tools has already shifted Geographic Information Science in this direction (Yu and Peuquet 2009; Zhang and Tsou 2009).

Geosimulation-like schemes are also being introduced in computer graphics research, specifically to endow synthetic characters in special effects and games with realistic behavioral geography (Pelechano et al. 2008). Thus far, the spatial intelligence afforded these synthetic (usually automata-based) characters has been relatively simple, but significant advances are being made, in ascribing them realistic vision (Terzopoulous et al. 1994), activity (Paris and Donikian 2009), behavior (Ulicny and Thalmann 2003), collective geography (Nieuwenhuisen et al. 2007), and even emotions (Badler et al. 2002). Cross-fertilization of ideas between computer graphics and geosimulation could catalyze significant gains in the computability of geosimulation models (which developers of computer graphics often excel at) while maintaining rich behavioral fidelity (which geographers often excel at). Several geographers have already made initial forays into this area from the perspective of Geographic Information Science, geovisualization (Crooks et al. 2009), and geosimulation (Torrens 2007a).

There also remains a relatively untapped potential for connecting urban geosimulation with geodemographics and related business intelligence. Geodemographics, as a field of study, concerns itself with classifying and grouping consumers based on the geography of their activity patterns and spending habits (Singleton and Longley 2009; Harris et al. 2005). It is used widely and practically in marketing and business analysis, for political polling, consumer testing, advertising, and actuarial analysis. Much of the spatial analysis used in geodemographics is relatively primitive, however, and would benefit substantially from the infusion of geosimulation, which would allow for more sophisticated models of individuals and their space-time activity and behavior to be developed (Kurose et al. 2001; Hui et al. 2009). Given the basis for geodemographics in data-collection and data-generation (Longley and Harris 1999), there also exists potential for calibration of geosimulation models. Of course, the potential for unwelcome uses of such systems and function creep beyond simple customer analysis is great (Dobson and Fisher 2003).

References

Ali, W., & Moulin, B. (2005). 2D-3D MultiAgent GeoSimulation with knowledge-based agents of customers' shopping behavior in a shopping mall. *Lecture Notes in Computer Science, 3693*, 445–458.

Allen, P. M., & Sanglier, M. (1979). A dynamic model of growth in a central place system. *Geographical Analysis, 11*(3), 256–272.

Arai, T., & Akiyama, T. (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), 65–84.

Arthur, W. B. (1990, February). Positive feedbacks in the economy. *Scientific American*, pp. 80–85.

Badler, N., Allbeck, J., Liwei, Z., & Byuan, M. (2002). Representing and parameterizing agent behaviors. In *Proceedings of computer animation* (pp. 133–143). Geneva: IEEE Computer Society.

Batty, M. (1997a). Cellular automata and urban form: A primer. *Journal of the American Planning Association, 63*(2), 266–274.

Batty, M. (1997b). Possible urban automata. *Environment and Planning B, 24*, 175–192.

Batty, M. (1998). GeoComputation using cellular automata. In P. A. Longley, S. M. Brooks, R. McDonnell, & B. Macmillan (Eds.), *Geocomputation: A primer* (pp. 96–129). Chichester: Wiley.

Batty, M. (1999). *Multi-agent approaches to urban development dynamics*. London: Centre for Advanced Spatial Analysis, University College London.

Batty, M. (2001). Polynucleated urban landscapes. *Urban Studies, 38*(4), 635–655.

Batty, M. (2005). *Cities and complexity: Understanding cities with cellular automata, agent-based models, and fractals*. Cambridge: The MIT Press.

Batty, M., Desyllas, J., & Duxbury, E. (2003). The discrete dynamics of small-scale spatial events: Agent-based models of mobility in carnivals and street parades. *International Journal of Geographical Information Science, 17*(7), 673–697.

Batty, M., & Jiang, B. (1999). *Multi-agent simulation: New approaches to exploring space-time dynamics within GIS* (CASA Working Paper 10). London: Centre for Advanced Spatial Analysis (CASA), University College London.

Batty, M., Jiang, B., & Thurstain-Goodwin, M. (1998). *Local movement: Agent-based models of pedestrian flows* (CASA Working Paper 4). London: Centre for Advanced Spatial Analysis (CASA), University College London.

Batty, M., & Xie, Y. (1994). From cells to cities. *Environment and Planning B, 21*, s31–s48.

Batty, M., & Xie, Y. (1997). Possible urban automata. *Environment and Planning B, 24*(2), 175–192.

Batty, M., Xie, Y., & Sun, Z. (1999). Modeling urban dynamics through GIS-based cellular automata. *Computers, Environment and Urban Systems, 23*(3), 205–233.

Bell, M., Dean, C., & Blake, M. (1999). A model for forecasting the location of fringe urbanisation, with GIS and 3D visualisation. In P. Rizzo (Ed.), *Proceedings of the sixth international conference on computers in urban planning and urban management (CUPUM)*, Venice, CD-ROM.

Benenson, I. (1998). Multi-agent simulations of residential dynamics in the city. *Computers, Environment and Urban Systems, 22*(1), 25–42.

Benenson, I., Birfur, S., & Kharbash, V. (2006). Geographic Automata Systems and the OBEUS software for their implementation. In J. Portugali (Ed.), *Complex artificial environments* (pp. 137–153). Berlin: Springer.

Benenson, I., & Torrens, P. M. (2004a). *Geosimulation: Automata-based modeling of urban phenomena*. London: Wiley.

Benenson, I., & Torrens, P. M. (2004b). Geosimulation: Object-based modeling of urban phenomena. *Computers, Environment and Urban Systems, 28*(1/2), 1–8.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic Web. *Scientific American, 284*(5), 34–44.

Blikstein, P., Abrahamson, D., & Wilensky, U. (2005). NetLogo: Where we are, where we're going. In M. Eisenberg & A. Eisenberg (Eds.), *Proceedings of the Annual Meeting of Interaction Design & Children*, Boulder, CO, June 8–10.

Blue, V., & Adler, J. (2001). Cellular automata microsimulation for modeling bi-directional pedestrian walkways. *Transportation Research Part B, 35*, 293–312.

Bouden, M., Moulin, B., & Gosselin, P. (2008). The geosimulation of West Nile virus propagation: A multi-agent and climate sensitive tool for risk management in public health. *International Journal of Health Geographics*, *7*(35), Online.

Brown, D. G., Page, S. E., Riolo, R. L., & Rand, W. (2003). Agent based and analytical modeling to evaluate the effectiveness of greenbelts. *Environmental Modelling and Software, 19*(12), 1097–1109.

Brown, D. G., Riolo, R., Robinson, D. T., North, M. J., & Rand, W. (2005). Spatial process and data models: Toward integration of agent-based models and GIS. *Journal of Geographical Systems, 7*(1), 25–47.

Chapin, F. S., & Weiss, S. F. (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.

Clarke, K. (1997). *Land transition modeling with deltatrons* [Online]. Department of Geography, University of California, Santa Barbara. Accessed January 1, 2000. Available from http://www.geog.ucsb.edu/~kclarke/Papers/deltatron.html

Clarke, K. C., Brass, J. A., & Riggan, P. J. (1994). A cellular automata model of wildfire propagation and extinction. *Photogrammetric Engineering and Remote Sensing, 60*(11), 1355–1367.

Clarke, K. C., & Gaydos, L. (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., Gazulis, N., Dietzel, C., & Goldstein, N. C. (2007). A decade of SLEUTHing: Lessons learned from applications of a cellular automaton land use change model. In P. Fisher (Ed.), *Classics in IJGIS: Twenty years of the International Journal of Geographical Information Science and Systems* (pp. 413–427). Boca Raton: CRC Press.

Couclelis, H. (1985). Cellular worlds: A framework for modeling micro-macro dynamics. *Environment and Planning A, 17*, 585–596.

Couclelis, H. (1997). From cellular automata to urban models: New principles for model development and implementation. *Environment and Planning B, 24*, 165–174.

Crooks, A., Hudson-Smith, A., & Dearden J. (2009). Agent street: An environment for exploring agent-based models in second life. *Journal of Artificial Societies and Social Simulation*, *12*(4). http://jasss.soc.surrey.ac.uk/12/4/10.html

Dibble, C. & Feldman, P. G. (2004). The GeoGraph 3D computational laboratory: network and terrain landscapes for RePast. *Journal of Artificial Societies and Social Simulation*, *7*(1). http://jasss.soc.surrey.ac.uk/7/1/7.html

Dijkstra, J., Timmermans, H. J. P., & Jessurun, A. J. (2000). A multi-agent cellular automata system for visualising simulated pedestrian activity. In S. Bandini & T. Worsch (Eds.), *Theoretical and practical issues on cellular automata* (pp. 29–36). London: Springer.

Dobson, J., & Fisher, P. (2003). Geoslavery. *IEEE Technology and Society Magazine, 22*(1), 47–52.

Doran, R. J., & Laffan, S. W. (2005). Simulating the spatial dynamics of foot and mouth disease outbreaks in feral pigs and livestock in Queensland, Australia, using a susceptible-infected-recovered cellular automata model. *Preventive Veterinary Medicine, 70*(1–2), 133–152.

Egenhofer, M. J. (2002). Toward the semantic geospatial web. In K. Makki & N. Pissinou (Eds.), *Proceedings of the tenth ACM international symposium on advances in Geographic Information Systems,* McLean, VA (pp. 1–4). New York: ACM Press.

Elwood, S. (2010). Geographic information science: Emerging research on the societal implications of the geospatial web. *Progress in Human Geography, 34*(3), 349–357.

Engelen, G., White, R., & Uljee, I. (2002). *The MURBANDY and MOLAND models for Dublin*. Maastricht: Research Institute for Knowledge Systems (RIKS) BV.

Engelen, G., White, R., Uljee, I., & Drazan, P. (1995). Using cellular automata for integrated modelling of socio-environmental systems. *Environmental Monitoring and Assessment, 30*, 203–214.

Epstein, J. M. (2006). *Generative social science: Studies in agent-based computational modeling*. Princeton: Princeton University Press.

Ermentrout, G. B., & Edelstein-Keshet, L. (1993). Cellular automata approaches to biological modeling. *Journal of Theoretical Biology, 160*, 97–133.

Evans, T. P., & Kelley, H. (2004). Multi-scale analysis of a household level agent-based model of landcover change. *Journal of Environmental Management, 72*(1–2), 57–72.

Gipps, P. G., & Marksjö, B. (1985). A microsimulation model for pedestrian flows. *Mathemathics and Computers in Simulation, 27*, 95–105.

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal, 69*(4), 211–221.

Guan, Q., Clarke, K. C., & Zhang, T. (2006). *Calibrating a parallel geographic cellular automata model*. Presented at AutoCarto 2006, Vancouver, WA, June, 2006.

Haklay, M., Singleton, A., & Parker, C. (2008). Web mapping 2.0: The neogeography of the GeoWeb. *Geography Compass, 2*(6), 2011–2039.

Hammam, Y., Moore, A., & Whigham, P. (2007). The dynamic geometry of geographical vector agents. *Computers, Environment and Urban Systems, 31*(5), 502–519.

Harris, R., Sleight, P., & Webber, R. (2005). *Geodemographics, GIS and neighbourhood targeting*. Chichester: Wiley.

Helbing, D., & Molnár, P. (1995). Social force model for pedestrian dynamics. *Physical Review E, 51*, 4282–4286.

Hui, S. K., Fader, P. S., & Bradlow, E. T. (2009). Path data in marketing: An integrative framework and prospectus for model building. *Marketing Science, 28*(2), 320–335.

Itami, R. (1988). Cellular worlds: Models for dynamic conception of landscapes. *Landscape Architecture, 78*(5), 52–57.

Itami, R. (2002). Mobile agents with spatial intelligence. In H. R. Gimblett (Ed.), *Integrating Geographic Information Systems and agent-based modeling for simulating social and ecological processes* (pp. 191–210). Oxford: Oxford University Press.

Kocabas, V., & Dragićević, S. (2006). Coupling Bayesian networks with GIS-based cellular automata for modeling land use change. *Lecture Notes in Computer Science, 4197*, 217–233.

Krumm, J., & Horvitz, E. (2007). Predestination: Where do you want to go today? *IEEE Computer, 40*(4), 105–107.

Kurose, S., Borgers, A. W. J., & Timmermans, H. J. P. (2001). Classifying pedestrian shopping behaviour according to implied heuristic choice rules. *Environment and Planning B, 28*(3), 405–418.

Laffan, S. W., Lubarsky, E., Ward, M. P., & Highfield, L. D. (2007). A Geographic Automata System for modelling disease outbreaks. In L. Oxley & D. Kulasiri (Eds.), *Proceedings of the MODSIM 2007 International Congress on Modelling and Simulation,* Canterbury, NZ (pp. 1252–1257).

Li, X., & Yeh, A. G.-O. (2000). Modelling sustainable urban development by the integration of constrained cellular automata and GIS. *International Journal of Geographical Information Science, 14*(2), 131–152.

Li, X., & Yeh, A. G.-O. (2002). Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographical Information Science, 16*(4), 323–343.

Lim, K., Deadman, P. J., Moran, E., Brondizio, E., & McCracken, S. (2002). Agent-based simulations of household decision-making and land use change near Altamira, Brazil. In H. R. Gimblett (Ed.), *Integrating Geographic Information Systems and agent-based modeling techniques for simulating social and ecological processes* (pp. 277–310). New York: Oxford University Press.

Liu, X. P., Li, X., Yeh, A., He, J. Q., & Tao, J. (2007). Discovery of transition rules for geographical cellular automata by using ant colony optimization. *Science in China Series D: Earth Sciences, 50*(10), 1578–1588.

Longley, P. A., Brooks, S. M., McDonnell, R., & Macmillan, B. (1998). *Geocomputation: A primer*. London: Wiley.

Longley, P. A., & Harris, R. J. (1999). Towards a new digital data infrastructure for urban analysis and modelling. *Environment and Planning B, 26*(6), 855–878.

Mandl, P. (2000). Geo-Simulation – Experimentieren und problemlösen mit GIS-modellen. In J. Strobl, T. Blaschke, & G. Griesebner (Eds.), *Angewandte geographische informationsverarbeitung XII: Beiträge zum AGIT-Symposium Salzburg* (pp. 345–356). Heidelberg: Herbert Wichmann Verlag.

Manson, S. M. (2000). Agent-based dynamic spatial simulation of land-use/land cover change in the Yucatán Peninsula, Mexico. In *Proceedings of fourth international conference on integrating GIS and Environmental Modeling (GIS/EM4): Problems, Prospects and Research Needs*, Banff, Alberta, Canada, September 2–8, 2000.

Manson, S. M. (2005). Agent-based modeling and genetic programming for modeling land change in the Southern Yucatan Peninsular Region of Mexico. *Agriculture, Ecosystems and Environment, 111*(1), 47–62.

Moreno, N., Ménard, A., & Marceau, D. J. (2008). VecGCA: A vector-based geographic cellular automata model allowing geometric transformations of objects. *Environment and Planning B: Planning and Design, 35*(4), 647–665.

Moreno, N., Wang, F., & Marceau, D. J. (2009). Implementation of a dynamic neighborhood in a land-use vector-based cellular automata model. *Computers, Environment and Urban Systems, 33*(1), 44–54.

Moulin, B., Chaker, W., Perron, J., Pelletier, P., Hogan, J., & Gbei, E. (2003). MAGS project: Multi-agent geoSimulation and crowd simulation. *Lecture Notes in Computer Science, 2825*, 151–168.

Nagel, K., & Schrekenberg, M. (1995). *Traffic jams in stochastic cellular automata*. Los Alamos: Los Alamos National Laboratory. 95ATS089.

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* (pp. 154–160). Montreal: Association Canadienne-Française pour L'Avancement des Sciences et Comité de Coordination des Centres de Recherches en Aménagement, Développement et Planification (CRADEP).

Nieuwenhuisen, D., Kamphuis, A., & Overmars, M. H. (2007). High quality navigation in computer games. *Science of Computer Programming, 67*(1), 91–104.

O'Sullivan, D. (2001). Exploring spatial process dynamics using irregular cellular automaton models. *Geographical Analysis, 33*(1), 1–18.

O'Sullivan, D. (2004). Complexity science and human geography. *Transactions of the Institute of British Geographers, 29*(3), 282–295.

O'Sullivan, D., & Haklay, M. (2000). Agent-based models and individualism: Is the world agent-based? *Environment and Planning A, 32*(8), 1409–1425.

Openshaw, S. (1983). *The modifiable areal unit problem*. Norwich: GeoBooks.

Openshaw, S., Charlton, M. E., Wymer, C., & Craft, A. (1987). A Mark 1 geographical analysis machine for the automated analysis of point data sets. *International Journal of Geographic Information Systems, 1*(4), 335–358.

Paris, S., & Donikian, S. (2009). Activity-driven populace: A cognitive approach to crowd simulation. *IEEE Computer Graphics and Applications, 29*(4), 34–43.

Park, S., & Wagner, D. F. (1997). Incorporating Cellular Automata simulators as analytical engines in GIS. *Transactions in GIS, 2*(3), 213–231.

Pelechano, N., Allbeck, J., & Badler, N. I. (2008). *Virtual crowds: Methods, simulation, and control*. San Rafael: Morgan & Claypool.

Phipps, M. (1989). Dynamic behavior of cellular automata under the constraint of neighborhood coherence. *Geographical Analysis, 21*, 197–215.

Phipps, M., & Langlois, A. (1997). Spatial dynamics, cellular automata, and parallel processing computers. *Environment and Planning B, 24*, 193–204.

Portugali, J. (2000). *Self-organization and the city*. Berlin: Springer.

Sakoda, J. M. (1971). The checkerboard model of social interaction. *Journal of Mathematical Sociology, 1*, 119–132.

Sanders, L., Pumain, D., Mathian, H., Guérin-Pace, F., & Bura, S. (1997). SIMPOP: A multiagent system for the study of urbanism. *Environment and Planning B, 24*, 287–305.

Schelhorn, T., O'Sullivan, D., Haklay, M., & Thurstain-Goodwin, M. (1999). *STREETS: An agent-based pedestrian model* (CASA Working Paper 9). London: University College London, Centre for Advanced Spatial Analysis.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology, 1*, 143–186.

Schweitzer, F. (1997). Active brownian particles: Artificial agents in physics. In T. Pöschel & L. Schimansky-Geier (Eds.), *Stochastic dynamics* (pp. 358–371). Berlin: Springer.

Semboloni, F. (1997). An urban and regional model based on cellular automata. *Environment and Planning B, 24*, 589–612.

Semboloni, F. (2000). The dynamic of an urban cellular automata model in a 3-D spatial pattern. In *Proceedings of XXI national conference aisre: Regional and urban growth in a global market*. Palermo, Italy, September 20–22.

Shen, Z., Kawakami, M., & Kawamura, I. (2009). Geosimulation model using geographic automata for simulating land-use patterns in urban partitions. *Environment and Planning B: Planning and Design, 36*(5), 802–823.

Singleton, A. D., & Longley, P. A. (2009). Geodemographics, visualisation, and social networks in applied geography. *Applied Geography, 29*(3), 289–298.

Stevens, D., & Dragićević, S. (2007). A GIS-based irregular cellular automata model of land-use change. *Environment and Planning B: Planning and Design, 34*(4), 708–724.

Strout, N., & Li, N. (2006). *What is agent analyst?* In ESRI (Ed.), Paper presented at 26th annual ESRI international user conference, San Diego, California, August 7–11.

Takeyama, M., & Couclelis, H. (1997). Map dynamics: Integrating cellular automata and GIS through Geo-Algebra. *International Journal of Geographical Information Systems, 11*(1), 73–91.

Terzopoulous, D., Tu, X., & Grzeszczuk, R. (1994). Artificial fishes: Autonomous location, perception, behavior, and learning in a simulated physical world. *Artificial Life, 1*(4), 327–351.

Tobler, W. (1959). Automation and cartography. *Geographical Review, 49*(4), 526–534.

Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*(2), 234–240.

Tobler, W. (1979). Cellular geography. In S. Gale & G. Ollson (Eds.), *Philosophy in geography* (pp. 379–386). Dordrecht: Kluwer.

Torrens, P. M. (2001). Cellular automata and urban simulation: Where do we go from here? *Environment and Planning B, 28*(2), 163–168.

Torrens, P. M. (2006a). Geosimulation and its application to urban growth modeling. In J. Portugali (Ed.), *Complex artificial environments* (pp. 119–134). London: Springer.

Torrens, P. M. (2006b). Simulating sprawl. *Annals of the Association of American Geographers, 96*(2), 248–275.

Torrens, P. M. (2007a). Behavioral intelligence for geospatial agents in urban environments. In T. Y. Lin, J. M. Bradshaw, M. Klusch, C. Zhang (Eds.), *Proceedings of IEEE Intelligent Agent Technology (IAT 2007),* Los Alamitos, CA (pp. 63–66). Washington, DC: IEEE.

Torrens, P. M. (2007b). A geographic automata model of residential mobility. *Environment and Planning B, 33*(2), 200–222.

Torrens, P. M. (2009). Process models and next-generation geographic information technology. In ESRI (Ed.), *GIS best practices: Essays on geography and GIS* (pp. 63–75). Redlands: ESRI Press.

Torrens, P. M. (2010). Geography and computational social science. *GeoJournal, 75*(2), 133–148.

Torrens, P. M., & Benenson, I. (2005). Geographic automata systems. *International Journal of Geographical Information Science, 19*(4), 385–412.

Torrens, P. M., & Nara, A. (2007). Modeling gentrification dynamics: A hybrid approach. *Computers, Environment and Urban Systems, 31*(3), 337–361.

Torrens, P. M., & O'Sullivan, D. (2000). Cities, cells and complexity: Developing a research agenda for urban geocomputation. In R. J. Abrahart (Ed.), *Proceedings of geocomputation 2000*, University of Greenwich, London, August 25–28, CD-ROM.

Turing, A. M. (1936). On computable numbers, with an application to the *Entscheidungsproblem*. *Proceedings of the London Mathematical Society*, Series 2, *42*(1), 230–265.

Turing, A. M. (1938). Correction to: On computable numbers, with an application to the *Entscheidungsproblem*. *Proceedings of the London Mathematical Society*, Series 2, *43*(1), 544–546.

Turing, A. (1950). Computing machinery and intelligence. *Mind, 49*, 433–460.

Ulicny, B., & Thalmann, D. (2003). Towards interactive real-time crowd behavior simulation. *Computer Graphics Forum, 21*(4), 767–775.

Wagner, D. F. (1997). Cellular automata and geographic information systems. *Environment and Planning B, 24*, 219–234.

Ward, M. P., Laffan, S. W., & Highfield, L. (2007). The potential role of wild and feral animals as reservoirs of foot-and-mouth disease. *Preventive Veterinary Medicine, 80*(1), 9–23.

Ward, D. P., Murray, A. T., & Phinn, S. R. (2000). A stochastically constrained cellular model of urban growth. *Computers, Environment and Urban Systems, 24*, 539–558.

Webster, C. J., & Wu, F. (1998). Simulations of urban growth with models of pollution property rights and subcentre formation. In A. Mumford (Ed.), *Graphics, visualisation and the social sciences: Report from workshop held May 8–9* (Advisory Group on Computer Graphics (AGOCG) Technical Report Series No. 33, pp. 113–122). Loughborough: Loughborough University Joint Information Systems Committee (JISC).

White, R., & Engelen, G. (1993). Cellular automata and fractal urban form. *Environment and Planning A, 25*, 1175–1199.
White, R., & Engelen, G. (1994). Urban systems dynamics and cellular automata: Fractal structures between order and chaos. *Chaos, Solitions & Fractals, 4*(4), 563–583.
White, R., & Engelen, G. (1997). Cellular automata as the basis of integrated dynamic regional modelling. *Environment and Planning B, 24*, 235–246.
Wu, F. (1996). A linguistic cellular automata simulation approach for sustainable land development in a fast growing region. *Computers, Environment and Urban Systems, 20*, 367–387.
Yu, C., & Peuquet, D. J. (2009). A GeoAgent-based framework for knowledge-oriented representation: Embracing social rules in GIS. *International Journal of Geographical Information Science, 23*(7), 923–960.
Zhang, T., & Tsou, M.-H. (2009). Developing a grid-enabled spatial Web portal for Internet GIServices and geospatial cyberinfrastructure. *International Journal of Geographical Information Science, 23*(5), 605–630.

Chapter 21
Applied Pedestrian Modeling

Anders Johansson and Tobias Kretz

Abstract With an increasing world population and with more cost effective transportation, mass gatherings become ever more frequent. The total size of such gatherings is often as large as millions of people. Furthermore, everyday life in cities becomes increasingly crowded with people. This development has prompted better solutions to mitigate crowded places and make them safer as well as more efficient in terms of travel time. One way to approach this crowd problem is to use crowd modeling tools to assess and optimize locations where pedestrian crowds move around. Within the last decade, crowd modeling has become a mature science and there now exist well calibrated pedestrian models that can reproduce empirically observed crowd features. In this chapter, we will introduce the field of crowd modeling, explain how crowd models can be calibrated with empirical data, and expand a bit on how navigation works in these models.

21.1 Introduction and Motivation

In the past, pedestrian simulations have mainly been used to *qualitatively* reproduce and understand various aspects of crowds. Nowadays however, neither the computing performance nor the amount and detail of available empirical data restrict us from aiming at reproducing crowd dynamics *quantitatively* as well.

One of the reasons why the microscopic simulation of pedestrians as a field of research has taken off as late as about 1985, and has gained pace only during the last

A. Johansson (✉)
Centre for Applied Spatial Analysis (CASA), University College London, London, UK
e-mail: a.johansson@ucl.ac.uk

T. Kretz
PTV AG, Stumpfstraße 1, D-76131 Karlsruhe, Germany
e-mail: tobias.kretz@ptv.de

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*, DOI 10.1007/978-90-481-8927-4_21,

decade, is the availability of computing power. The simulation of pedestrians and especially real-world applications related to simulation of pedestrians has been at the edge of available standard computational power and will remain there for some more years to come. Compared to vehicle simulations that are effectively 1 dimensional, crowd simulations are 2D or even 3D, and this increase in dimensionality results in more degrees of freedom, which requires more sophisticated models with higher temporal and spatial resolution. Therefore, computing power (in terms of speed as well as RAM) plays a more decisive role in the simulation of pedestrians.

This has at least three implications: (1) Pedestrian modelers have often implicitly or explicitly restricted their creativity in using mathematical tools to make the model results more realistic with respect to computational costs. (2) While refined models successively replace more coarse-grained models for medium-scale applications, applications such as iterative approaches in large-scale projects only recently became possible to carry out with spatially continuous microscopic models. (3) Using pedestrian modeling for large-scale urban planning and transport projects often turns out to be a highly challenging task, since the right balance has to be found between computational time, model complexity and scale, and accuracy of results.

21.2 Modeling Approaches

There are various different ways to approach pedestrian modeling, and among the first ideas to simulate interacting agents in a swarm-like way, was proposed by Reynolds (1987) with his *Boids* model. Different approaches to pedestrian modeling can be classified in various different ways, for example according to their level of abstraction:

- *Microscopic* models describe each pedestrian as a unique entity with its own properties.
- *Macroscopic* models delineate the average or aggregate pedestrian dynamics by densities, flows, and velocities as functions of space and time.
- *Mesoscopic* (gas-kinetic) models are in between the two previously mentioned levels, taking into account the velocity distribution. Mesoscopic models often include individual entities but model interactions between them with common fields.

Alternatively, models can also be classified by their respective detail of description:

- *Discrete space* models sub-divide the environment into a lattice, and the spatial resolution of the model is limited by the cell size of the lattice.
- *Continuous space* models describe the spatial resolution down to an arbitrary level of detail.

Also time in the model can be either discrete or continuous. The latter can be achieved if there is no fixed time step in the model. If instead time is advanced until the next *event* occurs, then that requires non-trivial calculations.

21.2.1 Agent-Based Models

A class of models which is especially popular in the computer science community is agent-based models (ABMs) (O'Sullivan and Haklay 2000; Musse et al. 1998). These models are characterized by a high level of autonomy of the simulated pedestrians, where each pedestrian is controlled by a set of rules (see Crooks and Heppenstall 2012 for an overview). The advantages with these kinds of models are that the motion can look very realistic and that the agents can be adaptive and possess a high degree of artificial intelligence, with emergent phenomena arising from simulations. This also makes ABMs suitable for crowd animation (Treuille et al. 2007; Popovic et al. 2003).

A disadvantage is that these kinds of models tend to be very complicated, which makes it hard to approach them analytically, and they typically also need a lot of computational effort. However, the separating line between ABMs and other types of microscopic models is not that clear, and in a sense, most models could be referred to or reformulated as ABMs.

21.2.2 Social-Force Model

The social-force model (Helbing and Molnar 1995; Helbing and Johansson 2009) is a microscopic model, which is continuous both in space and time. It is influenced by Newtonian mechanics, generalized to the motion of pedestrians. The forces consist of repulsive forces with respect to other pedestrians and boundaries, friction forces, attractive forces among group members, and driving forces related to desired velocities. A superposition of all these forces gives a resultant force which determines the acceleration of the pedestrians. Finally, by integrating over time, velocities and positions are obtained from the accelerations.

21.2.3 Cellular Automata Models

Another popular approach to pedestrian modeling is based on cellular automata (CA) (Bolay 1998; Blue and Adler 2000; Meyer-König et al. 2002; Batty et al. 2003; Nishinari et al. 2004; Kretz 2007; Iltanen 2012), which is a microscopic model, discrete both in time and space.

The exact specification of these models differs, but the common idea is to divide the walkable space into a lattice, where each cell has an area corresponding to the size of a human body projected onto the floor, approximately 40 × 40 cm. Each cell can either be occupied by *nobody* or by *one pedestrian*. The movements of pedestrians are carried out by iterating the time in steps in intervals of about 0.3–1.0 s. In each time step the pedestrians can move to unoccupied neighboring cells. However, even though the basic idea of CA models is simple, it often becomes complex with many rules for how the movement should be performed.

Since CA models are discrete in both time and space, and due to the fact that they use only local interactions, they are often used for simulating large crowds. One drawback of CA models, however, is the central role of the underlying lattice, which introduces artificial symmetries and tends to cause problems. An example is the tendency for deadlocks in counterflow situations at relatively low demand. The reason is that the grid structure promotes exact head-on movement. It is possible to solve this problem at the cost of giving up a part of the advantage of CA models, namely their computational efficiency. The grid structure itself poses a limit to the spatial precision. A bottleneck with a width of three cells can represent a real width just above 80 cm to just below 160 cm. Conversely, a real width of 100 cm can end up as a bottleneck with two or three cells in the model.

21.2.4 Fluid-Dynamic Models

When the crowd density is high, flows of pedestrians resembles fluid flows. Therefore, a macroscopic approach to crowd modeling is to use fluid-dynamic models (Helbing 1992; Hughes 2003) adapted to the simulation of pedestrian crowds.

An advantage of fluid-dynamic modeling of pedestrians is that it becomes possible to make analytical evaluations of changes in the infrastructure or changes in the boundary conditions.

21.2.5 Queuing Models

Queuing models (Watts 1987; Lovas 1994) make further simplifications to crowds. They are used to analyze how pedestrians are moving around in a network of modules, where the nodes and links can, for example, be doors and rooms, or intersections and roads. It is important to stress that the dynamics inside each node is not explicitly taken into consideration.

The idea is rather to grasp how the different modules are interacting with each other, by analyzing queues in the system. Each node has a certain 'service rate' and pedestrians move to the next queue as soon as they have been 'served'.

21.3 Calibration

No matter on which principles a pedestrian model is built, there is probably no model in existence without parameters. This opens the possibility and imposes the necessity to calibrate the models by comparison with empirical data. Calibration can be approached in at least three different ways: one is to measure pair-wise interactions of pedestrians in different situations, calibrate the model such that it

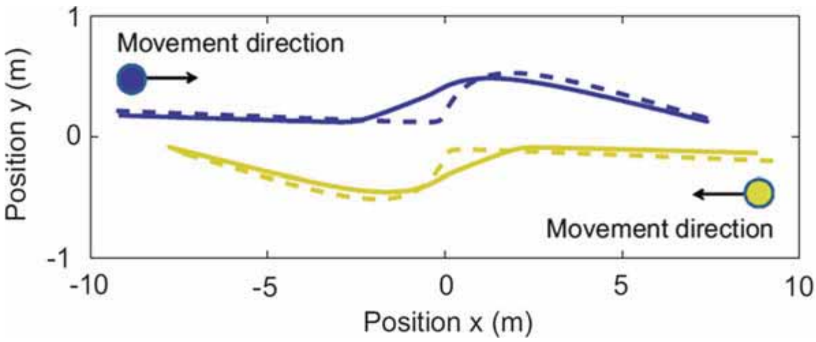


Fig. 21.1 Resulting trajectories from a simulation of two pedestrians who are approaching each other at a 180-degree angle. The simulation is carried out with the social-force model, with two different model specifications. The *dashed lines* are the resulting trajectories for an isotropic model and the *solid lines* are the resulting trajectories for an 'elliptical' anisotropic model, which gives smoother evading maneuvers and also a better fit to empirical data

reproduces these interactions and assume that the model with these parameters yields realistic results when pedestrians move in crowds (Johansson 2009). See Fig. 21.1.

The second approach is to measure aggregated macroscopic properties of moving crowds and calibrate the parameters according to these (Fischer 1933; Hankin and Wright 1958; Older 1968; Navin and Wheeler 1969; Fruin and Strakosch 1971; Predtechenskii and Milinskii 1978; Weidmann 1993; Virkler and Elayadath 1994; Muir et al. 1996; Hoogendoorn and Daamen 2005; Kretz et al. 2006a, b; Seyfried et al. 2009; Chattaraj et al. 2009; Seyfried et al. 2010a, b). The third approach is to calibrate the parameters for minimal deviation of individual trajectories of pedestrians moving in a crowd (Johansson et al. 2007; Hoogendoorn and Daamen 2009; Bauer and Kitazawa 2010), where the borderline between the second and third approach is fuzzy (Portz and Seyfried 2011).

These approaches are different methods of calibration, but they can also be combined, e.g. using method 3 for calibration and method 1 for validation to make sure that the model reproduces empirically obtained patterns on a macroscopic scale.

21.3.1 Shortest Path vs. Quickest Path

One aspect of pedestrian motion that has received very little attention is in terms of calibration work. It is this aspect which distinguishes most pedestrians from vehicles: pedestrians often choose between slowing down to walk a shorter path within a dense crowd or take some detour to keep the walking speed higher in a less dense region of the crowd (see Fig. 21.2). There is some theoretical and modeling work available on this issue (Kretz 2009a, b; Kirik et al. 2009; Steffen and Seyfried 2009; Dressler et al. 2010; Venel 2010; Rogsch and Klingsch 2010; PTV Planung Transport Verkehr AG 2010), but the empirical data are much sparser than for corridor movement.

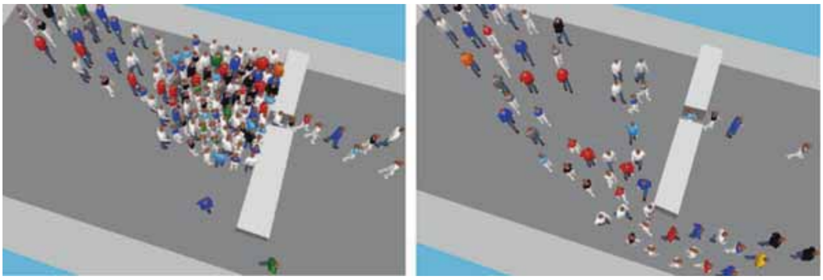


Fig. 21.2 A snapshot from a simulation with the social-force model. If the pedestrians take the shortest path (*left*), they gather in front of the bottleneck, and do not use the second path. When the pedestrians use the fastest path instead (*right*), they balance over the two possible routes

The current state of empirical crowd research is that it has even sometimes taken one step back and explicitly excluded this degree of freedom (evading vs. decelerating) in pedestrian motion by measuring the speed density relation for single file movement. This has yielded fruitful results (Chattaraj et al. 2009; Portz and Seyfried 2011; Seyfried et al. 2005, 2010b).

21.3.2 Principle of the Weakest Elements in Real-World Projects

Neither a model that perfectly describes single file movement nor a model that is perfectly calibrated for straight and wide corridor movement offers sufficient help to a project manager who is faced with a project that includes one or more corners and the movement around corners is corrupted in all models available to him, as the precision level of the entire project will normally be set by the worst precision of all elements of the simulation. Movement of a large crowd around a corner is the simplest situation one can think of where a pedestrian has to choose between trying to walk the quickest or the shortest path or something in between. Empirical efforts in the science of pedestrians need to and will turn to this aspect soon in the future.

21.3.3 Other Influences and Effects

The option set of travel time vs. travel distance can be generalized to a concept of generalized costs, as it has been done in a number of the utility-based models. Then not only travel time and travel distance can be combined to a utility for an individual pedestrian, but also, for example, the discomfort of walking on a bicycle lane or over muddy terrain or the comfort of walking shaded from rain or sunshine can be integrated in just the same manner. It is well known that the free speed (or desired speed) of pedestrians depends on their demographics, e.g. age, sex, air temperature, trip purpose (e.g. commuting or leisure time), time of day, trip circumstances (e.g. early for a train or late), culture and probably some more factors (Weidmann 1993; Chattaraj et al. 2009; Buchmüller and Weidmann 2006). Some of these factors are correlated (e.g. certain trip purposes occur not distributed equally over a day, air temperature has a typical course during a day, etc.). As the desired speed is relevant in any other movement situation, we may infer that these parameters also influence all other movement situations.

21.4 Navigation

A crowd simulation project is set up by defining the relevant boundaries: the spatial boundaries of the walkable area (i.e. the geometry of the model) and the boundary between external knowledge about the dynamics of the pedestrians and where the model needs to take over – in other words the localized demand, which is the inflow into the model, and how it varies over time. The last elementary definition a modeler needs to do is to set the destinations for the agents, which are set in the model.

The first and most basic element that a dynamics model then needs to include is a navigation or wayfinding method from the positions of the inflow ('sources') of the pedestrians to the given destinations ('sinks'). The remainder of the section will deal with that task.

One way to achieve shortest route navigation to the destination is to make use of a visibility graph (de Berg et al. 1997). Simply speaking, a visibility graph is the graph of mutually visible corners of obstacles of a pedestrian movement geometry. With Dijkstra's algorithm (Dijkstra 1959), the shortest path from the agent's current position to the closest corner point of the destination polygon can be found. When using this method, one is faced with the difficulty of where to place the navigational points exactly: if only individual agents are moving through the geometry in low density, the navigation points can be very close to the corners of the obstacles. If the agents are moving in large groups, then the navigation points need to be placed further away. Moreover, there has to be some minimum distance that allows agents to come close to these points such that an agent can proceed toward the next navigation point.

As an example in the social-force model (Helbing and Johansson 2009), the beeline direction from the current position of an agent toward the next navigation point would then typically use the direction of the desired velocity (the absolute value of the desired velocity is an external parameter).

A method that avoids this difficulty, but which requires more computational effort, is that of a floor field (also called "static potential"), which is a grid placed over the geometry, where the distance towards the destination (under consideration of the obstacles) is written to each grid point. Plainly spoken, it is a localized look-up table of distances. There are numerous methods to calculate this static potential. Typically the calculation time rises when the deviation from the Euclidean distance

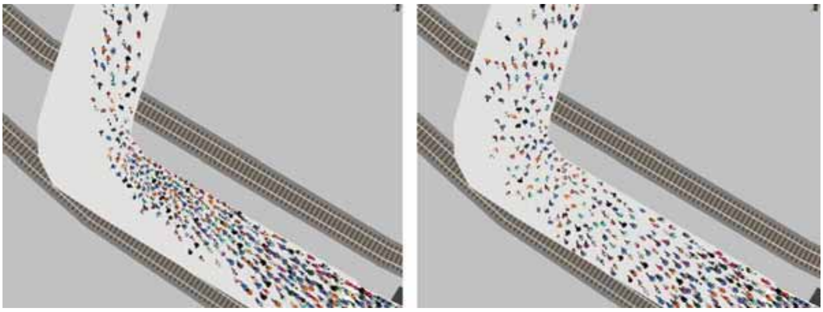


Fig. 21.3 A static potential field (*left*) compared to a dynamic potential field (*right*). The snapshots are taken at the same time instant in two identical simulation scenarios. Notice how pedestrians get stuck and delayed at the corner when a static potential field is used

is reduced (Kimmel and Sethian 1998; Jeong and Whitaker 2008; Kretz et al. 2010). The negative gradient of the static potential at the position of an agent gives the direction of the shortest path from that agent to the destination. Thus, the negative gradient of the static potential is used as the direction of the desired velocity. Recently a method has been proposed that directly and efficiently calculates the gradients without the need to calculate the static potential (Schultz et al. 2010).

The Fast Marching Method (Kimmel and Sethian 1998) and Fast Iterative Method (Jeong and Whitaker 2008) are well suited to also calculate a floor field, which contains the estimated remaining travel time from a grid cell to the destination (PTV Planung Transport Verkehr AG 2010). Contrast this with the method employed in Sect. 21.3.1. As the distribution of agents has a major impact on the estimated remaining travel time, and as the distribution of agents naturally changes in each simulation time step, such a floor field needs to be recalculated frequently. Therefore, it is called the 'dynamic potential'. In this way it is possible to make agents in the social-force model evade groups of other agents dwelling around or being jammed at a bottleneck or the inner side of a corner early on, already by the direction of their desired velocity (see Fig. 21.3). For the dynamics of the whole system this means that jams do not grow endlessly and that agents distribute better. Therefore this method can be seen as a kind of non-iterative assignment in two continuous spatial dimensions.

Let us assume that an agent wants to reach its destination as quickly as possible. In principle the 'bee line' would be the quickest way. An agent walking the shortest path under consideration of obstacles is modeled as someone, who accepts that inevitably obstacles prevent one from walking along the bee line. In principle the shortest path under consideration of obstacles would also be the quickest path under consideration of obstacles. An agent walking into the direction of the estimated quickest path under consideration of all other agents is modeled as someone who accepts that jams will inevitably cause delays and therefore might prevent the shortest path from being the quickest.

While real pedestrians can be assumed to have a very good comprehension of the situation around them, and while there are situations where it is safe to assume that early arrival is the single-most important movement criterion (passengers in a station who are late for a train), modeling pedestrians to walk on the shortest path under consideration of obstacles can nevertheless be justified in many situations. First there are situations in which the quickest path is not much different from the shortest and where inter-pedestrian forces can reproduce these differences.

Second, there can be situations, where the shortest path is valued more than the quickest. As has been stated above the quickest path/shortest path trade-off can be seen as a special case of a generalized cost. Instead of calculating a field of estimated remaining travel time to the destination, it is also possible to calculate a field of generalized cost to reach the destination associated with the field. This shows that by using the gradient of such a field as the direction of the desired velocity in the social-force model, it is possible to connect the force-based approach with the utility-based approach. This can be interpreted such that the information entering the direction of the desired velocity models the free planning process of an agent, while the forces act according to their name, and they force the agent to evade other agents at rather small distances to avoid collisions.

21.5 Conclusions

Pedestrian crowd modeling has emerged as a mature and active field of research, where models are challenged on their ability to reproduce empirically observed features. This has resulted in crowd simulation tools, both commercial and freely availably ones, that are routinely used in the planning of major events, and also for optimizing transport systems, assessing building evacuations, optimizing the organization of airports and train stations, etc. Some of the challenges ahead are to reach consensus as to which modeling approaches yield the most realistic results. Another ongoing challenge is to make crowd modeling tools more autonomous. Earlier crowd modeling tools relied heavily on the user to specify every single detail in the model scenario, whereas in more recent models, pedestrians find their way around complex spaces, they queue, and they even interact with and use public transport.

References

Batty, M., Desyllas, J., & Duxbury, E. (2003). Safety in numbers? Modelling crowds and designing control for the Notting Hill carnival. *Urban Studies, 40*(8), 1573–1590.

Bauer, D., & Kitazawa, K. (2010). Using laser scanner data to calibrate certain aspects of microscopic pedestrian motion models. In W. W. F. Klingsch, C. Rogsch, A. Schadschneider, & M. Schreckenberg (Eds.), *Pedestrian and evacuation dynamics 2008* (pp. 83–94). Berlin: Springer.

Blue, V. J., & Adler, J. L. (2000). Cellular automata microsimulation of bi-directional pedestrian flows. *Transportation Research Record, 1678*, 135–141. doi:http://dx.doi.org/10.3141/1678-17.

Bolay, K. (1998). *Nichtlineare Phanomene in einem fluid-dynamischen Verkehrsmodell*. Unpublished diploma thesis, University of Stuttgart, Stuttgart.

Buchmüller, S., & Weidmann, U. (2006). *Lehrstuhl für verkehrssysteme. Parameters of pedestrians, pedestrian traffic and walking facilities*. Zurich: Institute for Transport Planning and Systems (IVT) ETH Zurich.

Chattaraj, U., Seyfried, A., & Chakroborty, P. (2009). Comparison of pedestrian fundamental diagram across cultures. *Advances in Complex Systems, 12*(3), 393–405.

Crooks, A. T., & Heppenstall, A. J. (2012). Introduction to agent-based modelling. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 85–105). Dordrecht: Springer.

de Berg, M., van Krefeld, M., Overmars, M., & Schwarzkopf, O. (1997). *Computational geometry*. Berlin: Springer.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik, 1*(1), 269–271.

Dressler, D., Groß, M., Kappmeier, J. P., Kelter, T., Kulbatzki, J., Plumpe, D., Schlechter, G., Schmidt, M., Skutella, M., & Temme, S. (2010). On the use of network flow techniques for assigning evacuees to exits. *Procedia Engineering, 3*, 205–215.

Fischer, H. (1933). *Uber die Leistungsfähigkeit von Türen, Gängen und Treppen bei ruhigem, dichtem Verkehr*. Unpublished dissertation, Technische Hochschule Dresden, Dresden.

Fruin, J. J., & Strakosch, G. R. (1971). *Pedestrian planning and design*. New York: Metropolitan Association of Urban Designers and Environmental Planners.

Hankin, B. D., & Wright, R. A. (1958). Passenger flow in subways. *Operational Research Quarterly, 9*, 81–88.

Helbing, D. (1992). A fluid-dynamic model for the movement of pedestrians. *Complex Systems, 6*, 391–415.

Helbing, D., & Johansson, A. (2009). Pedestrian, crowd and evacuation dynamics. In R. A. Meyers (Ed.), *Encyclopedia of complexity and systems science* (Vol. 16, p. 6476). New York: Springer.

Helbing, D., & Molnar, P. (1995). Social force model for pedestrian dynamics. *Physical Review E, 51*, 4282–4286.

Hoogendoorn, S. P., & Daamen, W. (2005). Pedestrian behavior at bottlenecks. *Transportation Science, 39*(2), 147–159.

Hoogendoorn, S. P., & Daamen, W. (2009). A novel calibration approach of microscopic pedestrian models. In H. Timmermans (Ed.), *Pedestrian behavior: Models, data collection and applications* (pp. 195–214). Bingley: Emerald.

Hughes, R. L. (2003). The flow of human crowds. *Annual Review of Fluid Mechanics, 35*, 169.

Iltanen, S. (2012). Cellular automata in urban spatial modelling. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 69–84). Dordrecht: Springer.

Jeong, W.-K., & Whitaker, R. (2008). A fast iterative method for Eikonal equations. *SIAM Journal on Scientific Computing, 30*(5), 2512–2534.

Johansson, A. (2009). *Data-driven modeling of pedestrian crowds*. Saarbrücken: VDM Verlag Dr. Mueller.

Johansson, A., Helbing, D., & Shukla, P. K. (2007). Specification of the social force pedestrian model by evolutionary adjustment to video tracking data. *Advances in Complex Systems, 10*(4), 271–288.

Kimmel, R., & Sethian, J. A. (1998). Computing geodesic paths on manifolds. *Proceedings of the National Academy of Sciences of the United States of America, 95*(15), 8431.

Kirik, E. S., Yurgelyan, T. B., & Krouglov, D. V. (2009). The shortest time and/or the shortest path strategies in a CA FF pedestrian dynamics model. *Journal of Siberian Federal University. Mathematics & Physics, 2*, 271–278.

Kretz, T. (2007). Pedestrian traffic – Simulation and experiments. PhD thesis, Universitat Duisburg-Essen. urn:nbn:de:hbz:464-20070302a 120944-7

Kretz, T. (2009a). Pedestrian traffic: on the quickest path. *Journal of Statistical Mechanics: Theory and Experiment, 2009*(3), P03012.

Kretz, T. (2009b). The use of dynamic distance potential fields for pedestrian flow around corners. In *First international conference on evacuation modeling and management.* Delft: Technische Universiteit Delft.

Kretz, T., Bönisch, C., & Vortisch, P. (2010). Comparison of various methods for the calculation of the distance potential field. In W. W. F. Klingsch, C. Rogsch, A. Schadschneider, & M. Schreckenberg (Eds.), *Pedestrian and evacuation dynamics 2008* (pp. 335–346). Berlin: Springer.

Kretz, T., Grünebohm, A., Kaufman, M., Mazur, F., & Schreckenberg, M. (2006a). Experimental study of pedestrian counterflow in a corridor. *Journal of Statistical Mechanics: Theory and Experiment, 2006*(10), P10001.

Kretz, T., Grünebohm, A., & Schreckenberg, M. (2006b). Experimental study of pedestrian flow through a bottleneck. *Journal of Statistical Mechanics: Theory and Experiment, 2006*(10), P10014.

Lovas, G. G. (1994). Modeling and simulation of pedestrian traffic flow. *Transportation Research, 28B*, 429–443.

Meyer-König, T., Klüpfel, H., & Schreckenberg, M. (2002). Assessment and analysis of evacuation processes on passenger ships by microscopic simulation. In M. Schreckenberg & S. D. Sharma (Eds.), *Pedestrian and evacuation dynamics* (pp. 297–302). Berlin: Springer.

Muir, H. C., Bottomley, D. M., & Marrison, C. (1996). Effects of motivation and cabin configuration on emergency aircraft evacuation behavior and rates of egress. *The International Journal of Aviation Psychology, 6*, 57–77.

Musse, S. R., Babski, C., Capin, T., & Thalmann, D. (1998). Crowd modelling in collaborative virtual environments. In *Proceedings of the 1998 ACM symposium on Virtual reality software and technology* (pp. 115–123). New York: ACM.

Navin, P. D., & Wheeler, R. J. (1969). Pedestrian flow characteristics. *Traffic Engineering, 39*, 31–36.

Nishinari, K., Kirchner, A., Namazi, A., & Schadschneider, A. (2004). Extended floor field CA model for evacuation dynamics. *IEICE Transactions on Information and Systems, E87-D*, 726–732.

O'Sullivan, D., & Haklay, M. (2000). Agent based models and individualism: Is the world agent-based? *Environment and Planning A, 32*, 1409–1425.

Older, S. J. (1968). Movement of pedestrians on footways in shopping streets. *Traffic Engineering and Control, 10*, 160–163.

Popovic, J., Seitz, S. M., & Erdmann, M. (2003). Motion sketching for control of rigid-body simulations. *ACM Transactions on Graphics, 22*(4), 1034–1054.

Portz, A., & Seyfried, A. (2011). Analyzing Stop-and-Go Waves by Experiment and Modeling. In: R. D. Peacock, E. D. Kuligowski, & J. D. Averill (Eds) *Pedestrian and evacuation dynamics (*pp 577–586) Springer.

Predtechenskii, V. M., & Milinskii, A. I. (1978). *Planning for foot traffic flow in buildings*. Washington, DC: Amerind Pub. Co.

PTV Planung Transport Verkehr AG. (2010). *VISSIM 5.30 user manual*.

Reynolds, C. W. (1987). Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics, 21*(4), 25–34.

Rogsch, C., & Klingsch, W. (2010). Basics of software – tools for pedestrian movement identification and results. *Fire Technology*, (5 November 2010), 1–21.

Schultz, M., Kretz, T., & Fricke, H. (2010). Solving the direction field for discrete agent motion. In S. Bandini, S. Manzoni, H. Umeo, & G. Vizzari (Eds.), *Cellular automata – 9th international conference on cellular automata for research and industry (ACRI 2010)*. Berlin: Springer.

Seyfried, A., Boltes, M., Kuhler, J., Klingsch, W., Portz, A., Rupprecht, T., Schadschneider, A., Steffen, B., & Winkens, A. (2010a). Enhanced empirical data for the fundamental diagram and the flow through bottlenecks. In W. W. F. Klingsch, C. Rogsch, A. Schadschneider, & M. Schreckenberg (Eds.), *Pedestrian and evacuation dynamics 2008* (pp. 145–156). Berlin: Springer.

Seyfried, A., Passon, O., Steffen, B., Boltes, M., Rupprecht, T., & Klingsch, W. (2009). New insights into pedestrian flow through bottlenecks. *Transportation Science, 43*(3), 395–406.

Seyfried, A., Portz, A., & Schadschneider, A. (2010a). Phase coexistence in congested states of pedestrian dynamics. *Cellular Automata*, 496–505.

Seyfried, A., Steffen, B., Klingsch, W., & Boltes, M. (2005). The fundamental diagram of pedestrian movement revisited. *Journal of Statistical Mechanics: Theory and Experiment, 2005*(10), P10002.

Steffen, B., & Seyfried, S. (2009). Modelling of pedestrian movement around 90 and 180 bends. In B. H. V. Topping, Y. Tsompanakis (Eds.), *Proceedings of the first international conference on soft computing technology in civil, structural and environmental engineering* (Paper 34). Civil-Comp Press, Stirlingshire.

Treuille, A., Lee, Y., Popovic, Z. (2007). Near-optimal character animation with continuous control. *ACM Transactions on Graphics* 26(3). (SIGGRAPH '07 Conference Proceedings) Article 7.

Venel, J. (2010). Integrating strategies in numerical modelling of crowd motion. In W. W. F. Klingsch, C. Rogsch, A. Schadschneider, & M. Schreckenberg (Eds.), *Pedestrian and evacuation dynamics 2008* (pp. 641–646). Berlin: Springer.

Virkler, M. R., & Elayadath, S. (1994). Pedestrian density characteristics and shockwaves. In R. Akcelik (Ed.), *Proceedings of the second international symposium on highway capacity* (Vol. 2, pp. 671–684).

Watts, J. M. (1987). Computer models for evacuation analysis. *Fire Safety Journal, 12*, 237–245.

Weidmann, U. (1993). *Transporttechnik der Fußgänger – Transporttechnische Eigenschaften des Fußgängerverkehrs*. Schriftenreihe des IVT 90, ETH Zürich, 3. Zweite, ergänzte Auflage, in German.

Chapter 22
Business Applications and Research Questions Using Spatial Agent-Based Models

William Rand

Abstract Agent-based models (ABMs) provide a natural representation of large markets with many consumers interacting. As a result, business applications of these tools can provide powerful insights in to complex problems. When spatial and geographic modeling is added as well, these insights gain the ability to be transported to the real world, where challenging questions can be addressed. Most of the work that has been done in this area focusses on two different levels of spatial models: (1) the regional or macro-level, and (2) the small-scale or micro-level. Macro-level spatial ABMs are models which address the movement of individuals and the location of facilities across an entire region, such as spatial retail decisions, residential housing choices, or geographically extended supply chains. Micro-level spatial ABMs examine the movement of individuals within a constrained physical space, such as pedestrian modeling in a neighborhood, or consumer modeling within a retail location. We will discuss each of these levels of detail in turn and finish by discussing future applications of spatial ABMs to business.

22.1 Introduction

Despite some initial success,[1] the use of Agent-Based Models (ABMs) in business applications has only recently started to garner serious interest within the practitioner community (North and Macal 2007; North et al. 2010). One reason for this recent increase in interest may be that it has become increasingly clear that traditional modeling approaches are not able to handle many of the complexities and details involved in a modern market. Though the academic community has used ABMs for a long

[1]For example, Proctor and Gamble's use of ABM to redesign their supply chain, see Anthes (2003).

W. Rand (✉)
Department of Marketing and Center for Complexity in Business, University of Maryland,
3457 Van Munching Hall, College Park, MD, 20742, USA
e-mail: wrand@umd.edu

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_22,

time within managerial science, dating back at least to Cohen, March and Olsen's use of an agent-based modeling framework to understand organizational decision making (Cohen et al. 1972),[2] the recent interest by business experts in ABM has reinvigorated the academic community's research in to this unique use of ABM in a number of different managerial sub-disciplines from marketing (Rand and Rust 2011) to finance (Ehrentreich 2007) to managerial strategy (Davis et al. 2009).

This renewed interest is well-deserved. ABM has several natural advantages for modeling business situations: (1) *Multiscale and Complex Interaction Structures:* Most business applications deal with interactions between either multiple businesses of varying size, businesses and consumers, or individuals at different levels within an organization. Moreover, even when these agents are all of the same scale, often the interactions within a model need to occur on a complex interaction structure such as a social network; ABM provides the ability for multiple scales of agency and network-based interactions to be included within the same model. (2) *Heterogeneity:* Many business applications, such as consumer market models, require a large number of individuals who are very different from each other. These individuals could have different personal wealth status, different thresholds to innovation, or different preferences for a product. Though new approaches in equation-based/math modeling have allowed for more and more heterogeneity to be incorporated in to a model, ABM still provides a maximal level of potential heterogeneity. (3) *Adaptive Agents*: Most organizations and consumers do not operate according to the same rules of behavior throughout time; instead they learn from the past and change their behavior as a result of previous actions and consequences. This is especially true when there are large rewards to be gained for acting as optimally as possible. ABM is one of the few approaches that allows the modeler to construct agents which not only adapt the parameters of the rules by which they act, but to fundamentally alter the rules themselves. (4) *Rich Environments* – Many business applications occur distributed through a physical space; whether that be a city road network, a county development map, a neighborhood street, or within a store, the physical geography of these systems can dramatically alter the way the complex system unfolds. ABM allows for the relatively simple inclusion of this physical geography as an environment in which the agents operate and ABM has proven to be successful in helping to understand complex phenomenon related to geography and urban development in the past (Benenson and Torrens 2004; Batty 2005).

It is this last advantage that we will discuss in detail in this chapter. Of course, spatial models of business systems occur that do not involve ABMs (Longley and Clarke 1995), but GIS and spatial modeling techniques by themselves are not sufficient to capture the rich details necessary for some models. Static spatial models are a representation of pattern and describe very well how items of interest are distributed in space and relate to each other. However, static spatial models fail to describe temporal dynamics very well, and thus they lack a representation of process. ABM,

[2] Though Cohen, March and Olsen used what is clearly an ABM, they did not call their model, an "ABM", since that phrase was not yet in use.

on the other hand, describes processes very well; in fact, it might be argued that ABM is by its nature a process description, but without a rich environmental description, such as that provided by GIS. Without spatial modeling, ABM fails to capture the nuances of sophisticated spatially distributed patterns (Brown et al. 2005b). However, by combining ABM and GIS, researchers can build sophisticated tools that model both pattern and process simultaneously. Models of pattern and process are critical to several interesting business applications, since the discovery of spatially distributed patterns and how they evolve over time is key to their understanding. Therefore, spatial ABMs could prove a powerful tool within management science.

In order to explore this hypothesis in more depth, we will examine a number of different applications of spatial ABMs to questions of relevance within business applications. We will begin by examining macro-level models, or models where the scale of the model does not require detailed models of individual entity movement; examples of these types of models include residential housing decisions, retail location preferencing, and geographically extended supply chains. Then we will examine micro-level models, which are models in which the base representation is an individual moving around as a pedestrian; examples of these types of models include pedestrian traffic in a neighborhood, or even movement within a retail shop. We utilize this distinction since the types of data and models of behavior will differ significantly between these two levels. At the end of this chapter we will discuss how far spatial ABMs have advanced in business applications and potential avenues for future research.

22.2 Macro-Level Spatial ABMs

Large-scale spatial ABMs are useful in business application in which the business or phenomenon being modeled is large enough to extend spatially over an entire region. For instance, a delivery service that must manage a fleet of trucks could use ABM to maximize the routing of the trucks incorporating a traffic model that varies based on time of day (Hartenstein et al. 2001), or a chain of stores that are trying to cover a market ensuring that no store is located too far away from any consumer could use ABM to determine a new store location taking into account consumer commuting patterns (Lombardo et al. 2004), or a business that deals with a global distribution of goods and shipping logistics could use these technologies to track and examine the distribution of containers throughout the world (Sinha-Ray et al. 2003). Macro-level models are also useful in cases where a business competes with other firms and organizations and the needs of their customers are influenced by geography, even if the focal firm itself does not have multiple locations or provide services over a geography. For instance, a residential developer could use ABM to forecast future demand for new housing based upon individual decisions to locate in various regions (Brown et al. 2005a), or a gas station could use ABM to examine how to set its prices in comparison to its competitors (Heppenstall et al. 2005).

A significant barrier in the past to increased use of ABMs to model large-scale geographically dispersed systems was the lack of integration between ABM tools and GIS tools (Brown et al. 2005b). As eluded to in the introduction, an ABM, in its simplest form, represents dynamic aspatial processes, and so the basic concepts of an ABM are agents, timesteps, and behavior. A GIS, on the other hand, represents static spatial patterns, and so the basic concepts of a GIS are maps, projections, and spatial analysis. It is not intuitively clear how to get these two methodologies to work together in a way that facilitates the development of more complex models (Brown et al. 2005b). Should the ABM be the primary model, using the GIS simply as a storage platform? Or should the GIS be the primary model, using the ABM to update the spatial patterns? Or should both tools be rolled up in to one cohesive package? An ideal tool to address this problem has still not been developed, but there have been significant advances in tool development that have helped mitigate this problem. Specifically most of the most commonly used ABM platforms, such as NetLogo (Wilensky 1999) and RePast (North et al. 2005), have incorporated the ability to read and write GIS data. Though there are significant challenges in using spatial ABMs to address large-scale geographies, several research projects have successfully used this approach to develop new understandings to complex problems. In the next few sections we will highlight a couple of these projects and talk about them in more detail. This is not meant to be an exhaustive listing, but rather to provide a few illustrative examples.

22.2.1 Spatial Retailing Decisions

Almost all consumer retail decisions are spatially-influenced. In the end when it comes to purchasing from a brick-and-mortar retail location, whether it be out of preference (e.g., the ability to try on clothing, the convenience of lack of shipping, etc.) or necessity (e.g., there is no economically or practically feasible way to ship gasoline over the internet), consumers prefer to buy from locations that are spatially convenient to them, i.e., they aim to minimize some cost associated with the transaction (Hotelling 1929). This cost can result in firms locating near other firms since the same spatial location may wind up being the most convenient for a large part of the population. On the other hand, firms may want to locate as far away from each other as possible, in order to not compete directly with other firms nearby (d'Aspremont et al. 1979). Involved in all these decisions is not only where to locate, but what products to offer at what price; all of which are interrelated decisions (Hotelling 1929).

Despite the advances that have been previously made using game theory and equation-based modeling, more advanced models of these decision processes can be made by combining consumer-level behavioral rules with agent-based models. An example of this combination working well together is Heppenstall, Evans and Birkin's examination of petrol price setting in a spatially-influenced retail market (Heppenstall et al. 2005, 2006). This model examined the price setting behavior of individual petrol stations in the geographic area of West Yorkshire, UK (including Leeds, Bradford, Wakefield, Huddersfield and Halifax). In the model, petrol stations

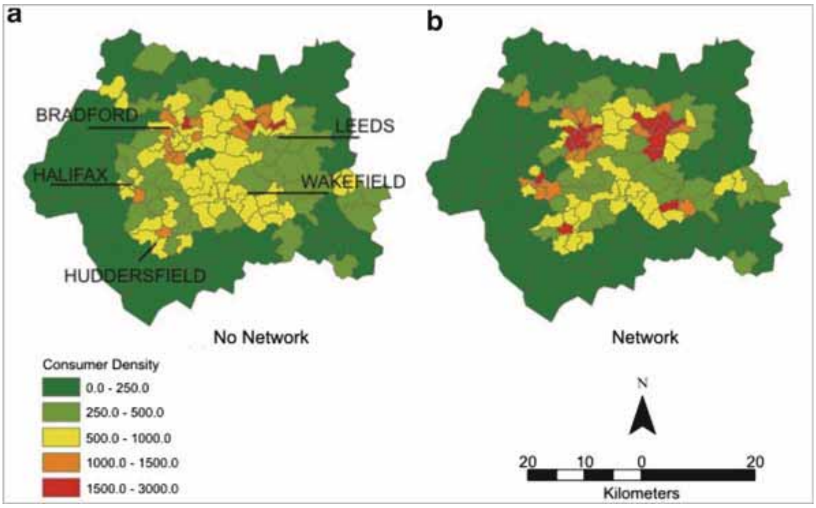


Fig. 22.1 Differences in the distribution of consumer density in the "hybrid" model (**a**) and the "network" model (**b**) (Reprinted with permission from Heppenstall et al. (2006))

were represented as full agents, while the consumers were represented by a spatial interaction model, since fully representing the consumers as agents as well was too computationally expensive. Within the model petrol stations set their prices based on local competition and profit earned in the last time period. Consumer demand was modeled using a spatial interaction model, where consumers decided which petrol stations to buy from on the basis of distance and price. In the initial "hybrid model" (it was called a hybrid model because it involved both an ABM and a spatial interaction model), consumers were always considered to exist at their home location. Heppenstall et al. (2005, 2006) went on to examine a "network" model, which modeled commutes to work as well and used this commuting data to re-examine the spatial model on the basis of where consumers would most likely be when driving. The difference between these two models in terms of their effect on consumer density within the ABM is illustrated in Fig. 22.1. After constructing these two models, they then compared and contrasted these two different models, performed robustness and validation checks, and showed that the model recreated real-world price and profit patterns.

This trade-off between the hybrid model, which models resident locations using their residency locations, and the network model, which approximates where residents will be when they are likely to refuel their cars is illustrative of design decisions that often need to be made when constructing spatial agent-based models at the macro-scale. The petrol stations that were modeled were well-defined and in fixed locations, so they were modeled at their present stations, but commuters are mobile, and so a decision must be made as to which level of mobility to model them. The hybrid model is an extreme case where the residents are modeled as existing at one fixed location, namely their homes. The advantage of this approach is that it is less computationally expensive, but the disadvantage is that the model is not as reflective of the real-world as a more fine-grained model that modeled actual transportation patterns. At the other end of the spectrum, commuters could be modeled as constantly moving, and their locations dynamically modeled at a minute-by-minute resolution. This has the advantage of being a better representation of the real-world, but also has the disadvantage of being extremely computationally expensive, not to mention the difficulties with obtaining the data necessary to model traffic data at that level of detail. The network model that Heppenstall, Evans and Birkin chose is a well-designed compromise that balances computational and data limitations with the fidelity of the model.

Once a model such as this has been constructed and validated, the results could be used to explore future chain-wide policy decisions, examine marketing strategies, or even determine the viability of current and future retail locations. One of the most intriguing aspects of this model is that there are many necessity goods for which a similar model could be built. Though petrol is uniquely dependent on the transportation system for its demand, in general people do like to shop in proximity to either their home or work (or someplace in-between) for most of their basic necessities. Thus, it would be possible to imagine similar models of prices that could be constructed for grocery stores (Schenk et al. 2007), big-box retail stores, restaurants, and health and beauty services, which are all retail locations that a commuter might visit once every few weeks.

22.2.2 Residential Location Preferencing

The petrol station model described in the previous section was mainly aimed at building a comprehensive model of the supply-side of the market, including retailers and competitors. The demand-side of the market was exogenous, assumed to operate using a fixed set of rules, and was not explicitly modeled as agents. However, demand-side modeling at the macro-level can also benefit from the use of spatial ABMs. As was mentioned previously, it is well-known that consumers make decisions based upon proximity and cost of travel (Hotelling 1929), so modeling where consumers will locate and the demand that they place upon a geographical area is critical to organizations and businesses that rely upon these forecasts.

For instance, within the housing market understanding where residents want move to is useful for contractors, developers, retailers, and even public policy analysts. One project that attempted to analyze this process using spatial ABM is project SLUCE (**S**patial **L**and **U**se **C**hange and **E**cological Effects)[3] and its successor SLUCE 2. The goal of which was to examine the relationship between the environment and suburban sprawl. Though the aim of project SLUCE was more on the side of providing advice to policymakers interested in reducing the negative effects of suburban sprawl, a model similar to the one that was constructed could also be used

[3] http://www.cscs.umich.edu/sluce/

from a business and organization perspective in understanding future demand patterns in a geographic area.

To address these concerns, Project SLUCE has built at least three different versions of the models to address different questions (Brown et al. 2008). The first model that was constructed, called SOME (**S**luce's **O**riginal **M**odel for **E**xploration), was built in the Swarm modeling toolkit[4] (Minar et al. 1996). The core agents of the SOME model are residents and service centers that interact on a complex landscape that includes roads and landscapes of aesthetic quality; both of which can either be generated or drawn from empirical GIS data. Later, a NetLogo[5] (Wilensky 1999) version of the SOME model was constructed. This model was a stripped down version that was used primarily to examine general patterns and as an educational tool. Finally, a new model was created, called DEED (**D**ynamic **E**cological **E**xurban **D**evelopment)(Brown et al. 2008), which was constructed using the RePast toolkit[6] (North et al. 2005). This model included several new agent types including: farmers, policy boards, and developers; as well as the residents present in the SOME model. This allowed more realistic models of exurban development to be constructed and different questions of policy impact to be explored; all of which are heavily dependent upon spatial interactions that exist between these various stakeholders.

Project SLUCE has used this suite of models to examine a number of different questions including: (1) the interaction of residents with policy constraints (e.g., greenbelts) (Brown et al. 2004) (2) the validation of the models against classically observed empirical patterns (Rand et al. 2003), (3) the development of new spatial validation techniques (Brown et al. 2005a), (4) the role of zoning in exurban sprawl (Zellner et al. 2009, 2010), and several other research questions (Brown et al. 2008).

A key aspect in investigating all of these different possibilities was that SLUCE had developed a suite of models rather than just one model. This allowed the project researchers to employ the right model at the right time. This approach, which has sometimes been referred to as full-spectrum modeling (Rand and Wilensky 2007), embraces the idea of choosing the right model for the right level of detail, backgrounding other decisions when need be, and at the same time bringing other details of interest to the front. Choosing the correct level of detail is often a concern in macro-level spatial modeling and something that should be considered whenever developing a spatial ABM with business applications. Sometimes there will be some agents that need to be emphasized, such as residents, where in other cases other agents will be more important, such as developers. The reason why this is particularly relevant when it comes to considering spatial ABMs at the macro-level is that there are many different types of consumers in any context. One of the arguments for the development of the DEED model by Project SLUCE was that for many policy questions regarding new housing, the developer, as opposed to the individual resident/homeowner, was a more appropriate unit of analysis, and so as Project

[4] http://www.swarm.org/

[5] http://ccl.northwestern.edu/netlogo/

[6] http://repast.sourceforge.net/

SLUCE transitioned to examining more of these types of questions, a new model was constructed. In the SOME model, the consumer being represented was the resident, consuming a housing location and making a residential location decision based upon the amenities offered, while in the DEED model the primary consumer was a developer who essentially consumed zoning policies and made development decisions based upon them. For researchers, and business practitioners interested in representing a large-scale market using a spatial ABM it may be useful to consider the possibility of constructing multiple models at different scales that represent these different levels of consumption and market interaction.

Macro-scale models of market demand and consumption could also be useful in other contexts as well. For instance, it is possible to build "heat maps" of potential future demand in different areas, and then use those to influence future retail construction decisions, or even future logistical needs for an organization that is a service provider to other industries. Moreover, consumers like to not only shop near where they live, but they also like to work near where they live. As consumers of employees, businesses should pay attention to how future housing trends are developing and attempt to locate where they will have the best access to highest quality labor. One of the findings of Project SLUCE (Brown et al. 2005a) was that often it is useful to examine not only individual runs of a model and "frequency" or "heat maps" which illustrate how often an area gets developed in the model, but also to break model forecasts up in to "variant" regions (regions that are sometimes developed and sometimes not under the model) and "invariant" regions (regions that are almost always developed or not developed in the model). This is useful because development tends to be path dependent, meaning that new development tends to follow recent development even if the recent development occurred at its location purely due to happenstance. Figure 22.2 illustrates how to explore these different effects using an ABM and teh comparison of these projections to an underlying reference map.

22.2.3 Other Macro-Level Models

There are many other business phenomenon that could be and have been modeled using spatial ABMs at the macro-level. For instance, geographically extended supply networks, where the physical distances between the locations is vital to understanding how the supply network operates could be modeled using macro-level spatial ABMs. Spatial ABMs can be used to examine dynamic supply networks that evolve and change in time (Emerson and Piramuthu 2004) taking in to account the information moving across these networks which has its own dynamics (Ahn and Lee 2004). Related to supply chains, is the question of vehicle routing. Given that you have a particular fleet of vehicles that need to deliver goods to a variety of locations what is the best way to do so, especially in the presence of changing traffic conditions, and potentially changing consumer demands (Kohout and Erol 1999). Moreover, and somewhat extending on both of the examples listed above, is the question of retail location placement (MacKay 1972). If we have a good representation of consumer

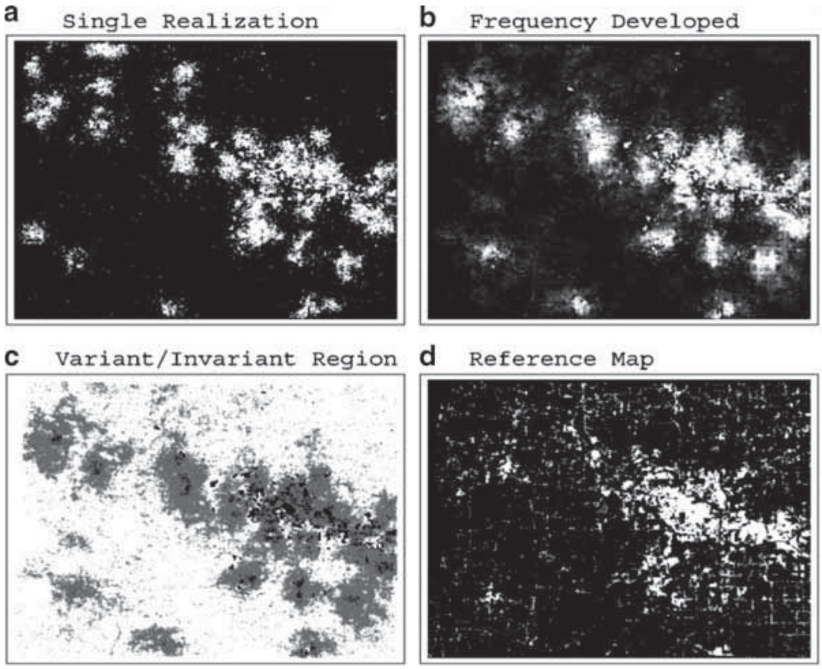


Fig. 22.2 Different ways of examining development models: (**a**) illustrates a single run of the model with white areas being developed land, and the black areas being undeveloped land, (**b**) uses grey scale to illustrate the frequency with which an area was developed over a large number of model runs, (**c**) breaks the map up into areas that are always developed (*black*), never developed (*white*) and sometimes developed (*grey*) for a large number of model runs, (**d**) is a reference map indicated what actually happened (Reprinted by permission from Brown et al. (2005))

movement in a market, and detailed information about competitor placement and prices, then we can ask the question: where should a firm locate its next store and should it close down any stores that it currently has open? (Lombardo et al. 2004) Such a model would provide a sophisticated decision support tool; but spatial ABMs could also be used to construct descriptive and exploratory models (Huang and Levinson 2008). For instance, they could be used to investigate why urban and retail concentrations occur in the first place? (Krugman 1996)

Finally, one interesting application of spatial ABMs at the macro-level does not involve any notion of physical movement but instead the movement of information. There has recently been considerable interest in understanding the diffusion of information across social media from a marketing perspective (Trusov et al. 2009; Domingos 2005). It might be thought that in these systems geography and space are unimportant, but in fact there is some evidence that geography still plays a role in diffusion of information (Goldenberg and Levy 2009). Thus, one macro-level spatial ABM that might be useful from a marketing and business perspective would investigate the spread of information across social networks and geography at the same time, and examine optimal strategies for seeding viral marketing strategies (Stonedahl et al. 2010) taking in to account both geography and network properties simultaneously.

22.3 Micro-Level Spatial ABMs

On the other end of the spectrum from macro-level spatial ABMs is micro-level spatial ABMs. Such models may still use GIS data to provide the environment within which the agents move, but the unit of focus here is on the individual constrained within a smaller geographic space. In particular these models require some model of pedestrian movement, whether it be within a retail location (Zacharias et al. 2005), through a mall (Bitgood and Dukes 2006), or around a neighborhood (Batty et al. 1998; Borgers and Timmermans 1986). Unlike the macro-level models described above, there has not been as much research applying these kinds of models specifically within the realm of business applications, though there is substantial research into the problem of pedestrian agent-based modeling in general (Batty 2003). However, recent advancements in the collection of data about consumers in business locations, using RFID-enabled shopping carts among other technologies (Larson et al. 2005), means that it may be possible to build micro-level spatial ABMs at a much greater level of fidelity than was previously possible. Pedestrian-level ABMs could be useful in understanding neighborhood foot traffic in order to develop retail locations (Borgers and Timmermans 1986), movement within a grocery store in order to optimally allocate store layout (Larson et al. 2005), or even model spatial interactions within an office building to improve organizational efficiency and interaction (Wineman 1982).

Despite the advances in new data collection techniques, there are still open research questions with regards to how to best integrate this data. Moreover, most ABMs are not built for physical interactions something that may be required to create sophisticated micro-level models. These are issues that deserve attention and research if micro-level spatial ABMs are to prove useful within the context of management science and business applications.

22.3.1 Neighborhood Pedestrian Modeling

Examinations in to agent-based pedestrian models have evolved over the years from random walk models to goal-oriented walking to flocking and object avoidance models (Batty 2003) and have been used to examine crowd movements at the Noting Hill Carnival (Batty et al. 2003) (see Fig. 22.3) and in the Tate Modern (Turner and Penn 2002). Models of pedestrian movement within a neighborhood could be useful within a number of different business and management science applications. For instance, it would be interesting to examine micro-level models of individuals moving around a neighborhood, and using this model of foot traffic to

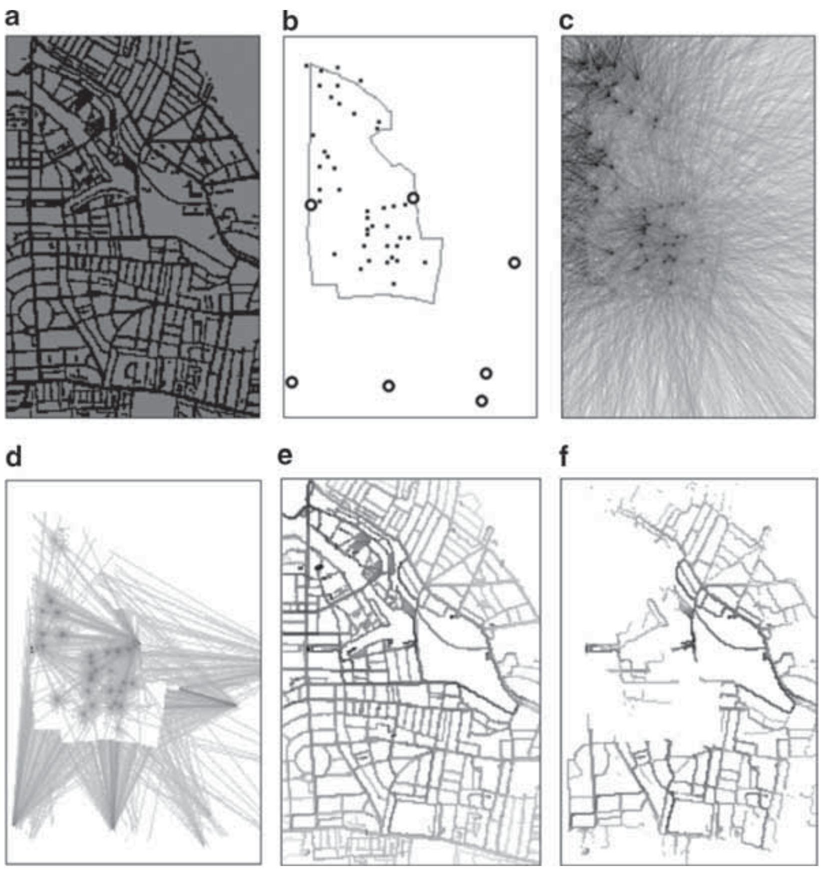


Fig. 22.3 Illustrations of data and analysis used in an ABM of pedestrian traffic around the Notting Hill Carnival: (**a**) street geometry, (**b**) parade route, sound systems, and tube stations, (**c**) all possible paths without street constraints, (**d**) shortest routes from tube stations without streets, (**e**) all possible paths with street constraints, (**f**) shortest paths with street constraints (Reprinted by permission from Batty et al. (2003))

determine retail location placement (Borgers and Timmermans 1986). In much the same way that a heat map of car traffic can be useful for understanding where to place a new retail location that is car accessible, pedestrian heat maps could be used to help examine how to place retail locations in pedestrian friendly areas, such as city centers or malls.

To some extent these models would not be very different than the macro-level models described above, but consumers who are moving through a pedestrian accessible landscape do have access to different information than those who are driving by a big box store. Pedestrians can partake in window shopping, and retailers can alter their displays to entice consumers, and even to compete with other stores or services nearby. Retail employees can interact with pedestrians, providing samples or coupons to entice them in to their store. This could be done adaptively in response to other spatially proximate competitors. This interaction between pedestrians, store employees and competitors creates a whole new level of marketing interaction that could be modeled and examined using spatial ABMs.

Of course, an issue that needs to be considered within the context of pedestrian modeling, especially if complexities such as window displays and sampling are to be included, is the lack of well-validated models at this level of detail. Though models of how pedestrians move have become more and more sophisticated, there currently is little or no research in to how these more goal-oriented pedestrian models would be affected by distractions and marketing interventions. This presents a number of interesting open research questions, and as data from closed-circuit TVs (Batty 2003) becomes more and more prevalent, it may be possible to build more sophisticated models of this complex consumer behavior. In the end, it would be useful to have a general model of pedestrian movement that could then be used to investigate a wide-range of questions from retail location placement to the effect of foot traffic on institutional environment design to questions of public safety (Kerridge et al. 2001), and steps have been taken in the past to create such models (Schelhorn et al. 1999), but a fully realized model that incorporates all of the concerns mentioned above has yet to be developed.

22.3.2 Retail Consumer Modeling

In general, the design of consumer retail environments (Babin and Darden 1995) and servicescapes, i.e., the built environment within which services operate (Bitner 1992), is critical to profitability and customer satisfaction. Despite the high impact of these decisions on business success, most guidelines about how to construct environments are based upon trial and error and past experience. There are very few ways to test out a design ahead of time and to determine if the physical layout of a retail environment will achieve the goals that management wants to achieve. Spatial ABMs provide a unique way to examine these environments before they are ever built and to try out multiple different layouts and investigate how well they perform. Of course such models require a sophisticated model of consumer movement within the space, and since each space is slightly different and that consumer interactions will vary from space to space, e.g., consumers act differently and have different goals in a bar than they do in a supermarket, this task can be quite difficult.

Though there has been some past work at examining in-store movement by consumers (Zacharias 2000, 2001; Zacharias et al. 2005), recent advances in the availability of RFID level data (Larson et al. 2005) (see Fig. 22.4 for an illustration of the kind of data obtainable from RFID tags), closed circuit television feeds (Batty 2003) and other technologies, have made it easier to construct and validate such models. One question that remains is at what level of detail to model the pedestrian movement. It is possible that every pedestrian entering a store could navigate in very different ways, but there do appear to be regular patterns of behavior, and research suggests that it may be possible to model consumer behavior within a shopping environment using a few simple rules (Zacharias et al. 2005).

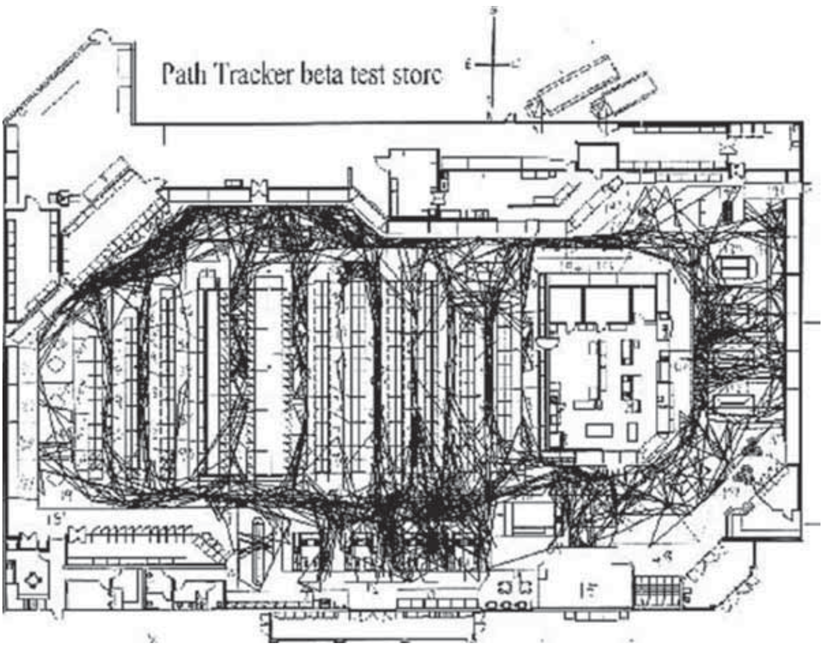


Fig. 22.4 RFID (PathTracker{©}) data from 20 random customers (Reprinted from Jeffrey et al. (2005), Copyright (2005), with permission from Elsevier)

One way to validate these models, would be to use the "virtual" models in combination with "real" data. For instance, Red Dot Square[7] is a company that creates photorealistic interactive retail "virtual" worlds, brings "real" consumers in to them and then examines what they do in order to provide advice to retailers about "real" store layouts. However, these real consumer experiments are very expensive. If there was a sophisticated enough model of consumer-level behavior either based on Red Dot Square's data or other data sources, then an ABM could be used to evaluate thousands of different store layouts using virtual consumers. After a smaller selection of store layouts were chosen from this larger set then they could be evaluated using either Red Dot Square's virtual environment and real consumers, or changes could actually be made in-store for an evaluative period.

Though it has been known for a long time that stores can increase sales through in-store marketing efforts (Chevalier 1975) and that the provision of different kinds of pricing comparisons can affect consumer information processing (Zeithaml 1982), these results could be built upon to create more sophisticated in-store retail ABMs. Moreover, applications of spatial ABMs to retail locations could even get down to the level of evaluating the effectiveness of displays with respect to the position and quantity of shelf-facings, i.e., how goods are actually laid out on a shelf. Though even simple models along these lines may garner new insights in to in-store marketing,

[7] http://www.reddotsquare.com

more advanced models could also be useful. For instant, recent research indicates that placing a product in an area of higher attention is sometimes not enough to drive additional purchases and that such models should include information about out-of-store marketing efforts and brand awareness (Chandon et al. 2009).

22.3.3 Other Micro-Level Applications

There are many other interesting questions within the management science context that could be examined using micro-level spatial ABMs. For instance, in the same way that diffusion of information is interesting at the macro-level, it is also interesting at the micro-level. Specifically, how does information diffuse within an organizational workspace? Clearly social networks and the chain of command play a role in organizational information diffusion (Cross et al. 2002), but physical proximity also plays a role. After all, the water cooler was the center of gossip for years not because individuals were socially tied to each other, but rather because the water cooler served as a central point where individuals interacted and exchanged information. To fully understand how information diffuses through an organization, it is necessary to take spatial process of movement with in the organizational space in to account. Models of this process could also be used to explore the effect of virtualization of organizations (Barrett et al. 2007).

Similarly, there has been a recent trend toward the development of "third" places, i.e., spaces that are neither home nor work, such as coffee bars and innovation cooperatives (Rosenbaum 2006). Clearly these new interaction environments will dramatically effect the diffusion of information both within an organization and across organizations as employees working in these third spaces interact with employees and entrepreneurs working in other organizations and industries. Potential research areas to consider include whether these spaces are useful for organizations, and how to design them to foster interactions at an appropriate level to reach some goal.

22.4 Discussion and Conclusion

The goal of this chapter has been to provide some examples of how agent-based models and spatial modeling can be combined to provide interesting insights in to business and management science applications. Throughout this chapter we have discussed not only applications of spatial ABM to business and management, but important issues that must be considered during this process. At the macro-level these issues usually revolve around what level of detail to create the model at, and whether it is better to create a suite of complementary models. At the micro-level there are fundamentally interesting issues that concern data integration, and to some extent how much fidelity to include in a model of pedestrian movement. As spatial ABMs continue to evolve, new issues and questions will arise, and new solutions and best practices to these already extant questions will be developed.

So far these models have been described as taking place at either the micro- or macro-level, and this distinction was drawn because the issues that must be examined at each level are often very different from each other. This distinction seems relatively valid since it is often possible to separate out micro- from macro-effects. For instance, it might be hypothesized that individuals who enter a grocery store may shop differently depending on where they had come from Herrmann and Beik (1968), but its not clear why a micro-level model would not be sufficient since the agents entering the store could be given different synthetic or empirically-derived histories. Alternatively if a store's contents varied based upon where individuals came from its not clear why this decision would also have to model the layout of the store at the micro-level. This is not to say that there will never be a need to blur this demarcation between scale boundaries but in many cases it may be more useful to build two complementary models instead.

In order to fully realize many of the goals and projects envisioned in this chapter, one aspect that need to be addressed is suitable tool development. As mentioned above, the integration of GIS tools into most of the modern ABM platforms has been a substantial assistance in terms of developing more integrated models, but unfortunately this integration is still far from seamless and requires effort on the part of the model developer. One possible approach that has been discussed elsewhere is a middleware approach (Brown et al. 2005b) that would allow a tighter integration between GIS and ABM and allow for a potentially platform independent model to be developed. Such an approach would also have the useful benefit of creating a somewhat more abstract language for talking about pattern (GIS) and process (ABM) in the same conceptual framework, a task which is currently hindered by the lack of a common language across these two methodologies. Since many phenomena of interest to modern business researchers and management science scholars contain both spatial and temporal elements, such a conceptual language would be very useful.

Of course such tools would be even more useful if they are teachable and explainable to management students. ABM has been used for management education in the past by combining it with participatory modeling (a version of ABM where some agents are played by actual people). The classic example is the "beer game" in supply chain management education (Sterman 1992), where managers control part of a supply chain and learn about how time lags affect supply chain operations. Though the beer game is relatively aspatial, spatial ABMs have also been combined with participatory simulation in other contexts to help determine local land-use policies (Castella et al. 2005; D'Aquino et al. 2003), and it seems clear that such techniques could also be employed in management science.

One of the main benefits of ABM in general is that the model representation is easily explainable to a non-technical stakeholder. As a result, spatial ABMs have the potential to be very convincing in business applications because of the relatively close relationship between their ontology and the ontology of the real business world. It seems like a natural combination then to bring together spatial ABMs with participatory modeling to develop sophisticated models of business applications and situations. The simulation could even use rules for the non-human agents that were inferred from previous observations of human behavior, allowing management students to compete again "human-trained" computational agents. And why stop there? The computational agents could even observe human participant behavior in real-time and adapt their strategies in response. Clearly, there are many possibilities and research opportunities in the application of spatial ABMs to business and management science education and research.

References

Ahn, H., & Lee, H. (2004). An agent-based dynamic information network for supply chain management. *BT Technology Journal*, *22*(2), 18–27.

Anthes, G. (2003, January 27). Agents of change. *Computer World*, 26–27.

Babin, B., & Darden, W. (1995). Consumer self-regulation in a retail environment. *Journal of Retailing*, *71*(1), 47–70.

Barrett, M., Davidson, E., Silva, L., & Walsham, G. (2007). Virtualization and institutions. *Virtuality and Virtualization*, 369–372.

Batty, M. (2003). Agent-based pedestrian modelling. In *Advanced spatial analysis: The CASA book of GIS* (p. 81). Redlands: ESRI Press.

Batty, M. (2005). *Cities and complexity*. Cambridge, MA: MIT Press.

Batty, M., Jiang, B., & Thurstain-Goodwin, M. (1998). Local movement: Agent-based models of pedestrian flows. *Working paper*. Centre for Advanced Spatial Analysis (UCL), London.

Batty, M., Desyllas, J., & Duxbury, E. (2003). Safety in numbers? Modelling crowds and designing control for the Notting Hill Carnival. *Urban Studies*, *40*(8), 1573.

Benenson, I., & Torrens, P. (2004). *Geosimulation: Automata-based modeling of urban phenomena*. Hoboken: Wiley.

Bitgood, S., & Dukes, S. (2006). Not another step! Economy of movement and pedestrian choice point behavior in shopping malls. *Environment and Behavior*, *38*(3), 394.

Bitner, M. (1992). Servicescapes: The impact of physical surroundings on customers and employees. *The Journal of Marketing*, *56*(2), 57–71.

Borgers, A., & Timmermans, H. (1986). City centre entry points, store location patterns and pedestrian route choice behaviour: A microlevel simulation model. *Socio-Economic Planning Sciences*, *20*(1), 25–31.

Brown, D., Page, S., Riolo, R., & Rand, W. (2004). Agent-based and analytical modeling to evaluate the effectiveness of greenbelts. *Environmental Modelling & Software*, *19*(12), 1097–1109.

Brown, D., Page, S., Riolo, R., Zellner, M., & Rand, W. (2005a). Path dependence and the validation of agent-based spatial models of land use. *International Journal of Geographical Information Science*, *19*(2), 153–174.

Brown, D. G., Riolo, R., Robinson, D. T., North, M., & Rand, W. (2005b). Spatial process and data models: Toward integration of agent-based models and GIS. *Journal of Geographical Systems*, *7*(1), 25–47.

Brown, D., Robinson, D., An, L., Nassauer, J., Zellner, M., Rand, W., Riolo, R., Page, S., Low, B., & Wang, Z. (2008). Exurbia from the bottom-up: Confronting empirical challenges to characterizing a complex system. *Geoforum*, *39*(2), 805–818.

Castella, J., Trung, T., & Boissau, S. (2005). Participatory simulation of land-use changes in the northern mountains of Vietnam: The combined use of an agent-based model, a role-playing game, and a geographic information system. *Ecology and Society*, *10*(1), 27.

Chandon, P., Hutchinson, J., Bradlow, E., & Young, S. (2009). Does in-store marketing work? Effects of the number and position of shelf facings on brand attention and evaluation at the point of purchase. *Journal of Marketing*, *73*(6), 1–17.

Chevalier, M. (1975). Increase in sales due to in-store display. *Journal of Marketing Research*, *12*(4), 426–431.

Cohen, M. D., March, J. G., & Olsen, J. P. (1972). A Garbage Can model of organizational choice. *Administrative Science Quarterly*, *17*(1), 1–25.

Cross, R., Borgatti, S., & Parker, A. (2002). Making invisible work visible: Using social network analysis to support strategic collaboration. *California Management Review*, *44*(2), 25–46.

D'Aquino, P., Le Page, C., Bousquet, F., & Bah, A. (2003). Using self-designed role-playing games and a multi-agent system to empower a local decision-making process for land use management: The SelfCormas experiment in Senegal. *Journal of Artificial Societies and Social Simulation*, *6*(3).

d'Aspremont, C., Gabszewicz, J., & Thisse, J. (1979). On Hotelling's "stability in competition". *Econometrica: Journal of the Econometric Society*, *47*(5), 1145–1150.

Davis, J. P., Eisenhardt, K. M., & Bingham, C. B. (2009). Optimal structure, market dynamism, and the strategy of simple rules. *Administrative Science Quarterly*, *54*(3), 413–452.

Domingos, P. (2005). Mining social networks for viral marketing. *IEEE Intelligent Systems*, *20*(1), 80–82.

Ehrentreich, N. (2007). *Agent-based modeling: The Santa Fe institute artificial stock market model revisited (Lecture notes in economics and mathematical systems)* (1st ed.). Berlin/New York: Springer.

Emerson, D., & Piramuthu, S. (2004). Agent-based framework for dynamic supply chain configuration. In *Proceedigns of the 37th annual Hawaii international conference on systems sciences*, Big Island. IEEE Computer Society.

Goldenberg, J., & Levy, M. (2009). Distance is not dead: Social interaction and geographical distance in the internet era. Arxiv preprint arXiv:0906.3202.

Hartenstein, H., Bochow, B., Ebner, A., Lott, M., Radimirsch, M., & Vollmer, D. (2001). Position-aware ad hoc wireless networks for inter-vehicle communications: The Fleetnet project. In *Proceedings of the second ACM international symposium on mobile ad hoc networking & computing* (p. 262). New York: ACM.

Heppenstall, A., Evans, A., & Birkin, M. (2005). A hybrid multi-agent/spatial interaction model system for petrol price setting. *Transactions in GIS*, *9*(1), 35–51.

Heppenstall, A., Evans, A., & Birkin, M. (2006). Using hybrid agent-based systems to model spatially-influenced retail markets. *Journal of Artificial Societies and Social Simulation*, *9*(3), 2.

Herrmann, R., & Beik, L. (1968). Shoppers' movements outside their local retail area. *The Journal of Marketing*, *32*(4), 45–51.

Hotelling, H. (1929). Stability in competition. *The Economic Journal*, *39*(153), 41–57.

Huang, A., & Levinson, D. (2008). An agent-based retail location model on a supply chain network. *Working papers*. University of Minnesota.

Kerridge, J., Hine, J., & Wigan, M. (2001). Agent-based modelling of pedestrian movements: The questions that need to be asked and answered. *Environment and Planning B*, *28*(3), 327–342.

Kohout, R., & Erol, K. (1999). In-time agent-based vehicle routing with a stochastic improvement heuristic. In *Proceedings of the national conference on artificial intelligence* (pp. 864–869). Wiley.

Krugman, P. (1996). Urban concentration: The role of increasing returns and transport costs. *International Regional Science Review*, *19*(1–2), 5.

Larson, J., Bradlow, E., & Fader, P. (2005). An exploratory look at supermarket shopping paths. *International Journal of Research in Marketing*, *22*(4), 395–414.

Lombardo, S., Petri, M., & Zotta, D. (2004). Intelligent Gis and retail location dynamics: A multi agent system integrated with ArcGis. *Computational Science and Its Applications–ICCSA 2004*, *3044*, 1046–1056.

Longley, P., & Clarke, G. (1995). *GIS for business and service planning*. Cambridge, MA/New York: Wiley.

MacKay, D. (1972). A microanalytic approach to store location analysis. *Journal of Marketing Research*, *9*(2), 134–140.

Minar, N., Burkhart, R., Langton, C., & Askenazi, M. (1996). The Swarm simulation system: A toolkit for building multi-agent simulations. *Working papers*.

North, M., Howe, T., Collier, N., & Vos, J. (2005). The repast simphony runtime system. In *Proceedings of the Agent 2005 conference on generative social processes models and mechanisms* (pp. 151–158).

North, M. J., & Macal, C. M. (2007). *Managing business complexity: Discovering strategic solutions with agent-based modeling and simulation*. Oxford: Oxford University Press.

North, M. J., Macal, C. M., Aubin, J. S., Thimmapuram, P., Bragen, M., Hahn, J., Karr, J., Brigham, N., Lacy, M. E., & Hampton, D. (2010). Multiscale agent-based consumer market modeling. *Complexity*, *15*(5), 37–47.

Rand, W., Brown, D., Page, S., Riolo, R., Fernandez, L., Zellner, M., et al. (2003). Statistical validation of spatial patterns in agent-based models. In *Proceedings of agent based simulation* (Vol. 4). Montpellier.

Rand, W., & Rust, R. (2011). Agent-based modeling in marketing: Guidelines for Rigor. Technical report, University of Maryland.

Rand, W., & Wilensky, U. (2007). Full-spectrum modeling: From simplicity to elaboration and realism in urban pattern formation. In *North American Association Computational Social and Organization Sciences conference (NAACSOS) 2007*.

Rosenbaum, M. (2006). Exploring the social supportive role of third places in consumers' lives. *Journal of Service Research*, *9*(1), 59.

Schelhorn, T., O'Sullivan, D., Haklay, M., & Thurstain-Goodwin, M. (1999). STREETS: An agent-based pedestrian model. *Working paper*. University College, London.

Schenk, T., Löffler, G., & Rauh, J. (2007). Agent-based simulation of consumer behavior in grocery shopping on a regional level. *Journal of Business Research*, *60*(8), 894–903.

Sinha-Ray, P., Carter, J., Field, T., Marshall, J., Polak, J., Schumacher, K., Song, D., Woods, J., & Zhang, J. (2003). Container world: Global agent-based modelling of the container transport business. In *Fourth international workshop on agent-based simulation* (pp. 161–165). Monpellier.

Sterman, J. (1992). Flight simulators for management education. *OR/MS Today*, *19*(5), 40–44.

Stonedahl, F., Rand, W., & Wilensky, U. (2010). Evolving viral marketing strategies. In *Proceedings of the 12th annual conference on genetic and evolutionary computation* (pp. 1195–1202). New York: ACM.

Trusov, M., Bucklin, R., & Pauwels, K. (2009). Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site. *Journal of Marketing*, *73*(5), 90–102.

Turner, A., & Penn, A. (2002). Encoding natural movement as an agent-based system: An investigation into human pedestrian behaviour in the built environment. *Environment and Planning B*, *29*(4), 473–490.

Wilensky, U. (1999). *Netlogo*.

Wineman, J. (1982). Office design and evaluation. *Environment and Behavior*, *14*(3), 271.

Zacharias, J. (2000). Shopping behavior at Alexis-Nihon plaza in Montreal. *Journal of Shopping Center Research*, *7*(2), 67–79.

Zacharias, J. (2001). Pedestrian behavior pedestrian behavior and perception in urban walking environments. *Journal of Planning Literature*, *16*(1), 3.

Zacharias, J., Bernhardt, T., & De Montigny, L. (2005). Computer-simulated pedestrian behavior in shopping environment. *Journal of Urban Planning and Development*, *131*, 195.

Zeithaml, V. (1982). Consumer response to in-store price information environments. *Journal of Consumer Research*, *8*(4), 357–369.

Zellner, M., Page, S., Rand, W., Brown, D., Robinson, D., Nassauer, J., & Low, B. (2009). The emergence of zoning policy games in exurban jurisdictions: Informing collective action theory. *Land Use Policy*, *26*(2), 356–367.

Zellner, M., Riolo, R., Rand, W., Brown, D., Page, S., & Fernandez, L. (2010). The problem with zoning: Nonlinear effects of interactions between location preferences and externalities on land use and utility. *Environment and Planning B: Planning and Design*, *37*(3), 408–428.

Chapter 23
Using Agent-Based Models for Education Planning: Is the UK Education System Agent Based?

Kirk Harland and Alison J. Heppenstall

Abstract Agent-Based Modelling (ABM) is a relatively new spatial modelling technique. The ability of ABM to simulate a real world system, the UK education market, is explored in this chapter. It is shown how a simple ABM incorporating common sense rules can provide acceptable results with over 60% of pupils being allocated to the correct schools and 75% of schools containing at least 50% of correct pupils when compared to observed data. The exploration outlined here highlights that the education has a good deal to offer researchers in the ABM field. Possibly more importantly, the real potential of ABM as a technique for simulating real world systems and delivering appreciable benefits to the general population is demonstrated.

23.1 Introduction

The world is a complex place; the systems that make up the environment in which we live are both diverse and interactive. In a social context, the complexities of interactions between different peoples' lives have been observed and explored by filmmakers for many years, a good example being 'Love Actually'. In this film all of the ongoing sub-stories are interwoven by a network of friends, family and work colleagues, many of whom are unaware of many of the other characters in the film; yet the decisions they make have wide ranging impacts. A global story emerges from the micro, 'individual', level interactions.

However, in the real world micro level interactions produce emergent level macro behaviours or events. A real world example of this would be the protests observed in Tunisia, Egypt and Jordan in February–March 2011. These protests

K. Harland • A.J. Heppenstall (✉)
School of Geography, University of Leeds, Leeds, UK
e-mail: k.harland@leeds.ac.uk; a.j.heppenstall@leeds.ac.uk

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_23,

start with interactions at a micro level and result in events that make news headlines throughout the world. Agent-Based Modelling (ABM) is a relatively new modelling technique used to mimic micro scale interactions to observe what the macro level outcomes are (see Crooks and Heppenstall 2012 for an overview). To date many ABM applications have been theoretical, most notably SugarScape (Epstein and Axtell 1996) and discussions by authors such as O'Sullivan and Haklay (2000) have asked questions as whether the 'world really is agent based'? Despite the questions raised as to whether ABM can actually represent the real world, some ABMs are now starting to have real world applications, such as the crowd control applications for the Notting Hill Carnival (Batty et al. 2003), modelling infectious diseases (Epstein 2009), petrol price modelling (Heppenstall et al. 2006) or more recently crime simulation (Malleson et al. 2010). The world is made up of many different discrete objects that interact given a particular set of rules or laws, governments interact, companies interact, clubs and social groups interact, individual people interact, chemicals interact even down to the particle physics world of protons, electrons and quarks etc. all interact. Therefore, the question shouldn't be whether the world is Agent-Based, but whether we have the computational power and ability to model it. All things considered, it comes down to finding a suitable scale at which to model interactions where the processing time and complexity can be balanced against gathering fruitful and useful results.

This chapter will explore the application of ABM to the simulation of events in a real world system, the education system in England (UK). Section 23.2 will present contextual background with Sect. 23.3 presenting the case for the importance of education planning. The model is presented in Sect. 23.5 with results discussed in Sect. 23.6 and concluding thoughts given in Sect. 23.7.

23.2 A Brief History of the English Education System

Over the past one hundred years education policy and provision has evolved significantly. There have been periods of unprecedented development, but also periods of unrivalled contradiction and controversy. As the education sector developed in the early part of the twentieth century, momentum gathered. This momentum carried into a period of substantial development leading to reform and the introduction of the 1944 Education Act. This was widely recognised as the Act that "laid the foundation for the modern education system" (Statham et al. 1991, p. 42). It abolished the Board of Education and replaced it with the Ministry of Education, with the Minister having a much more proactive role in education policy formation. Robert Butler became the first Minister of Education and was the main proponent of the Act spearheading it through Parliament in controversial circumstances. "[I]t was overseen by a Conservative MP (Butler), taking advice from Labour MPs ([James] Chuter Ede, [Ernest] Bevin and [Clement] Attlee amongst others), and with a civil service department in agreement that the time for reform was at hand. It was passed by a Coalition government in direct opposition to the Prime Minister [Winston Churchhill]" (Langley 1997, p. 38).

These sweeping reforms and the distinctive shift of power away from the schools and towards the Local Education Authorities (LEAs) have been described as "the single most important piece of legislation to be passed between 1939 and 1945." (Chitty 2004, p. 18). However, omissions from the Act would prove troublesome over the next 44 years. Firstly, the provision of religious instruction for primary and secondary schools was the only curriculum requirement of the 1944 Act, allowing schools to develop individual curricula. A second major omission that would prove a particular point of contention was the lack of any specific framework for the structure of new compulsory secondary schools. Would the system be a tripartite system as recommended in the Spen Report (Lawrence 1992) or a comprehensive system favoured by the Hadow Report (Armytage 1970)? This question would prove inhibiting to the smooth implementation and running of the secondary education system until the next round of major reforms in 1988. In the following years political polarisation exploited the gaps in the 1944 'Butler' Act and left the education sector with a legacy of school types, many still in existence today and each having different characteristics such as admission policies.

In 1988 the Education Reform Act (ERA) opened up the education market place and closed up the loop holes in the 'Butler' Act. This piece of legislation forms the cornerstone of the education system in operation today. It laid the groundwork for information collection, competition between schools based on performance and school inspections. Subsequent, legislation has built on provisions in the 1988 ERA. Today schools operate in a quasi-competitive market, with competition between schools for pupils who in turn have a choice of the school they wish to attend. The ubiquity of the educational product supplied by schools has been diversified after Key Stage 3, with the introduction of specialist schools that focus on particular vocational themes.

23.3 Why Is Education Planning Important?

Recent demographic trends have provided education planners with considerable challenges. For the first time since significant development of the education sector was undertaken, declining pupil numbers have meant that school rationalisation has been required. Surplus school places are recommended to be no more than 10% within an LEA and no more than 25% in any single institution (Audit Commission 2006) with current fiscal pressures underlining the need to keep surplus school places to a minimum while ensuring educational the requirements of the population are met. The challenge to ensure that school places are available at the institutions preferred by pupils and parents has fallen to the LEAs as the 'commissioners' of education. Over the preceding century, control of the education sector has been shifted from schools to the LEAs, and then recently, from the LEAs up to Government, with the setting of the National Curriculum, and back down to schools, with the advent of 'Trust schools'. LEAs sit in the middle layer with a great deal of responsibility for the planning of education provision, and ensuring that education is supplied fairly for all sections of society, but with much reduced control over their local area.

Accurate school roll forecasting has become increasingly important for education planning professionals because of the dynamic nature of population demographics. The level and distribution of demand for education constantly change overtime as the pupil population either increases or decreases over space. It is not just population change that is demanding more sophisticated projection techniques, life style changes have altered the way that people operate in the spaces in which they live. The school run has become part of a multi-purpose journey which includes other functions, such as a journey to or from work plus, perhaps, a shopping trip (Pooley et al. 2005). Therefore, family convenience influences school choice decisions. Additionally, concerns over child safety during their daily commute to school have become a significant concern to parents when selecting a school for their children to attend (Valentine and McKendrick 1997) over and above the conventional attractiveness factors, such as teaching quality, attainment levels and proximity to home. All of these factors create a more complicated environment in which education planners seek to keep the supply of education commensurate with demand.

Information in the education sector has become more abundant in recent years. The introduction of the Pupil Level Annual School Census (PLASC) dataset and its incorporation into a National Pupil Database (NPD) in 2001/2002 made detailed pupil level information available to education planners. The information contained in the NPD, and in particular PLASC, is not only detailed but also longitudinal, providing a resource of immense value to LAs and education planners. Unfortunately, projection methods used by most LAs have not reflected the changes in the education market or the increased availability of information in this sector. The projection method of choice for most education planners is still a basic cohort progression model where the underlying demand is based on either the previous years demand directly or on a weighted average of a number of previous years demand. In the 1970s the cohort progression model was adopted into the education planning process (Simpson and Lancaster 1987), at a time when the education sector was still experiencing considerable expansion. However, the cohort progression model, although easy to understand and apply, lacks the sophistication required to respond to rapidly changing pupil populations and in turn school network restructuring tasks required in the modern education market.

In recent years, there has been a growing interest in various aspects of the education sector by geographers. Social and ethnic segregation in the education sector has been the focus of a good deal of research including notable texts by Gibson and Asthana (2000b), Gorard (1999), Gorard and Fitz (1998a), Goldstein and Noden (2003), Harris et al. (2007) and Johnston et al. (2006). Other aspects of academic investigation into the education sector have examined the effects of school performance on local house prices (Cheshire and Sheppard 2004; Leech and Campos 2003; Croft 2004), links between school roll size and academic achievement by pupils (Bradley and Taylor 1998), competition and performance between schools and the resulting effect of parental choice after the 1988 ERA (Gereluk 2005; Pooley et al. 2005; Bradford 1990, 1991). However, there has been little academic research into pupil daily commuting patterns and the journey to school. Pupil commuting patterns do result from the school selection process. Equally, school selection is influenced by geographical factors, such as school proximity and, more importantly how connected a school and the pupil's home location are. Therefore, school selection behaviour and pupil commuting patterns are inter-connected.

The planning demands of a dynamic population, a competitive education system with open parental choice and Government policy changes, requires that education planning professionals develop more sophisticated and effective methods of assessing and supporting the planning decisions they make. A Spatial Education Model framework consisting of a number of layered spatial interaction models simulating pupil movements interacting with schools represented as agents provides a series of promising results and provides a potential resolution to the current lack of a sophisticated planning tool for use in the education sector (Harland and Stillwell 2010). However, frameworks such as these do not handle individual pupil characteristics well. For example, single sex schools are not easily serviced when the demand (pupils) side of the model is serviced by an aggregate model such as a spatial interaction model. This type of issue seems ideally suited to the application of an ABM and with the abundance of individual level information available within the education sector the model can be based on and measured against a real world social system.

23.4 Data in the Education Sector

The National Pupil Database (NPD) is a relatively new dataset created in 2002 and contains individual pupil records for all state educated school children (Ewens 2005). It is updated on an annual basis with additions in excess of eight million individual pupil records collected by each Local Authority (LA) in England and Wales and is maintained by the Department for Education (DfE formerly known as the Department for Children Schools and Families (DCSF)) (Jones and Elias 2006). Access to the NPD has recently been provided through a central gateway funded jointly by the DfE and the Economic and Social Research Council (ESRC) and managed by the Centre for Market and Public Organisation (CMPO) at the University of Bristol where the PLASC/NPD User Group (PLUG) is based (Burgess et al. 2006). The NPD is stored in a relational database structure with several different datasets capable of being linked together using either a Unique Pupil Number (UPN) or a unique establishment identification number to allow for both temporal and cross-sectional analysis, creating a powerful information resource for policy formulation (Jones and Elias 2006).

Completion of the Pupil Level Annual School Census (PLASC) is statutory for all state maintained primary, secondary and special schools under section 537A of the Education Act 1996 (Jones and Elias 2006). The DfE began collection of the data in 2002 and it now forms the cornerstone of the NPD. Individual schools are required to submit a PLASC return to the LA on the third Thursday of January each year. The return consists of entries for every pupil on roll with data such as home postcode, ethnicity, Special Education Need (SEN) status and Free School Meals (FSM) eligibility, plus information relating to the school and its staff (for more detail on the complete contents of the PLASC dataset and the structure of the NPD see Harland and Stillwell 2007). In actual fact, the data collection of pupil information is no longer referred to as PLASC because a tri-annual data collection procedure called the School Census with a modular structure was introduced in 2006 for secondary schools and in 2007 for primary schools (Department for Education and Skills 2006b). One of the three data collections is still carried out in January, with two further collections on the third Thursday in May and the third Thursday in September augmenting the January collection (Department for Education and Skills 2006a). The tri-annual data collections coincide with the three school terms and enables more effective tracking of pupil migrations, moves between homes and moves between schools, throughout the year.

Ewens (2005, p. 4) comments that "the National Pupil Dataset is amongst the most important national innovations in data collection in the recent past. Its potential is considerable and the scope for development is also considerable." These comments made by Ewens are true in more ways than one. The collection of pupil data is critical for the evaluation education policy and progression in raising the standards of education provision. Moreover, the collection of such datasets assists education planners in their efforts to align the supply of education with demand. A relatively self contained system, such as education, with a rich supply of complete real world data, where a great deal is known about the individual, is surely a candidate to construct an ABM and test how applicable this relatively new modelling paradigm is at simulating real world situations.

23.5 Model Construction

23.5.1 Model Structure

The ABM applied here is constructed using the Java object orientated programming language and built into the Flexible Modelling Framework developed at the University of Leeds to assist in the application of social science modelling studies. Figure 23.1 below shows the basic design of the ABM.

The top level class 'Agent' is an abstract class that contains common attributes that all agents within the model will require, such as location coordinates. The two classes

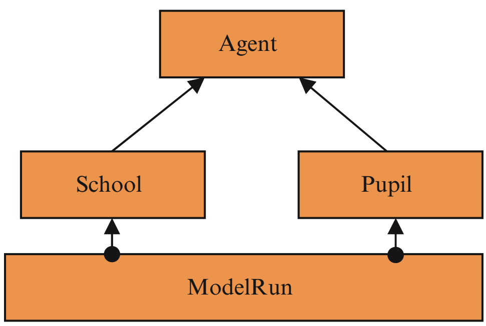


Fig. 23.1 Basic design diagram of the education ABM

below ‘Agent’, ‘School’ and ‘Pupil’, inherit from ‘Agent’ and so contain the common attributes, but they also contain attributes specific to the real world subjects for which they are a software representation. ‘School’ contains attributes such as whether it is a single sex institution, the maximum number of admissions that can be accepted and performance information such as the average point score for pupils examined while attending that institution. ‘Pupil’ contains information relevant to pupils such as their gender, whether they are eligible to FSM, ethnicity and a derived Catholic religion attribute discussed in more detail below.

When the model runs, the ‘ModelRun’ object is invoked and a collection of pupil agents is created and a collection of school agents are created. Once created the pupil agents are iterated over and according to the implemented rules (outlined in more detail below) they select their preferred school, it is worth noting that all pupil agents have perfect knowledge of all schools agents. Once all pupil agents have expressed a preference of school agent the school agents are iterated over and they accept pupil agents that have expressed an interest in the school agent according to the implemented rules (outlined in more detail below). This process is repeated three times or until all pupil agents have be accepted by a school agent. If at the end of three iterations of the complete model some pupil agents have not been accepted by school agents they are allocated to the closest school agent which is not full and offers education suitable to the pupil agent.

The reason the model has been constructed in this particular way is to mimic as closely as possible the pupil / school application / admissions process observed in the study area. Certain types of school can choose to apply an alternative admissions policy to that of the LA, so long as they are within the mandatory requirements of the ‘School Admissions Code’ (Department for Education and Skills 2007). In the Leeds study area the main alternative admissions policy is implemented by Voluntary Aided schools and incorporates some aspect of prioritising by religious denomination. However, the majority of schools in the study area apply the overarching admissions policy for the area which is:

- “Priority 1 –
  - A. Children with a statement of Special Education Need
  - B. Any child deemed by Education Leeds to benefit significantly by admission to the preferred school.
- Priority 2 – Siblings
- Priority 3 –
  - A. All children are offered a place if there are enough places.
  - B. In cases of oversubscription places are offered to nearest children measured as a straight line distance with priority to -
    1. Preference of nearest Leeds school to home address.
    2. Preference school other than nearest Leeds school to the home address.
  - C. If parental preferences cannot be met by these criteria then a place is guaranteed at the nearest community school.” (Education Leeds 2004, p. 2–3).

Table 23.1 Order of rules applied to the model

Rule #	Rule
1	Random selection
2	Closest school
3	Admissions limits
4	Single sex schools
5	Catholic schools
6	Network distances
7	Affluence with school performance

23.5.2 Model Rules

As demonstrated above, the model structure is a good approximation of the real world school admissions process in the study area. Further rules are implemented at both the 'School' agent and 'Pupil' agent levels to define the model more realistically. Each of these rules is introduced progressively so that the impact of each rule can be analysed. These rules are shown in Table 23.1.

The first model is a baseline model which assigns pupil agents to school agents randomly. This model can be used to estimate how much of the agreement between the model results and observations in the data can be explained by random chance allocation. The following two rules are applied to the school agent to simulate individual school characteristics more closely. The first of these, rule 2, simulates the school admissions policy, priority 3, by assigning pupil agents to their closest school agent. Rule 3 introduces the concept of school agents having a finite capacity to accept education agents, as in real life schools can only admit a particular number of pupils which is dictated by complex calculations taking into consideration school characteristics as diverse as the amount of common space in the school, area of teaching space and number of teachers.

The following two rules are applied to both school agent and pupil agent. Rule 4 ensures that single sex school agents will only accept applications from pupil agents of the correct gender. This rule is also applied so that pupils of will make applications to single sex schools if they accept applications from pupils of their own gender. Rule 5 is similar in application, however, rather than a strict yes / no rule where a male pupil agent will not apply to an all female school agent, and would not be successful if they did, pupil agents who attended a catholic primary school will seek out a catholic school agent and the catholic school agent will favour applications from a catholic pupil agent.

How is the likelihood of a pupil agent seeking out a catholic school agent arrived at, and how is the probability of a school agent accepting the application of a non-catholic pupil agent over a catholic school agent calculated? And indeed why have this rule in the first place? The answer lies in the previous research and empirical analysis. Religion is, quite rightly, considered to be an important school choice driver by Pooley et al. (2005). Schools can be selective on religious grounds and, in 2005/2006, there were eight selective primary schools in Leeds having a 'SEL4' code in the PLASC dataset, one Jewish, three Church of England and four Catholic, and two selective secondary schools, one Church of England and one Catholic. However, there are many more schools that prioritise a particular religion, but are not shown to have a selective admissions policy in the PLASC dataset (for detailed school admissions policies for 2009/2010 see Education Leeds (2008)). It follows that parents and pupils of a particular religious denomination will be more inclined to select a school that prioritises their religion and less inclined to select a school that is orientated to an alternative religion. The problem is that the religion of school pupils cannot be identified. PLASC returns contain information on the ethnic origin of each pupil but there is no information on the religious denomination of pupils.

However, during the transfer between primary and secondary school, it is possible to calculate the proportions of children progressing between different selective schools, or schools identified to be of a specific religious denomination. Of the 8,141 pupils moving between primary and secondary schools in Leeds in 2005/2006, 83% of those moving from primary schools identified as Catholic went to Catholic secondary schools, and 90% of the intake of all secondary schools identified as Catholic in this year originated from Catholic primary schools. These statistics highlight the importance of religion, especially Catholicism, in school selection by parents and pupils, and present an argument for between school moves of this type to have a rule associated when modelling is undertaken. However, of the 8,141 pupils moving between primary and secondary schools only 7.5% were Catholic. Although, other religious denominations are prevalent in the Leeds study area, Catholic pupils are the most identifiable, and also display the most selective behaviour. For example, the one Leeds Church of England secondary school shown to be selective in the PLASC dataset in 2005/2006 could only have 7.5% of its intake identified as originating from Church of England primary schools. In contrast, the one Catholic secondary school shown to be selective in the PLASC dataset had an intake consisting of 86% of pupils originating from Catholic primary schools. Therefore, despite religion clearly being an important factor in school choice, the extent to which it can be used in modelling the interactions between pupils and schools is limited to prominent religious denominations that can be identified, such as Catholic pupils in the Leeds study area. Rule 5 reflects the empirical analysis and Catholic pupil agents will actively prefer a school if it is Catholic 80% of the time and any school 20% of the time, with the school accepting Catholic pupil agent applications 85% of the time.

Rule 6 considers the use of network distances rather than Euclidean distances when pupil agents select schools agents. This again is further added realism to the model. The influence of school accessibility and the presence of physical barriers on school selection are considered by both Pooley et al. (2005) and Parsons et al. (2000), but quantifying the presence of a physical barrier is not easily achieved. However, the significant effect of physical barriers on primary school territories has been identified by Harland (2008), and although difficult to quantify, they must be considered. One method of doing this is to calculate the distance to school for pupils using the road network rather than Euclidean distance. Although the use of network distance calculations does provide a method for introducing connectivity and incorporating physical barriers to some extent, it also introduces further issues. Pupils walking to school do not necessarily follow road networks, making use of short cuts between roads and crossing greenfield sites such as parks. Therefore, the use of road networks in analysis and modelling could be beneficial for some social groups, but for others it could prove detrimental.

The final rule, rule 7, combines the assessment of school performance by pupil agents with a proxy for affluence. As shown in Harland (2008), the use of school performance in school selection by pupils and their families is related to social status and the education level of the parents. In order to reflect this observation in the model a proxy for social status or affluence is used. When this proxy is of a type that would indicate a pupil agent that would consider school performance, a density function, incorporating both the distance to the school agent and the performance of the school agent is applied to find the pupil agents preferred school agent. In these circumstances this rule overrides rule 2, where the pupil agent type is not one that would find school performance important rule 2 persists.

23.6 Results

In order to compare the affect of the different rules on the model outcomes, each rule has been applied and the model executed. The model results are then compared to the observed data, and a percentage of pupils that end up attending the correct school is calculated, a simple but accurate effectiveness measure. Table 23.2 below shows the results for each model run. It is important to note that, with the exception of rule 1, each rule builds on the next, that is to say that each model builds on the previous one. To exemplify this, if the model run with rule 5 is considered, this model run incorporates rule 2, 3, 4 and 5. The baseline model, rule 1, simply assigns pupil agents to randomly selected school agents. This model is stochastic, the results will vary with each run, therefore the model is run 1,000 times and the percentage result show the average result from all run results. The model runs for rules 2, 3 and 4 are deterministic, there is no stochastic element and as such the results will be exactly the same with each run so long as the input data remains constants. Rule 5 introduces stochastic elements to the model and as such the results are the average over 1,000 model runs.

Table 23.2 Progressive model results

Rule #	Rule	% pupils correct
1	Random selection	2.75
2	Closest school	50.59
3	Admissions limits	49.98
4	Single sex schools	50.19
5	Catholic schools	55.43
6	Network distances	54.26
7	Affluence with school performance	60.06

Table 23.3 Spread of correct results (%)

	2	3	4	5	6	7
Low	3.85	10.14	0.00	0.00	0.00	1.93
High	84.06	82.55	83.49	83.18	82.87	88.79
Difference	80.21	72.41	83.49	83.18	82.87	86.86

It can be seen that the results of assigning the pupil agents to their closest school agent using a Euclidean distance, rule 2, is a vast improvement over the baseline random allocation model with over 50% of pupil agents being admitted to the correct school agent. When admission limits are applied to the school agents in rule 3 the percentage of correct admissions drops by a little over half a percent. The introduction of single gender institutions in rule 4 improves the model fit slightly but a larger improvement is gained from the introductions of Catholic school and pupil agents in rule 5. Applying network distances in Rule 6 decreases the overall fit of the model, which is consistent with research performed by Harland (2008) demonstrating that distance was a more ubiquitous consideration in primary education with impacts limited to less affluent families in secondary education. The introduction of rule 7 significantly improves the model fit to over 60%.

Examining each model run results from this high level vantage point shows a steady and gradual improvement as rules are introduced. But is this improvement homogeneous throughout the model? Table 23.3 shows the lowest, highest and difference between the percentages of pupil agents admitted to the correct school agents. It is clear that the relatively simple model, rule 3, which allocates pupils agents to the closest school and applies a school admission limit presents as the most consistent model. It has the highest low value, however it also has the lowest high value. This would suggest consistency within the model. Considered in context with the overall percentage correct value, which is the lowest of all model combinations with the exception of the baseline random selection model, this suggests that the model is relatively consistent but also relatively consistently incorrect.

To examine the internal distribution of pupil agents admitted to the correct school agents the percentage of correctly assigned pupil agents in each school agent is banded into five groups 0–20%, 20–40%, 40–60%, 60–80% and 80–100%. The counts of the number of school agents falling within each band are displayed in Fig. 23.2 below. The results summarised in Fig. 23.2 show that a high number of school agents have quite low percentages of pupils agents correctly assigned in model '2'. This situation improves in models '3' and '4' and then again in models '5' and '6'. However, it is model '7' that shows a distinct shift to the higher percentage bands demonstrating that this model contains the greatest proportion of school agents with correctly assigned pupil agents. In fact, model '7' has an average of 77.5% of school agents admitting at least 50% of the correct pupil agents and an average of 25% of school agents admitting at least 80% of the correct pupil agents.

Figure 23.3 shows the spatial distribution of average percentage of correctly allocated pupil agents to school agents for model configuration '7'. Most of the school agents with low percentages of correctly assigned pupil agents are contained

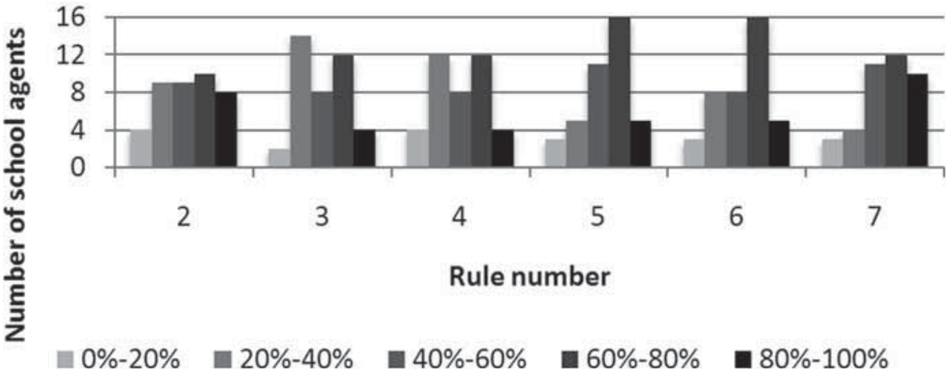


Fig. 23.2 Percentage of pupil agents admitted to correct school agents by model rule

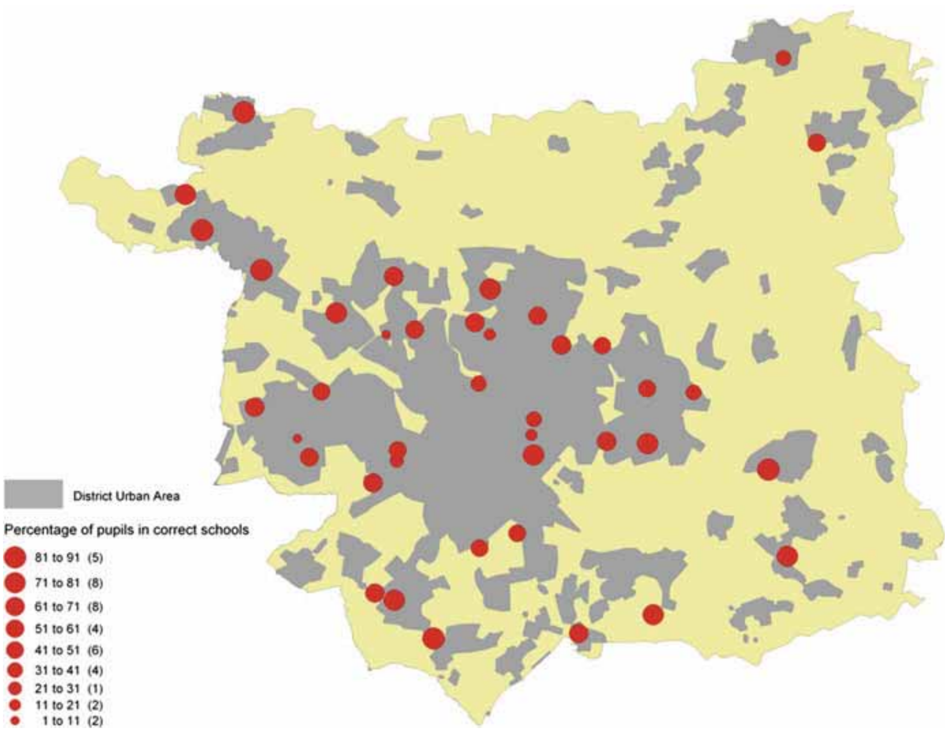


Fig. 23.3 Percentage of pupil agents admitted to correct school agents for model configuration '7'

in the inner urban area of the city, a traditionally less affluent area with a relatively high density of secondary schools. Within this area of the city there is a wide range of school performance with a little over 3.5 km distance between the lowest and highest performing schools in the city. However, we know from education research that less affluent pupil and their families are less likely to consider performance of a school when making school selection. Financial considerations are possibly more pressing with the option of pupils catching public transport to travel greater distances to attend a better performing school is potentially not financially practical. Therefore it follows that more less affluent pupils are likely to attend the closest school to their home location. The pattern observed in Fig. 23.3 with many urban schools not attracting the correct pupil agents is potentially a manifestation of the limitations associated with the proxy used for social status, Free School Meals (FSM) eligibility. Social stratification is more complex than can be assessed from a yes or no answer to whether a child is eligible for aid with school meals. Variables such as, but not limited to, parental education levels, occupation type, access to private transport and tenure all have an impact on the social experience and opportunities available to a child. There is also the possibility that changes in the benefit system or to the criteria for assessing FSM eligibility can change a pupil's eligibility status without the child's living circumstances altering (Burgess et al. 2006). Furthermore, if a family is eligible for, but does not claim, certain state benefits because of either pride or ignorance to their eligibility, this will influence a pupil's eligibility for FSM. It is therefore highly likely that pupil agents representing only the very poorest pupils are identified using this proxy for social status and the resulting pattern of under representation of correctly allocated pupil agents in the inner city area is a facet of this limitation.

Another limitation with the use of FSM as a proxy for social status is that there is no way to identify the opposite end of the social spectrum, the most affluent. A process of indirect selection which is commonly referred to in the education literature as 'selection by mortgage' (Leech and Campos 2003), where more affluent parents can afford to move closer to a perceived good school to increase the chances that their child will secure a place at their chosen school remains undetectable using only the FSM social status proxy. There is a consensus in previous research, both in the UK and internationally, that perceived good schools do influence house prices in the surrounding area. A study by Cheshire and Sheppard (2004), found a premium of up to 34% or £42,550 on houses in close proximity to perceived good schools in Reading. Given the lack of affordable housing in Leeds, the same is likely to be true and makes for a substantive indirect selection criterion, insurmountable to less affluent families. However, the decision making process that leads to a home move is complex and isolating one particular motivation for moving, such as moving closer to a desirable school, difficult. Although, the influence of perceived good schools on house prices cannot be ignored, unravelling the intricate motivations for residential movement is complicated, and is an area where a great deal more research is required in order to identify the effect more accurately. Such a selection criteria would manifest itself as more affluent pupils attending the closest school to their home location, simply because family relocation would ensure that the 'desirable' school for the pupil would be the closest to home.

This means that the group of pupils where school performance criteria would be a large selection factor, from a modelling perspective, would be the mid-range social groups. A further limitation of the modelling structure utilised here, and likely to be reflected in the results, is the ability of a pupil's parents to 'play the system'. Parents with higher education attainment are much more likely to have the confidence and experience to challenge school allocation procedures or read papers and literature where school performance information is published. Additionally, Parental education levels are suggested to be important influencing factors in the selection of a school and in the eventual performance of a child at school. Bradley and Taylor (2004) discovered strong correlations between education attainment of pupils and parental occupation variables with pupils having parents in professional occupations much more likely to achieve higher grades than those with parents in unskilled, semi-skilled or manual work. Dustmann (2004) draws conclusions from his study on the influence of parental background on the educational track of children in Germany which supports this assertion. However, Dustmann notes that the relatively young age of 10 at which the educational track is chosen in Germany differentiates this study from education markets like the UK and USA where the track choice is taken much later. In contrast to these studies, Feinstein and Symons (1999) conclude that the most important influencing factor on pupils' educational attainment is parental interest. However, they find high correlation levels between parental educational level and parental interest and between social class and parental interest, suggesting that parental interest is, at least in part, a culmination of these two variables. The influence of parental status is not considered in this model structure whatsoever.

23.7 Conclusions

In the introduction we considered, with a somewhat unorthodox example from the film industry, how micro level events manifest themselves into macro level stories / behaviour. This research has gone on to apply ABM technology to build a bottom-up model of the secondary education sector in Leeds. The application of this model is not conventional in terms of ABM literature. We have not been looking for emergence per se but rather creating a simulation model capable of assigning the correct pupil agents to the correct school agents based on rules derived from both the overarching admissions policy published by the study area of Leeds and from the education research literature. Accurate spatial models are required to assist education planners in their effort to align investment with demand, this is particularly important in the current challenging financial climate. It is necessary that any models produced are capable of being used to assess the impacts of alterations to school networks, neighbouring education authority provision or the pupil population size and complexion; to achieve this they first have to be able to simulate the current situation in a robust and scalable manner. This research has demonstrated that ABM technology can be used in this type of application. Furthermore, it has shown how the implementation of simple common sense rules observed in the real world can be used to construct an Agent Based Simulation Model. Moreover, it has become apparent that ABM technology excels at representing pupil level attributes such as gender, religion or ethnicity and can equally well represent different school attributes such as whether a school has a religious admissions policy etc. These are issues that are difficult to address in traditional aggregate spatial models.

However, this work has also shown that there are many complexities of the education system that are not well represented by this simplistic ABM. Mostly these complexities are centred around the decision making process for school selection and there are techniques that can be applied here that could bolster the models performance. The selection process that was employed here was a very simple probabilistic function, however, much more complex behavioural simulation techniques are available, two such behavioural simulation models are Physical conditions, Emotional States, Cognitive Capabilities and Social Status, known as PECS (Urban 2000) and Beliefs, Desires and Intentions more commonly referred to as BDI (Rao and Georgeff 1995; Müller 1998). Other areas where this simple model could be improved include the inclusion of more datasets to augment the rich PLASC data. School preference data is collected nationally by all education authorities and would prove an invaluable resource in developing a more accurate behavioural school selection model. The social status of the pupil agents could be derived from their location through the use of geodemographic databases and would surely be a much improved source of information over the single binary indicator of FSM eligibility.

The education sector is rich in data. However, it is not only rich in data but rich in individual level data. This is a sector that agent based modellers must move towards to refine their approach and transition ABM technology from a research tool into an applied modelling method with real world applications and quantifiable impacts with tangible benefits for the general population. A concerted effort to develop ABM technologies for the education sector can have no other effect than beneficial. Beneficial to agent based model researchers through pushing their methods further into the main stream; beneficial to education planners by providing them with better insight; beneficial to the pupil population because a better planned education system is a better understood education system which will provide better education at the point of need; beneficial to central government, with better planning comes increased financial efficiency. In all ABM has an important role to play, none more so than in the education sector.

References

Armytage, W. H. G. (1970). *Four hundred years of english education* (2nd ed.). Cambridge: Cambridge University Press.

Audit Commission. (2006). Higher standards, better schools for all: More choice for parents and pupils department for education and skills. *Consultation response*, London: Audit Commission.

Batty, M., Desyllas, J., & Duxbury, E. (2003). Safety in numbers? Modelling crowds and designing control for the Notting Hill carnival. *Urban Studies, 40*(8), 1573–1590.

Bradford, M. (1990). Education, attainment and the geography of choice. *Geography, 76*(1), 3–16.

Bradford, M. (1991). School performance indicators, the local residential environment and parental choice. *Environment and Planning A, 23*, 319–332.

Bradley, S., & Taylor, J. (1998). The effect of school size on exam performance in secondary schools. *Oxford Bulletin of Economics and Statistics, 60*(3), 291–324.

Bradley, S., & Taylor, J. (2004). Ethnicity, education attainment and the transition from school. *The Manchester School*, *72*(3), 317–346.

Burgess, S., Briggs, A., McConnell, B., & Slater, H. (2006). *School choice in England: Background facts* (Working Paper 06/159). Bristol: Center for Market and Public Organisation, University of Bristol.

Cheshire, P., & Sheppard, S. (2004). Capitalising the value of free schools: The impact of supply characteristics and uncertainty. *The Economic Journal, 114*(499), F397–F424.

Chitty, C. (2004). *Education policy in Britain*. Basingstoke: Palgrave Macmillan.

Croft, J. (2004). Positive choice, no choice or total rejection: The perennial problem of school catchments, housing and neighbourhoods. *Housing Studies, 19*(6), 927–945.

Crooks, A. T., & Heppenstall, A. J. (2012). Introduction to agent-based modelling. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 85–105). Dordrecht: Springer.

Department for Education and Skills. (2006a). *2006 School census: Maintained secondary schools technical specification*. Last Accessed March 8, 2006. URL: http://www.teachernet.gov.uk/doc/9096/School Census 2006 Specification School and Pupilv1.91.doc

Department for Education and Skills. (2006b). *National pupil database: The future*. Last Accessed September 23, 2008. URL: http://www.bris.ac.uk/Depts/CMPO/PLUG/userguide/catherine.ppt#258,2,National Pupil Database: The Future

Department for Education and Skills. (2007). *School admissions code*. Norwich: The Stationary Office.

Dustmann, C. (2004). Parental background, secondary school track choice, and wages. *Oxford Economic Papers, 56*, 209–230.

Education Leeds. (2004). *Review of admissions policies* (Internal Report). Leeds: Education Leeds.

Education Leeds. (2008). *Directory of school admissions policies in Leeds 2009 2010*. Last Accessed September 23, 2008. URL: http://www.educationleeds.co.uk/Admissions/Current year/2009 to 2010 directory body.pdf

Epstein, J. M. (2009). Modelling to contain pandemics. *Nature, 460*, 687.

Epstein, J. M., & Axtell, R. (1996). *Growing artificial societies: Social science from the bottom up*. Washington, DC: Brookings Institution Press.

Ewens, D. (2005). *The National and London Pupil Datasets: An introductory breifing for researchers and research users* (DMAG Breifing 2005/8). Data Management and Analysis Group, Greater London Authority, City Hall, The Queen's Walk, London, SE1 2AA.

Feinstein, L., & Symons, J. (1999). Attainment in secondary school. *Oxford Economic Papers, 51*, 300–321.

Gereluk, D. (2005). Communities in a changing educational environment. *British Journal of Educational Studies, 53*(1), 4–18.

Gibson, A., & Asthana, S. (2000). What's in a number? Commentary on Gorard and Fitz's "Investigating the determinants of segregation between schools". *Research Papers in Education, 15*(2), 133–153.

Goldstein, H., & Noden, P. (2003). Modelling social segregation. *Oxford Review of Education, 29*(2), 225–237.

Gorard, S. (1999). 'Well. That about wraps it up for school choice research': A state of the art review. *School Leadership and Management, 19*, 25–47.

Gorard, S., & Fitz, J. (1998). The more things change … the missing impact of marketisation? *British Journal of Sociology of Education, 19*, 365–376.

Harland, K. (2008). *Journey to learn: Geographical mobility and education provision*. Ph.D. thesis, School of Geography, University of Leeds, Leeds.

Harland, K., & Stillwell, J. (2007). *Commuting to school in Leeds: How useful is the PLASC?* (Working Paper 07/02). Leeds: School of Geography, University of Leeds.

Harland, K., & Stillwell, J. (2010). Commuting to school: A new spatial interaction modelling framework. In J. Stillwell, O. Duke-Williams, & A. Dennett (Eds.), *Technologies for migration and population analysis: Spatial interaction data applications*. Pennsylvania: IGI Global.

Harris, R., Johnston, R., & Burgess, S. (2007). Neighborhoods, ethnicity and school choice: Developing a statistical framework for geodemographic analysis. *Population Research and Policy Review, 26*(5–6), 553–579.

Heppenstall, A. J., Evans, A. J., & Birkin, M. H., (2006). Application of multi-agent systems to modelling a dynamic, locally interacting retail market. *Journal of Artificial Societies and Social Simulation*, *9*(3).

Johnston, R., Burgess, S., Wilson, D., & Harris, R. (2006). School and residential ethnic segregation: An analysis of variation across England's local education authorities. *Regional Studies, 40*(9), 973–990.

Jones, P., & Elias, P. (2006). *Administrative data as a research resource: a selected audit* (Economic & Social Research Council Regional Review Board Report 43/06). Coventry: Warwick Institute for Employment Research.

Langley, R. (1997). *The use and development of geographical information systems (GIS) and spatial modelling for education planning*. Ph.D. thesis, School of Geography, University of Leeds, Leeds.

Lawrence, I. (1992). *Power and politics at the Department of Education and Science*. London: Cassell.

Leech, D., & Campos, E. (2003). Is comprehensive education really free?: A case-study of the effects of secondary school admissions policies on house prices in one local area. *Journal of the Royal Statistical Society Series A (Statistics in Society), 166*(1), 135–154.

Malleson, N., Heppenstall, A., & See, L. (2010). Crime reduction through simulation: An agent-based model of burglary. *Computers, Environment and Urban Systems*, *31*(3), 236–250.

Müller, J. P. (1998). Architectures and applications of intelligent agents: A survey. *The Knowledge Engineering Review, 13*(4), 252–280.

O'Sullivan, D., & Haklay, M. (2000). Agent-based models and individualism: Is the world agent-based?. *Environment and Planning A, 32*(8), 1409–1425.

Parsons, E., Chalkley, B., & Jones, A. (2000). School catchments and pupil movements: A case study in parental choice. *Education Studies, 26*, 33–48.

Pooley, C., Turnbull, J., & Adams, M. (2005). The journey to school in Britain since the 1940s: Continuity and change. *Area, 37*(1), 43–53.

Rao, A. S., & Georgeff, M. P. (1995, June). BDI agents: From theory to practice. In V. Lesser & L. Gasser (Eds.), *Proceedings of the first international conference on multi-agent systems* (ICMAS-95), San Francisco: MIT Press.

Simpson, S., & Lancaster, D. (1987). *Forecasting pupil numbers for educational planning: A guide to school roll forecasting in the planning process of local authorities in Britain* (No. 71 in Sheffield Papers in Education Management). Sheffield: Sheffield City Polytechnic, Centre for Education Management and Administration.

Statham, J., Mackinnon, D., Cathcart, H., & Hales, M. (1991). *The education fact file*. London: The Open University.

Urban, C. (2000). PECS: A reference model for the simulation of multi-agent systems. In R. Suleiman, K. G. Troitzsch, & N. Gilbert (Eds.), *Tools and techniques for social science simulation* (Chap. 6, pp. 83–114). Heidelberg: Physica-Verlag.

Valentine, G., & McKendrick, J. (1997). Children's outdoor play: Exploring parental concerns about children's safety and the changing nature of childhood. *Geoforum, 28*, 219–235.

Chapter 24
Simulating Spatial Health Inequalities

Dianna M. Smith

Abstract Simulation models are historically vital to epidemiology and have recently become more common in social sciences such as geography, where the aim is to understand underlying causes of population health inequalities. Traditionally the methods used to estimate health outcomes at the local level rely heavily on regression techniques and are restricted by data availability. One possible way forward is to take the best elements of current methods to model interactions between individuals at multiple scales, and extend such models to predict changes in health over time. Complex systems dynamics models and agent-based models in particular are one methodological improvement that would realise both of these goals.

24.1 Introduction

Little is known about population health at the local level. Although there is knowledge of trends towards changing health behaviours and outcomes in the national population as a result of regular surveys, the focus on health at such a coarse scale can mask local variation. For instance, the UK is known to have very high rates of child and adult obesity as a nation, but there are likely to be areas where rates are far above or below the mean values (Moon et al. 2007). Understanding how these rates might change and what might be driving the change among different populations in disparate areas are questions often asked by epidemiologists and health professionals. Social and spatial variation in health has remained a focus of health geography over the previous decades, particularly in wealthier nations as the gaps between the wealthy and poorer members of society has become clearer. Recent research addressed this issue in The Widening Gap, a book which clearly illustrated the

D.M. Smith (✉)
Department of Geography, Queen Mary, University of London, London, UK
e-mail: d.smith@qmul.ac.uk

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_24,

evolving geographies of health inequalities in Britain through extensive data analysis and detailed maps of health outcomes over time and space (Shaw et al. 1999). This work and similar publications (Macintyre et al. 1993) have highlighted the importance of geography, and the potential influence of local environments, in understanding public health variation within a country.

The fascination with social and spatial variation in health has extended beyond the exploration of historical patterns to include present-day trends in non-contagious and infectious disease, health behaviours and the introduction of predictive models that aim to estimate how the spread of disease or health-related outcomes may adjust over time in relation to a changing population. The predictive aspect of many models is especially relevant in countries where health care is provided by the state rather than through individual health insurance schemes. If health professionals and policy makers have an idea of the current and future patterns in smoking, obesity, or cardiovascular disease, then they are better prepared to allocate resources to areas of greatest demand.

The utility of dynamic models is clear in the wake of recent infectious disease outbreaks across the world (H1N1, H5N1 flu) as discussed in a recent article appearing in Nature (Epstein 2009) and later in this book (Simoes 2012). Both authors show how agent-based models (ABMs) of the disease diffusion can be used to formulate policy response to current and future infectious disease outbreaks at the macro scale. This chapter will outline the development of public health models in epidemiology and the social sciences such as geography, and focus particularly on the microspatial, local-level element of any models. The current options available for static models, which estimate health characteristics of populations for one point in time, will first be outlined to give readers an overview of the various techniques and algorithms used by researchers and health organisations to model public health. The chapter concludes with a discussion of the advancement towards dynamic models, which consider population change and observed predictors of disease/behaviours to estimate future public health trends. ABMs, already suggested by epidemiologists to be the best way forward in modelling public health (Auchincloss and Diez Roux 2008), are one strong alternative to the traditional regression-based models.

24.2 Individual Level Models of Health Outcomes

Few geographical health-oriented models deal with individuals; most are prevalence models which look at aggregate local area population characteristics (from a population census) to identify the likelihood of various health outcomes occurring at the population level. Often these models are simply identifying the population-level attributes which are known to influence the disease or health outcome of interest, for instance, the risk of type 2 diabetes increases with age so is more prevalent among retirees than university students. Implicitly all of the models outlined in the following section will consider the relationship between geographic place and health, the intersection of context and composition. This place/person relationship has become a central interest in the study of spatial inequalities; if you remove people from one environment and place them in another, will it impact their health? How can these associations be identified and quantified?

Epidemiologists and geographers have developed several modelling approaches which use some type of regression equation to derive probabilities of behaviours or disease in relation to the local population and/or environment. Most methods are a type of direct standardisation, where the predictors of a disease/outcome in a representative population (such as a national level health survey like the annual Health Survey for England [HSE] or the periodic New Zealand Health Survey [NZHS]) are statistically identified through regression analysis to estimate the likelihood of individuals with certain predictive traits to experience a health condition. The regression may take into account only the individual-level (compositional) characteristics, or it may be extended to include area-level (contextual) attributes. The probabilities created from the regression process can then be applied to the local population.

Indirect standardisation methods take the opposite approach, which is to look at the predictors of a heath outcome in a sample population such as a local health survey and apply it upwards to the national population. This is much less frequently used in spatial public health models due to the prohibitive cost associated with carrying out comprehensive local level health surveys which provide the health data for this technique.

The focus in this chapter is to consider models with finer spatial scale, and usually these models may be categorised into one of the following groups: epidemiological, synthetic population estimation (multilevel or spatial microsimulation) or Empirical Bayesian. The next section outlines these main types of static estimation models and gives examples of their application within the United Kingdom. There is an evolution from the earliest estimation models as computational power and data collection has improved, as will be shown later in this section, where the line between more traditional static models has begun to blur into the dynamic microsimulation models (Wu and Birkin 2012; Portz and Seyfried 2011) that can be seen as predecessors of ABMs.

24.2.1 Multilevel Models

One of the most inherently geographical approaches to creating local-level estimates of health outcomes or behaviours is the use of multilevel, or hierarchical models, to develop local prevalence estimates (Moon et al. 2007; Pearce et al. 2003; Twigg and Moon 2002). The structure of multilevel models is described in the name; people are 'nested' within multiple area levels, such as neighbourhoods, schools or work environments. Multilevel models have gained substantial popularity in the social sciences as they allow researchers to quantify the magnitude of the influence that place-based characteristics might have on population health. For example, how might neighbourhood deprivation affect mental well-being? (Fagg et al. 2006).

Prior to the implementation of a multilevel model, relevant predictors for the health outcome need to be identified. Each of the predictors need to be relevant to the health outcome and present in both the survey dataset and the small-area population data (Twigg and Moon 2002). Logistic regression models are preferable in situations where the outcome is a dichotomous value (not diabetic/diabetic; non-smoker/smoker) and the predictors are either a continuous scale (such as age) or categorical (such as ethnic groups) (Gatrell 2002).

One example of the multilevel modelling framework is the creation of nationwide probabilities for smoking status based on data (age, sex, smoking status, home Output Area [OA]) from the Scottish Health Survey and the 1991 Census (Pearce et al. 2003). Each of the 13,784 respondents are grouped into 12 age-sex bands to calculate the probability of smokers in each band; the age-sex distribution is available from the 1991 Census. Using the known smokers/non-smokers and their area of residence alongside age and sex bands at the individual level, the authors were able to estimate additional data from the census about each OA's population characteristics, including 16 'person' variables (including % unemployed) and 9 'household' variables (e.g., % owner occupied households). There were also two variables at the next largest area (pseudo postcode sector), deprivation and an Office of National Statistics (ONS) Ward classification (Pearce et al. 2003).

After testing a series of multilevel models and identifying the significant variables influencing smoking at the individual and area level, the parameter estimates from the final multilevel model were used to calculate new probabilities for smoking in each of the age/sex groups, based on several new variables; these probabilities were then applied to all of the output areas (where all of the predictive variables were available) across Scotland. The results showed a wide range of smoking prevalence, but the predictor variables which proved most significant were consistent with previous studies (Pearce et al. 2003). The combination of small-area data with survey responses is very similar to the epidemiological modelling approach; however, the multilevel framework allows researchers to clearly identify significant predictors at more than one scale. In addition, the inclusion of cross-level interactions between predictor variables adds greater accuracy to the resulting estimates (Twigg and Moon 2002).

One limitation of the multilevel modelling framework for the creation of prevalence estimates is the need for data on predictor variables to be available at the geographic scale for the resulting estimates. As will be explained later in this section, spatial microsimulation techniques are not as limited by data to create estimates. Where the multilevel modelling approach assigns the parameter estimates to individuals matching a multifaceted profile (for instance, white males aged 30–39 years in social class AB), the microsimulation method assigns probabilites for behaviours iteratively to each of the four attributes in turn (ethnicity, sex, age and social class).

The results from this prevalence estimation approach can be tested for accuracy by comparing the outputs against known local-level surveys. Previous results have indicated that the method is quite robust when used for tobacco smoking estimation, although less reliable in accurately predicting alcohol consumption (Twigg and Moon 2002).

24.2.2 Epidemiological Models

The primary difference between epidemiological models and the alternative options for static estimation processes is the use of multiple datasets to generate probabilities. The challenge with this type of model is that the user is limited to only estimate outcomes on typically small-scale studies for derivation of reference rates. However, one particular application of this method has been used extensively by the National Health Service (NHS) as a way of firstly estimating the national (English) prevalence of type 2 diabetes, which is often undiagnosed, and to also create these estimates at a more local level (Forouhi et al. 2006). Because these models are dependent on relatively small local surveys, they may use data from sources that are far apart in time and place. In the case of the model of Forouhi et al. (2006) model, they used reference rates based on age, sex and ethnicity from six datasets ranging from 1986 to 2000. The authors created a set of time and place adjustments to correct for differences between the study populations and locations.

Once the reference rates are created from the epidemiological datasets, they can be applied to crosstabulated 2001 Census data (age-sex-ethnicity) at the smallest area level where such crosstabulations are available. The benefit of this modelling approach is that all the data are based on a variety of real-world datasets. However, users are constrained by the need for crosstabulated census data to build up the estimates. In the case of diabetes, the lack of data flexibility meant that socioeconomic status was not used as a predictor in the model, although this variable is known to influence diabetes incidence (Connolly et al. 2000; Evans et al. 2000). Unlike the multilevel modelling framework, there is no scope for adding area-level predictors such as land use mix.

A different type of synthetic population estimation similar to the epidemiological models described above, but with greater flexibility in how the predictor variables are included in the model, is through the incorporation of Bayesian estimates. As with the epidemiological method, the models are designed to be used at a scale where crosstabulations of the necessary attributes are available. This method has been used to estimate coronary heart disease (CHD) and diabetes in England (Congdon 2006, 2008).

The diabetes estimates created in this way are similar to the epidemiological model described above, but the initial data come from the 1999 and 2003 HSE to calculate age by sex by ethnic group specific prevalence rates for both type 1 and 2 diabetes. The estimated rates are then applied to the 2001 Census wards, where the age-sex-ethnic group population distributions are known. The Bayesian methods employed by Congdon include a 1999 diabetes risk factor to create accurate predictions of diabetes prevalence and confirm the probabilities for diabetes created from the regression of 2003 data. There is significant overlap in the modelling techniques between Congdon's model and those implemented using a multilevel approach or epidemiological method.

24.2.3 Microsimulation

Spatial microsimulation techniques offer the ability to link non-spatial datasets, such as national health surveys, with spatial data such as sociodemographic attributes from the population census. Unlike the other approaches, the microsimulation model is not dependent on having cross-tabulated data at each area level where the estimates are being created. Instead, the purpose of spatial microsimulation is to iteratively replicate known characteristics of the population which predict the health outcome of interest reliably at the local level. There are several different computational algorithms for spatial microsimulation, which are outlined elsewhere in this book (Birkin and Wu 2012). Deterministic reweighting has been used in a suite of models for health behaviours and outcomes including smoking, diabetes and obesity (Tomintz et al. 2008; Smith 2007). Other options include stochastic methods such as simulated annealing and combinatorial optimization. With deterministic reweighting, a probability for each person who responded to the non-spatial survey to live in each local area is calculated, based on a reweighting algorithm that takes each of the predictive variables in turn (Smith et al. 2009; Ballas et al. 2006). The sums of all the probabilities generated for each area will add up to the census-based population total. These probabilities can be used to generate prevalence estimates as they will give an indication of the proportion of the population affected by the health outcome/behaviour.

Where microsimulation differs strongly from the alternative methods outlined above is that multiple outcomes or behaviours may be estimated for a local population at one time rather than creating a series of outcome-specific models which have to be re-run for every desired characteristic. For example, if the prevalence rates of adult obesity and type 2 diabetes were created using a multilevel modelling framework, this would require two separate modelling runs for each health condition rather than only one with spatial microsimulation. However, the lack of specificity in the synthetic population creation from microsimulation may mean that resulting estimates are not as accurate as alternative methods because different health conditions may be best predicted by very different sociodemographic characteristics. The predictors of smoking behaviour and high levels of physical activity are quite different, so it is unlikely that one model might provide the most accurate estimation of both outcomes. If the conditions are predicted by similar characteristics, such as with obesity and diabetes, then the use of one model is appropriate.

As with the other static prevalence models, validation of prevalence estimates is difficult due to the lack of real-world data. Options to test the reliability of the model predictions can include comparing the model estimates against a related outcome with known prevalence at the same scale, or aggregating the estimates up to a geography where the prevalence is known (Tomintz et al. 2008; Congdon 2008). All of the models are only estimating health based on observed relationships between the modelled health outcome and the local populations' sociodemographic profile that is associated with that outcome (Moon et al. 2007).

Static models, like the dynamic models described in the next section, are limited by the available data that can be included in them. One of the biggest challenges with any type of prevalence estimation is the use of older data for the baseline population (to include attributes such as age, sex, ethnicity). The UK census of population takes place every 10 years but the tables with population characteristics are not immediately available for researchers, so the models are never based on real-time population characteristics. Depending on the country, the larger/national health surveys, which can be used to create the estimates, may not be collected every year; the Health Survey for England is annual but the Scottish Health Survey has only recently been conducted each year (2008 through 2011).

24.2.4 Dynamic Models

Dynamic models attempt to create health outcomes not only for one point in time but also for future populations, by taking into account potential population changes such as an aging population. The utility of predictive models for future planning is particularly important for countries where health care is funded by the government and future budgets must be allocated in advance. Dynamic models can take the form of the regression analysis described above (dynamic microsimulation) or may be based on more intricate relationships, like the complex systems dynamics models which consider individual and environmental level interactions. The systems dynamics models are iterative in nature, building on the baseline data and creating new data as the populations evolve and interact; one specific example of this type of model is an ABM.

24.2.5 Dynamic Microsimulation Modelling

Dynamic microsimulation modelling is described in detail elsewhere in this book (Birkin and Wu 2012). Briefly, this method is an advance beyond the simple static models outlined earlier, often including a stochastic element to the population generation process. Similarly to the static models, health outcomes are estimated based on previous observed associations with demographic characteristics in a type of regression analysis. However, with the dynamic models, the baseline populations are allowed to change in line with expected demographic evolution within an area. For example, aging populations or migration of different ethnic groups between areas will affect the model's estimated outcomes, as will possible changes to government policy related to the behaviour, such as tobacco taxation and smoking policies.

Dynamic microsimulation models have already been created to estimate the future prevalence of obesity (Kopelman et al. 2007). However, the models are still constrained by linear relationships defined by regression analysis. Using obesity prevalence as an example, these models may fail to accurately represent how real people would react to a variety of influences such as less expensive food, better access to fitness facilities, or increased education about the risks associated with obesity.

24.2.6 Complex Systems Dynamics Models

Newer methods in disease estimation approach the health outcome as a complex system, with the aim of including as many potential influences as possible. Gatrell has recently acknowledged the difficulty of accurately modelling health outcomes using predictive models. However, the means of dealing with complexity is not strongly developed in public health applications (Gatrell 2005). Many of the issues raised by Gatrell are intuitive: the inability of the models to account for interactions between variables (beyond the simplistic methods in multilevel models); the simplified, linear nature of the models that are unable to account for non-linear relationships (which, arguably, are widespread in health research); the inherently complex nature of relationships between people and place; the idea of epidemiology as a 'web' of inter-connected mechanisms, which uniquely combine in individual lives (Gatrell 2005).

A recent issue of the American Journal of Public Health was devoted to exploring potential approaches to modelling complex systems, with several authors who echo Gatrell's call for improved models. One of the models, developed in the United States to estimate the impact of various governmental policy on diabetes prevalence, is created using systems dynamics (Jones et al. 2006). This is perhaps the closest that researchers have come to acknowledging the true complexity of public health. However, the model is currently only feasible at the national scale. This particular model, created by health planners at the Centers for Disease Control in the United States, was designed specifically to understand population dynamics related to diabetes. The intention was to inform public health strategy by predicting the future prevalence of diabetes through 2050. The model incorporated factors such as death rates, health insurance, diabetes diagnosis and medication. This model, along with others currently under development at the CDC, promises to improve health planning by better predicting the effects of interventions on public health (Jones et al. 2006).

There appears to be a trade-off in terms of the level of complexity allowed in a model and the unit of geographical analysis for which it can estimate disease prevalence. As complexity studies continue to gain momentum (and computational powers increase), this 'choice' may be resolved, leading to more robust models which can accurately depict current and future health trends at a finer spatial scale.

Complex systems dynamic models are a general category of advanced simulation models that includes ABMs (see Crooks and Heppenstall 2012). The benefit of this family of models is their ability to incorporate multiple scales of influence (like a multilevel model) as well as considering the changing relationships between influences on agents' health within the model. The inherent complexity in person-environment interactions is best modelled using this type of approach because the agents (people) in the model are allowed to react to changes in causal factors for disease from the local environment or each other. The environment may not be such an obvious causal factor in non-communicable disease as it is for illness such as malaria or Dengue fever, but much of the recent work that aims to investigate the increasing trends in obesity suggests that a person's local environment plays a key role (Egger and Swinburn 1997).

An ABM has the unique ability to combine multiple scales/types of influence as well as interactions and feedback loops to ideally replicate interactions that cannot be represented in regression-based models (Auchincloss and Diez Roux 2008). The dynamic nature of ABMs is a great asset to health planning; people vary over time and are influenced by any number of factors at different ages, and this method is the best way to address such complexity. 'Agents' in the models do not necessarily have to be individuals but this is the most common configuration. Attributes and behavioural rules are assigned to the agents based on available data (commercial data, qualitative studies) to begin the simulation, and there is the option to add a random element to the evolving interactions that will dictate how the agents may respond to different situations. The model is then run numerous times to generate a variety of outcomes (Auchincloss and Diez Roux 2008). The model is usually created in a computer programming language like Java, but there are several ready-made programmes such as Recursive Porous Agent Simulation Toolkit (REPAST) that may be adapted by individual users who have less programming experience.

Obesity is a good example of how ABMs can move the epidemiological research forward (Galea et al. 2009). With the wealth of research devoted to studying obesity-promoting (obesogenic) influences at the personal and area level, the complexity of obesity aetiology is well documented (Kopelman et al. 2007). A recent example of agent-based modelling of BMI with regards to local stores and varied strength of an individual's social networks gave one illustration of a possible policy scenario (Galea et al. 2009). In this simplistic model, created in a ready-made ABM framework, the results suggested that people with weaker social network ties had a greater decrease in BMI. However, they were also more likely to have an increase from baseline BMI after the food stores had returned to normal.

A more ABM to predict the evolution of BMI at local levels would likely incorporate much more data. A good basis for a comprehensive ABM for obesity would be to include the obesogenic environment framework outlined by Egger and Swinburn (1997). Their ecological model of obesity breaks the 'environment' into four distinct types (physical, economic, political and sociocultural) and further subdivides these types into the micro (i.e., neighbourhoods, schools, homes) and macro (transport, health regulatory system). Then the individual factors could be introduced in the model (age, sex, ethnicity, social class, marital status, educational attainment, etc.). All of these individual attributes and their relative importance in predicting obesity may be identified from the same types of surveys used to inform the regression-based models. It would be best to isolate aspects of the different influences to understand the relative importance of certain parameters on different people. For instance, women may be less likely to use parks for physical activity than men, or men may make less healthy choices with regards to available food.

ABMs, as one of the complex systems dynamics models, are clearly a big step forward for epidemiological research. However, as with all methods, there are limitations to be considered. The rules that govern agent behaviour are often influenced by the assumptions of the researchers creating the model, or may be overly simplistic. The parameters that are included in the model may not be based on large samples of observed data, particularly with regards to interactions (Galea et al. 2009).

24.3 Conclusion

Increasing computational power has changed the available methods and allowed for the evolution of complex models to more accurately capture the behaviours that contribute to health outcomes. While early prevalence models were restricted in power to a static population, the new developments in systems dynamic models and agent-based modelling have led to more flexible and powerful choices for social scientists and policy analysis.

As discussed elsewhere in this book, advanced computational methods are valuable in predicting the spread of infectious disease and have historically been used by many governments and health organisations to this end. The increasing ability to capture population health dynamics for non-communicable disease may have a significant role in protecting public health in the future as limited funds and resources may be allocated to areas of greatest need. Alternatively, the models will enable users to test the efficacy of various policies to reduce the prevalence of tobacco use, binge drinking or obesity among heterogeneous populations in disparate areas.

Although there are clear challenges to the use of a systems dynamic or agent-based approach to the simulation of population-level spatial health outcomes, the advancement beyond regression based models is a significant addition to the toolbox available for public health and social science. With careful consideration for the data included in the models, including rules of behaviour for the agents, ABMs provide a great improvement from previous methods that took little or no account of individual variation and interactions (Galea et al. 2009). Researchers are encouraged to be aware of limitations to this method. As with any new approach, the outputs must be interpreted with an understanding of the underlying processes that are used to generate them. However, the shift towards complex systems dynamics modelling is a move towards true individual-based modelling in non-infectious epidemiology.

References

Auchincloss, A. H., & Diez Roux, A. V. (2008). A new tool for epidemiology: The usefulness of dynamic-agent models in understanding place effects on health. *American Journal of Epidemiology, 168*(1), 1–8.

Ballas, D., Clarke, G., Dorling, D., Rigby, J., & Wheeler, B. (2006). Using geographical information systems and spatial microsimulation for the analysis of health inequalities. *Health Informatics Journal, 12*, 65–79.

Birkin, M., & Wu, B. (2012). A review of microsimulation and hybrid agent-based approaches. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 51–68). Dordrecht: Springer.

Congdon, P. (2006). Estimating diabetes prevalence by small area in England. *Journal of Public Health, 28*, 71–81.

Congdon, P. (2008). Estimating CHD prevalence by small area: Integrating information from health surveys and area mortality. *Health & Place, 14*, 59–75.

Connolly, V., Unwin, N., Sherriff, P., Bilous, R., & Kelly, W. (2000). Diabetes prevalence and socioeconomic status: A population based study showing increased prevalence of Type 2 diabetes mellitus in deprived areas. *Journal of Epidemiology and Community Health, 54*, 173–177.

Crooks, A. T., & Heppenstall, A. J. (2012). Introduction to agent-based modelling. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 85–105). Dordrecht: Springer.

Egger, G., & Swinburn, B. (1997). An 'ecological' approach to the obesity pandemic. *British Medical Journal, 315*, 477–480.

Epstein, J. M. (2009). Modelling to contain pandemics. *Nature, 460*, 687.

Evans, J. M. M., Newton, R. W., Ruta, D. A., McDonal, T. M., & Morris, A. D. (2000). Socio-economic status, obesity and prevalence of Type 1 and Type 2 diabetes mellitus. *Diabetic Medicine, 17*, 478–480.

Fagg, J., Curtis, S., Stansfeld, S., & Congdon, P. (2006). Psychological distress among adolescents, and its relationship to individual, family and area characteristics in East London. *Social Science & Medicine, 63*, 636–648.

Forouhi, N. G., Merrick, D., Goyder, E., Ferguson, B. A., Abbas, J., Lachowycz, K., & Wild, S. H. (2006). Diabetes prevalence in England, 2001 – estimates from an epidemiological model. *Diabetic Medicine, 23*, 189–197.

Galea, S., Riddle, M., & Kaplan, G. A. (2009). Causal thinking and complex system approaches in epidemiology. *International Journal of Epidemiology*. doi:10.1093/ije/dyp296.

Gatrell, A. C. (2002). *Geographies of health: An introduction*. Oxford: Blackwell.

Gatrell, A. C. (2005). Complexity theory and geographies of health: a critical assessment. *Social Science & Medicine, 60*, 2661–2671. Burning issues – Selected papers from the 10th international symposium in medical geography, Manchester, 2003.

Jones, A. P., Homer, J. B., Murphy, D. L., Essien, J. D. K., Milstein, B., & Seville, D. A. (2006). Understanding diabetes population dynamics through simulation modeling and experimentation. *American Journal of Public Health, 96*, 488–495.

Kopelman, P., Jebb, S. A., & Butland, B. (2007). Executive summary: FORESIGHT tackling obesities: Future choices project. *Obesity Reviews, 8*, vi–ix.

Macintyre, S., Maciver, S., & Sooman, A. (1993). Area, class and health: Should we be focusing on places or people? *Journal of Social Policy, 22*, 213–234.

Moon, G., Quarendon, G., Barnard, S., Twigg, L., & Blyth, B. (2007). Fat nation: Deciphering the distinctive geographies of obesity in England. *Social Science & Medicine, 65*, 25–31.

Pearce, J., Boyle, P., & Flowerdew, R. (2003). Predicting smoking behaviour in census output areas across Scotland. *Health & Place, 9*, 139–149.

Portz, A., & Seyfried, A. (2011). Analyzing Stop-and-Go Waves by Experiment and Modeling in R.D. Peacock, E.D. Kuligowski, J.D. Averill, (eds) Pedestrian and Evacuation Dynamics, Springer, pp 577–586

Shaw, M., Dorling, D., Gordon, D., & Davey Smith, G. (1999). *The widening gap: Health inequalities and policy in Britain* (Studies in poverty, inequality and social exclusion). Bristol: Policy Press.

Simoes, J. (2012). *An agent-based approach to spatial epidemics through GIS*. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 591–610). Dordrecht: Springer.

Smith, D. M. (2007). *Potential health implications of food access*. Unpublished PhD thesis, School of Geography, University of Leeds, Leeds.

Smith, D. M., Clarke, G. P., & Harland, K. (2009). Improving the synthetic data generation process in spatial microsimulation models. *Environment and Planning A, 41*, 1251–1268.

Tomintz, M. N., Clarke, G. P., & Rigby, J. E. (2008). The geography of smoking in Leeds: Estimating individual smoking rates and the implications for the location of stop smoking services. *Area, 40*, 341–353.

Twigg, L., & Moon, G. (2002). Predicting small area health-related behaviour: A comparison of multilevel synthetic estimation and local survey data. *Social Science & Medicine, 54*, 931–937.

Wu, B. M., & Birkin, M. H. (2012). Agent-based extensions to a spatial microsimulation model of demographic change. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 347–360). Dordrecht: Springer.

Chapter 25
Agent-Based Modelling of Residential Mobility, Housing Choice and Regeneration

René Jordan, Mark Birkin, and Andrew Evans

Abstract Dynamics in the housing market can be simulated using agent-based modelling. Focusing on the theme of urban regeneration, we present a housing market model framework which explores the causal relationships that occur in this market.

25.1 Introduction

The housing market is a dynamic system of intricately woven interdependent processes. It is affected by the volatility in the financial markets and the conditions of this market affect discriminatory individual level behaviour. Like other applications, agent-based models (ABMs) can be used to simulate activity in this market with a view to gaining a better understanding of how the market works as well as to realise causal relationships that occur.

The terms residential mobility and housing choice are standard within housing market research and can be found across the housing studies literature (Kim et al. 2005; Tu and Goldfinch 1996). These terms encapsulate the movement questions which lead households to decide whether to move and subsequently choose a new home. One of the most influential factors which affect these processes is the family life cycle, in collaboration with income (Dieleman 2001). Changes in the family life cycle affect location choice whilst overall government policy affects demand and supply of the housing stock. The linkages between these processes are important. We can use ABMs to answer this question.

R. Jordan (✉) • M. Birkin • A. Evans
Centre for Spatial Analysis and Policy, School of Geography, University of Leeds, Leeds, UK
e-mail: r.j.jordan@leeds.ac.uk; m.h.birkin@leeds.ac.uk; geoaje@leeds.ac.uk

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_25,

In this chapter, we first build the foundation of such a model by taking a look at the theory of residential mobility and housing choice with particular emphasis on the family life cycle, income, location choice and government policy. The link between this theory and ABMs will be made while exploring how ABMs have been used to simulate aspects of the housing market. A modelling framework is presented illustrating how housing market behaviour can be represented programmatically. Results of the model are then presented followed by a discussion of the usefulness of this technique for land use planners and housing study practitioners.

25.2 Residential Mobility and Housing Choice

Residential location decisions are strongly influenced by national housing policies, local planning constraints, and by a wide range of fiscal and social policies. In addition to government policy, household characteristics often initiate movement and in turn influence where the household will move. In a more basic sense, households progress through the family life cycle with various levels of income at their disposal. Changes in either of these attributes – the family life cycle or income – are likely to trigger residential mobility. These are the main links between the household's characteristics and residential mobility. While some households may choose not to move because of limitations in income and/or supply on the market, other, less constrained households will try to find a new house. By taking a closer look at these processes, we can build an understanding of how the housing market operates.

25.2.1 Family Life Cycle, Income and Location Choice

Figure 25.1 illustrates the general progression of the family life cycle. Typically, the cycle begins with a single individual and advances through varying household formations. This process is punctuated by significant events such as job losses or gains, marriage, births, deaths, divorce or separation, retirement and adult children leaving the parental home. As these significant events occur, changes in household attributes can be observed. For example, events such as marriage and births result in larger families while the converse is true for deaths. Thus the need for larger or smaller homes is likely to be triggered.

The financial budget is by far one of the most important factors when the decision to move is considered (Boehm 1982). There is a direct positive correlation between the amount of disposable income that can be used for housing and the cost of the house. For example, a promotion may make more money available and can trigger a move to a more expensive house. The converse of this statement is also likely to be true.

Therefore, income has a knock-on effect for the type of property which households can afford, the size of the house and the housing tenure. In the case of the

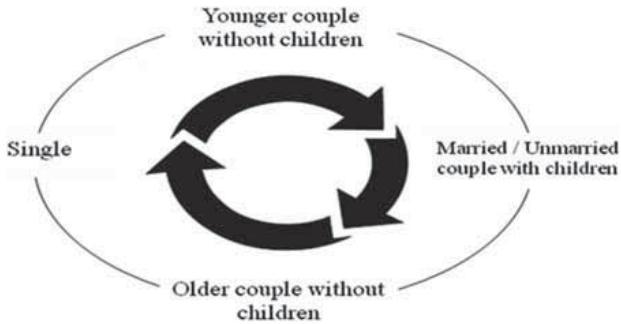


Fig. 25.1 The family life cycle

framework that we develop in this chapter, we use type to indicate the accommodation type, i.e., whether a house is detached, semi-detached, terrace, a flat or maisonette. Size is represented by the number of rooms in the house and tenure type indicates whether a house is on the private rental market, the public rental market or up for full ownership. If we consider every combination of these three variables, we can quickly come to the conclusion that when these combinations are compared, stratification in the housing market can be observed. For example, a detached house, up for ownership, with ten rooms is likely to be more expensive than a publicly rented house with five rooms.

In addition to these factors, the influence of neighbourhood quality must be underscored. The physical conditions of the neighbourhood, and amenities such as shops, school quality, security, transport connections, green spaces, and the proximity to built up areas, are characteristics which can alter house prices. Understandably this further increases the complexity of household preferences and choice. The choice of a new home can be conceptualized as a choice between a weighted combination of dwelling alternatives. The strength of each weight may be based on the changing needs of the household. These weights are discussed when the modelling framework is introduced.

25.2.2 Housing Policy

None of the determinants of movership work in isolation; they are interrelated and interdependent. They are also largely affected by market conditions. A household's choice of dwelling is constrained by the demand for and the requisite supply of houses as well as the availability and accessibility of financial instruments such as mortgages. Each of these factors is influenced by government policies in general. More specifically, changes in housing policy can affect the supply of new houses. Demographic changes can affect demand. Economic policy can alter interest rates, which can affect house prices and access to lending instruments. Therefore, households will choose homes based on the extent to which government policies affect them.

Although there are numerous government and housing policies that can be mentioned here, we focus on urban regeneration policy and its effect on public housing or council tenants.

25.2.2.1 The Evolution of Urban Regeneration Policy

Council housing in the UK has gone through significant changes over the last century. Also known as public housing, it was initially created to improve dwelling conditions for those without the means to do so for themselves. Government-owned housing transitioned to a market which provided another housing option for anyone regardless of income (Mullins and Murie 2006). Over time the housing market changed; many of the houses in good condition were sold off leaving an inventory of poor quality houses in the council housing market. Poor quality housing and low-income tenants entangled in the cycle of poverty added to the melee of problems that plagued this sector (Power and Mumford 1999). It was in this context that Urban Regeneration policies were conceptualized.

Aimed to encourage social mixing, policies focusing on urban regeneration were introduced in the 2007–2008 period. Proponents of such policies believe that communities of mixed socio-economic status can encourage social development among disadvantaged households (Tunstall 2003). In theory, different classes of people have the potential to attract a wider range of new businesses and new residents in communities affected by social problems. This type of cross tenure community is likely to comprise of a range of people at different socio-economic levels, with different lifestyles, values and attitudes where the more productive socio-economic groups are thought to positively influence the other groups (Bridge 2002). The government believes that regeneration efforts are likely to create stable communities and disadvantaged households can experience reduced financial dependence on the public purse as a result of increased aspirations and availability of jobs.

25.2.2.2 The EASEL Case Study

One example of the implementation of Regeneration Policy can be observed in the East and South East district of Leeds in the UK (EASEL). Home to approximately 36,000 households, the area is noted to have some of the worst deprivation statistics (Index of Multiple Deprivation) in the country as reported by the Office of National Statistics. Issues of poor housing, high unemployment rates and low educational attainment are some of the negative characteristics that add to the stigma associated with this area. With an aim to improve these statistics, Leeds City Council has initiated intense regeneration efforts to the tune of £90 million (Leeds City Council 2007). Improvements in housing stock quality and quantity are designed to provide more affordable homes nestled in new mixed tenure communities.

Policies such as these appear legitimate in theory, but in reality their practical impact is questionable. Regeneration policy attributes the inherent social problems in low-income communities to the fact that these communities are segregated. Although there is evidence to support this, it is unlikely that households with sufficient disposable income would choose to live in neighbourhoods that are badly stigmatised. Furthermore, the early work of Thomas Schelling suggests that, if people are allowed to exercise slight demographic preferences when relocating, they will cluster together based on these preferences (Schelling 1969). This is one of the major challenges to mixed communities, one which can be explored through the implementation of an ABM.

25.3 Where Do the Agents Come In?

We can take these theoretical observations and use them to build our agent-based simulation. Households and houses can be classified as agents while housing policy can be simulated by altering the housing stock available.

We may associate various attributes such as age, number of children, number of cars, socio economic status, accommodation type and tenure type with each household. This list is not exhaustive but captures the type of information that is useful when modelling the household. One may choose to model the entire household as one agent or model every possible person while using the unit household, to represent an aggregation of individuals dependent on the sophistication in the life-cycle model to be used. Where houses are concerned, we can use attributes such as accommodation type, value, vacancy status, etc. A collection of houses may form a district or some other aggregated unit used in the real world.

The work of Schelling (1969), Aguilera and Ugalde (2007), Laurie and Jaggi (2003) and Yin (2009) illustrate how these entities can be used to simulate housing market activity via ABMs. Schelling (1969, 1971) examined the role of preferences in an artificially created community and illustrated how individual behaviour can create significant collective results not directly intended by the individual. Schelling (1969) demonstrated this by using only one rule, i.e., all agents preferred to live among at least 33% of agents of the same ethnic group as themselves. The result was total segregation.

Aguilera and Ugalde (2007) attached house prices to each space on a lattice grid. Individuals were rated by socio-economic status and income. House prices were strongly related to the type of neighbourhood each house was located in and evolved in such a way that prices fluctuated at times. Agents moved momentarily in order to match their status with house prices by exchanging their location with other agents. This inequality in income among agents was noted to be the factor strongly, positively correlated to segregation. In other words, the more unequal a neighbourhood is, the more segregated it becomes.

The work of Laurie and Jaggi (2003) also used the basic segregation model as proposed by Schelling (1969, 1971) in examining the role of vision in effecting segregation. Vision is used in this sense to describe the number of neighbours the agent assesses in determining whether or not they wish to move. The model illustrated that with only slight preferences and increasing vision, the society became more and more integrated. The converse is true when vision and the level of tolerance are decreased, i.e., society becomes more and more segregated.

Yin (2009) increased the dynamics in his model by devising a social simulation based on the City of Buffalo in the United States. In his model, the issue of race and social class are examined as they relate to residential segregation. Yin's research builds on Schelling's theory and illustrates how factors such as race and economic constraints, when exercised as a part of the housing choice process, can cause segregation of varying degrees at the aggregate level. However, Yin illustrated that when housing policies were implemented, this segregation could be reduced once racial sensitivity was low.

These and other models such as those by Pancs and Vriend (2007), Zhang (2004a, b) and Benenson (2004) give us an idea of how agents can be used to simulate and test phenomena in the housing market. Notice how each of these models focus on a different aspect of the housing market. Whether ethnicity preference, socio-economic status, neighbourhood distance, house prices, or integration policy, the dynamics are diverse and yield varying results. We can build on these examples by introducing other dynamics in order to create a complete agent-based market model.

25.4 The Model Framework

Using the entities, households, houses and the surrounding environment, a modelling framework is presented here to recreate the housing market. Fundamentally, we know that amidst a list of households, some choose to move while others do not. We also know that once these households choose to move, they need to find a house to relocate to.

25.4.1 How Do We Know Which Households Want to Move?

For ABMs, individual records are ideally needed to represent households. In the UK, these households can be derived from data sources such as the Household Sample of Anonymised Records (H-SAR). The H-SAR contains attributes such as age, ethnicity, accommodation type, tenure type and the propensity to move statistic. The propensity to move is a migration indicator which dictates if the household moved within the last year of the recorded census (CCSR 2010). It is very important for our model as it is used to determine which households need to find a new house. A more extensive discussion of the use of this variable can be found in the paper by Jordan et al. (2011).

Other data such as Output Areas, roads, and significant buildings are also used. This data is stored in shapefile format and can be obtained for the UK through data providers such as Edina UK Borders and Ordnance Survey MasterMap.

25.4.2 Where Do These Households Go?

Seven rules are used to determine where households will move to. The rules are defined as follows:

1. *Households move to areas where the ethnic makeup is tolerable.*
   When a household desires to move, the search for a new house begins. A house is deemed favourable if at least 33% of the surrounding households are of the same ethnicity type as the household wanting to move. The rule is augmented for minority groups with strong religious ties; i.e., the new house must also be within close proximity (~5 miles) to a religious centre, for example, a mosque (Johnston et al. 2002; Phillips 1998, 2007; Peach 1999; Schelling 1969).
2. *Households look for a new house within known areas.*
   Communities where households frequent for the purpose of work and other activities can be characterized as known areas. Since this simulation focuses on housing, for each household, we create memory by storing all the districts in which the household may have lived. The proximity of the surrounding community is thought to be between 6 and 20 miles. For public renters this distance is close to 6 miles while for private sector households, the distance from the previous house may vary between either extremes of this range (Cho et al. 2004; Cho 2004). Therefore, when a new, vacant house is found, its location is checked to ensure that it is within a known area.
3. *Households move to houses where the size of the house is adequate.*
   Ensuring that the size of the house is acceptable is important (Dieleman 2001). This can be determined by trying to find a house with the desired number of rooms for each household. This variable is derived from the H-SAR, i.e. number of rooms required.
4. *Households move to areas where schools are accessible.*
   If the household contains school aged children, the proximity to schools is taken into consideration. Desirable schools are generally thought to be within a 5 mile radius of the home although this distance may increase with secondary school aged children (Gibbons and Machin 2003; Black 1999; Strand 2002). This distance measure is used in the current implementation.
5. *Neighbourhood quality plays a role in influencing household choice.*
   The determinants of neighbourhood quality include amenities such as shops, schools, green spaces and security. Households often take this into account when choosing a new home (Tu and Goldfinch 1996). Using the Index of Multiple Deprivation (IMD), the neighbourhood (Output Areas) in which a house exists is compared to that of the current house. The IMD is a statistic which ranks Output Areas across the UK. It is made up of average statistics pertaining to crime rates, employment deprivation, education and barriers to housing. A new house is thought to be more favourable if the IMD within the new Output Area is higher than that of the current house.
6. *The socio-economic status of a household influences the type of house chosen.*
   Households in higher socio-economic brackets are likely to be owner occupiers while households in lower socio-economic brackets are likely to be private or

public renters (Cho et al. 2004; Cho 2004). For example, a manager in a larger organisation is likely to live in an owner occupied home. On the contrary, an unskilled worker is likely to live in the public or private rental market.

7. *Households will move to areas where transport routes are accessible.*

Although this rule applies to all households, it is especially important for households without cars. As a means of ensuring that the journey to work is manageable, this rule ensures that a major road is found within a 1 mile radius of the new house (Böheim and Taylor 1999; Gjessing 2009). Therefore, for a household without cars, a new house is favourable if it is located within a 1 mile radius of a major road. For households without cars, it is not important that this criterion be satisfied.

25.4.2.1 Ranking the Ruleset

The model is initiated with a realistic proportion of vacant houses. As each household is interrogated with the movement questions, each vacant house is taken through a process of ranking. For example, in Rule 1, a house found within a neighbourhood where the ethnic mix is not tolerable (less than 33%) would be ranked with a value of 0 while for a tolerable mix the house would be ranked with a value of 1. Similarly, for the socio-economic status by tenure (Rule 7), if a public rented house is found and the household is deemed to be low-income, then this council house will be ranked higher than an owner occupied house. For a household with a managerial role, a council house will be ranked lower than an owner occupied house.

The choice of a new house is a combination of dwelling alternatives (Tu and Goldfinch 1996; Dieleman 2001). Using this process of ranking, the profile of the household is compared to the characteristics of all available houses. The house with the highest combined ranking is chosen.

25.4.2.2 Advancing the Model in Yearly Timesteps

In order to move the model from year to year, we create a threshold based on the percentage movers in any given year. This statistic can be calculated using the original data source, i.e., the number of people who have moved in a given year divided by the total number of people. Once this threshold is met, another year begins. Note that yearly time spans measured in timesteps may not be equal from year to year as this threshold condition must first be met.

With the framework in place, we can begin to explore some of the results of the model.

25.5 Results and Application

Let us use a series of scenarios to examine the model results. Scenario 1 is the result of running Rule 4 in isolation. Scenario 2 is the result of running all the rules simultaneously. In this instance, the number of vacant houses is severely constrained.

Fig. 25.2 Initialised EASEL area display before execution of Rule 4. Here the *coloured dots* are used to represent households of varying ethnic types while the *small rectangular polygons* are representative of houses. The *larger green polygons* are used to represent Output Areas

Table 25.1 Details of the starting conditions

Initial state	
# Households	465
# Houses	490
# Iterations	Unspecified; not linked to actual time

Households are limited by a vacancy rate of 4% of the total houses. Scenario 2 is executed a second time with a significant increase in vacant houses.

25.5.1 Scenario 1 – Rule 4: Households Move to Areas Where Schools Are Accessible

Figure 25.2 is a pictorial representation of the initial distribution of households before Rule 4 is executed and Table 25.1 contains the starting conditions. Notice how the households are reasonably distributed across the Output Area zones after initialisation.

Assuming that only households which move contain school aged children, Fig. 25.3 is the result of this rule. Notice the shift in the households toward the lower left area of the diagram. This is because there is a school in the vicinity of this area as shown above.

In executing each rule separately, similar trends are realised. In the case of Rule 1 – Ethnicity, households begin to cluster around others who are of the same

Fig. 25.3 Resultant EASEL area display after execution of Rule 4 (~40 timesteps). Here the system has reached a point of stability, i.e. movement within the model is negligible indicating that the majority of the households are satisfied with their current location

ethnic type as themselves. Rule 2 – Known Areas, households choose a new house in areas previously known to them; this limits the number of potential new locations where they may choose to live. In the case of Rule 6 – Socio-economic Status, households begin to cluster together based on their socio-economic status. This is largely because tenure types are clustered together in a similar manner.

These trends are most pronounced when each rule is run separately. However, when the ranking system is employed, the individual trends become less distinct.

25.5.2 Measuring Diversity

We can extend our interpretation of Scenario 1 to include an index by which diversity is measured. In the case of the previous scenario, the question can be asked: how does the schools rule affect the demographic makeup of Output Areas once implemented? In the case of Rule 6 – Socio-economic Status, we can query how a household's social class can affect the demographic makeup of Output Areas once implemented. If the goal of urban regeneration is to create mixed communities, then such a statistic can be used to inform us of the extent of this mixing over time. This can be illustrated by using the Index of Diversity (Blau 2000).

The index of diversity is a statistical indicator that can be used to examine the relative diversity of households within each Output Area. Diversity can be measured based on any variable of interest. Thus, ethnic diversity, socio-economic diversity and demographic diversity are some of the scenarios that can be examined. The index is defined as follows:

$$D = 1 - \sum_{i=1}^{N} p_i^2$$

Here p_i is the proportion of households in Output Area i of a specific type, e.g. ethnicity, N is the total number of Output Areas and D returns values between 0 and 1. Values closer to 0 indicate that the Output Area is not very diversified while values closer to 1 are indicative of heterogeneous Output Areas, i.e., where communities are mixed. Let us explore the usefulness of this indicator using Scenario 2.

25.5.3 Scenario 2 – All Rules

We can explore the results of our model when all of the rules are implemented together. Using the initial states shown in Table 25.2, the model is first executed for a period of 50 years with a 6% vacancy rate. The model is then executed a second time with a less constrained vacancy rate as detailed in the table below. The results are shown in Fig. 25.4.

Table 25.2 Details of the starting conditions. Approximately 625 houses are used with a vacancy rate of 6%

Initial state	Constrained vacancy	Less constrained vacancy
# Households	587	587
# Houses	625	875
# Iterations	1,089	1,086
# Years	50	50

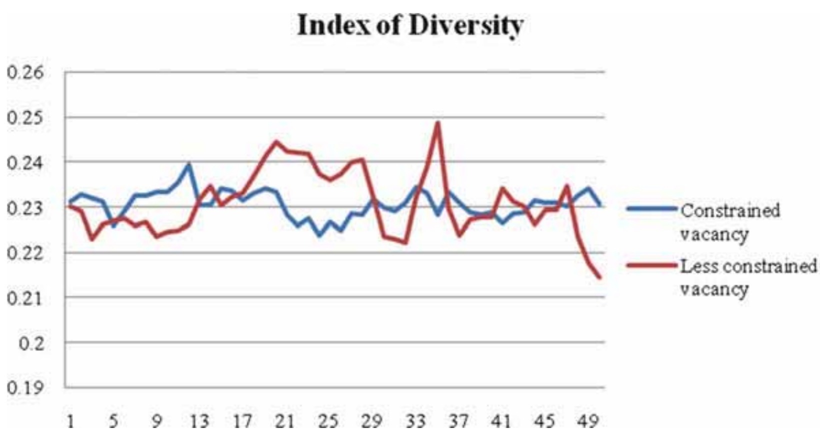


Fig. 25.4 Index of diversity

Figure 25.4 shows the result of the simulation executed under two different conditions and uses the Index of Diversity to analyse the variation in diversity. Although the variation in the statistics is limited, varying from 0.21 to 0.25, we can still observe a difference in the two simulation runs. When the simulation is executed with a 6% vacancy rate, the variation is more limited than when the simulation is executed with a higher vacancy rate. One may link this contrasting pattern to the fact that under constrained conditions, households have fewer housing options and may compromise on their ideal home, an effect of the ranking system mentioned in Sect. 25.4.2.1. When housing options increased, households have a wider selection of vacant houses to choose from and are more likely to choose homes that are ideal for their circumstance, if available. Thus, more variation in diversity can be seen.

25.6 Discussion and Conclusions

The ABM framework, which has been introduced in this chapter, can be used to give insights into the dynamics of the housing market in relation to urban regeneration plans. What is interesting about the rules introduced is that they are all limited by parameters. Rule 1 dictates that households will search for homes where the ethnic makeup is tolerable. This is limited to a percentage of at least 33%. Rule 2 dictates that households look for houses in known areas. A distance measure of 6–20 miles is used in this rule. A similar distance measure of 5 miles is used in Rule 4 (Schools).

In general, these parameters are used to constrain the model each time it is executed. They can be used to explore how different combinations of parameters affect the model outcome. Thus as shown in Scenario 2, a higher vacancy rate increases the potential range of new homes which households can occupy and leads to greater homogeneity. Other scenarios can be created to examine lower tolerance levels in terms of the ethnicity rules while distance measures can be increased or decreased again to examine their effects.

This is how we can analyse the effects of regeneration policy. When new houses are built, this increases the previously limited housing stock, therefore increasing the housing options for households. If new schools are built or schools are demolished, the distance to schools is affected. Socio-economic status can be monitored as these changes are made, and overall, the change in diversity in the study area can be measured over time.

Even with these considerations, further work is proposed for this model. The model variables need to be updated from year to year, parameters need to be calibrated and the correctness of the model should be validated.

As mentioned in Sect. 25.2, changes in the family life cycle cause changes in housing needs. Therefore, it is important that family characteristics be altered from time to time to reflect reality. At the very least, mortality and fertility rates should be a consideration. Another consideration may be the fact that households may move out of the region of interest. These changes are important as they affect demand and supply of the housing stock and the distribution of households.

Calibration can be used to find the most suitable combination of parameters that replicate reality, which may be determined using a genetic algorithm to analyse the performance of various combinations of parameters. In turn, the validity of the model can be tested by comparing yearly diversity indices generated by the model with yearly diversity indices generated from known datasets such as the Pupil Level Annual School Census (PLASC) data available in the UK. Such a data source contains details on ethnicity at the Output Area level. The distribution of school age children in the PLASC dataset can then be compared to the distribution of households with school aged children in our model.

We have created a framework to explore and analyse the dynamics of the housing market and urban regeneration. Such a framework allows us to unpack the building blocks of the housing market with a view to understanding not only how the market works but also the effects that changing parameters can have on housing market outcomes. This allows us to explore the extent to which urban regeneration schemes can result in the creation of mixed communities.

Acknowledgements Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

References

Aguilera, A., & Ugalde, E. (2007). A spatially extended model for residential segregation. *Discrete Dynamics in Nature and Society*. doi:10.1155/2007/48589.

Benenson, I. (2004). Agent-based modeling: From individual residential choice to urban residential dynamics. In M. F. Goodchild & D. G. Janelle (Eds.), *Spatially integrated social science: Examples in best practice* (pp. 67–95). Oxford: Oxford University Press.

Black, S. (1999). Do better schools matter? Parental valuation of elementary education. *Quarterly Journal of Economics, 114*, 578–599.

Blau, J. (2000). Group enmity and accord: The commercial press in three American cities. *Social Science History, 24*(2), 395–413.

Boehm, T. (1982). A hierarchical model of housing choice. *Urban Studies, 19*(1), 17–31.

Böheim, R., & Taylor, M. P. (1999). *Residential mobility, housing tenure and the labour market in Britain* (ILR working papers 035). London: Institute for Labour Research.

Bridge, G. (2002). *The neighbourhood and social networks* (Centre for Neighbourhood Research, Paper, 4). CNR: London.

CCSR (The Cathie Marsh Centre for Census and Survey Research). (2010). *The Samples of Anonymised Records (SARS)* [online]. Accessed June 8, 2010. Available from http://www.ccsr.ac.uk/sars/2001/hhold/variables/

Cho, Y. (2004). Who moves and where? A comparison of housing association tenants in London and northern regions. In *Housing corporation sector studies 40*. London: Housing Corporation.

Cho, Y., Lyall Grant, F., & Whitehead, C. (2004). Affordable housing in London: Who expects to move and where? In *Housing Corporation Sector Studies 39*. London: Housing Corporation.

Dieleman, F. (2001). Modelling residential mobility: A review of recent trends in research. *Journal of Housing and the Built Environment, 16*(3), 249–265.

Gibbons, S., & Machin, S. (2003). Valuing English primary schools. *Journal of Urban Economics, 53*, 197–219.

Gjessing, M. (2009). *Interview with René Jordan*. December 11, 2009. Leeds.

Johnston, R. J., Poulsen, M. F., & Forrest, J. (2002). Are there ethnic enclaves/ghettos in English cities? *Urban Studies, 39*, 591–618.

Jordan, R., Birkin, M., & Evans, A. (2011). Agent-based simulation modelling of housing choice and urban regeneration policy. In T. Bosse, A. Geller, & C. Jonker (Eds.), *Multi-agent-based simulation XI* (pp. 152–166). Berlin: Springer.

Kim, J., Pagliara, F., & Preston, J. (2005). The intention to move and residential location choice behaviour. *Urban Studies, 42*(9), 1621–1636.

Laurie, A., & Jaggi, N. (2003). The role of 'vision' in neighbourhood racial segregation: A variant of the Schelling segregation model. *Urban Studies, 40*(13), 2687–2704.

Leeds City Council. (2007). *East and south east Leeds area action plan: Leeds local development framework*. Leeds: Leeds City Council.

Mullins, D., & Murie, A. (2006). *Housing policy in the UK*. Basingstoke: Palgrave Macmillan.

Pancs, R., & Vriend, N. (2007). Schelling's spatial proximity model of segregation revisited. *Journal of Public Economics, 91*, 1–24.

Peach, C. (1999). London and New York: Contrasts in British and American models of segregation. *International Journal of Population Geography, 5*, 319–351.

Phillips, D. (1998). Black minority ethnic concentration, segregation and dispersal in Britain. *Urban Studies, 35*(10), 1681–1702.

Phillips, D. (2007). Ethnic and racial segregation: A critical perspective. *Geography Compass, 1*(5), 1138–1159.

Power, A., & Mumford, K. (1999). *The slow death of great cities? Urban abandonment or urban renaissance*. York: York Publishing Services Limited.

Schelling, T. (1969). Models of segregation. *The American Economic Review, 59*(2), 488–493.

Schelling, T. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology, 1*(2), 143–186.

Strand, S. (2002). Pupil mobility, attainment and progress in Key Stage 1: A study in cautious interpretation. *British Education Research Journal, 28*, 63–78.

Tu, Y., & Goldfinch, J. (1996). A two-stage housing choice forecasting model. *Urban Studies, 33*(3), 517–537.

Tunstall, R. (2003). 'Mixed tenure' policy in the UK: Privatization, pluralism or euphemism? *Housing, Theory and Society, 20*, 153–159.

Yin, L. (2009). The dynamics of residential segregation in Buffalo: An agent-based simulation. *Urban Studies, 46*(13), 2749–2770.

Zhang, J. (2004a). A dynamic model of residential segregation. *Journal of Mathematical Sociology, 28*(3), 147–170.

Zhang, J. (2004b). Residential segregation in an all-integrationist world. *Journal of Economic Behavior & Organization, 54*(4), 533–550.

Chapter 26
Do Land Markets Matter? A Modeling Ontology and Experimental Design to Test the Effects of Land Markets for an Agent-Based Model of Ex-Urban Residential Land-Use Change

Dawn C. Parker, Daniel G. Brown, Tatiana Filatova, Rick Riolo, Derek T. Robinson, and Shipeng Sun

Abstract Urban sprawl is shaped by various geographical, ecological and social factors under the influence of land market forces. When modeling this process, geographers and economists tend to prioritize factors most relevant to their own domain. Still, there are very few structured systematic comparisons exploring how the extent of process representation affects the models' ability to generate extent and pattern of change. This chapter aims to explore the question of how the degree of representation of land market processes affects simulated spatial outcomes. We identify four distinct elements of land markets: resource constraints, competitive bidding, strategic behavior, and endogenous supply decisions. Many land-use-change models include one or more of these elements; thus, the progression that we designed should facilitate analysis of our results in relation to a broad range of existing land-use-change models, from purely geographic to purely economic and from reduced form to highly structural models. The description of the new agent-based model, in which each of the

D.C. Parker (✉) • S. Sun
School of Planning, Faculty of Environment, University of Waterloo, Waterloo, ON, Canada
e-mail: dcparker@uwaterloo.ca; s27sun@uwaterloo.ca

D.G. Brown
Center for the Study of Complex Systems, School of Natural Resources and Environment, University of Michigan, Ann Arbor, MI, USA
e-mail: danbrown@umich.edu

T. Filatova
Centre for Studies in Technology and Sustainable Development, University of Twente, Enschede, The Netherlands
e-mail: t.filatova@utwente.nl

R. Riolo
Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI, USA
e-mail: rlriolo@umich.edu

D.T. Robinson
School of Geosciences, University of Edinburgh, Edinburgh, UK
e-mail: derek.t.robinson@ed.ac.uk

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_26,

four levels of market representation can be gradually activated, is presented. The behavior of suppliers and acquirers of land, and the agents' interactions at land exchange are discussed in the presence of each of the four land-market mechanisms.

26.1 Introduction

Growing concern regarding the development of fragmented patterns of land conversion at the urban-rural fringe ("sprawl") has lead to the development of a wide variety of fine-scale spatial models of land-use change at the urban-rural fringe. Models developed using a wide variety of techniques, including those based on cellular automata (CA), neural networks, spatial econometrics, and agents, have successfully replicated the fragmented patterns of development that occur at the urban-rural fringe (Verburg et al. 2006).

These models differ greatly in terms of their level of detail and representation of real-world processes. At one end of the spectrum, CA models calibrated based on the historical spatio-temporal pattern of land-cover change can be characterized as highly inductive, pattern-based geographic approaches. At the other end of the spectrum, detailed agent-based models (ABMs) explicitly model socioeconomic processes, with correspondingly higher data demands for model parameterization and/or calibration. While these differences in modeling approaches are widely acknowledged, few structured comparisons have been undertaken to explore how the extent of process representation affects the models' ability to generate extent and pattern of change.

In this paper, we focus on a small subset of this spectrum of models – ABMs of land-use change at the urban-rural fringe driven by open space amenity values – to explore the question of how the degree of representation of land market processes affects model spatial outcomes. Land market factors such as credit availability, interest rates, the strength of demand relative to supply, and institutional details of land market function can be significant drivers of land-use change. In addition, interdisciplinary research is often conducted with the goal of producing policy recommendations for market-based mechanisms (e.g. subsidies, taxes, quotas, and insurance). Yet, these market factors have not been included in the majority of land-use change models. To our knowledge, few formal comparisons have been conducted to explore how the representation of land markets within land-use models affects projected land-use patterns.

We describe a series of model extensions to a simple land-change model with minimal market mechanisms that create a land-change model that has a simple but complete land market. We further describe a series of structured, comparative experiments that progressively introduce important aspects of land market interactions – including economic resource constraints, competition for land, strategic behavior, and endogenous land supply decisions – that seek to answer the question, "Do land markets matter?" for the spatial outcomes of land-use ABMs. In short, does a land-use change model that incorporates a process-based, land market model grounded in spatial economics produce more realistic spatial patterns of land development than a model based on reduced form representations of these influences? The paper elaborates the design of model versions with progressive introduction of market mechanisms. Future papers will present the experimental results.

We define a land market as the series of transactions and exchange of land between buyers and sellers in a bounded region. The number of buyers and the price at which they would acquire a given land parcel – their willingness to pay (WTP) – provide demand for the resource or product (i.e. land). The supply side of the market is defined by sellers' decisions to offer land for sale on the market and the price that they would accept for the land they are offering – their willingness to accept (WTA). The aggregate average price of land, the amount of land available (i.e. supply), the number of buyers and their willingness to pay (i.e. demand), the factors of production (i.e. inputs to the land and its biophysical and geographical characteristics), and the opportunity costs of taking part in land transactions versus other commodity transactions create what we call the land market and its associated dynamics over time. These land market dynamics also influence choices about factors of production (e.g. through land use and land management).

The following general conceptual questions frame our experimental approach:

- To what degree does the incorporation of constraints and competitive bidding (the focus of our first set of experiments) alter development patterns, gains from trade (the difference between WTP and WTA), and agent utility?
- Do models that exclude market mechanisms include sufficient proxies for market mechanisms to be considered reduced-form versions of fuller models, replicating results of fuller models in many circumstances? Or rather, do these models exclude important processes that influence land market outcomes in a significant way?
- Even when market outcomes are modeled, many economic models are forced to make simplifying assumptions, such as agent and environmental homogeneity, modeling only transaction prices rather than WTA/ask and WTA/bid price formation, and not modeling strategic behavior, for the sake of analytical tractability. What are the implications of these simplifying assumptions for the ability of these models to project the extent and pattern of land-use change?
- To what extent do the effects of including or excluding market processes depend on the particular socioeconomic circumstances modeled? More specifically, are there some sets of parameter settings for which inclusion of market mechanisms has a large effect, and some for which effects are relatively small? If so, what would be the real-world interpretation of the conditions that these parameter settings represent?
- How does the incorporation of heterogeneity (i.e. in agent preferences and resource constraints) and level of information (i.e. the number of sites evaluated) influence these outcomes?

This modeling exercise is part of the SLUCE II project, an interdisciplinary, multi-university project funded by the US NSF Dynamics of Coupled Natural-Human Systems program. Our new model builds from two existing models by including new economic elements in each. The first, ALMA (Parker and Filatova 2008; Filatova et al. 2009a, b), focuses on land market interactions and the microeconomic determinants of WTP and WTA. The second, SOME (Brown and Robinson 2006; Brown et al. 2008; Robinson and Brown 2009), uses survey and spatial data to develop empirically-based utility/suitability measures for residential agents, and then uses these measures to sequentially allocate land-use change events in the landscape, in the tradition of spatial statistical modeling. Our new model builds from these two existing models, expanding on both existing models to include additional economic elements.

26.2 Conceptual Overview

Our initial goal for this effort was to compare a land-use change model with and without a “land market,” with the idea that a land market is a single, comprehensive concept that can be switched on or off. However, given the diversity of modeling approaches and institutional environments in which land is traded globally, it quickly became clear that “land market” does not have a single, comprehensive definition. Our discussions led us to break down the operation of land markets into several distinct elements that progressively add four important aspects of markets: *Resource Constraints, Competitive Bidding, Strategic Behavior,* and *Endogenous Supply Decisions*. Many land-use-change models include one or more of these elements; thus, the progression that we designed should facilitate analysis of our results in relation to a broad range of existing land-use-change models, from purely geographic to purely economic and from reduced form to highly structural models. We describe four levels of representation for modeling land markets (Table 26.1) and use these levels to design experiments to explore the answers to our questions (above).

The contents and motivation for each of these levels are as follows.

- **Level 0:** Level 0 is essentially a “first-come, first-served” sequential allocation model, i.e. a demand-driven model. A new land manager (or in many cases, simply a new land use) is selected for the parcel based on a utility ranking or suitability score, with parcels with highest utility or suitability selected first. This utility function reflects preferences for land attributes, a key building block of

Table 26.1 Degrees of market representation: model levels and their definitions

Level 0	Level 1	Level 2	Level 3	Level 4
No LM	Add resource constraints	Add competition	Add strategic behavior	Add endogenous supply decisions
No resource constraints, competitive bidding, strategic behavior, or endogenous supply decisions	Level 0 plus resource constraints for buyers and sellers	Level 1 plus allocation via competitive bidding	Level 2 plus strategic bid/ ask price formation	Level 3 plus modeled decision to sell rural parcel

demand in any market model. However, the utility function is the only "market" element in the model. Sequential allocation models are generally constrained by a top-down quantity of change, either total or categorical, although models can also be driven by a total population that needs to be allocated across the landscape. This type of land allocation mechanism is used with a variety of parcel-scale land-use-change models, including CA, statistical regression models, and ABMs (Verburg et al. 2006).

- *Limitations*: Level 0 models generally do not explicitly incorporate the damping and sorting effects of economic resource constraints. A corollary to this point is that these models lack any explicit land prices, even exogenous prices. They may also assume unlimited land acquisition budgets on the part of acquiring agents. As a result, economically implausible land transactions may occur, such as land being acquired by an agent who in reality lacks the economic resources to accomplish that acquisition. In the case of several land- use types, a vacant parcel can be occupied by a land use that would be outcompeted in reality by a higher-value use.
- **Level 1:** The level 1 model overcomes the limitations found in Level 0 by adding *parametric, exogenous buyer and seller land values and land budgets* to the model. These limit the ability of a buyer to acquire the highest utility parcel. A buyer can then acquire a parcel only if the parcel is affordable under her budget constraint, and a seller will accept her bid only if it is higher than his WTA. Thus, each transaction, in theory, will generate positive or neutral gains from trade.
- *Limitations*: In the level 1 model, although the acquiring agent may be able to afford the parcel, that parcel may be of higher value to another agent. The Level 1 model does not allow for competitive bidding, thus potentially preventing the higher-valuing agent from acquiring the parcel. As a result, land may not be allocated to the highest privately valued use. Although in the real world a variety of factors might mean that land is not necessarily allocated to its highest valued use, if a land market were allowed to operate, an agent with a higher value should, in theory, outbid an agent with a lower value.
- **Level 2:** The level 2 model allocates parcels via *competitive bidding*, rather than sequentially, giving the short-term opportunity for an agent with the highest valuation to acquire the parcel. This competitive bidding process also creates an endogenous land price or land rent, one that in theory reflects the highest valued use of the land. This approach is taken in several agent-based models and analytical models of land markets (see Parker and Filatova (2008) for a recent review).
- *Limitations*: In the level 2 model, buyers and seller each reveal their true valuation for the parcel (their WTP and WTA). In real land markets, these values are carefully guarded, and relative bargaining power may influence actual bid and ask prices, final transaction prices, and the final distribution of gains from trade.
- **Level 3:** The Level 3 market model adds bid and ask prices to the level two model. Once competitive bidding is introduced into a market model, buyers and sellers have an incentive to behave strategically in order to capture the highest possible amount of surplus (the difference between a seller's willingness to accept and a buyer's willingness to pay). This strategic behavior is expressed through setting of ask prices and bid prices that are respectively higher and lower than the

sellers' and buyers' WTA and WTP. The ways in which these strategic decisions are made in land markets have not been thoroughly investigated as of yet, but such decisions are clearly a function of expectations regarding future trajectories of land prices and the participation decisions of other agents in the land market. Parker and Filatova (2008) lay out possible influences on such expectations, and Filatova et al. (2009a, b) implement a simple version based on the proportion of buyers and sellers active in the market.

- *Limitations*: Although the Level 3 model allocates land to its highest valued use, it does not allow feedbacks between this highest valued use and the supply of economically scarce land for conversion to the market. Just as market conditions may lead to strategic setting of bid and ask prices, market conditions may also lead to strategic incentives regarding when to supply a parcel to the market.
- **Level 4:** The level 4 model *endogenizes land supply decisions*. Expectations regarding sales prices have a strong influence on the decision to supply land to a market, and thus, the probability of finding a parcel on the market will be higher closer to the city center, where property values are higher.

26.3 Behaviors of Suppliers and Acquirers of Land in Each Model

We now describe the detailed assumptions for the behavior of suppliers of land, acquirers of land, and land-exchange mechanisms for each of the models. We follow a slightly modified version of the MR POTATOHEAD template for land-use-change models described in Parker et al. (2008). Each models' mechanisms are summarized in Tables 26.2–26.4. MR POTATOHEAD terminology is in *italics*. Our experimental design strives, as much as possible, to keep most of the elements the same between model levels, and progressively changes one or few elements at each level.

26.3.1 Suppliers of Land

The *suppliers of land* (Table 26.2) in our first, simple model implementation are rural sellers, who are assumed in each case to put the single parcel that they own up for sale.

For Levels 0–3, the *motivation for supply* is not explicitly modeled. Motivation for supply also essentially describes the *event sequencing/triggers for land transfer* that are part of the Exchange Rules section of the Land Exchange class of the MR POTATOHEAD model. A simple rule, consistent with the approach taken in the base versions of the SLUDGE, SOME, and ALMA models, will be used to determine which rural parcels are put on the market in each time period in Levels 0–3.

Table 26.2 Suppliers of land details for different levels of market representation

MP element	Agent/role/scale	No land market (Level 0)	Add resource constraints (Level 1)	Add competition (Level 2)	Add strategic behavior (Level 3)	Add endogenous supply decision (Level 4)
Suppliers of land	Rural owner/ seller/parcel	(No resource constraints, comp. bidding, or strategic behavior)	(Level 0 plus resource constraints for buyers and sellers)	(Level 1 plus allocation via competitive bidding)	(Level 2 plus strategic bid/ask price formation)	(Level 3 plus decision to sell parcel is modeled)
Motivation for supply		All/random (potentially dependent on distance to center)	Same as level 0	Same as levels 0 and 1	Same as levels 0, 1, and 2	Decision to supply based on expected price or price differential between WTA and expected bid price (endogenizes location differential)
Parcels supplied		Entire parcel	Entire parcel	Entire parcel	Entire parcel	Entire parcel
Resource constraint		0	In WTA	In WTA	In WTA	In WTA
WTA		Opportunity cost- exogenous parameter, set to zero	Opportunity cost- exogenous parameter, set with a positive value	Opportunity cost- exogenous parameter, set with a positive value	Opportunity cost- exogenous parameter, set with a positive value	Opportunity cost- exogenous parameter, set with a positive value
Terms offered		Opportunity cost (Ask = WTA, zero)	Opportunity cost (Ask = WTA, positive)	Opportunity cost (Ask = WTA, positive)	Ask price – endogenous dependent on price/ market expectations	Ask price – endogenous dependent on price/ market expectations

Table 26.3 Acquirers of land details for different levels of market representation

MP element	Agent	No land market (Level 0)	Add resource constraints (Level 1)	Add competition (Level 2)	Add strategic behavior (Level 3)	Add endogenous supply decision (Level 4)
Acquirers of land	Resident/buyer/parcel	(No resource constraints, comp. bidding, or strategic behavior)	(Level 0 plus resource constraints for buyers and sellers)	(Level 1 plus allocation via competitive bidding)	(Level 2 plus strategic bid/ask price formation)	(Level 3 plus decision to sell rural parcel is modeled)
Motivation for acquisition		Incoming buyer	Incoming buyer	Incoming buyer	Incoming buyer	Incoming buyer
Proximity influence		Disutility of commuting	Disutility of commuting, travel cost	Disutility of commuting, travel cost	Disutility of commuting, travel cost	Disutility of commuting, travel cost
Neighborhood influences		Open space amenities from undeveloped land	Open space amenities from undeveloped land	Open space amenities from undeveloped land	Open space amenities from undeveloped land	Open space amenities from undeveloped land
Parcels hope to acquire		Max utility from sub-sample	Max utility from sub-sample	Max utility from sub-sample	Max utility from sub-sample	Max utility from sub-sample
Resource constraint		None	Housing/transport budget	Housing/transport budget	Housing/transport budget	Housing/transport budget
WTP		Set to zero	Function of utility and budget constraint	Function of utility and budget constraint	Function of utility and budget constraint	Function of utility and budget constraint
Terms offered		Bid = WTP	Bid = WTP	Bid = WTP	Strategically set bid price	Strategically set bid price

Table 26.4 Initial conditions and land exchange mechanisms for different levels of market representation

MP element	Agent	No land market (Level 0)	Add resource constraints (Level 1)	Add competition (Level 2)	Add strategic behavior (Level 3)	Add endogenous supply decision (Level 4)
Land exchange	Initial sellers	All	All	All	All	Endogenously determined
	Initial buyers	Constrained to Level 2 total	Constrained to Level 2 total	Endogenously determined	Endogenously determined	Endogenously determined
Allocation mechanism		Sequential allocation	Sequential allocation	Competitive bidding	Competitive bidding	Competitive bidding

Initially, all rural parcels are initially available for acquisition.[1] The Level 4 model will endogenize the supply decision, with the potential supplier of land forming an estimate of the price at which they are likely to sell their parcel. If the difference between this expected price and the potential seller's WTA reaches a certain threshold, the parcel will be put up for sale. In each case, the *parcels supplied* will include the entire rural parcel (no subdivision), although in later work, rural parcels can be subdivided by developers.

Suppliers of land for the Level 0 model will have no *resource constraint*, where their resource constraint sets a minimum threshold for compensation for their parcel. In this first set of experiments (Level 1), we essentially define this constraint in terms of a willingness to accept (WTA), highlighting the fact that in the Level 0 model, their parcel will be supplied without compensation. In economic terms, this basically assumes no land scarcity; an economically scarce resource is defined as one for which, if the resource were available without cost, more would be demanded than is supplied. Again to keep our experiments simple, we assume that the WTA is defined by a parametrically set opportunity cost; the value of the land in its current use.[2] This value will be set to zero for the Level 0 model.

For Levels 0–2, the *terms offered* (or ask price in standard economic terminology) will simply be equal to the WTA. The actual compensation received may still be above the WTA of the supplier, as the differential will depend on the bidding rules and the number of buyers. For Levels 3–4, the ask price will be strategically set based on the expectations of market conditions, following Parker and Filatova (2008) and Filatova et al. (2009a, b).

26.3.2 Acquirers of Land

At each level, the *acquirers of land* (Table 26.3) will represent new buyers entering the region. Each will seek to acquire a single residential parcel (their *motivation for Acquisition/trigger for market participation*).

In all models, acquirers of land will base their decision to acquire land on a simple utility function, whose representation will stay fixed between experiments. They will gain utility from two factors: the *proximity influence* as a function of distance to a city/service center and the *neighborhood influences* of open-space amenities, which, for simplicity, will be generated by undeveloped open space in the local neighborhood of the parcel.[3] For each case, the *parcel they hope to acquire*

---

[1] We debated whether or not to make parcel availability stochastic, with the probability of parcel supply being higher for parcels closer to the city center. We may explore the effects of these alternative algorithms in future work.

[2] In later versions of the Level 3 model, the WTA could also be a function of additional resource constraints (credit and debt constraints) and of expected sales price.

[3] Later versions of the models will have much more sophisticated definitions of open space amenities based on land management as well as land use.

will be the highest utility parcel based on this utility function. In some experiments, agents will have heterogeneous preferences, so their utility for the same parcel will potentially differ. For model tractability, it is possible that this highest utility parcel may be selected from a sub-sample of all available parcels, constrained by budget if applicable.

In the Level 0 model, acquirers of land will face no *resource constraints*; they will be able to acquire their highest utility parcel at no cost. They will also face no constraint on their transportation budget. In Level 1–4 models, acquirers of land (buyers) will have both housing and transportation budget constraints, following Filatova et al. (2009a, b). Total income and transport costs will be set parametrically and varied experimentally. It follows that in the Level 0 model, acquirers' willingness to pay will be set to zero to reflect the fact that they are assumed to be able to acquire any parcel at no cost. For Levels 1–4, the willingness to pay will be a function of the utility gained from the parcel and the budget constraint. To avoid the need to model price expectations in the Level 1–2 models, the functional transformation used in Filatova et al. (2009a, b) will be used to create a WTP function that has the behavioral properties of a standard economic demand function.

The *terms offered* (bid price, in economic terms) for Levels 0–2 will simply be the WTP, in parallel with the setting of ask = WTA for suppliers of land. Again, in cases where the acquirers' WTP is above the suppliers' WTA, the actual transaction price may lie below the WTP. Bid prices at Levels 3–4 will be strategically set, dependent on market conditions, and will follow the approaches outlined in Parker and Filatova (2008) and Filatova et al. (2009a, b).

Model initialization and exchange rules: Here we describe some key elements of the various levels in the form of both the *Exchange Rules* subclass of the *Land Exchange* and *Model Operation* classes in MR POTATOHEAD (Parker et al. 2008).

Initial agent numbers, types, and locations: As mentioned above, we will initialize our landscape with a single active seller located on each parcel. Our initial experiments will focus on comparison of the Level 0–2 models. One of the biggest challenges in model design was the decision of what macro-scale constraint would drive land development in the model. In the real world, land development is driven by a combination of such factors as migration and changes in employment and population structure. In land-use modeling, these influences are often represented by proxies that make assumptions regarding a quantity of cells/parcels that should change land in each time period. In an ABM, these influences could be represented through the assumption that a fixed number of agents enter (or leave) the market in each time period. In our longer-term modeling, we intend to have progressive in-migration of new buyers, at rates consistent with the land change dynamics of our study area, and we also intend to endogenously model relocation decisions of currently settled agents (see "extensions," below). However, in order to be able to draw a broad set of conclusions, relevant to the many other previously developed land-use change models that fall under our different model levels, the first version will fix some concept of "quantity of change" in order to facilitate comparison between models.

Executing experiments with comparable levels of change is difficult because one of the major differences between models with sequential allocation vs. land market exchange is whether they have a fixed vs. endogenous quantity of change. A first-come/first-served allocation method, such as in Level 0, requires that the quantity of change be limited, or the entire landscape will be converted. In contrast, a model with resource constraints (Level 1 and above) will limit conversion to only those parcels where the willingness to pay of the buyer is above the willingness to accept of the seller. With positive transport costs, these constraints will also lead to clustering of development around city/service centers. Once competitive bidding is introduced, the process of land-use change will be further constrained, because a buyer not only needs a higher level of resources than a seller, but also a higher bid than other potential buyers who strive to acquire the same parcel.

The implication of these differences in model mechanisms is that, for a fixed population of agents who differ only in their resource constraints, (i.e., moving from a Level 0 to a Level 1 model), less land-use conversion should occur in the Level 1 model than the Level 0 model. Furthermore, even less land conversion should occur in the Level 2 model than the Level 1 model. (Filatova et al. (2009a, b) demonstrate that more conversion may occur when strategic bidding is introduced). We will find appropriate parameter settings for the model (a homogeneous total population less than the number of cells on the landscape, and a combination of utility and budget parameters that imply that not all agents in the Level 2 model will seek to buy) in order to run a baseline model that verifies these results. If our model behaves as we anticipate, we will then fix the number of agents in at each level to an amount that produces the same number of land exchange events. This means that the number of participating agents may be different at each level, but that the total amount of land conversion will be the same. This will allow us to run experiments that examine the effects of the model runs on the extent and pattern of land conversion, holding the number of converted cells fixed.

Land allocation mechanisms, event sequencing, and scheduling: For the Level 0 model, acquirers of land will be allowed to sequentially select and acquire their most preferred parcel. For the Level 1 model, this acquisition will be limited to the parcels that are affordable under the buyer's budget constraint, accounting for both the purchase price (the WTA of the seller) and the transport costs to city/service center. In each case, for any model run in which there are positive open space amenities and/or any heterogeneity in agent characteristics or resources, multiple model runs with different stochastic draws need to be run to account for stochastic elements. For the Level 2 model and above, initially all buyers will put a bid on their highest valued parcel, and sellers will then review bids and accept the highest valued bid, if it lies above their WTA. Buyers who do not succeed in acquiring a parcel, and sellers who do not succeed in selling their parcel, will participate in a next round of trade. Rounds of trade will continue until no more trades occur. Again, multiple model runs will be needed in most cases, since different agents may have equal utility and WTP for a given parcel.

Experiments: Our initial experiments will compare the Level 0–2 models. Our first goal is to identify sets of parameter values that demonstrate the extremes of model outcomes. In other words, we will search for a set of parameter values that lead to the smallest effects of including market mechanisms as well as those that lead to the largest effects. Consistent with our previous work, we plan experiments that vary distributions of agent characteristics in terms of preferences for proximity and open space amenities and resource constraints.

Although we plan to run baseline models that set open space amenities to zero for verification purposes, we are essentially interested in land-use change models that explore the effects of open space amenities on the pattern and extent of land conversion in ex-urban settings. Given the wide variety of models that explore similar questions that have been developed using a variety of modeling methods (including those based on CA, spatial econometrics, neural nets, and agents), we believe that a set of experiments that incorporates open-space amenities will still be quite generally informative with respect to investigating the importance of land market mechanisms on land-use change models of this type. Therefore, the bulk of our experiments will include positive open space amenities.

Hypotheses: How do we expect inclusion of the land market mechanisms to affect patterns? We are still developing these hypotheses, and of course, one reason for building simple simulation models of complex systems is to help develop theoretical hypotheses for systems for which simple intuition and/or mathematics fail. However, an initial hypothesis is that a model that excludes market mechanisms (Level 0) may predict more expansion and sprawl than a model that includes them. Including representations of positive open space amenities, disamenities from commuting without transport costs, and some prior development, a parcel that is relatively distant from the city center will provide relatively higher utility than a closer in parcel, which will likely have more highly developed neighborhood density. If an acquirer of land is not constrained by a housing or transport budget, they will easily acquire that higher utility parcel. Having located there, they then decrease the utility of that location for another potential resident with a high preference for open-space amenities, leading to path dependence in which the next acquirer occupies a parcel even further out than they would have had they not been able to acquire that parcel.

This path-dependent, leapfrog-generating location incentive will be present in any of our model runs that have positive open-space amenities. However, our hypothesis is that the pull towards the city center – whether from transportation cost constraints or from the need to outbid other buyers – will be stronger for the model that include market mechanisms. Thus, the constrained development path may be progressively more compact for the Level 1 and 2 models. This result, however, could be dependent on relative parameter values.[4] It may also no longer hold, or may be dampened, in models that allow endogenous relocation by residents (see Sect. 26.4). Finally, based on our previous work with heterogeneous agents, we anticipate that differences between model outcomes will be magnified as the degree

[4]Note that this hypothesis could break down if buyers had sophisticated expectations regarding future paths of development. However, in a complex environment, even the most intelligent boundedly rational agent would likely fail to anticipate exact future patterns of local development. Modeling of such expectations, in any case, will be an interesting topic for future work.

of agent heterogeneity – in terms of preferences for open space amenities and proximity and resource constraints – increases.

A second set of questions relate to the degree to which the modeled landscape produces an economically efficient allocation, where economic efficiency is measured by the sum of economic surplus (the difference between WTP and WTA) generated by the landscape. Economic efficiency is, in theory, characterized by Pareto optimality, under a very narrow set of conditions that include a "no externality" condition. When open-space amenities are present, every landscape pattern/quantity allocation outcome will be characterized by a potentially different pattern of external costs and benefits. Thus, for the majority of our experiments, we do not expect the land market allocation to be economically efficient.

Yet, given the limitations of models that omit market mechanisms that we discuss above, we are interested in the economic surplus generated in each of our experimental outcomes, since it should reflect the success of the competitive allocation algorithms. We plan to evaluate the economic efficiency of each landscape outcome for a baseline, no open space amenity model, relative to a baseline random allocation model. Since calculations of economic efficiency depend on WTP and WTA, economic efficiency for each outcome will be calculated using the WTP and WTA for the level 2 model. One hypothesis is that, in terms of relative orders of magnitude, the level 0 model (which bases allocation on utility-metric preferences) will lead to the highest relative increase in the economic efficiency of the generated landscapes, relative to the level 1 and 2 models. A counter hypothesis is that only the level 2 model, which most closely resembles the traditional market models on which economic efficiency theorems are based, will lead to a significantly more efficient landscape. From the perspective of economics, these comparisons will shed light on the question of whether land markets matter from a formal theoretical perspective.

26.4 Model Extensions

We have described a series of incremental steps to add market mechanisms (Levels 0–4) and evaluate the effects of including a market on spatial development patterns. A number of additional mechanisms could be added to (1) extend the levels of market complexity and (2) include additional mechanisms that may alter the effects of our current set of market mechanisms. In this section we focus on the second of these two types of extensions. We provide a brief discussion on endogenous price expectation in supply that could lead to relocation by residents and the incorporation of developers into the land-change system as mechanisms that could alter our model behavior and provide increased realism desired by policy and decision makers.

Endogenous relocation by residents: Spatial economics suggests that there are several main reasons for the migration and relocation of households (Clark and van Lierop 1986; van der Vlist et al. 2002; Clark et al. 2003). Employment opportunities elsewhere are a main driver of inter-urban migration. Intra-urban migration occurs when households become dissatisfied with the neighborhood or home they live in and they find more attractive housing options elsewhere (potentially due to a change in life-cycle stage). Housing bundle theory identifies three components that influence the attractiveness of a particular property for a buyer: (1) housing structure (2) neighborhood quality, which includes both social and environmental components and (3) accessibility to public and private services (influenced by transport costs and geographic/institutional restrictions) (Adams 1984). The relocation process is largely determined by the demand and supply of these components. Relocation creates two important feedbacks. In the short run, relocation can change neighborhood quality; and in the long run, it also changes the quality and cost of public services. Given the focus of our modeling work on open-space amenities, we are most interested in how these short-run feedbacks may trigger a subsequent cascade of endogenous relocation.

Although a lack of endogenous relocation has been put forward as a criticism of land change models that lack a land market (Polhill et al. 2005), endogenous relocation can be modeled even in the absence of resource constraints, competitive bidding, and strategic behavior. If new residents influence the quality and character of natural or neighborhood amenities, then the utility/value that an agent holds at a location may change. In a non-market context, an agent will have a utility-based incentive to move. In a market context, a parcel that was initially allocated to its highest valued use may no longer be.

In later work we will evaluate the influence of endogenous relocation on model output for each of the model levels described above. For Levels 0 and 1, agents will be able to relocate when the expected utility (constrained by budget where relevant) of relocation exceeds the expected utility of remaining on the current parcel. For Level 2 and beyond, this relocation would, under most circumstances, be contingent on the ability of the current agent to sell the current parcel and make a gain from trade. Endogenous relocation should facilitate evolution of a more dynamic landscape. For the market model variants, it will allow the highest valued use to evolve over time.

We hypothesize that the inclusion of endogenous relocation may increase residential sprawl. Preferences for natural amenities by households have been increasing over time as evidenced by the rate of exurban development, which has outpaced that of population growth in the conterminous USA between 1980 and 2000 (Theobald 2005). We anticipate that as neighborhood density increases, there will be increasing incentive for agents to relocate to locations with higher in natural amenities or to locations that have socio-economic characteristics that are more preferable but are constantly changing over time.

Modeling developers: Conventional urban economic models typically assume that agricultural land is transferred to households directly (Alonso 1964). In some cases, these models omit the direct modeling of agricultural land owners and focus instead on transactions between developers and households (Henderson and Thisse 1999). In others, the transactions occur between agricultural sellers and developers

(Asami and Teraki 1991). Rarely are all three actors, i.e. agricultural owners, developers and households explicitly represented in a spatially explicit model.[5] Perhaps the largest void exists with respect to research on developers. However, developers significantly influence land change. It is the developer who, in a free market, generally determines which agricultural lots to convert to residential land use, at what structural density (subject to government constraints) and at what price to offer parcels in the land market. These decisions and the market transactions between developers and farmers, and further between developers and residential households are important processes influencing land market dynamics (land patterns and land prices). As a next step of model development we plan to include developer agents in the land market. As with other agents we would like to consider three levels of economic behavior of a developer based on two criteria: resource constraints and land acquisition via competitive bidding. In future work we will evaluate the influence of developers on model output for each of the model levels (0–4) described above.

26.5 Discussion and Conclusions

Our focus has been on evaluating the incremental inclusion of land-market mechanisms, using a suite of ABMs, on spatial settlement patterns and market dynamics. ABMs of land-use use virtual agents to provide computational representation of the actions and decision-making strategies used by real-world actors. The forms of interaction among agents in the absence of a market are typically through substitution (i.e. the acquisition of one property alters the selection choices available to other agents), by constraint (i.e. a township invokes a land-use policy that excludes a specific type of development action from occurring), or through neighborhood effects (i.e. the evaluation of a settlement location by residential household agents involves comparing its location preferences with those of possible neighbors at the evaluated site). However, with the inclusion of market mechanisms, the degree of interaction is increased through competitive bidding, strategic behavior, and endogenous supply decision making. Anderson (1972) notes that 'more is different' with respect to the degree of interaction among agents. Therefore we speculate that through the increasing degrees of interaction brought about by market mechanisms, our results will illustrate that markets do influence settlement patterns.

Ultimately we are interested in the role of land development dynamics and patterns on ecosystem function(s) through land-use and land-cover change. Several market interactions are relevant to this question. First, land cover patterns intervene in the

[5] Analytical non-spatial models that account for the behavior of all three categories in a land market exist (Kraus 2006). However, within each group (developers, agricultural land owners and households) all agents are assumed to be homogeneous, perfectly rational and, with constant returns to scale. Space is assumed to be homogeneous except for the distance to the center, and enters economic models as travel costs and amount of spatial good acquired (sq foot). Location specifics and neighborhood externalities remain unexplored.

choices of land for development and residence, through the individual preferences of residents for particular landscape characteristics and the perception of those preferences by developers. Second, to the degree that residents and developers are concerned about the market value of their land, residents' and developers' perceptions of the influence of landscape characteristics on the choices of other residents could influence their choices about landscape management activities, regardless of their own landscape preferences. In later stages of our modeling, we will explore such questions in detail.

Acknowledgements Funding from NWO-ALW(LOICZ-NL) Project No. 014.27.012 and the U.S. National Science Grants No. 041406, 813799, and 0119804 is gratefully acknowledged.

References

Adams, J. S. (1984). The meaning of housing in America. *Annals of the Association of American Geographers, 74*(4), 515–526.

Alonso, W. (1964). *Location and land use: Toward a general theory of land rent*. Cambridge: Harvard University Press.

Anderson, P. W. (1972). More is different. *Science, 177*(4047), 393–396.

Asami, Y., & Teraki, T. (1991). On sequential negotiation procedures: Optimal negotiation orders and land prices. *Regional Science and Urban Economics, 20*(4), 537–556.

Brown, D. G., & Robinson, D. T. (2006). Effects of heterogeneity in residential preferences on an agent-based model of urban sprawl. *Ecology and Society, 11*(1), 46.

Brown, D. G., Robinson, D. T., An, L., Nassauer, J. I., Zellner, M., Rand, W., Riolo, R., Page, S. E., Low, B., & Wang, Z. (2008). Exurbia from the bottom-up: Confronting empirical challenges to characterizing a complex system. *Geoforum, 39*(2), 805–818.

Clark, W. A. V., & van Lierop, W. F. J. (1986). Residential mobility and household location modelling. In P. Nijkamp (Ed.), *Handbook of regional and urban economics* (pp. 97–132). Amsterdam: Elsevier.

Clark, W. A. V., Huang, Y., & Withers, S. (2003). Does commuting distance matter? Commuting tolerance and residential change. *Regional Science and Urban Economics, 33*(2), 199–221.

Filatova, T., Parker, D., & van der Veen, A. (2009a). Agent-based urban land markets: Agent's pricing behavior, land prices and urban land use change. *Journal of Artificial Societies and Social Simulation*, *12*(1), 3, http://jasss.soc.surrey.ac.uk/12/1/3.html

Filatova, T., van der Veen, A., & Parker, D. C. (2009b). Land market interactions between heterogeneous agents in a heterogeneous landscape: Tracing the macro-scale effects of individual trade-offs between environmental amenities and disamenities. *Canadian Journal of Agricultural Economics, 57*(4), 431–457.

Henderson, J. V., & Thisse, J. F. (1999). On the pricing strategy of a land developer. *Journal of Urban Economics, 45*(1), 1–16.

Kraus, M. (2006). Monocentric cities. In R. Arnott & D. McMillen (Eds.), *A companion to urban economics* (pp. 96–109). Malden, USA: Blackwell.

Parker, D. C., & Filatova, T. (2008). A conceptual design for a bilateral agent-based land market with heterogeneous economic agents. *Computers, Environment and Urban Systems, 32*(6), 454–463.

Parker, D. C., Brown, D. G., Polhill, J. G., Deadman, P. J., & Manson, S. M. (2008). Illustrating a new "conceptual design pattern" for agent-based models of land use via five case studies – the MR POTATOHEAD framework. In A. L. Paredes & C. H. Iglesias (Eds.), *Agent-based modeling in natural resource management* (pp. 22–53). Valladolid: INSISOC.

Polhill, J. G., Parker, D. C., & Gotts, N. M. (2005). *Introducing land markets to an agent based model of land use change: A design*. Paper read at representing social reality: Pre-proceedings of the third conference of the European Social Simulation Association, Koblenz, Germany.

Robinson, D. T., & Brown, D. G. (2009). Evaluating the effects of land-use development policies on ex-urban forest cover: An integrated agent-based GIS approach. *International Journal of Geographical Information Science, 23*(9), 1211–1232.

Theobald, D. M. (2005). Landscape patterns of exurban growth in the USA from 1980 to 2020. *Ecology and Society, 10*(1), 1–32.

van der Vlist, A., Gorter, C., Nijkamp, P., & Rietveld, P. (2002). Residential mobility and local housing-market differences. *Environment and Planning A, 34*(7), 1147–1164.

Verburg, P. H., Kok, K., Pontius, G. R., & Veldkamp, A. (2006). Modeling land-use and land-cover change. In E. F. Lambin & H. J. Geist (Eds.), *Land-use and land-cover change: Local processes and global impacts* (pp. 117–135). Berlin: Springer.

Chapter 27
Exploring Coupled Housing and Land Market Interactions Through an Economic Agent-Based Model (CHALMS)

Nicholas R. Magliocca

Abstract Land markets are characterized by spatially distributed exchanges of heterogeneous goods and decision-making by heterogeneous, adaptive participants. Land market dynamics influence and are influenced by spatially varying demands for residential housing through housing markets. This chapter describes a spatially disaggregated, economic agent-based model for exploring ex-urban growth patterns emerging from coupled interactions between housing and land markets (CHALMS). CHALMS simulates the conversion of farmland to housing development over time, through the actions of the agents in the land and housing markets. Three types of agents—consumers, farmers and a developer—make decisions based on microeconomic principles, and use stylized expectation formation models to adapt to dynamic market conditions. The location, price, and density of housing are represented explicitly, as are the location, price, and productivity of individual farms. The possibility of many possible system states, due to agent and landscape heterogeneity, stochastic processes, and path-dependence, requires multiple model runs, as does the expression of the spatial distribution of housing types, overall housing density, and land prices over time in terms of the most likely, or 'average', patterns. CHALMS captures stylized facts of diminishing population density and land prices at greater distances from the center city, increasing land prices over time, and dispersed leapfrog patterns of development evident in most suburban areas of the U.S.

N.R. Magliocca (✉)
Department of Geography and Environmental Systems, University of Maryland, Baltimore, MD, USA
e-mail: nmag1@umbc.edu

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_27,

27.1 Introduction

As urban sprawl and other undesirable development patterns become more prevalent, policy-makers and researchers alike are coming to grips with the complexity of forces that generate such patterns (Atkinson and Oleson 1996; Brown et al. 2005; Brown and Robinson 2006; Magliocca et al. 2011). Land-use patterns on the urban fringe emerge from many individual landowners' conversion decisions in response to changing economic opportunities and landscape features. With such complexity, land-use change simulation models have become valuable tools to understand processes of land conversion and development, and for analyzing the effects of land use policies. Various modeling methodologies have been applied in a wide range of disciplines, such as urban planning, landscape ecology, geography, and economics, to build the current understanding of land-use change (Irwin 2010; Parker et al. 2003; Veldkamp and Verburg 2004).

However, most models represent either spatially detailed development patterns or individual-level decision-making—rarely integrating both elements explicitly. For example, models that provide spatially explicit representations of land-use patterns may lack an equally rigorous representation of agent decision-making processes (Fernandez et al. 2005; Parker and Filatova 2008; Parker et al. 2012). Models that take into account microeconomic agent decision-making, on the other hand, may fail to capture the full heterogeneity of those agents and inadequately describe spatial characteristics of model outcomes (Irwin 2010).

This chapter describes the Coupled Housing and Land Markets model—CHALMS. It is an economic agent-based model (ABM) of housing and land markets that captures the conversion of farmland to residential housing of varying densities over time in a hypothetical, growing ex-urban area. The primary goal of this study is to develop some general theoretical insights into the individual-level processes that drive regional development patterns. CHALMS is unique among ABMs of land-use for its integration of: (1) microeconomic decision-making rules for consumer, farmer, and developer agents in a spatially explicit framework; (2) representation of heterogeneous agent characteristics and spatial goods (e.g. land productivity and housing sizes and densities); and (3) direct linkages between adaptive price expectations and demand and supply decisions of developer and farmer agents through housing and land markets. The model demonstrates how patterns in development density and land prices predicted by traditional urban economic theory can be reproduced in an ABM framework. In addition, it shows how disconnected, leapfrog development patterns emerge from the simulation of individual price expectations and market transactions.

Section 27.2 reviews the capabilities and limitations of current land-use modeling approaches, and describes how integrating the insights from recent economic and non-economic ABMs can provide a more complete representation of the processes driving urban growth. Section 27.3 details the structure of CHALMS, agent representations, and market interactions. Section 27.4 presents baseline results and outcomes of preliminary sensitivity analysis. Finally, Sect. 27.5 concludes with a discussion of model capabilities and limitations and directions for future research.

27.2 Some Previous Land-Use Modeling Approaches

27.2.1 Spatial Equilibrium Economic Models

Economic models of urban land use are typically built on the assumption of spatial equilibrium. These models assume that over the long run housing rents will reach equilibrium and offset differences in spatially heterogeneous attributes such as transportation costs to the central business district (CBD), neighborhood amenities, and access to employment. Early models in the urban economics literature used a monocentric city framework in which location is defined purely by distance to a CBD where all jobs are located (Alonso 1964; Muth 1969; Mills 1972). Decreasing housing rent and density gradients are a feature of these monocentric models—i.e., rents and housing density fall as distance to the CBD increases. The basic monocentric framework has been expanded to incorporate growth and uncertainty, include environmental and open space amenities, evaluate zoning and other regulations, and study a variety of other issues (Capozza and Helsley 1990; Mills 2005; Wheaton 1974; Wu and Plantinga 2003). In recent years, economists have relaxed the monocentricity assumption (e.g. Epple and Sieg 1999; Walsh 2007).

Although spatial equilibrium models have many desirable features—a rigorous representation of agent behavior and capitalization of spatial differences in amenities and other factors into land values (Irwin 2010)—several strong assumptions are made to ensure analytical tractability. First, spatial equilibrium is a particularly restrictive assumption, because out-of-equilibrium dynamics, such as path dependence of development location, are important drivers of urban systems (Arthur 2006; Brown et al. 2005; Irwin 2010; Tesfatsion 2006). Second, in order to ensure analytical tractability, agent heterogeneity is typically quite limited.[1] More detailed discussions of the limitations of these assumptions are available elsewhere for a wide range of applications (Arthur et al. 1997; Arthur 2006; Axtell 2005; Kirman 1992; Filatova et al. 2009; Irwin 2010; Parker and Filatova 2008; Tesfatsion and Judd 2006). Since the intent here is to investigate the spatial and temporal dynamics of housing density patterns, a framework that can account for both agent preferences for spatially heterogeneous goods and idiosyncratic differences in decision-making processes is necessary. Path dependence of land-use patterns can then be explicitly linked to individual-level motivations of land conversion decisions.

27.2.2 Agent-Based Models

Agent-based modeling (ABM) has emerged as an alternative method for modeling urban growth and land use change (see Crooks and Heppenstall 2012 for an

[1] Some models include more heterogeneity than others. See, for example, Anas and Arnott (1991) and Epple and Sieg (1999) for models with heterogeneous consumers.

overview). Parker et al. (2003) provide a detailed review of the different types and applications of ABMs for modeling land use change. Although ABMs differ widely in their focus, assumptions, and formalizations of agent interactions (e.g. Benenson and Torrens 2004; Ettema 2010; Filatova et al. 2007, 2009; Ligtenberg et al. 2004; McNamara and Werner 2008; Otter et al. 2001; Parker and Filatova 2008; Robinson and Brown 2009) they all rely on interactions between many distributed agents to form emergent larger-scale patterns (Manson 2001). Thus, microeconomic fundamentals can be incorporated into individual agents' decision-making rules to simulate emergent trends in a spatially explicit framework.

However, examples of incorporating microeconomic decision-making rules into ABMs are few. Filatova et al. (2009) and earlier papers (Filatova et al. 2007; Parker and Filatova 2008) present the fullest, economically-based implementation of an agent-based land market to date. The authors relax the conventional spatial equilibrium assumption by explicitly modeling decentralized, bilateral transactions between land buyers and sellers. Transaction prices for land are determined by specifying a buyer's and seller's willingness to pay and willingness to accept, respectively, which are then adjusted to form bid and asking prices accounting for different market power scenarios (Filatova et al. 2009; Parker and Filatova 2008). The authors have provided valuable insights into methods for relaxing spatial equilibrium assumptions and incorporating microeconomic decision-making into the ABM framework. However, their model lacks a housing market and cannot capture the feedbacks between land and housing markets that influence spatial rent structures.

Ettema (2010) presents an economic ABM of a housing market, which explicitly simulates relocation and price setting processes. Housing prices are produced through bilateral transactions between a buyer and seller, and are constrained by the agents' perceptions of market conditions and by the buyer's budget constraint and housing preferences. The buyer's opportunity costs are explicitly considered by comparing utility derived from housing dwellings available in the current period to the maximum expected utility of potential housing in the future. Expectation formation, executed using Bayesian updating, is a key advance from this model design. However, the expectation formation process only accounts for price changes driven by changing consumer preferences attributed to life cycle effects. For the purposes of simulating spatially explicit development patterns—which the author acknowledges is beyond the scope of his current model—the model's design cannot accommodate spatial characteristics of housing goods or the formation of spatially heterogeneous price expectations.

Robinson and Brown (2009) present a detailed spatial representation of regional development patterns in a GIS-based ABM named dynamic ecological exurban development (DEED). Land and housing markets are integrated by the conversion of farm parcels to residential subdivisions of different densities by developers, and the acquisition of deeds to subdivision lots by residential household agents. In addition, township agents are able to specify zoning and land acquisition polices to alter development patterns. However, land conversion events are not based on microeconomic decision-making. Farm and residential parcel sales probabilistically occur on the basis of land or lot characteristics. No markets are represented in which competing land uses can be valued, and the economic constraints or opportunity costs of the acting agents are not considered. Although the authors make a valuable contribution towards empirically grounding ABMs, this approach makes it difficult to gain general insights into the underlying economic forces that drive land conversion decisions.

CHALMS builds upon the above ABMs by integrating many of their innovations into one framework capable of simulating development density patterns through coupled housing and land markets. Similar to Robinson and Brown (2009), housing and land markets are linked through the supply and demand functions of the developer and consumer households, respectively; however, our agents respond directly to and create market prices subject to economic constraints. Mechanisms of land and housing transactions in CHALMS are built upon the bilateral transaction framework developed by Parker and Filatova (2008), but are expanded to link the developer's rent expectations in the housing market to his bid prices in the land market. Price expectations play a similar role in CHALMS as they do in Ettema's model (2010). Adaptive expectations of future prices and market conditions are used to compare the utility of present and potential future transactions—directly influencing the timing of transactions. In addition, our agents' price expectation models are designed to capture spatially dependent price trends that directly affect the location of housing and land sales. These advances allow us to investigate both the supply- and demand-side forces driving spatial patterns of land conversion and development density over time.

27.3 Model Description

27.3.1 Model Structure

A growing exurban area is represented in which land is converted from farming to residential housing of varying densities over time. Farmland differs randomly in its productive capacity across farms, and farmers differ in how they form expectations about future prices of their land. Farmers compare the returns from farming to expected profit from selling their land to a single representative developer and make the decision each period whether to continue farming or enter the land market. Inequality between farmers' total supply and the developer's demand for land establishes the bargaining power of farmers, which influences land transaction prices.

The developer determines the profitability of different types of housing that vary by both structure and lot size. He sells a housing good (i.e. a combination of a given house and lot size) to consumers who prefer to be close to the urban area to minimize transport costs, and are differentiated by both income and preferences over different housing types. CHALMS tracks development over time incorporating elements of path dependence and stochastic uncertainty that determine spatial development. A schematic of agent decision-making and market interactions, along with the sequence of events, is shown in Fig. 27.1. Price prediction models for farmers and the developer are used to form expectations of future land and housing prices, respectively, and are described in detail in the Appendix.

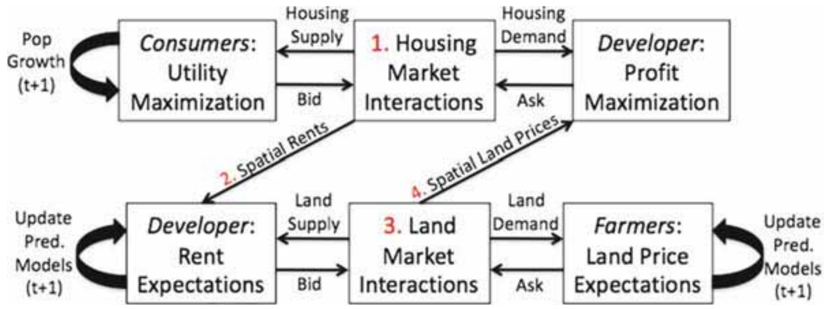


Fig. 27.1 Conceptual map of agent and market interactions in CHALMS. The *numbers* indicate the (counter-clockwise) sequence of events within one simulated time period (t). Agents (*italics)* are labeled with the underlying conceptual model that governs their behavior. Inter-temporal processes (t+1) shown include updating developer's rent prediction models, updating the farmers' land price prediction models, and exogenous growth of the consumer population (Taken from Magliocca et al. 2011)

27.3.2 Formation of Agent Price Expectations

27.3.2.1 Consumer Utility, Willingness to Pay (WTP), and Willingness to Bid (WTB)

A consumer *c* calculates standard Cobb-Douglas utility derived from the consumption of a general consumption good and a housing good. Each housing good can be considered a 'bundle' of 1 of 18 different housing types, which are distinguished by different combinations of three different house sizes (*h*)—1,500, 2,000, and 2,500 square feet—and six different lot sizes (*l*)—¼, ½, 1, 2, 5, and 10 acre; these lot and house sizes are meant to represent a typical ex-urban area. Consumer *c*'s utility function is assumed to have a Cobb-Douglas form:

$$U(c,n) = \left(I_c - P_{ask|n} - \psi_n\right)^{\alpha_c} h_n^{\beta_c} l_n^{\gamma_c} \tag{27.1}$$

where I_c is income, ψ_n is the travel cost from the location of house *n* to the CBD, and β_c and γ_c are the consumer's idiosyncratic preferences for house and lot sizes, respectively. $P_{ask/n}$ is the developer's asking price for house *n*, which is determined by Eqs. 27.15 or 27.16 below, depending on whether the house is being re-sold or is newly constructed, respectively (see Sect. 27.3.4.1).

The WTP of consumer *c* for any given house *n* is then equal to the portion of the consumer's income that he/see is willing to pay for housing as given by the Cobb-Douglas structure:

$$WTP(c,n) = \left(I_c - \psi_n\right)\left(\beta_c + \gamma_c\right) \tag{27.2}$$

Although this functional form for the utility function implies that consumers would pay the same amount for all housing net of transportation costs, consumers identify the housing option with the greatest utility and adjust their bids on other houses relative to this most preferred option. First, the maximum utility possible across all houses, U^*, is found. Holding U^* constant for all housing options, the rent, R^*, that would produce the same utility to the consumer as the most preferred choice (i.e. an optimal rent such that the consumer would be indifferent among housing options) is calculated for each housing option.

$$R^*(c,n) = I_c - \psi_n - \left(\frac{U^*}{h_n^{\beta_c} l_n^{\gamma_c}} \right)^{\frac{1}{\alpha_c}} \tag{27.3}$$

Second, the difference between the rent being asked by the developer, $P_{ask|n}$, and the optimal rent, R^*, is used to form a willingness to bid (WTB) from WTP for each house.

$$WTB(c,n) = WTP(c,n) - \left(P_{ask|n} - R^*(c,n) \right) \tag{27.4}$$

Consumers therefore bid more or less than the constant share of income for housing depending on their income and idiosyncratic preferences for house and lot size, and on the seller's asking prices for the houses actually available at a point in time. It is important to note that the full heterogeneity of consumer preferences is captured, and bids reflect the relative utility of each housing option offered.

27.3.2.2 Developer's Rent and Return Projections and Willingness to Pay (WTP) for Land

The developer is assumed to use housing information, such as incomes and utilities of residents and records of past housing prices, to form rent expectations, which in reality would be available from a 'real estate agent' or similar source. Housing information is recorded in discrete 'zones' of five by five blocks of cells, which segment the entire simulated landscape. This information includes the average expected rent, lot size, house size, number of bidders before sale, percent that sale price was above/below the original asking price, the number of houses of each type in the zone, and residents' income and utility levels for all houses in each zone. For any given house, the developer uses financial prediction models (see Appendix, Eqs. A.1–A.6) to form a rent expectation (R_{expt}) for that house in $t+1$ given past price information from the neighboring zones. Based on rent expectations for existing housing, the developer makes spatially explicit rent projections for all housing types for all undeveloped cells.

Rent projections are made by one of the three different methods described below. Projected rents are a combination of weighted local and regional (city-wide) rent information. For a given housing type to be built in a given location, a similar housing type within a local geographic area provides rent information from which a direct extrapolation can be made based on distance and local price trends. However when a similar housing type does not occur locally, the developer must rely on rent prediction methods that draw from similar housing types in a larger geographic region. In this ‘regional case’, rent predictions are less direct than in the ‘local case’. Thus, the appropriate rent projection method is adopted based on the amount of rent information available in a given area. For each undeveloped cell, a rent for each housing type is projected taking into account the distance of the given cell from the CBD and associated travel costs.

For a given undeveloped cell, the distance to every other grid cell is calculated and mapped. The specified parameter, n_{close}, sets the number of closest cells to be considered as a local search area for rent information. Using n_{close} developed cells, a distance-from-the-CBD-weighted average rent is calculated for each housing type present. This subset of local houses, n_{close}, is the basis of rent projections so that high demand in particular areas (e.g. due to desirable housing types and/or a relative shortage of housing in close proximity to the CBD) can be capitalized into rents that may exceed what is predicted based on only the travel cost gradient. Depending on whether the housing type for which a rent projection is being made is present in n_{close} search cells, one of the following methods for projecting rent is used:

1. If the housing type for which a projection is made *is present* in the n closest cell:

$$R_{proj}^{loc}(i,lt) = R_{lt}^{loc} - mcD^{loc}(i,lt) \tag{27.5}$$

   where R_{lt}^{loc} is the local distance-weighted average rent for housing of type lt within he closest developed cells, mc is the travel cost per cell (converted from \$/mile), and $D^{loc}(i,lt)$ is the distance from the cell i to the closest developed cell of the same housing type lt.

$$R_{proj}^{reg}(i,lt) = R_{lt}^{reg} - mc\left(D_i - D_{lt}^{reg}\right) \tag{27.6}$$

   where R_{lt}^{reg} is the regional average rent for housing type lt, D_i is the distance from the CBD of cell i, and D_{lt}^{reg} is the average distance from the CBD of all housing of type lt in the region. The resulting rent projection is given by:

$$R_{proj}(i,lt) = w_{loc}R_{proj}^{loc}(i,lt) + w_{reg}R_{proj}^{reg}(i,lt) \tag{27.7}$$

   where w_{loc} and w_{reg} are local and regional weights of 0.3 and 0.7, respectively.
2. If the lot type for which a projection is being made *is not present* in the n closest cells, but exists somewhere in the city, the rent projection is solely based on regional rental information and is given by Eq. 27.6 for $R_{proj}^{reg}(i,lt)$.
3. If the lot type for which a projection is being made *is not present* in the n closest cells, and it *does not exist* anywhere else in the city, then rent projections are made based on average utilities:

$$R_{proj}^{loc}(i,lt) = I_n^{loc} - \psi_i - \left(\frac{U_n^{loc}}{h^{\beta_n} l^{\gamma_n}}\right)^{\frac{1}{\alpha_n}} \tag{27.8}$$

Table 27.1 Selection of model parameters

Mean (Std. dev) farm size, in acres	128 (70.67)
Mean (Std. dev) agricultural return, in \$/acre	\$2,486 (\$249)
Building cost per square foot	\$85–\$165
Infrastructure costs per housing unit[a]	
One acre lots or smaller	\$6,000–\$17,000
2 acre lots	\$11,000–\$20,000
5+ acre lots	\$13,000–\$25,000
Share of income on housing expenditure, ß + γ	
Low income	.35–.42
Middle income	.27–.34
High income	.18–.26
Proportion of housing expenditure on land, γ/(ß + γ)	.10–.90
Transportation costs (costs/mile)	
Time[b]	\$1.30
Out of pocket (BTS 2007)	\$0.54
Exogenous rate of population growth	10%

[a]Based on Frank (1989) and Fodor (1997)

[b]We assumed time costs to be a function of average road speed (30 mph), average number of workers per house (2), average wage per person (\$30/h), value of time as a percent of wage (50%), and the road network indirectness coefficient (0.3) (this is the ratio of network distance to the Euclidean distance)

where I_n^{loc} is the average income (available from zonal housing information, see above in Sect. 27.3.2.2) households located in the n closest cells, and U_n^{loc} is the average utility of households located in the n closest cells.

$$R_{proj}^{reg}(i,lt) = I^{reg} - \psi_i - \left(\frac{U^{reg}}{h^{\beta} l^{\gamma}}\right)^{\frac{1}{\alpha}} \tag{27.9}$$

where I^{reg} and U^{reg} are the average household income and utility, respectively, over the entire region. The rent projection for housing type lt in cell i is then given by Eq. 27.7.

Based on projected rents, potential returns are calculated for every housing type in every undeveloped cell by subtracting the costs of construction and infrastructure (Table 27.1), which vary by housing type, and the price of land for the given cell. The maximum return for each cell is calculated as the housing type with the maximum return over all possible housing types (subject to zoning constraints) for the given cell. Maximum returns are then projected onto the gridded landscape to be used by the developer to determine the type and location of housing construction that maximizes profit across all vacant holdings.

Given the rent projections for every undeveloped cell, the rent associated with the housing type that produces the maximum return in each cell i of farm F is specified as $R_{max|i}$. The developer's WTP for a given farm F is the average $R_{max|i}$ over the extent of the farm:

$$WTP(F,t) = \frac{\sum_{j=F_i} R_{\max|j}}{A_F} \tag{27.10}$$

where A_F is the total acreage of farm F.

27.3.2.3 Formation of Farmer's Willingness to Accept (WTA)

Farmer expectations of land prices are formed using a randomly allocated set of 20 prediction models. Each prediction model uses one of six different methods for forming predictions based on up to 10 years of past land prices from which to extrapolate the next period's price expectation (Eqs. A.1–A.6 in Appendix). A farmer's decision to sell to a developer or continue farming is based on the expected return from selling his farm relative to the value of the farm's agricultural return per acre in perpetuity, V_{agr}. The projected land price for cell i on farm F, $P_{Lproj|Fi}$, which consists of spatially discounted (Eq. A.10 in Appendix) and predicted (Eqs. A.1–A.6 in Appendix) price components, is compared to the farmer's baseline WTA.

$$WTA(F_i,t) = \max\left\{P_{Lproj|F_i}, V_{agr|F_i}\right\} \tag{27.11}$$

The farmer's WTA is dynamically set to the greater of the two values. This enables the farmer to capture speculative gains from sale of his/her land when development pressure is high, while enforcing a rational threshold below which the farmer would be better-off farming.

27.3.3 Land Market Interactions

27.3.3.1 Bargaining Power

If the developer's WTP for a given farm is greater than the farmer's WTA for his land, then the two enter into bilateral negotiation to determine the final transaction price of each parcel. Bargaining power in the land market, ε, is adapted from Parker and Filatova (2008) and captures differences in the developer's demand for and the farmers' supply of land at the initial WTP of the developer.

$$\varepsilon = \frac{\left(d_{Land} - A_{F^*}\right)}{\left(d_{Land} + A_{F^*}\right)} \tag{27.12}$$

where d_{Land} is the acreage demanded by the developer and A_{F*} is the acreage supplied by participating farmers. $F*$ is the subset of all farmers for which the condition

WTP > WTA is true. If the developer demands more land than farmers supply, ε is positive and farmers bid above their WTA (see Sect. 27.3.2.3). If farmers supply more land than is demanded by the developer, ε is negative and the developer will bid below his initial WTP (see Sect. 27.3.2.1). Bargaining power is dynamic because the amount of land supplied by farmers depends on the initial WTP of the developer. Also, the developer's WTP for a given farm depends on the level of rents in the housing market. Thus, housing and land markets are explicitly linked.

27.3.3.2 Formation of Farmer's Asking Price ($P_{ask|L}$)

After bargaining power is observed (Sect. 27.3.3.1), farmers participating in the market (F^*, i.e. WTP > WTA for their farm) form an asking price in response to market conditions to maximize their gains from trade (Parker and Filatova 2008).

$$P_{ask|L}(F_i^*, t) = \max\left\{WTA(F_i^*, t) * (1+\varepsilon), V_{agr|F_i^*}\right\} \tag{27.13}$$

The asking price of the market-participating farmer, F_i^*, is equal to or greater than the value of his land in agriculture. If the developer demands more land than farmers supply, each farmer will mark up his asking price to potentially maximize gains from trade.

27.3.3.3 Formation of the Developer's Bid Price ($P_{bid|L}$)

After bargaining power is observed (Sect. 27.3.3.1), the developer forms a bid price for each farm for which the condition, WTP > WTA, is true.

$$P_{bid|L}(F_i^*, t) = \min\left\{WTP(F_i^*, t) * (1+\varepsilon), WTP(F_i^*, t)\right\} \tag{27.14}$$

The developer's bid price for the farm of a market-participating farmer (F_i^*) is equal to or less than his initial WTP for the farm. If farmers supply more land than the developer demands, the developer will mark down his bid price for each farm to maximize both gains from trade and profit from sales of houses in that location.

27.3.4 Housing Market Interactions

27.3.4.1 Formation of Asking Prices for Houses ($P_{ask|H}$)

Houses enter the housing market as either new construction or as pre-existing, recently vacated houses. For existing housing, the asking price equals the developer's expected rent, which is formed using the price expectation models described in Sect. 27.3.2.2 and the Appendix (Eqs. A.1–A.6). For newly constructed houses, the asking price equals the developer's projected rent subject to varying levels of rent information, as described in Sect. 27.3.2.2 and specified by Eqs. 27.5–27.9.

27.3.4.2 Housing Market Competition

The set of houses on which consumer *c* bids, H_j, are identified by the criteria:

$$\left\{H_j \in H_n : WTB(c,j) \geq P_{ask|j}\Omega_{lt}\right\} \tag{27.15}$$

Consumer *c* will bid on houses for which his *WTB* is greater than or equal to the developer's asking prices, $P_{ask|j}$, multiplied by the bid level, Ω_{lt} for housing type *lt*. The bid level is the running average percentage that sale prices have been above/below the original asking prices for houses of type *lt* in the past.

The housing market competition factor, *HMC*, describes the competition for housing that each consumer faces in the housing market. It is calculated by comparing the number of houses consumer *c* will bid on to the number of other consumers bidding on the same houses:

$$HMC_c = \frac{(NC - NH)}{(NC + NH)} \tag{27.16}$$

where *NH* is the number of houses in H_j and *NC* is the number of other consumers bidding on H_j.

27.3.4.3 Formation of Consumer Bidding for Housing

After *HMC* is observed (Eq. 27.16), consumer *c* sets his bid price for each particular house *j* in the set H_j in relation to his optimal rent for that house, $R^*(c, j)$, in response to market conditions:

$$P_{bid}(c,j) = R^*(c,j) + HMC_c\left[WTP(c) - P_{ask}(j)\right] \tag{27.17}$$

If HMC_c is positive, competition for housing for consumer *c* is high and his bids will be set above his optimal rents. If HMC_c is negative, competition for housing for consumer *c* is low and his bids will be set below the asking prices. If HMC_c is zero, the number of consumers bidding on consumer *c*'s set of houses is the same as the number of houses *c* is bidding on, and his bids will equal his optimal rents. The adjustment of the consumers' bid prices in response to market conditions allows consumers to try to simultaneously maximize their gains from trade and the likelihood that they will be the highest bidder.

27.3.4.4 Rules for Matching Consumers with Houses

After the bidding process is completed, the highest bidder on each house is identified. Consumers possessing at least one 'winning bid' are put into a subset of 'winning bidders'. For each consumer in the set of winning bidders, the set of houses for which the consumer owns the highest bid is identified. The consumer's utility is recalculated (using Eq. 27.1) for each of these houses using his winning bid instead of the initial asking price. Given these new levels of utility, the consumer is matched with the house for which he is the highest bidder and derives the highest utility. Once a consumer is matched with a house, both the consumer and house are removed from the market. The matching process is repeated with the remaining bids (which are kept constant) until all consumers are matched, all houses are occupied, or all positive bids are exhausted. This process ensures consumers are matched to houses that generate their maximum possible utility levels given competitive bids from other consumers and discrete housing options provided by the developer.

27.4 Model Experiments

CHALMS was run on an 80 × 80 gridded landscape with each cell representing an acre for a total region of 6,400 acres, or 10 square miles. The CBD was set in the middle of the top row at coordinates (1,40) with an established ex-urban developed area shown as the dark blue half-moon at the top of Fig. 27.2. Although CHALMS was able to replicate 18 different housing types, where type is defined by lot and housing size, initial development only consisted of randomly placed housing types

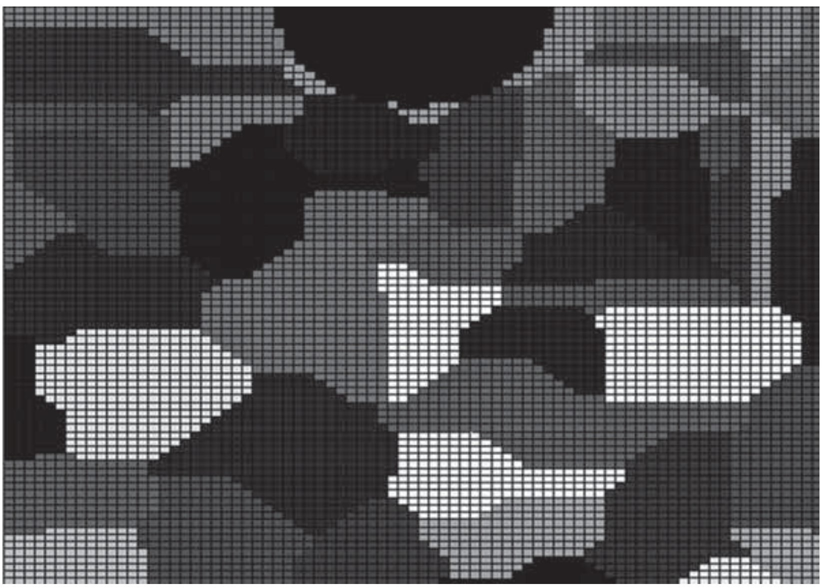


Fig. 27.2 Initial landscape configuration. Each polygon represents the location of one of 50 farms. The semi-circle (top center) represents the initial 'city' location

Table 27.2 Number of lots by type of house/lot combination, at t=20

Housing type	Lot size (acres)	Housing type description	Mean number of lots	Std. dev.	Mean annual rents (2007 $)	Std. dev.
1	¼ ac lots	Small house	87	58	7,737.35	797.10
2		Medium house	51	41	12,156.92	883.71
3		Large house	104	77	14,502.24	788.30
4	½ ac lots	Small house	144	110	9,252.51	1,426.78
5		Medium house	173	124	12,382.07	1,430.72
6		Large house	155	76	15,946.08	981.74
7	1 ac lots	Small house	429	185	12,218.53	689.40
8		Medium house	231	110	14,786.07	605.53
9		Large house	141	76	18,559.56	856.78
10	2 ac lots	Small house	475	88	19,653.39	629.50
11		Medium house	358	77	21,342.20	653.40
12		Large house	183	40	24,739.68	716.58
13	5 ac lots	Small house	0	0	–	–
14		Medium house	0	0	–	–
15		Large house	0	0	–	–
16	10 ac lots	Small house	30	32	30,461.25	4,374.33
17		Medium house	12	26	32,581.86	3,424.89
18		Large house	1	3	33,047.47	2,959.10

1 through 12 (see Table 27.2 for a description of the housing types). Fifty farms surrounded the initial development and are shown as different colored patches in Fig. 27.2.[2] Initially, 334 consumers participated in the housing market, and an exogenous growth rate of 10% a year was assumed. Incomes of incoming households are assumed to vary from $20,000 for the lowest quintile to $200,000 for the highest quintile.[3] Travel costs for households were assumed to depend both on time and monetary costs (Table 27.1). As new households moved to the region, they demanded housing; a single developer for the region responded by buying land from farmers and building houses. Thus, farmland was gradually converted to developed uses over time.

CHALMS was run 30 times[4] and each run tracks growth over a 20-year period. Farmers' locations and agricultural returns were held constant across all runs, as were the distribution and location of housing types in the initial city. Draws from income and consumer preference distributions and the initial assignment of all prediction models (i.e. for farmers' price predictions and distance discounting, and

[2] Colors are used in Fig. 27.2 to delineate the farms but have no other meaning.

[3] These data were based on median household incomes for suburban counties in the Mid-Atlantic region (Delaware, Maryland, Pennsylvania, and Virginia) from the 2000 Census. In general, the model is meant to represent a hypothetical community on the urban fringe in one of these states; we parameterize the model using data from this region.

[4] Thirty runs were determined to be a sufficient sample size as given by $n = z_\alpha^2\left(\frac{\sigma^2}{\delta^2}\right)$ for estimates of mean rents and number of lots (Table 27.2) at the 95% confidence level.

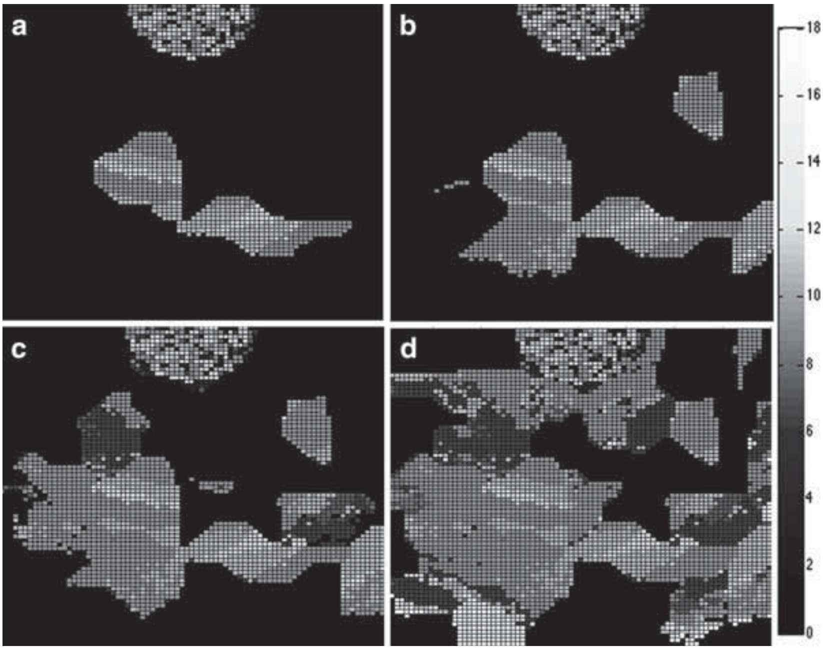


Fig. 27.3 'Average' development pattern maps for time steps (**a**) 5, (**b**) 10, (**c**) 15, and (**d**) 20. Housing types are color-coded from 1 (*dark*) to 18 (*light*)

developer's price predictions) were allowed to vary randomly across each of the 30 runs. Holding landscape features constant across runs eliminates sources of geographic variability, while exploring the effects of path-dependence and stochastic processes on development patterns that result from agent heterogeneity.

Stochastic elements in CHALMS (i.e. random draws from consumer income and preference distributions and assignment of prediction models) limit the insight of any single model realization. Instead, maps of the most likely, or 'average', development patterns were constructed (Fig. 27.3a–d). For each time step displayed, the development pattern consists only of cells that were developed above a threshold frequency, which was calibrated to produce an 'average' development pattern that closely approximated the calculated average percent-developed area and dispersion across 30 runs. Within each of those cells, the housing type with the highest probability of occurrence is mapped. In addition, Fig. 27.4 shows the probability of development at any density occurring across 30 runs.

27.4.1 Results

Table 27.2 provides a description of housing and lot sizes associated with each housing type, and summary statistics of final outcomes across 30 model runs. Even though the initial landscape configuration was held constant across runs, the housing

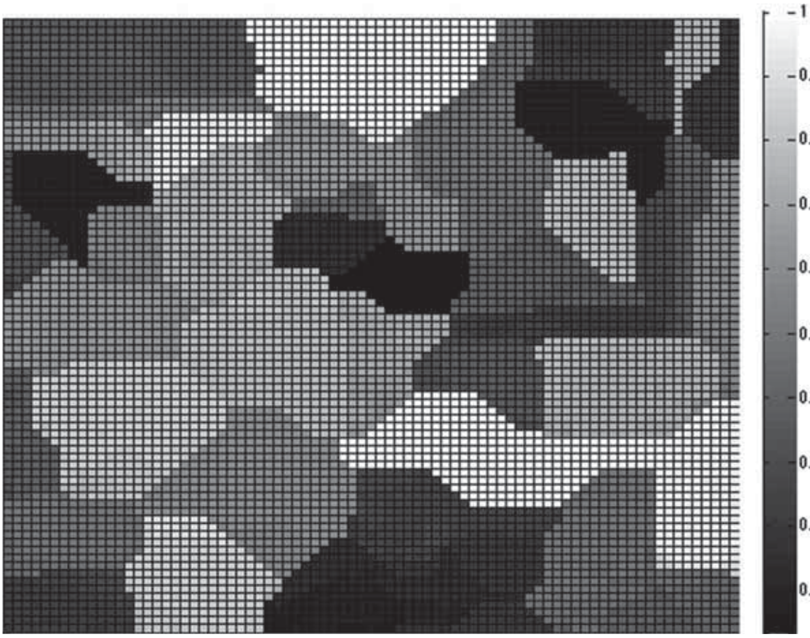


Fig. 27.4 Probability of final development patterns of any density occurring at t = 20 across 30 runs. Land that was always developed is color-coded as *white*, while land that had a low probability of developing is color-coded as *black*

types built across runs showed a good deal of variation. This variation reflected the importance of heterogeneity in consumer demand. The most frequently developed housing types were those with small or medium sized houses on 1- and 2-acre lots, which were affordable for most consumers. No 5-acre lots were built over the entire period, but 10-acre lots show up. The absence of 5-acre lots was due to the combined effects of high construction costs relative to expected rents, and the wealthiest consumers demanding houses on 10-acre lots.

The results exhibit a general development pattern that is consistent with urban economic theory: as shown in Fig. 27.5, housing density tends to decrease, and average lot sizes increase, as distance from the CBD increases (Mills 1972; Brueckner and Fansler 1983). Also consistent with urban economic theory, land prices tend to decrease with distance from the CBD (Fig. 27.6) and increase over time as population grows and demand for land increases (Fig. 27.7). The results also show a pattern that is typical of urban "sprawl" (Fig. 27.8): a divergent relationship over time between the number of lots and acreage developed (Heimlich and Anderson 2001).

In the first five time steps, development density and location were primarily driven by consumer demand and relative farm productivity. From the initial housing stock, consumers generally derived higher utility from 1- to 2-acre lots than from other lot sizes. This resulted in strong competition for those housing types and a subsequent bidding up of their rents. Relatively high rent levels prompted the developer to purchase land and capitalize on the strong demand for 1- and 2-acre lots.

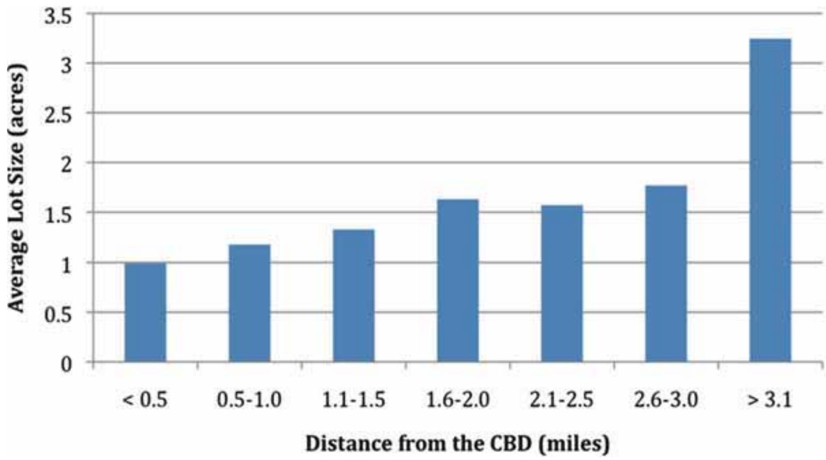


Fig. 27.5 Mean density by zone after 20 time steps. Zones form concentric circles at equal intervals away from the CBD. Rounded interval values are shown in miles

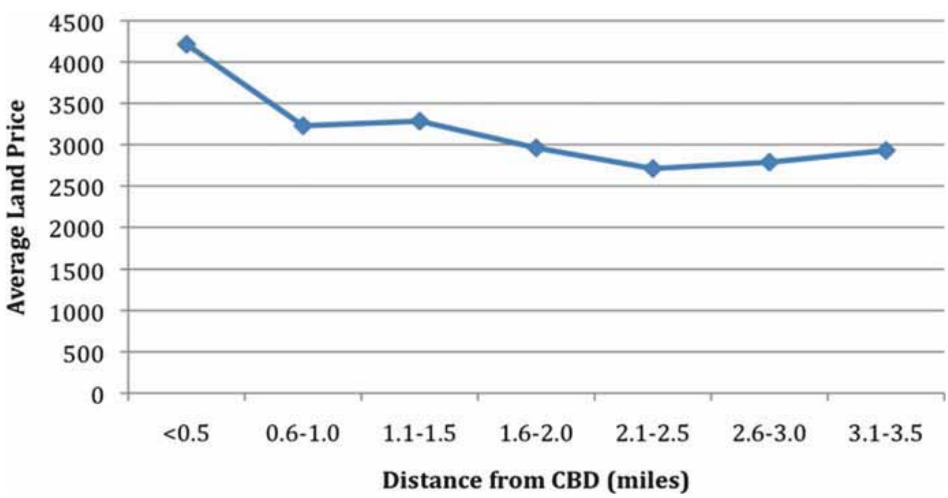


Fig. 27.6 Average price of farmland sold for development over 30 runs, at any time step, as a function of distance from the CBD

The development pressure filtered through to the land market, where farmers adjusted their WTA levels upward attempting to capture gains from sale above their return from agriculture. This price signal was strongest close to the initial development, resulting in high land prices that decreased with distance. Thus, the first farms sold were those with relatively low asking prices, distant from initial development (weak price signal) and comparatively low productivity (low initial asking price). As a result, early development progressed in a 'leapfrog' pattern (Fig. 27.3a, b) with

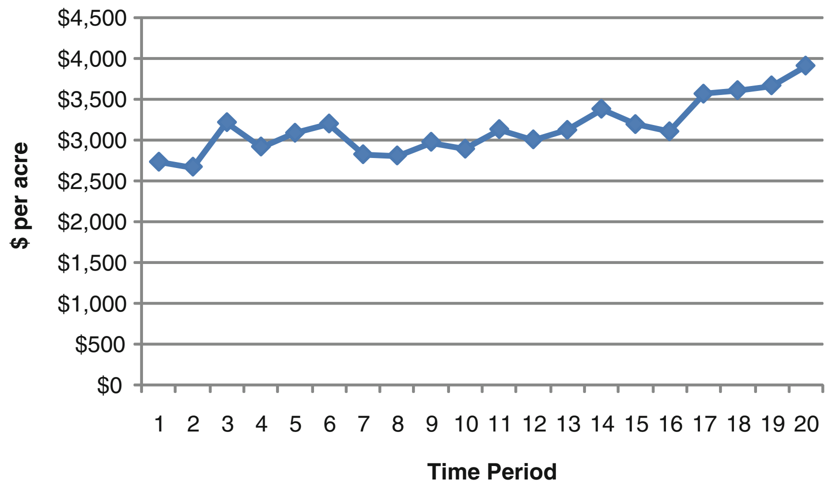


Fig. 27.7 Average price of farmland sold for development over 30 runs in each time period

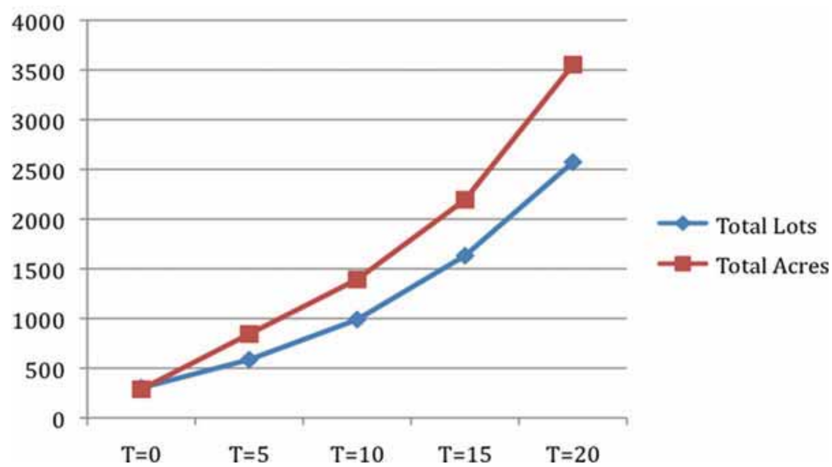


Fig. 27.8 Comparison of number of lots versus acreage developed over time

farms far from initial development sold first. Furthermore, because of strong demand and high returns net of land prices and construction costs, 1- and 2-acre lots were built on the first farms sold.

As time progressed, increased land prices coupled with consumer demand prompted the construction of houses on comparatively small (1-acre or less) or large (10- acre) lot sizes. Rents for these relatively scarce housing types rose faster than those of other housing types in the existing housing stock and prompted a shift in construction. Concurrently, development pressure and land scarcity drove land prices upward as population growth spurred competition for housing and farmers reacted to an upward trend in past land prices. Faced with higher asking prices from farmers and consumer demand for scarce housing options, the developer shifted lot sizes and location. Generally, smaller lots—i.e., higher density housing—occurred on expensive land closer to the initial 'city', while lower density housing was built on remaining land far from the initial 'city'. Spatial and temporal variability around this general pattern were due to heterogeneity in farmers' expectations of selling prices and consumer housing demands and the resulting profitability in each particular location.

Figure 27.4 provides a sense of the probability distribution of spatial outcomes across model runs. Comparisons between Figs. 27.3a, b and 27.4 demonstrate that several farms have a greater than 85% probability of being developed early in any given run. Those farms have relatively poor land and sufficiently low expectations of land prices to prompt early development consistently across runs. After time step 10, however, the remaining farms shown as developed in Fig. 27.3c, d generally have much lower probabilities of being developed in any given model run. At this point in the simulations, land prices are determined more by the agents' price expectation models than by differences in agricultural land productivity. Thus, development patterns are less dependent on landscape features and become more directly influenced by stochasticity inherent in agents' price expectation models (Brown et al. 2005). This leads to increased stochasticity in development patterns in the last half of the simulations.

27.5 Discussion and Conclusions

CHALMS is an ABM of urban growth and land-use that integrates microeconomic fundamentals into a framework capable of capturing full heterogeneity and spatially explicit development patterns. Optimizing behavior of heterogeneous consumers, farmers, and a developer, a spatially differentiated landscape, population growth, and a variety of housing and lot types are included as part of the development process. At the same time, bounded rationality, or the lack of perfect foresight, is assumed on the part of all agents. CHALMS describes the dynamics and spatial outcomes of the development process in a hypothetical ex-urban locale.

CHALMS as it currently exists has some limitations. The current version is simulated on a simplified landscape that lacks natural features such as water bodies, topography, or soil quality, which would influence a particular location's attractiveness for development and/or suitability for agriculture. In the real world, many of the features that constitute good agricultural land are often favorable for development, which compounds their influence on development patterns. In addition, proximity-based valuation of natural amenities or publicly provided goods by consumers is not represented, which has been shown to significantly influence development patterns (Filatova et al. 2007; Irwin and Bockstael 2002; Wu and Plantinga 2003). Future model iterations will incorporate more detailed natural landscape features and associated proximity-based valuation to explore their effects on development patterns.

Another limitation is the representation of only one developer. This was a simplification made to ease interpretation of simulated interactions and outcomes in both the land and housing markets. The introduction of competition between developers may change current development patterns. Although incorporating the above elements into the model's structure would likely improve its realism, such elements would also add further complexity into an already complex model. Moreover, the existing framework allows us to establish baseline development patterns subject to heterogeneous consumer preferences and incomes and farm productivity without the added complexity of a more detailed landscape. Thus, further testing of model sensitivities and outcomes will take priority before additional landscape features are introduced.

Our results demonstrate qualitative behaviors consistent with urban economic theory that emerge from explicitly coupling housing and land markets in the ABM framework. The interplay between markets and agents' heterogeneous preferences and perceptions reproduces many trends predicted by conventional urban economic models but also shows a dispersed, "leapfrog" development pattern that is common in ex-urban areas. This has three important implications. First, CHALMS demonstrates that housing and land markets influence and are influenced by one another. Thus, simulating feedbacks that emerge between markets is critical for understanding the forces that drive urban growth patterns. Second, our formalization of economic agents shows that microeconomic decision-making can be incorporated into an ABM framework to reproduce regional patterns consistent with those produced by conventional spatial equilibrium approaches. Finally, by simulating urban growth from the 'bottom-up', ABMs allow the researcher to represent full agent and environment heterogeneity and build an individual-level understanding of the dynamics of growing urban systems—a combination of advantages unique to the agent-based modeling approach.

Acknowledgements This research was supported by a U.S. Environmental Protection Agency Science to Achieve Results (STAR) grant to Resources for the Future. Partial support of the first author was provided by NSF Grant #0549469 as an UMBC IGERT trainee.

27.6 Appendix

27.6.1 Prediction Models

27.6.1.1 Financial Prediction Models

Developers and farmers make pricing decisions informed by expectations of future housing and land prices, respectively. Adapted from price expectation models used in agent-based financial literature (e.g. Arthur 1994, 2006; Axtell 2005), agents try to predict the next period's price based on current and past price information. An agent is given a set of 20 prediction models. Each prediction model may use one of six different prediction methods, and there may be more than one model applying the same prediction method in the agent's set of 20 models. Some of these prediction methods map past and present prices (P) into the next period using various extrapolation methods.

1. *Mean model*: predicts that P(t+1) will be the mean price of the last *x* periods.

$$P(t+1) = \frac{\sum_{i=t-x:t} P(t_i)}{x} \tag{A.1}$$

2. *Cycle model*: predicts that P(t+1) will be the same as *x* periods ago (cycle detector).

$$P(t+1) = P(t-x) \tag{A.2}$$

3. *Projection model*: predicts that P(t+1) will be the least-squares, non-linear trend over the last *x* periods.

$$P(t+1) = aP(t_s)^2 + bP(t_s) + c; \tag{A.3}$$

   where t_s is the time span of $t-x$ to t, and a, b, and c are coefficients of fit. Other methods translate changes from only the last period's price to next period's price.
4. *Mirror model*: predicts that P(t+1) will be a given fraction ξ of the difference in this period's price, P(t), from last period's price, P(t−1), from the mirror image around half of P(t).

$$P(t+1) = 0.5P(t) + \left[0.5P(t) - (1-\xi)\left(P(t) - P(t-1)\right)\right] \tag{A.4}$$

5. *Re-scale model*: predicts that P(t+1) will be a given factor ζ of this period's price bounded by [0,2].

$$P(t+1) = \zeta P(t) \tag{A.5}$$

6. *Regional model*: predicts that P(t+1) is influenced by regional price information coming from neighboring agents.

For farmers, land prices are a function of land scarcity as measured by the number of remaining farmers, N_f, in the region at time t.

$$P(t+1) = P(t)\left(1 + \frac{1}{N_f}\right) \tag{A.6}$$

For developers, the expected price of house types with size, h, on lot size, l, in a given neighborhood, N_b, is the mean of the prices of the houses and lots of the same sizes in adjacent neighborhoods, N_{nei}. N_{nei} are neighbors in the cardinal directions.

$$P\left(N_{b|hl}, t+1\right) = mean\left\{P\left(N_{nei|hl}, t\right)\right\} \tag{A.7}$$

All models in the agent's set of prediction models are used to predict the price in the next time period ($P(t+1)$). In time $t+1$, the actual price is known and an error squared is calculated for each model by squaring the difference between the predicted price and the actual price. The prediction model with the least error is used to make the agent's pricing decisions in the current period. This same process of prediction and evaluation is used every period so that the most successful prediction model is used every time.

27.6.1.2 Developer's 'New Consumers' Prediction Models and Demand for Land

Adapted from Arthur's (1994) "El Farol Problem", the developer attempts to predict the population at time t using past population information from the last 10 years. Population information for time t is not known until new consumers bid for houses on the housing market (Sect. A.2). Just as agents are allocated 20 financial prediction models, developers are allocated 20 population prediction models. However, instead of receiving six different predictions methods, developers receive only the first five prediction methods listed above in Sect. A.1.1. For trends in population from time $t-x$ to $t-1$ (where x ranges from 2 to 10 years in the past), developers attempt to predict how many new consumers will enter the market in time t.

The developer uses this prediction as the number of new consumers in time t, which corresponds to the number of new houses that need to be supplied in time t for new consumers, N_{new}. In addition, the developer observes the number of consumers who bid on houses but were not the highest bidder on any house in $t-1$ and therefore did not locate in the region, N_{old}. By combining the number of houses needed for new consumers (N_{new}) and consumers from the last period that did not locate (N_{old}), the number of new houses that need to be constructed in the current period (H_{new}) is calculated.

$$H_{new}(t) = N_{new}(t) + N_{old}(t-1) \tag{A.8}$$

Based on the developer's rent projections (Sect. A.1.2), the H_{new} most profitable houses are chosen for construction later in the period. Given this housing set and the associated land required to build each, the developer calculates how much land will be needed in the current period. The developer's demand for land is then the difference between the amount of land needed for new construction and the amount of vacant land already owned by the developer from previous land purchases (if any). For example, if the developer calculates ten new houses are needed in time t and the ten most profitable houses require 2 acres each, but the developer already owns 5 acres that are vacant, then the developer's demand for land in the current period will be 15 acres.

27.6.1.3 Farmer's Spatial Discounting Models

Land is an immobile good with spatially heterogeneous attributes, thus land prices vary in space and time. Farmers observe the price and location of one or more land transactions through time. A farmer then attempts to discount the observed transaction price(s) based on the distance from his location. The spatially discounted price(s) accounts for spatially variable land values and enables an adjustment of land prices based solely on trends in the market land price.

A coefficient of spatial discounting is predicted using a genetic algorithm that enables the farmer to 'learn' the best coefficient over time. Initially, each farmer is allocated a 'population' of 100 random coefficients bounded by [−200, 200]. After the transaction price(s) is observed, it is discounted using each coefficient in the farmer's 'population' of the coefficients and compared to the farmer's current asking price to evaluate the 'fitness' of each coefficient.

$$\chi_i(t) = \chi_i(t-1) + \left| \left(\frac{P_{ask|F}(t) - P_{L|F}(t)}{\bar{D}_F} \right) - \beta_i \right|; \tag{A.9}$$

where the fitness, χ_i, of coefficient β_i is the absolute value of the difference between the current asking price of farmer F, $P_{ask|F}$, and the average of the transaction price(s), $\bar{P}_L$, divided by the average distance, $\bar{D}_F$, of the observed transaction price(s) from farmer F. 'Fitness' is measured as such so that the 'most fit' coefficient will be the one with the least error. The 'most fit' coefficient is designated as 'active' and is used as β_L in Eq. A.10 to spatially discount observed transaction prices.

The farmer spatially discounts the observed transaction price(s) by predicting the coefficient of spatial discounting in a linear extrapolation to give the spatially discounted price, $P_{L|F}$, faced by farmer F.

$$P_{L|F}(t) = \beta_L \bar{D}_F + \bar{P}_L(t); \tag{A.10}$$

The coefficient of spatial discounting, β_L, represents the marginal discount of the observed transaction price(s) per cell away from farmer F. The spatially discounted price, $P_{L|F}$, is then given as an input into the farmer's financial prediction models (Sect. A.1.1).

27.6.2 *Housing Market Competition Factor*

The housing market competition factor has several characteristics that demand further explanation. *HMC* can change over time based on the income distribution of new consumers and the type and price of new houses that come onto the market. Holding incomes of existing and new consumers constant, if relatively more expensive homes are introduced to the market, the number of consumers that can afford to bid on the most expensive housing is reduced. This affects the amount of competition faced by consumers of varying income levels. For example, the wealthiest consumers would face reduced competition (i.e. *HMC* is only slightly positive or even negative), because fewer consumers can afford to bid on the most expensive houses. Conversely, lower income consumers would face increased competition for the remaining houses (i.e. *HMC* becomes increasingly more positive), because there are comparatively fewer houses per lower income consumer to bid on.

This interaction between housing prices and incomes can occur with a change in the distribution of incomes too. Holding housing prices constant, if the income distribution skews towards lower incomes, competition for housing would increase for lower income consumers, but it would not change for higher income consumers. If the income distribution skews towards higher incomes, competition for housing would increase for all consumers. Wealthier consumers would experience comparatively more competition for the most expensive houses. Lower income consumers would also experience increased competition for housing, because higher income consumers that were not the highest bidders on more expensive houses would likely outbid most lower income consumers for the remaining housing.

References

Alonso, W. (1964). *Location and land use*. Cambridge, MA: Harvard University Press.

Anas, A., & Arnott, R. J. (1991). Dynamic housing market equilibrium with taste heterogeneity, idiosyncratic perfect foresight, and stock conversions. *Journal of Housing Economics, 1*, 2–32.

Arthur, W. B. (1994). Inductive reasoning and bounded rationality. *American Economic Association Papers and Proceedings, 84*, 406–411.

Arthur, W. B. (2006). Out-of-equilibrium economics and agent-based modeling. In L. Tesfatsion & K. L. Judd (Eds.), *Handbook of computational economics: Agent-based computational economics* (pp. 1551–1564). Amsterdam: Elsevier.

Arthur, W. B., Durlauf, S. N., & Lane, D. (1997). *The economy as an evolving complex system II* (Santa Fe institute studies in the science of complexity, Vol. XXVII). Reading: Addison-Wesley.

Atkinson, G., & Oleson, T. (1996). Urban sprawl as a path dependent process. *Journal of Economic Issues, 30*(2).

Axtell, R. (2005). The complexity of exchange. *The Economic Journal, 115*(504), F193–F210.

Benenson, I., & Torrens, P. (2004). *Geosimulation: Automata-based modeling of urban phenomena*. London: Wiley.

Brown, D., & Robinson, D. (2006). Effects of heterogeneity in residential preferences on an agent-based model of urban sprawl. *Ecology and Society, 11*(1).

Brown, D. G., Page, S., Riolo, R., Zellner, M., & Rand, W. (2005). Path dependence and the validation of agent-based spatial models of land use. *International Journal of Geographical Information Science, 19*(2), 153–174.

Brueckner, J., & Fansler, D. (1983). The economics of urban sprawl: Theory and evidence on the spatial sizes of cities. *The Review of Economics and Statistics, 65*(3), 479–482.

Bureau of Transportation Statistics. (2007). Average cost of owning and operating an automobile (Assuming 15,000 Vehicle-Miles per Year. *Research and Innovative Technology Administration*, URL: www.bts.gov/publications/national_transportation_statistics/html/table_03_14.html

Capozza, D., & Helsley, R. (1990). The stochastic city. *Journal of Urban Economics, 28*(2), 187–203.

Crooks, A. T., & Heppenstall, A. J. (2012). Introduction to agent-based modelling. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 85–105). Dordrecht: Springer.

Epple, D., & Sieg, H. (1999). Estimating equilibrium models of local jurisdictions. *Journal of Political Economy, 107*(4), 645–681.

Ettema, D. (2010). A multi-agent model of urban processes: Modeling relocation processes and price setting in housing markets. *Computers, Environment and Urban Systems*. doi:10.1016/j.compenvurbsys.2010.06.005.

Fernandez, L. E., Brown, D. G., Marans, R. W., & Nassauer, J. I. (2005). Characterizing location preferences in an exurban population: Implications for agent-based modeling. *Environment and Planning B: Planning and Design, 32*(6), 799–820.

Filatova, T., Parker, D. C., & van der Veen, A. (2007). Agent-based land markets: Heterogeneous agents, land prices, and urban land use change. In *Proceedings of the 4th Conference of the European Social Simulation Associations (ESSA'07)*, Toulouse.

Filatova, T., Parker, D. C., & van der Veen, A. (2009). Agent-based urban land markets: Agent's pricing behavior, land prices and urban land use change. *Journal of Artificial Societies and Social Simulation*, *12*(13). URL: http://jasss.soc.surrey.ac.uk/12/1/3.html

Fodor, E. (1997). The real cost of growth in Oregon. *Population and Environment*, *18*(4), 373–388.

Frank, J. E. (1989). *The costs of alternative development patterns: A review of the literature*. Washington, DC: Urban Land Institute.

Heimlich, R., & Anderson, W. (2001). *Development at the urban fringe and beyond: Impacts on agriculture and rural land* (Agricultural Economic Report No. 803). Washington, DC: Economic Research Service, U.S. Department of Agriculture.

Irwin, E. (2010). New directions for urban economic models of land use change: Incorporating spatial heterogeneity and transitional dynamics. *Journal of Regional Science, 50*(1), 65–91.

Irwin, E., & Bockstael, N. (2002). Interacting agents, spatial externalities, and the evolution of residential land use patterns. *Journal of Economic Geography, 2*, 31–54.

Kirman, A. P. (1992). Whom or what does the representative individual represent? *Journal of Economic Perspectives, 6*(2), 117–136.

Ligtenberg, A., Wachowicz, M., Bregt, A. K., Beulens, A., & Kettenis, D. L. (2004). A design and application of a multi-agent system for simulation of multi-actor spatial planning. *Journal of Environmental Management, 72*, 43–55.

Magliocca, N. R., Safirova, E., McConnel, V., & Walls, M. (2011). An economic agent-based model of coupled housing and land markets (CHALMS). *Computers, Environment and Urban Systems, 35*(3), 183–191.

Manson, S. M. (2001). Simplifying complexity: A review of complexity theory. *Geoforum, 32*, 405–415.

McNamara, D. E., & Werner, B. T. (2008). Coupled barrier island–resort model: 1. Emergent instabilities induced by strong human-landscape interactions. *Journal of Geophysical Research-Earth Surface, 113*(F1), F01016.

Mills, E. S. (1972). *Urban economics*. Glenview: Scott Foresman.

Mills, E. S. (2005). Why do we have urban density controls? *Real Estate Economics, 33*(3), 571–585.

Muth, R. F. (1969). *Cities and housing*. Chicago: University of Chicago Press.

Otter, H. S., van der Veen, & A. de Vriend, H. J. (2001). ABLOoM: Location behavior, spatial patterns, and agent-based modeling. *Journal of Artificial Societies and Social Simulation*, *4*(4), 2. URL: http://www.soc.surrey.ac.uk/jasss/4/4/2.html

Parker, D. C., & Filatova, T. (2008). A conceptual design for a bilateral agent-based land market with heterogeneous economic agents. *Computers, Environment and Urban Systems, 32*, 454–463.

Parker, D. C., Manson, S. M., Janssen, M. A., Hoffmann, M., & Deadman, P. (2003). Multi-agent systems for the simulation of land-use and land-cover change: A review. *Annals of the Association of American Geographers, 93*(2), 314–337.

Parker, D. C., Brown, D. G., Filatova, T., Riolo, R., Robinson, D. T., & Sun, S. (2012). Do land markets matter? A modeling ontology and experimental design to test the effects of land markets for an agent-based model of ex-urban residential land-use change. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 525–542). Dordrecht: Springer.

Robinson, D. T., & Brown, D. G. (2009). Evaluating the effects of land-use development policies on ex-urban forest cover: An integrated agent-based GIS approach. *International Journal of Geographical Information Science, 23*(9), 1211–1232.

Tesfatsion, L. (2006). Agent-based computational economics: A constructive approach to economic theory. In L. Tesfatsion & K. L. Judd (Eds.), *Handbook of computational economics: Agent-based computational economics* (Vol. II, pp. 831–880). Amsterdam: Elsevier.

Tesfatsion, L., & Judd, K. L. (2006). *Handbook of computational economics: Agent-based computational economics* (Vol. II). Amsterdam: Elsevier.

Veldkamp, A., & Verburg, P. H. (2004). Modelling land use change and environmental impact. *Journal of Environmental Management, 72*, 1–3.

Walsh, R. (2007). Endogenous open space amenities in a locational equilibrium. *Journal of Urban Economics, 61*, 319–344.

Wheaton, W. C. (1974). A comparative static analysis of urban spatial structure. *Journal of Economic Theory, 9*, 223–237.

Wu, J., & Plantinga, A. J. (2003). The influence of public open space of urban spatial structure. *Journal of Environmental Economics and Management, 46*(2), 288–309.

Part V
Linking Agent-Based Models to Aggregate Applications

Chapter 28
Exploring Urban Dynamics in Latin American Cities Using an Agent-Based Simulation Approach

Joana Barros

Abstract The high rates of urban growth in Latin America during the 1960s and 1970s produced rapid urbanization and housing problems. In developing countries, planning policies as well as the research community have approached urban growth as a static problem rather than as a spatial form that emerges from the urban development process and that is part of a constant dynamic process. This paper focuses on a specific kind of urban growth that happens in Latin American cities, called 'peripherisation'. This is characterized by the formation of low-income residential areas in the peripheral ring of the city and a perpetuation of a dynamic core-periphery spatial pattern. The dynamics of growth and change in Latin American cities are explored using agent-based simulation. The objective is to increase the understanding of urban spatial phenomena in Latin American cities, which is essential to providing a basis for future planning actions and policies. The first part of the chapter presents a brief overview of urban growth and dynamics in Latin American cities. The Peripherisation Model is introduced, and its implementation and evaluation described. Simulation exercises were used to revisit assumptions about urbanization issues in Latin American cities and investigate important aspects of growth and change in these cities. These exercises allowed the problem of urban growth in Latin American cities to be unfolded through their dynamics, relating these dynamics to urban morphology, and thus presenting a new and important perspective on the phenomenon.

J. Barros (✉)
Department of Geography, Environment and Development Studies,
Birkbeck, University of London, London, UK
e-mail: j.barros@bbk.ac.uk

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_28,

28.1 Introduction

Rapid urbanization has been the main theme of urban studies in Latin America since the explosion of rates of growth in the 1960s and 1970s. While studies predicted an unprecedented rate of growth in these cities by the year 2000, the speed of development has been blamed as the cause of spatial inequalities and problems in these cities. Yet, the actual rates of growth have slowed since the 1980s and studies suggest that the tendency is that the rates will remain as they are. The urban problems, however, have not disappeared in the last two decades, and, despite lower rates of population growth, cities keep growing and developing in the very same way. Hence, the principal problem of urban growth in Latin American cities is no longer the high rates of population growth and rural-urban migration. Rather, it is the spatial pattern of growth and its underlying dynamics of change, the peripherisation process, which enlarges the peripheral rings of cities and metropolises despite the reduction in overall urban growth rates.

There have been a large number of studies of urbanization issues in these countries, mainly focusing on the rapidity of growth of cities and the social inequalities in urban space produced by this process. Most of these studies have taken a sociological and political approach, often discussing either the role of the poor and spontaneous settlements, or the State in the context of economic and urban development. Hence, while studies of demographic trends, housing, urban poor and urbanization proliferated during the 1970s and 1980s, very few studies have been devoted to the morphology and dynamics of Latin American cities to date. Planning policies as well as the research community have approached urban growth as a static problem rather than as a spatial form that emerges from the urban development process and that is part of a constant dynamic process.

The present study looks at issues related to the growth of Latin American cities by investigating the dynamics of this mode of urban growth and change. The objective is to increase the understanding of urban spatial phenomena in Latin American cities, which is essential to providing a basis for future planning actions and policies.

The study of urban dynamics requires tools that allow for the exploration of change phenomena in time and space. Urban modeling techniques have been traditionally used to explore issues in urban dynamics, and automata models like cellular automata (CA) and agent-based models (ABMs) seem to be a particularly suitable approach for this kind of study. Therefore, an ABM was used in order to unfold the problem of urban growth of Latin American cities through their dynamics. ABMs are based on the understanding that human decision-making plays a major role in urban processes and urban change. Their framework allows interactions between agents and their landscape to be explicitly represented. Hence, this kind of model permits the analysis of dynamic processes that link spatial development with social issues, which is of fundamental importance when dealing with cases of strong social differentiation, as is the case of urban dynamics in Latin American cities.

The present chapter describes an exploratory use of ABMs, which conceives the ABM simulation as a laboratory where existing theories of urban phenomena can be explored and further developed. The simulation model is, thus, seen as part of a theory-building process. Simulation exercises were used to revisit assumptions on urbanization issues in Latin American cities and investigate important aspects regarding growth and change in them. These exercises allowed the problem of urban growth in Latin American cities to be revealed through their dynamics, relating these dynamics to urban morphology, and thus presenting a new and important perspective on the phenomenon.

28.2 Urban Dynamics in Latin American Cities

While the problem of urban growth in Europe and North America has been formulated in terms of sprawl in the Third World and, more specifically, in Latin America the main focus has been the rapidity of the growth of cities. Indeed, it has been the fastest urban growth in history. During the period between 1950 and 1980 growth rates were very high (Hall 1983; Valladares and Coelho 1995) and, based on these data, studies anticipated continuing high rates of growth. It was believed that many areas would double in population and a few would triple, creating urban areas that by the year 2000 would be without parallel in history (Hall 1983). Latin American countries went from being predominantly rural to predominantly urban in a few decades, with high concentrations of urban population in cities with more than one million inhabitants (UNCHS 1996). This rapid urbanization produced various kinds of social problems, especially in terms of housing since the governments of such countries did not manage to provide enough housing and urban infrastructure to house the influx of migrants to the cities.

However, this population change has shown marked signs of change since 1980. After decades of explosive urbanization, urban growth rates have slowed, the rate of metropolitan growth has fallen and fertility rates have declined (Valladares and Coelho 1995). Moreover, rural to urban migration has come to have a much smaller role in urban population growth and, most recently, the pace of urban expansion has been maintained by births in the cities. These new trends have been detected in the period between 1980 and 1990, and have been confirmed by recent censuses.

Thus, the principal problem of urban growth in Latin American cities is no longer the high rates of population growth and rural-urban migration. Rather, it is the spatial pattern of growth, the peripherisation process, which enlarges the peripheral rings of cities and metropolis despite the reduction in the overall urban growth rates. The peripherisation phenomenon is becoming an increasingly significant issue, particularly in the larger cities of Latin America. In those cities, the demographic growth rate has slowed right down, migration has taken second place to natural increase, and the bulk of the housing stock now consists of upgraded (or in the process of upgrading) low-income residential areas, with a large number of spontaneous settlements.

The phenomenon of peripheral growth, which has been recognized by Latin American researchers and planners and termed 'peripherisation', can now be considered as an established process of growth of most Latin American cities. Peripherisation can be defined as a kind of growth process characterized by the expansion of the borders of the city through the massive formation of peripheral settlements, which are, in most cases, low-income residential areas. These areas are incorporated to the city by a long-term process of expansion in which some of the low-income areas are recontextualised within the urban system and occupied by a *higher* economic group while new low-income settlements continue to emerge on the periphery.

Peripherisation is an urban spatial problem which has strong effects in social and economic terms, a problem that is unlikely to fade away without strong planning action. The peripheral ring of Latin American cities consists mostly of low-income housing including large spontaneous settlements, which usually lack urban services of any kind. As such, peripherisation clearly constitutes a social problem. However, it is not only a problem in the sense of the extreme social inequalities that appear in the city in a very concrete spatial form. Rather, the problem is the *perpetuation* of such a form in space and time and, in this sense, peripherisation is a social problem of spatial order.

In terms of urban planning policies, the peripherisation phenomenon is seen from a static point of view. The focus of the governments' interventions is still on the local/housing scale, upgrading settlements and providing housing tracks for the low-income groups. There has been no focus either on the dynamics of the process or on the linkage between local and global scales. The overall growth of the cities has so far been seen as a mere result of a demographic phenomenon.

Peripherisation, like urban sprawl, is a suburbanization phenomenon. Whilst urban sprawl has been studied in detail and its main features seem to be broadly understood, in Latin America's case the understanding of the peripherisation process remains a central issue. Contrary to sprawl, an inherently spatial problem, urban peripherisation is essentially a *social* problem with *spatial* characteristics. From a social point of view, peripherisation is not a simple problem to solve. As a spatial problem, and more specifically as an urban development problem, the phenomenon still needs to be further investigated.

Like urban sprawl, peripherisation is fragmented and discontinuous development. It also presents problems related to urban sustainability, transportation and the cost of infrastructure and urban services. Studies from the 1970s suggest that the lowest densities in Latin American cities are found in high-income residential areas, with the highest densities in middle-class areas and the densities of spontaneous settlements are somewhere between these two (Amato 1970a). Finally, an interesting difference between urban sprawl and peripherisation is related to the fact that, while urban sprawl is directly related to the preference of people for suburban settings, peripherisation is not a direct consequence of locational preference. On the contrary, people who move to the city's border do not wish to live there but are compelled to.

28.3 The Peripherisation Model

The Peripherisation Model is an exploratory ABM for urban growth in Latin American cities that explores agent-landscape relationships only, and was elaborated in such a way that the behavioural rules are as simple as possible.

The simulation model was developed by adding features to a simple logic, or in other words, increasing the model's complexity step-by-step so that the understanding of its behaviour was not lost during the development process. Each of these four successive modules adds up to a model that simulates different aspects of urban growth and change in Latin American cities. Thus, module one focuses on the peripherisation phenomenon. Module two consists of the peripherisation module supplemented by a consolidation rule, and focuses on the formation of spontaneous settlements. Module three examines inner city processes, which are also added to the peripherisation logic. Finally, module four introduces spatial constraints on the simulation.

The Peripherisation Model was written in the JAVA Programming Language, using RePast (REcursive Porous Agent Simulation Toolkit). This is a software framework for creating agent-based simulations using the JAVA language developed at the University of Chicago (2003).

The model's logic and implementation will be detailed below using ODD Protocol (Grimm et al. 2006; Polhill et al. 2008; Grimm and Railsback 2012), i.e. purpose, entities and behaviour, and variables will be explained where appropriate for each module.

28.3.1 Module One: The Peripherisation Module

The main process behind the peripherisation phenomenon has similar dynamics to those proposed in the succession and expansion model of Burgess (1925). In Latin American cities, as the growth of the city passes over low-income areas, many of their original inhabitants move further out, while new peripheral rings are created on the border of the city. The process of peripherisation consists of the expansion of the city borders through the formation of peripheral low-income settlements that are then incorporated into the city by a long-term process of expansion in which some of the low-income areas are recontextualised within the urban system and occupied by *higher* economic groups while new low-income settlements keep emerging on the periphery.

28.3.1.1 Purpose

The main objective of the peripherisation module is to investigate the dynamics of formation and continuity of the core-periphery pattern. This module reproduces the process of expulsion and expansion by simulating the residential locational processes of distinct economic groups (see Fig. 28.1). In the model, the population is

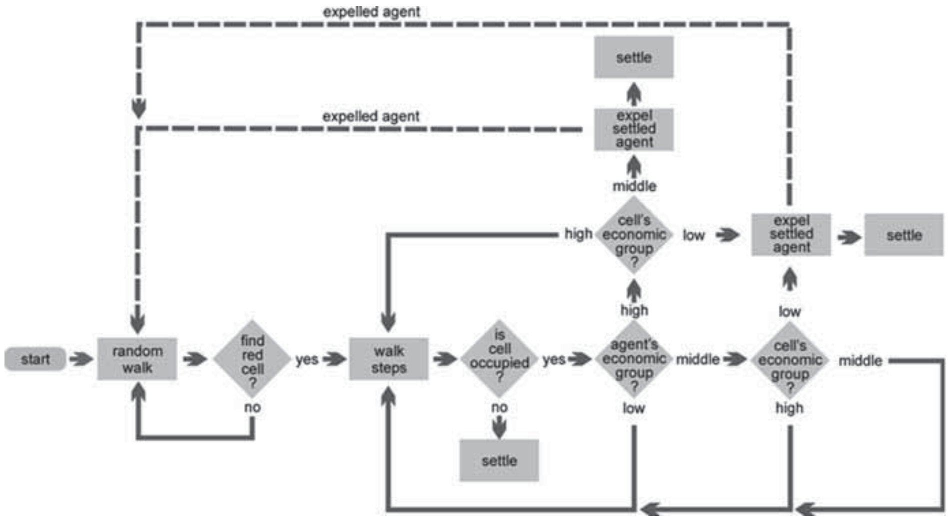


Fig. 28.1 Flowchart of agent's rules for the peripherisation module

divided into three economic groups according to the pyramidal model of distribution of income in Latin American countries. This model suggests that the society is divided into three economic groups (high, medium and low income) that can be represented by a pyramid where the high-income group is the minority at the top, the middle-income group is the middle part of the triangle and the low-income group is the majority, at the bottom of the triangle.

28.3.1.2 Entities and Behaviour

The model has mobile agents which walk randomly over a grid space. The entities (agents) in the model are proxies to individual households. However, the peripherisation phenomenon is studied independently of the spatial scale, and therefore the size of the cell might not correspond to the scale of a single plot.

Agents are divided into three economic groups (breeds) in a proportion based on the division of Latin American society by income. The main difference between agents lies in their economic group.

The simulation model is underlain by an economic logic, although it is not an economic model. It assumes that, despite the economic differences, all agents have the same locational preferences, which means that they all want to locate close to the areas served by infrastructure, with nearby commerce, job opportunities and so on. Since in Third World cities these facilities are found mostly close to the high-income residential areas, agents look for a place close to a high-income group residential area. The behaviours of the three income groups are differentiated by the restrictions imposed on their economic power. Thus, the high-income group (represented in the model in medium-grey) is able to locate in any place of its preference. The medium-income group (in light-grey) can locate everywhere except where the high-income group is already located; and, in turn the low-income group (in dark-grey) can locate only in otherwise vacant space. Since some agents can occupy other agents' cells, this means that the latter are 'evicted' and must find other places to settle. A detailed diagram of agents' rules of behaviour can be found in Fig. 28.2.

28.3.1.3 State Variables

Two main parameters define the behaviour of the Peripherisation Module: steps and proportion of agents per income group. '*Steps'* is the number of cells that the agent walks before trying to settle in a place (cell). This parameter represents how far people are willing to settle from their ideal location. The proportion of agents per income group is a percentage of the total number of agents belonging to each economic group. It is important to note that in reality, the proportion of agents per economic group differs from country to country and even from city to city, and that the proportion in the model represents a relative proportion only, as there is no definition of 'economic group' implied in the model.

28.3.2 Module Two: Spontaneous Settlements

28.3.2.1 Purpose

This module is intended to simulate the process of formation and consolidation of spontaneous settlements as part of the urban growth dynamics of the Latin American city. The module combines the original peripherisation logic with a consolidation rule. This rule refers to a process in which spontaneous settlements are gradually upgraded, and, as time passes, turn into consolidated favelas or, in other words, spontaneous settlements that are immune from eviction, as detailed in Fig. 28.2. As a result of the introduction of the consolidation logic, the spontaneous settlements module generates a more fragmented landscape than the homogeneous concentric-like spatial distribution of classes in which consolidated spontaneous settlements are spread all over the city.

28.3.2.2 State Variables

The consolidation process is built into the model through a '*cons*' variable. This cons variable has its value increased at each iteration of the model and, at a certain threshold ('*consLimit'*), the low-income cell turns into the consolidation state. If a high-income or medium-income agent tries to settle on the low-income cell in a stage previous to the consolidation threshold, the low-income cell is replaced by the respective new occupant's economic group. Otherwise, consolidated cells are 'immune' from eviction.

28.3.3 *Module Three: Inner City Processes*

28.3.3.1 Purpose

In this module, features were added to the model in order to simulate these other aspects of Latin American urban development, such as re-occupation and regeneration of older housing in attractive inner city districts. These features attempt to reproduce some of the main dynamic processes in cities: inner city decay, movement of elites towards the city edge and gentrification by the process of location and relocation of individual agents from different income groups.

The model simplifies these dynamic processes using a set of very simple spatial interaction rules and allows the modeller to examine how these rules produce contrasting and complex spatial patterns in different kinds of cities. The aim of this module is to study the nature of inner city dynamic processes, and examine how these dynamics produce global spatial residential patterns.

28.3.3.2 Entities and Behaviour

Three sets of rules compose the inner city processes module: transition from a higher to lower-income group; transition from a lower to higher-income group, and movement of higher-income groups towards the suburbs.

Hence, the processes of filtering and gentrification are translated in the model into the change from one economic group to another. Simplifying the inner city change processes in this form, filtering is translated into the model as the occupation of housing stock by a lower income group than that which previously occupied it (moving down the social scale) while gentrification is the opposite, an occupation of a housing stock by a higher economic group. The detailed schema of the agent behavioural rules in these terms is illustrated in Fig. 28.2, below.

In short, agents walk randomly through the grid in search of a place to settle. As in the peripherisation module, all agents have the same spatial preferences but react according to different restrictions. Each agent settles on a place based on local knowledge only, such as neighborhood density, their own income-group, and the income-group occupying their desired location.

28.3.3.3 State Variables

The housing transition from higher to lower income groups is introduced in the model by adding two variables: '*age*' which refers to occupied cells, and '*density*' which is the neighborhood density. Two thresholds for these variables were also added: a parameter '*decayStartPoint*' which is the threshold value for '*age*' in which the decay is activated, and a parameter d which corresponds to the maximum neighborhood density value.

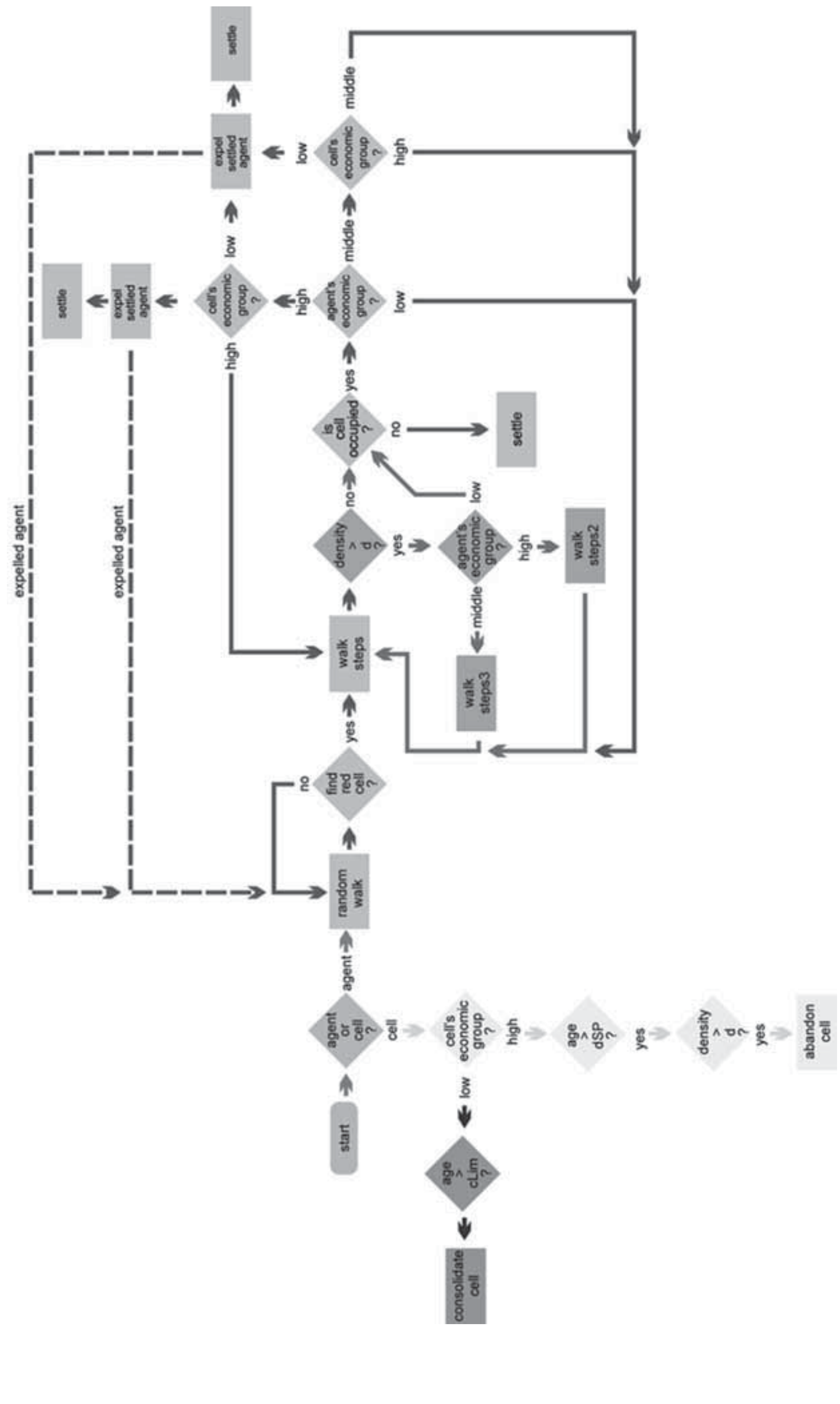


Fig. 28.2 Flowchart of the agent's rules (modules one, two and three)

At every iteration, age value is increased in the simulation for all cells occupied by high-income agents. When age gets to a certain limit (parameter '*decayStartPoint*'), a percentage of cells occupied by high-income agents may start to decrease through 'inoccupation' as these places become available to lower-income groups (their color turns green, which corresponds to empty cells that can only be occupied by low-income groups). Also, if the '*age'* value of a high-income cell is higher than the parameter '*decayStartPoint*', and density is higher than '*d'*, the cell becomes available to the lower-income agents who wish to settle there, and the high-income agent must look for a place further out. A second part of the rule determines that low-income cells with an age value higher than the parameter '*consolidationLimit'* will consolidate, i.e. they will no longer be subject to eviction.

28.3.4 Module Four: Spatial Constraints

28.3.4.1 Purpose

The objective of module four is to introduce spatial constraints to the simulation model. These spatial constraints represent bodies of water, steep slopes, or any other area where urbanisation is not possible.

Spatial constraints are implemented by the introduction of areas where agents are not allowed to settle or even walk as initial conditions. In the code, for every movement that agents make towards new cells, it checks if the new position belongs to a spatial constraint area or not and, if it does, the agent returns to their previous position and changes direction in order to avoid returning to the same cell.

28.4 Model Verification and Validation

As part of the model's development process, an evaluation of the model was performed. This included both comparisons of the model outputs with the modeled real-world system and understanding the sensitivity of the model to its internal parameters (Turner et al. 2001).

These two evaluation steps, more commonly referred to as *verification* and *validation,* concern, respectively, the correctness of model construction and truthfulness of a model with respect to its problem domain. To perform verification, a sensitivity analysis of relationships between a model's parameters and its outputs was performed.

Validation concerns how well the model outcomes represent the real system behavior and its methodology depends on the objectives of the model. Where accurate predictions are the main goal, measuring the accuracy of spatial outcomes is necessary. However, where the goal is to represent a process and explain general patterns observed across a variety of situations, as is the case with the present model, validation requires evaluating how well a model reproduces critical system properties in terms of spatial and temporal dynamics (Brown et al. 2004b; Rand et al. 2003). It is important to stress that this process "involves judgments about how well a particular model meets the modeler's goals, which in turn depends on choices about what aspects of the real system to model and what aspects to ignore" (Brown et al. 2004a, p. 2). Ngo and See (2012) provide an overview of calibration and validation of models in relation to land cover change.

For the Peripherisation Model, a set of sensitivity analysis tests was performed in order to study the relationships between input and output of the model. In other words, the study of the effects of changes in the parameters on the model's output (Saltelli et al. 2000) which is, in this case, the spatial pattern. Since grasping a whole set of aspects of an ABM's behaviour usually requires a series of different analyses, a mixture of techniques were used to establish confidence in the results of the model.

The tests presented here (see Fig. 28.3) are only a sample of the ones developed for this research. Each of the parameters of the model was tested and its impact on the model's behaviour analyzed. These tests should serve to demonstrate the typical behaviour of the model and establish the relationship between inputs (parameters and initial conditions) and outputs (spatial pattern). The tests were conducted with each of the model's module parameters and initial conditions to improve the understanding of the model's behaviour, identifying the role of each parameter and the typical behaviour of the model (Fig. 28.4).

One of the most important outcomes of the sensitivity analysis tests is the acknowledgement of the typical behaviour of the model and the effect of changes in the parameter values in the simulation outcomes. This led to a reasonable set of parameter settings for the model, which are used when it is applied to explore aspects of the real world, to be presented in the next section.

Tests also demonstrated that, despite the specific roles played by each parameter within the model, the final outcomes are strongly defined by path dependence and random process effects. This was further confirmed by tests which revealed that one of the striking features of the model behaviour is related to the time-scale. Tests suggest the time-scale is mainly defined by the probability of settlement of high-income agents, and therefore is highly vulnerable to stochastic processes and path dependence effects.

In summary, the sensitivity tests allowed the modeler to understand the typical behaviour of the model by distinguishing the features that are inherent in the model's behaviour (result from the change in parameters) from those that can be observed as proxies for the real behaviour of the system modeled. This is particularly important when the model is built as an instrument to help understand and question aspects of the real system.

28.5 Simulation Exercises

Simulation exercises were used to explore a number of aspects of dynamic change and question some of the main assumptions of urban growth in Latin American cities. These exercises consist of analyses of the model's outcomes. On the basis of

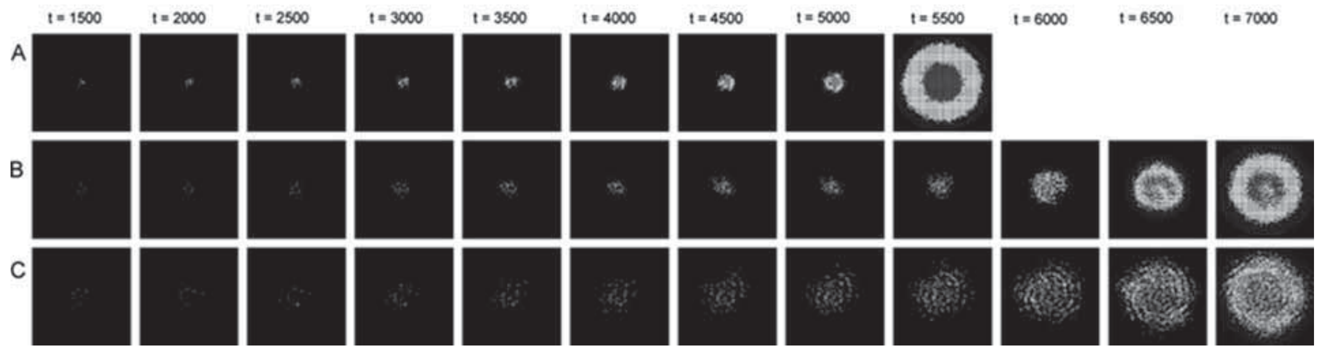


Fig. 28.3 Example of sensitivity test: sequences of snapshots testing the parameter *step*, where sequence *A* uses *steps* = 2, sequence *B* *steps* = 4, and sequence *C* *steps* = 8

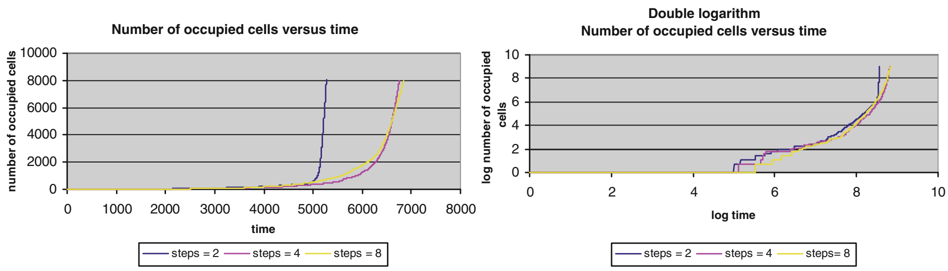


Fig. 28.4 Example of sensitivity test: charts testing the steps parameter, where sequence *A* uses *steps* = 2, sequence *B steps* = 4, and sequence *C steps* = 8

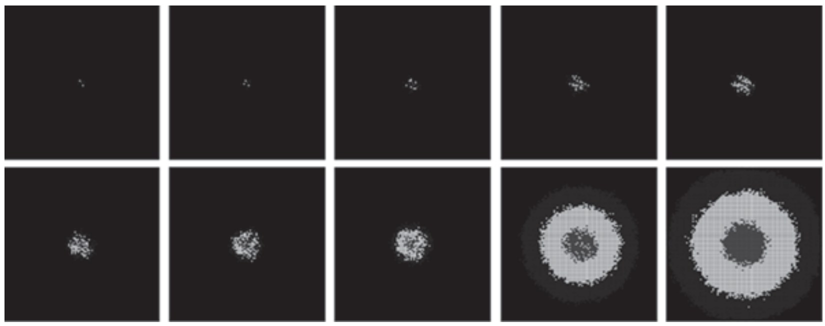


Fig. 28.5 Spatial pattern produced with the peripherisation module

these results, we attempt to provoke discussions and draw conclusions about some aspects of the reality of Latin American cities.

Four sets of simulation exercises were performed, each exploring one of the Peripherisation Model simulation modules. For the purposes of the present chapter, only the first exercise is presented. This exercise is based on the peripherisation module and discusses general issues of urban growth in Latin American cities.

The peripherisation module's agents rule base is a very simplistic one and produces the spatial patterns presented in Fig. 28.5, which shows the spatial development of the simulation run through a sequence of snapshots. The parameters used for this simulation run were *steps* equal to 2 and *proportion of agents per economic group* 10% high-income, 40% middle-income, and 50% low-income.

The final spatial pattern shown in the sequence of snapshots in Fig. 28.5 consists of three concentric rings where medium-grey represents high-income settled agents, light-grey represents middle-income agents, and dark-grey represents low-income agents. Although the simulation starts with a mixed set of patches, the final outcome is a clearly segregated pattern where each of the three economic groups congregates, forming a single large patch completely separate from one another.

This outcome is not 'unpredictable' and makes sense within the rules of the model. However, there is nothing in the rule base to suggest that the spatial outcome of the model would be a segregated pattern, and nothing suggesting high-income groups should be located in the centre surround by buffering rings of middle and

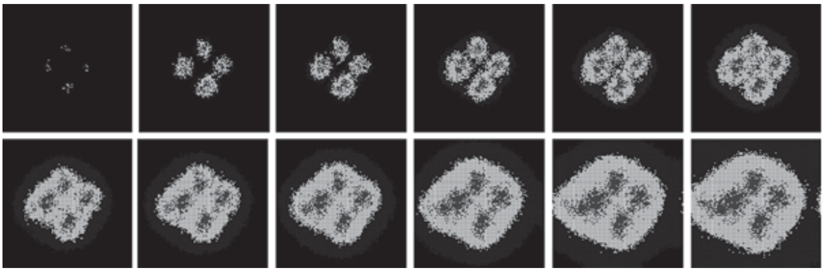


Fig. 28.6 Spatial pattern produced with the peripherisation module using multiple seeds as initial condition

low-income cells. Therefore, it is possible to conclude that the pattern is emergent, since it is the product of the local interactions of agents only.

Although very simplistic, the pattern produced by the model appears to approximate the spatial structure found in the residential locational pattern of Latin American cities. Similarly, the pattern produced by the model when using multiple initial seeds, resembles certain characteristics of metropolitan areas.

Figure 28.6 presents a sequence of snapshots using four seeds, all equidistant from the centre of the grid. The simulation was conducted using the same set of parameter values used for the previous experiment. This sequence approximates to the development of metropolitan areas, which are the result of the combination of several cities or villages that end up as a single spatial area because of their proximity. It is interesting to note how the spatial development starts with a very mixed structure, and as time passes, the core-periphery structure emerges. As in reality, this spatially segregated pattern is consolidated in the model, and as the simulation progresses, the spatial development expands, maintaining the core-periphery structure.

It is particularly striking to see how the high-income areas of the spatial patterns become slowly linked to each other, and the light-grey and dark-grey rings simply follow the shape imposed by high-income areas, acting as buffering zones for high-income areas, following the pattern described by Amato (1970b) for Bogotá, Colombia. This suggests that the model reproduces not only a final spatial pattern that is consistent with reality, but also that the evolutionary process shaping this pattern is relevant too.

The simple spatial pattern produced by the Peripherisation Model resembles in essence the spatial pattern and dynamics of urban growth in Latin American cities. As discussed previously, Latin American cities are characterized by their high rates of growth, and their spatial patterns are the result of this fast process of urban development. It is understood that high rates of urban growth have overwhelmed the capacity of urban governments to provide either adequate services or infrastructure and, therefore, are the main cause of the urban inequalities found in Latin American cities.

The simulation experiments provide material with which to discuss this assumption using the model as a *tool to think with,* and to examine the context of the rate of

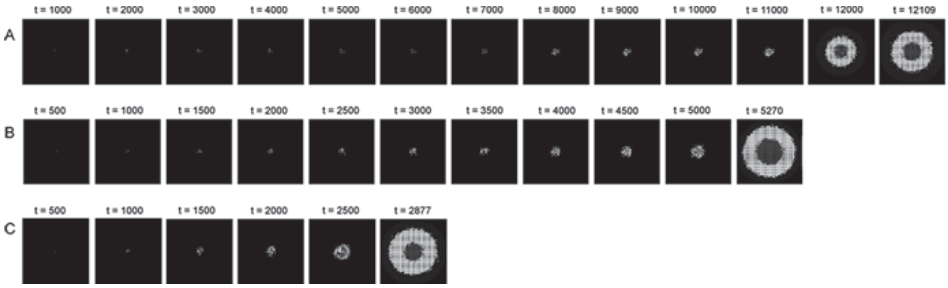


Fig. 28.7 Sequences of snapshots testing different numbers of agents

development within the simulation. In the Peripherisation Model, the presence of a great percentage of high-income agents, for instance, results in larger and faster development. This is implicit in the model's rules since high-income cells act as a catalyst for urban development. In reality it seems that the presence of high-income groups has a similar effect. The richer a city is, the more attractive it is and, therefore, more people migrate to it causing a higher rate of growth.

Whenever urban growth in Latin American cities is encountered, the first factor to be mentioned is the high rate of growth. In the literature, this rate is seen as an essential cause of the resulting spatial patterns. The present simulation exercises make clear that if the rules of the model are related in any sense to the way locational decision takes place in reality, then speed has little, if any, influence on the generation of the core-periphery spatial pattern. In the model, speed can be manipulated by, for example, increasing the number of agents within the simulation. This does not affect the spatial pattern at all, as can be observed in Fig. 28.7, which shows snapshots of simulation using different numbers of agents.

These exercises showed how some assumptions about urban growth in Latin American cities should be questioned and further investigated. The simulation exercises made it evident that the research community must review the causes driving the spatial patterns of those cities, and this knowledge must be fed back to urban planning practice.

28.6 Comparison with Reality

This section presents simple maps built from the Census 2000 dataset for São Paulo. Although these maps are static representations, the patterns of income concentration combine with the simulation model to help demonstrate the locational pattern generated by peripherisation in Latin American cities.

Figure 28.8 shows maps of income distribution per census sector in the metropolitan area of São Paulo, Brazil. The data used here are the average of the head of household monthly income per census sector (enumeration district/census block), which are part of the Census 2000 dataset provided by the Brazilian Institute of

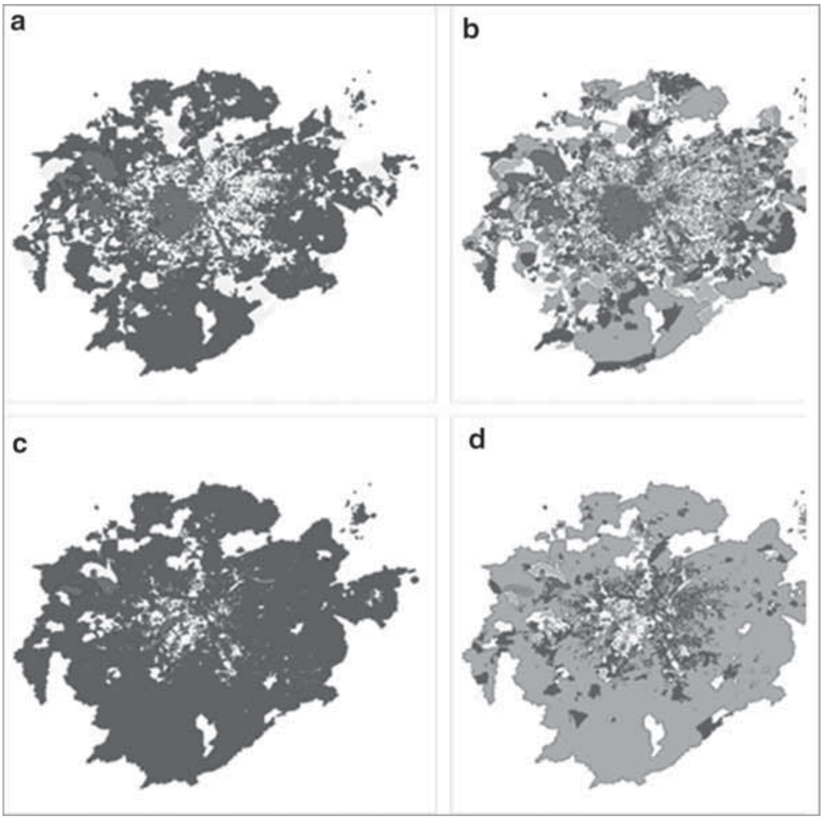


Fig. 28.8 Maps of São Paulo showing distributions of income in the urban area. Maps (**a**) and (**b**) were built, respectively, using 3 and 6 quantile breaks and maps (**c**) and (**d**) were built using, respectively, 3 and 6 natural breaks

Geography and Statistics (IBGE). This variable was chosen because of its similarities to the rules of the Peripherisation Model, which is based on the division of agents into economic groups.

The aggregated data per urban census sector were normalized by the number of householders in each sector and then classified into three ranges (maps A and C in Fig. 28.8) or six ranges (maps B and D). The maps use medium-grey for the higher income groups, light-grey for middle-income groups, and dark-grey for the lower income groups as in the simulation model, to aid comparison. One can easily identify a concentric pattern in the maps, in which the high-income groups are concentrated towards the centre of the urban area and thus the concentration decreases towards the urban periphery.

It should be noted that we have not used established definitions of income groups either in the simulation model or in the maps shown above, and our focus is only on the relative locational pattern of these groups within the city. As such, the actual number in each income group is not relevant for the present study.

When comparing the spatial pattern produced by the Peripherisation Model to the maps in Fig. 28.8, the first noticeable conclusion is that the spatial pattern in reality is not as concentric as the patterns produced by the simulation model. This is due to various factors such as initial conditions, topography, the presence of bodies of water, etc. In particular, the topography of these areas has strong influences on the spatial development of these cities.

A second very clear difference is that high-income groups are not all concentrated in the (historical) centre of the city, but may reach towards the city's outskirts. Similarly, middle-income groups are at times located on the city edge and in more central areas surrounded by low-income areas. These suggest that there are more dynamic processes in action than those simulated in module one of the model. Some of the shortcomings highlighted by these exercises have led towards the next stages in the development of the model, which were added to the model through the modules presented previously.

28.7 Conclusions

The main idea of the Peripherisation Model is concentrated in module one, which, therefore, forms the main stream of this research. This is because the module is based on the rather simple assumption that residential locational patterns in Latin American cities can be explained by essentially two concepts. The first is the idea that the composition of society, or how society is divided in groups, has a great impact on spatial development. The second is that restrictions rather than preferences generate the spatial pattern. Once these two factors are established, urban development appears to become locked in a vicious circle with high-income groups located in the best locations, while low-income groups are pushed away from all urban facilities.

The kind of government housing typically provided in Latin American cities consists mainly of housing tracts for low-income groups located in the vast majority of cases on the urban fringe. It is evident that governments have acted without knowledge of the global dynamics of the urban system and, reinforcing the current dynamics, have attracted more low-income groups to the outskirts of the city. As such, the need for centrally located housing for low-income groups was not understood. Similarly, government policies for spontaneous settlements have disregarded the dynamics of the global process of which their location and evolution is part. Because of the absence of available housing in the central area combined with the lack of good transport systems, spontaneous settlements continue to be a reasonably good alternative for low-income citizens. Yet urban interventions continue to approach the problem from a static and local point of view, and although upgrading interventions by urban governance have supported the natural upgrading process, those intervention programs have not dealt with the problem as a whole, and new spontaneous settlements continue to proliferate on the urban fringes and in other unoccupied areas within the city. From this point of view, it seems that rather than not being able to cope with the housing demand, urban governments do not have the appropriate knowledge to deal with this situation.

It seems the major planning problem in Latin American cities is how to stop such a process once it has been initiated. The role of the present investigation is not to answer this question, but to attempt to raise alternative points of view and speculate about urban development in those cities on the basis of the simulation experiments. It is important to note that the simulation exercises provide insights provoking debate, not only when the simulation results accord with reality, but also when they do not. In this case, the modeller is obliged to look for further explanations. Hence, the findings from this chapter are neither conclusive nor proven. Rather, they draw attention to gaps in our knowledge in the urban development of Latin American cities that deserves further investigation

The dynamic modeling exercises presented in this chapter have helped to further develop an understanding of the rapid urbanization process and its dynamics. They have changed the perspective on the problem from a demographic and static viewpoint to a dynamic and morphological one. The findings of this chapter have thus taken a step in the direction of bringing a new perspective to an old problem.

In this light, a question arises as to what in fact generates the segregated spatial pattern found in most Latin American cities; is it the rate of growth that was so uncontrollable that planners could not contain or control it, or is it perhaps a simple product of the inequalities of a segregated urban society? If it is assumed that the roots of a segregated spatial pattern can be largely explained by the unequal division of urban society and its economic power, then the role played by speed in the formation of the spatial pattern must be questioned.

The Peripherisation Model seems to be a good example of an exploratory simulation model, and the simulation exercises seem to be an effective way to explore aspects of reality. In addition, agent-based simulation proved to be a suitable technique to explore urbanization issues at the conceptual level, and allowed spatial patterns, dynamics and social issues to be handled within the same conceptual and modeling framework.

The need for an increased understanding of urban spatial phenomena in cities of the Third World is essential in order to provide a basis for future planning actions and policies. The approach outlined in this study has taken a step in this direction. This study provides evidence that urban modeling tools can provide an appropriate basis for research on Latin American urban processes, and makes clear the need to approach the problem by relating morphology and dynamics, for which dynamic modeling provides the appropriate means.

References

Amato, P. (1970a). A comparison: Population densities, land values and socioeconomic class in four Latin American cities. *Land Economics, 46*, 447–455.

Amato, P. (1970b). Elitism and settlement patterns in the Latin American city. *Journal of the American Institute of Planners, 36*, 96–105.

Brown, D. G., Page, S. E., Riolo, R., & Rand, W. (2004a). Agent based and analytical modeling to evaluate the effectiveness of greenbelts. *Environmental Modelling and Software, 19*(12), 1097–1109.

Brown, D. G., Walker, R., Manson, S. M., & Seto, K. (2004b). Modeling land-use and land-cover change. In B. L. Turner & G. Gutmann (Eds.), *Land change science: Observing, monitoring, and understanding trajectories of change on the land surface*. Dordrecht: Kluwer.

Burgess, E. W. (1925). The growth of the city: An introduction to a research project. In R. Park, E. W. Burgess, & R. D. McKenzie (Eds.), *The city* (pp. 47–62). Chicago: University of Chicago Press.

Grimm, V., & Railsback, S. F. (2012). Designing, formulating and communicating agent-based models. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 361–377). Dordrecht: Springer.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., et al. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling, 198*(1–2), 115–126.

Hall, P. (1983). Decentralization without end? A re-evaluation. In J. Patten (Ed.), *The expanding city: Essays in honour of professor Jean Gottmann* (pp. 125–155). London: Academic.

Ngo, T. A., & See, L. M. (2012). Calibration and validation of agent-based models of land cover change. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 181–196). Dordrecht: Springer.

Polhill, J. G., Parker, D., Brown, D. G., & Grimm, V. (2008). Using the ODD protocol for describing three agent-based social simulation models of land-use change. *Journal of Artificial Societies and Social Simulation, 11*(23).

Rand, W., Brown, D. G., Page, S. E., Riolo, R., Fernandez, L. E., & Zellner, M. (2003). *Statistical validation of spatial patterns in agent-based models.* Paper presented at the proceedings of agent based simulation 2003, Montpellier.

Saltelli, A., Chan, K., & Scott, E. M. (2000). *Sensitivity analysis*. New York: Wiley.

Turner, M. G., Gardner, R. H., & O'Neill, R. V. (2001). *Landscape ecology in theory and practice: Pattern and process*. New York: Springer.

UNCHS. (1996). *An urbanizing world: Global report on human settlements*. Oxford: Oxford University Press for the United Nations Centre for Human Settlements (Habitat).

University of Chicago. (2003). RePast 2.0 (Software) from http://repast.sourceforge.net/

Valladares, L., & Coelho, M. P. (1995). Urban research in Brazil and Venezuela: Towards an agenda for the 1990s. In R. Stren (Ed.), *Urban research in the developing world*. Toronto: University of Toronto.

Chapter 29
An Agent-Based/Network Approach to Spatial Epidemics

Joana A. Simoes

Abstract This study focused on the development of a spatially explicit agent-based model (ABM) for simulating the spreading of infectious diseases; furthermore it was assumed that the diseases spread in non-homogeneous mixed populations, distributed in an irregular space. We believe that in large scale simulation models, a realistic description of space and movement is as essential to the spreading of the disease as the description of the infectious process itself. Therefore a lot of effort was put on the development of a realistic movement model. The final version of this ABM comprehends both a movement and an infection model, which are sequentially applied at each time step. In order to be applied an tested, the model was implemented as a standalone software using the Object Oriented Paradigm (OOP), integrating the modelling algorithms with Geographic Information Systems (GIS) concepts and functionalities.

29.1 Introduction

This chapter covers the description of a spatially explicit agent-based model (ABM) for simulating the spreading of infectious diseases in non-homogeneous mixed populations, irregularly distributed in space. We developed this work based on the assumption that in large scale simulation models, a realistic description of space and movement is **as** essential to the spreading of the disease as the description of the infectious process itself. For this reason, the description of movement is a strong component of the model. The complete model comprehends two sub-models – for describing movement and infection – which are sequentially applied at each time step.

J.A. Simoes (✉)
Centre for Advanced Spatial Analysis, University College of London, London, UK

e-GEO, FCSH, Lisbon, Portugal
e-mail: j.simoes@ucl.ac.uk

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_29,

The use of computer simulations allows the precise study of the dynamical consequences of models which cannot be solved by analytical methods (Nowak and Lewenstein 1996) – as is the case of ABM – and therefore the actual implementation of the model is an essential part of the modelling process. In this study we adopted the Object Oriented Paradigm (OOP), which we believe is pertinent for representing the entities and processes in ABM. The key difference between the object oriented and the traditional structured approach is that objects are entities and functions rather than procedures, and in this way they tend to closely reflect natural systems in the real world (Montgomery 1994). There was also an effort to fully integrate the temporal and spatial components of the model, the latter one strongly relying on Geographic Information Systems (GIS) concepts and functionalities. Finally it is important to state that being a computational model, the conception of the model is so tied to the software design, that sometimes we will describe parts of the software, in order to describe parts of the model.

The model description that we present in the next sections, follows the ODD (Overview, Design concepts, Details) protocol for describing individual- and agent-based models (Grimm et al. 2006; Grimm and Railsback 2012). We believe the application of this method can bring redundancies in some of the sections, as sometimes things are explained more than once across the different points; however, it has the value of trying to implement a standard approach and therefore it can make the contents of the model more easily perceived by the reader.

Finally, as recommended by the ODD specification, Sect. 29.9 covers the application of the model to a case study: a mumps outbreak in continental Portugal between 1993 and 1996. This allowed us to gain an insight into the strengths and limitations of this particular kind of simulation model when dealing with real-world data.

29.2 Purpose

The purpose of this model is to simulate the spreading of a non-vector disease, a disease that spreads directly from one host to another without the intervention of an organism, such as a mosquito, for carrying the disease-causing microorganisms. The majority of work reviewed in this field (e.g. Mansilla and Gutierrez 2000; Fuks and Lawniczak 2001) follow the assumption of random mixing of the population; this can be a way of overcoming the lack of data or lack of knowledge about the mechanisms of movement. However, we think it is a weak assumption for human populations, as we believe humans follow routines with structured movement patterns which are unlikely to be random. One of the main concerns of this model was to put aside this simplistic assumption and develop a realistic movement model. On the other hand, we also tried to build a realistic infection model, and although so far it has only been applied to mumps (Simoes 2007), it has a flexibility of parameters that allows it to capture the behaviour of different kinds of diseases.

29.3 Entities, State Variables and Scales

As a peer-to-peer model, the main entity of this model is the individual, which for large scale simulations can be seen as a generalization of a group of individuals. The other entity of the model, which abstractly represents the geographic environment, is the region. In general terms, we can say that this model covers the interaction between individuals and between individuals and regions, within regions.

The case study for this model, mumps (infectious parotitis), is an acute infectious disease that is highly contagious and affects many children aged 5–9 years, usually before the age of 15 (Simoes 2007). Since we targeted this age group as the population of the model, and since mumps affects both genders, gender and age are not state variables of the entities (individuals).

The variable that describes the state of the individual and that distinguishes it from other individuals and from itself, at other points in time, is its health condition, as it is going to be described with more detail in the infection sub-model (see Sect. 29.8.2).

The SIR model (Susceptible-Infectious-Removed) formulated by Kermack and McKendrick in 1927 (Epstein 1997) expresses the relations between different population states. These states are:

- Susceptibles (S) – individuals capable of acquiring the disease.
- Infected (I) – individuals who can transmit the disease.
- Removed (R) – individuals who are either dead, recover or become immune from the disease.

Therefore. each individual is characterized by a variable "infection status", that can assume the exclusive values: susceptible, infected or removed. As a network inspired model, we also store the contacts of each individual. These connections are a state variable of the entity individual, which stores pointers to other individuals.

The entity "region" is a geographical area with well defined borders, i.e. it is taken as having a high degree of mixing inside it (despite the fact that it might not be homogeneous) and a lower degree of mixing with other regions. This assumption, which is inspired by the household model (Deijfen 2000), targets the representation of cities where the geographical aggregation imposes some kind of identity that assembles individuals into a common domain. The state variables for a region account for the size and current state of its population, in terms of the number of susceptible, infected and removed individuals. Moreover, it is important to say that individuals also store a pointer to their current region (which changes in time as they travel), and it can therefore also be considered as another of their state variables.

Finally, the simulation process itself can be seen as an entity of the model since it controls the general flow of the model. In this case it was implemented as a thread, which is characterized by a Boolean variable describing its state: *running* or *not running*.

Figure 29.1 contains a UML diagram with a more accurate description of these classes: "cidade" (region), "turtle" (individual) and "processo" (process). These classes

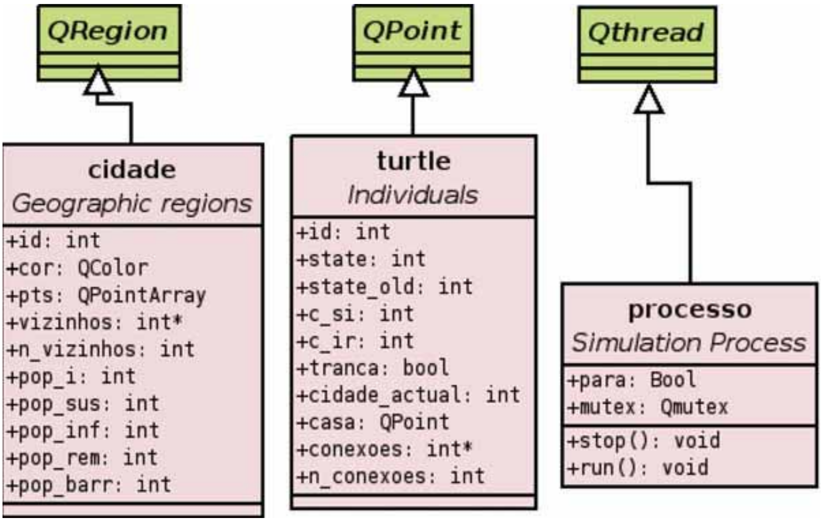


Fig. 29.1 Classes that implement the ABM: "cidade", "turtle" and "processor"

illustrate the suitability of the OOP to ABM, a framework which is totally based on the behavior of singular entities with a common structure.

The linking of different spatial scales is a key feature of the movement model. To summarize, we can say that there are three nested levels (or scales) involved in the movement model: the global scale, regions (split into two levels: inter and intra region), and neighborhoods as we illustrate schematically in Fig. 29.2. At some level, all these scales relate to the concept of regions, as described earlier on this section.

The time scale in this model is mostly related to the movement patterns that we want to describe. Since the model describes daily activities (such as going to work, etc.), we assumed that the finest division of time (time step) should be daily, and therefore we can simulate a succession of days that can describe weeks, months or even years. Again, it is possible to do some generalization, sacrificing the representation of the principles of the model, and assume that each time step is a greater temporal unit.

29.4 Process Overview and Scheduling

The core of the program is the method "run", represented on Fig. 29.3. Discarding all procedures related to the User Interface (UI) and storing of information for the GIS, the simulation can be described purely as a set of two nested cycles: one related to the number of steps and the other to the number of simulations, which call the sub-models (movement and infection) sequentially.

As time was introduced as a discrete unit in this model, the steps are accumulated in order to represent a certain time span. If we assume, as stated in the previous

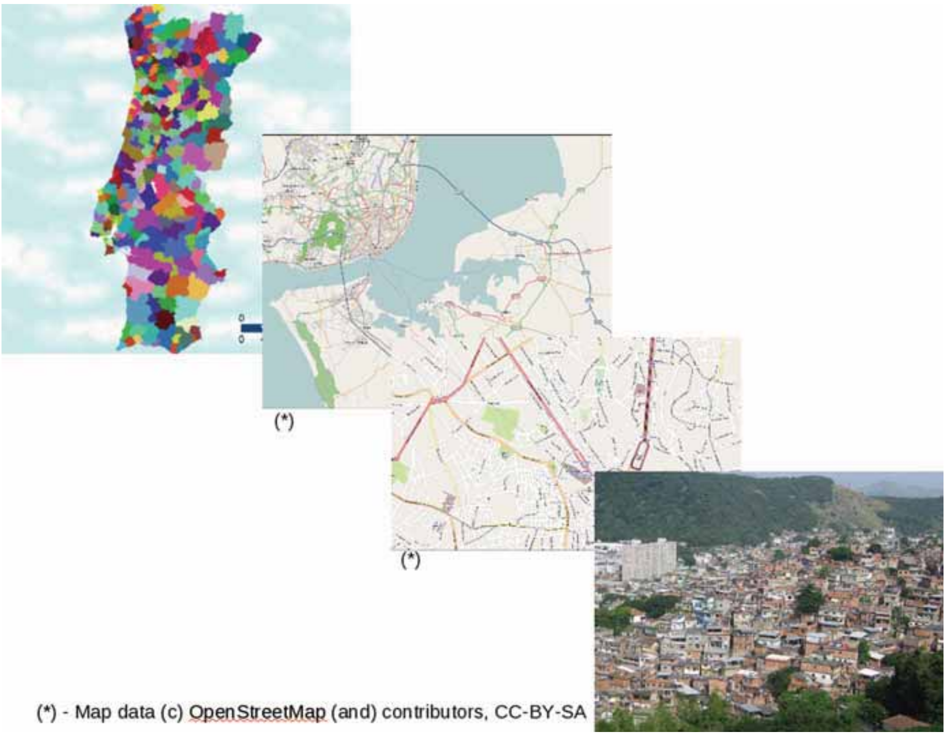


Fig. 29.2 Different scales involved in the movement model. From *top* to *bottom*: neighborhood, intra-region, inter-region and global

section, that each time step represents a day, this span can be weeks, months, or even years. The outside loop, that controls the number of simulations, was introduced as a matter of convenience. As ABM simulations are not reproducible (the "law of large numbers"), we may want to run a high number of simulations and look at the average values, rather than using one single run of the model.

The function "Mov()" is shown in Fig. 29.4. This function reads the probabilities assigned on the initial conditions and loops through all individuals to animate them with movement.

The way that movement is assigned to each individual is: generating a random variable, and comparing this value with the probabilities of each component of movement, through a chain of nested conditions.

This system is perhaps easier to understand, by looking at Fig. 29.5, where A to D correspond to the different components of movement. Once one component of movement is selected, it calls the corresponding algorithm of movement to displace the individual.

The infection model, is represented in Fig. 29.6. All individuals are members of the class "turtle" (Fig. 29.1) and its state is controlled by the variable "state", which can assume the values of: 0 (susceptible), 1 (latent), 2 (infected) and 3 (removed). Variables "state_old" and "tranca" are in the program to assure the synchronous update of all individuals.

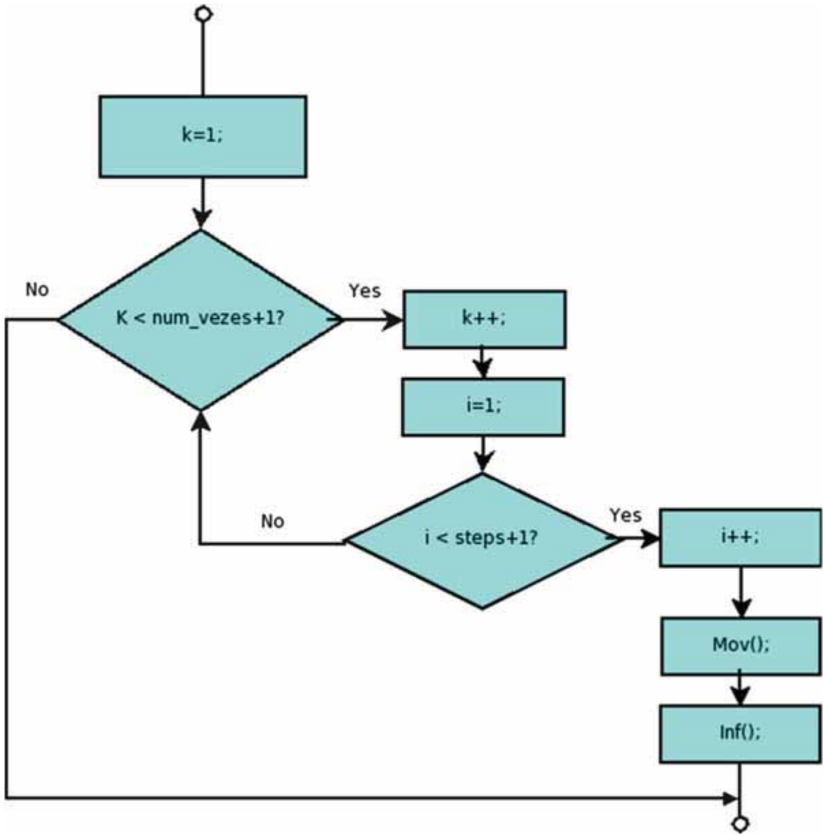


Fig. 29.3 The method "run" implementing the simulation process

The function loops through the whole population and targets the latent and infected individuals. The latent individuals can change their state if the latent period is over. The infected individuals can infect any individual contained in a "buffer" around them that matches certain conditions, and can change state if the infectious period is over.

29.5 Design Concepts

Human epidemics are strongly related to the dynamics of human populations and therefore to the network of social contacts. These kinds of systems present **self organization**, a behavior in which order may arise from low level interactions without any supervision from higher-order structures (Nowak and Lewenstein 1996). Self organizing systems typically display emergent behavior, where **emergence** is the process of complex pattern formation from simpler rules.

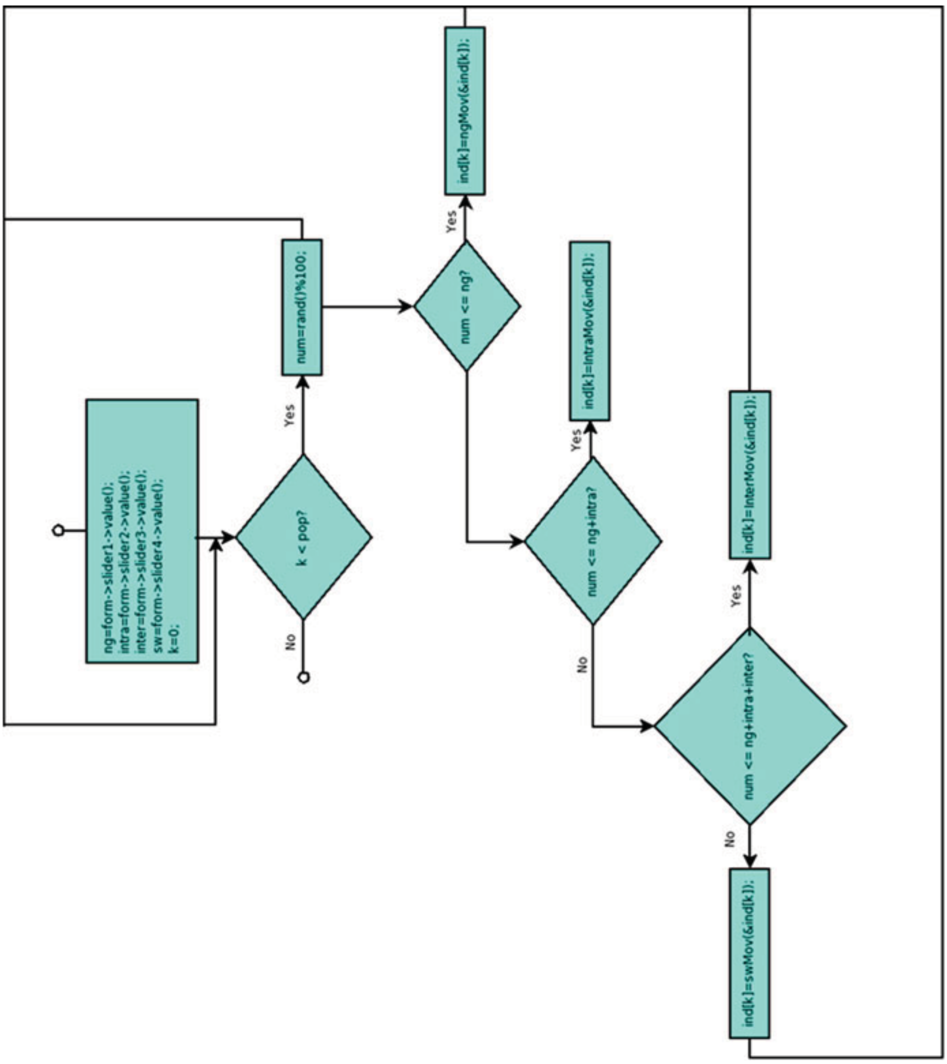


Fig. 29.4 Flowchart of the function "Mov()"

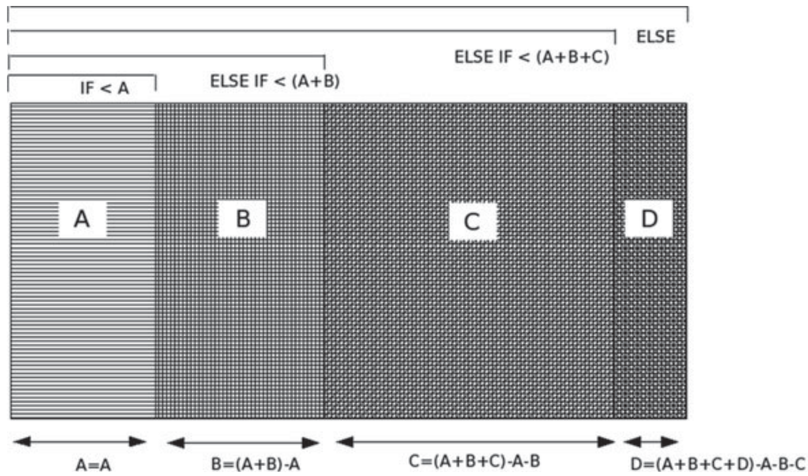


Fig. 29.5 Algorithm for random selection of the movement component, according to the pre-defined probabilities

In the case of this model, we expect two complex phenomena to emerge: the movement of individuals and the spreading of epidemics. These systems are not totally independent, as the latter one depends on the first one, and although this was not considered in the current model, the first one could also depend on the second one. However, contrary to what happens in the traditional approach (e.g., statistical models, systems of differential equations), retroactions (feedbacks) and variable dependencies are not a problem in bottom-up models.

Since this is not an evolutionary model, the concepts of "learning" and "adaptation" are not really represented; however, "stochasticity" is a strong component that is present at several stages of the simulation process.

The movement model is a network approach, inspired by the Small World graph (Newman 1999). There are several ways to generate this graph, and in the Watts and Strogatz (1998) model, this can be achieved by adding some links randomly, which will produce the shortcuts responsible for the short average path length. Following this approach, we introduced a random component in the movement model, which can be explained in terms of mobility, as a small number of individuals that travel large distances over the domain, e.g. politicians, football players or academics. In the sensitivity analysis of the model (Simoes 2007), we found that this component, even in a small proportion, has a major impact on the outcome of the simulations, since it is able to provide "shortcuts" in the network that are totally outside our structured movement patterns, thus reformulating all the accessibilities.

In the infection model, a stochastic variation of the infection force was introduced, which can represent the mutation of a virus in reaction to external factors such as antibiotics or even environmental conditions. Although for the moment we treated this variation in the infection force as a purely stochastic process, we believe there would be room to introduce some truly evolutionary traits to represent gene mutation for antibiotic resistance.

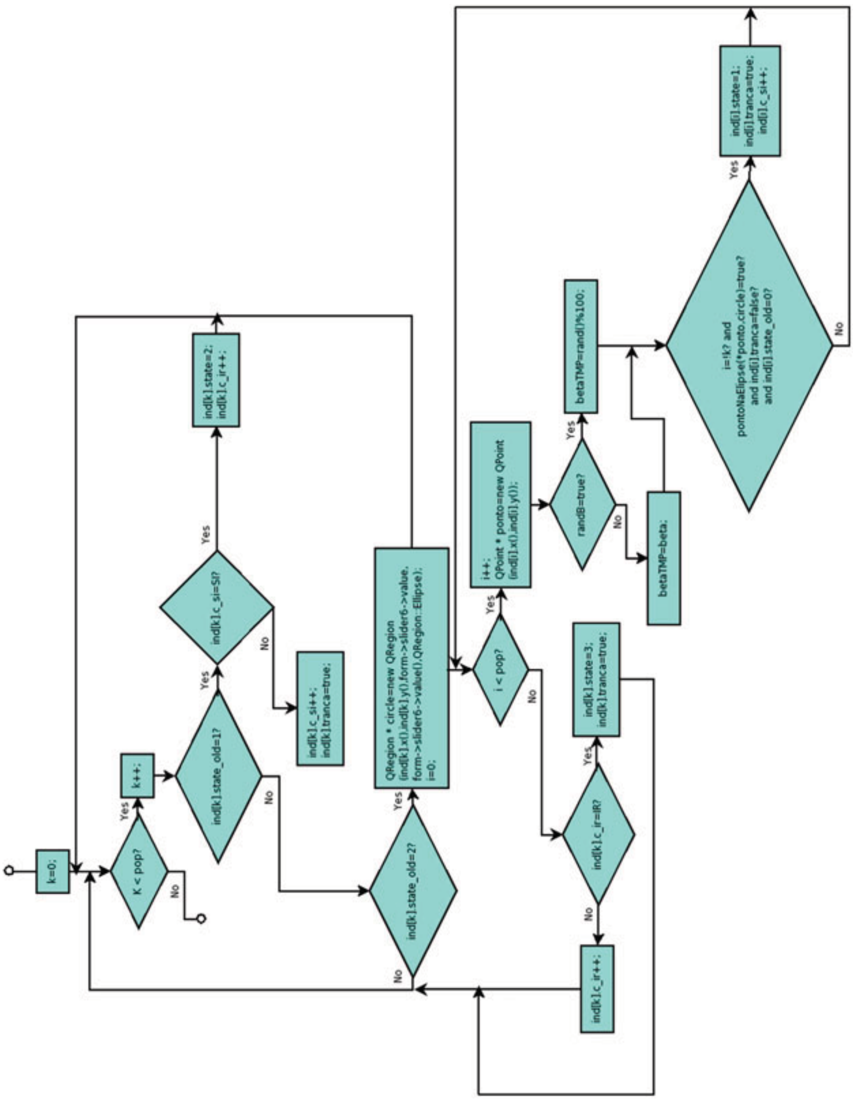


Fig. 29.6 Flowchart of the function "Inf()"

An obvious consequence from the previous paragraphs is that the stochastic nature of the model makes the simulations unique and irreproducible.

As in any ABM, interaction between entities is a crucial part of the model. There is an explicit interaction between individuals stated in the (infection) model rules and an implicit interaction between individuals and regions, since regions constrain the individual mobility within its borders and influence the number of contacts with its size and the composition of its neighborhood. As regions are different from each other, so is their influence on the individuals within them, and perhaps this is a striking difference from the grid representation of space that we see in Cellular Automata (CA) models (Margolus 1987).

Finally, it is important to say that there is no concept of collectives in this model, and the behavior of the population and of the country as aggregates, emerges directly from the behavior of the entities: individuals and regions.

29.6 Initialization

The initial conditions of the model are constituted by the model parameters (Fig. 29.7), which are all of the variables that are considered to be constant during the simulation, and therefore **exclude the state variables** (as described in Sect. 29.3).

Apart from the population parameters (which are tied to a "population model" described in the next section), the parameters can be divided into two groups corresponding to the two sub-models: movement and infection.

Each component of the movement model is described in terms of its proportion of the overall movement or the probability of occurrence. This choice regulates the actual behavior of the movement model (and therefore the spreading of infection) and it should not be made by chance. After a careful sensitivity analysis, we recommend the choice of parameters on Figs. 29.7 and 29.8, which follows a distance decay law: an individual is more likely to stay where it is than to travel over large distances. Therefore, movements with a longer range (like inter region or random) have a lower probability than movements with a shorter range (like intra region or neighborhood).

The neighborhood radius and the infection radius are spatial parameters, which are numerical quantities that obviously reflect the scale, but ultimately they control the "closeness" of the community (how large is the area we call "neighborhood") and the infectiousness of disease (how close an individual has to be from an infectious individual to become potentially infected).

The other parameters of the infectious model are the latent and morbid periods (in terms of time steps) and the strength of the infection force (from 0 to 1), which can also be random. The choice we see in Fig. 29.7 reflects the behavior of mumps, our case study, but these values can easily be changed to accommodate other diseases.

Finally, the time steps and number of simulations (as mentioned in Sect. 29.4) are also input parameters of the model.

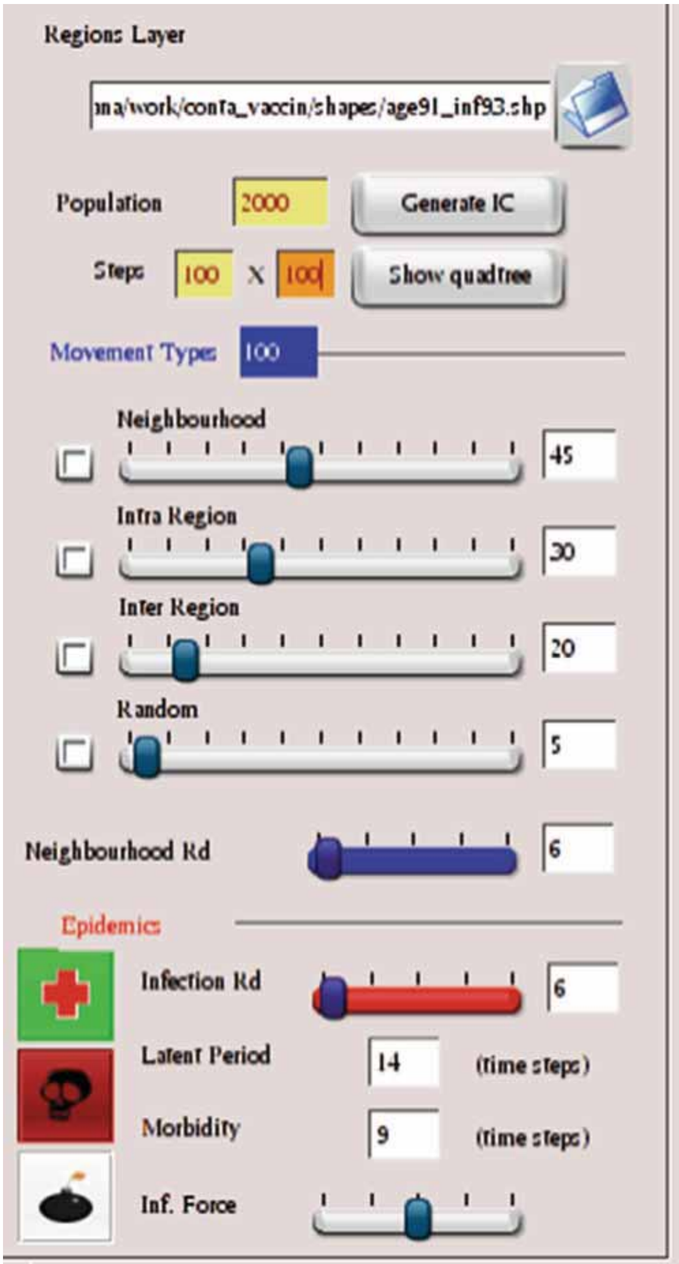


Fig. 29.7 Parameters of the model, as presented in the user interface

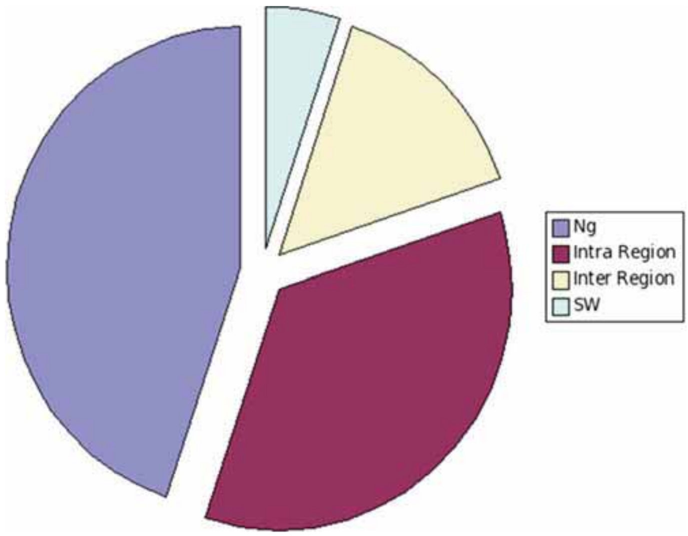


Fig. 29.8 Probabilities for each component of the movement

29.7 Input Data

As a spatial model, the input data for this model must be spatially distributed, and as the format is vector, s*hapefiles* were chosen. A shapefile is a working and interchange format promulgated by ESRI for simple vector data with attributes.[1] This is also an open format that is fully described in ESRI's whitepaper (ESRI 1998).

The shapefile prepared for the model must contain the definitions of the regions (represented as polygons), and the attribute table must contain the susceptible population – allocated by region – and may also contain the infected population.

Prior to running the simulation, a simple model is run that allocates the individuals randomly within the regions.

The initial infected population can also be introduced manually, by clicking on the screen, after activating an input mode (Fig. 29.9). This can be particularly useful in sensitivity analysis where we want to assess the reaction of the model to specific configurations of the initial infected individuals, e.g., a single individual or a "ring" of individuals around a region.

29.8 Submodels

29.8.1 Movement

Although we did not find many studies focused specifically on movement networks, there are analogue studies on social networks, which we used as a basis for the movement model.

[1] http://www.esri.com/

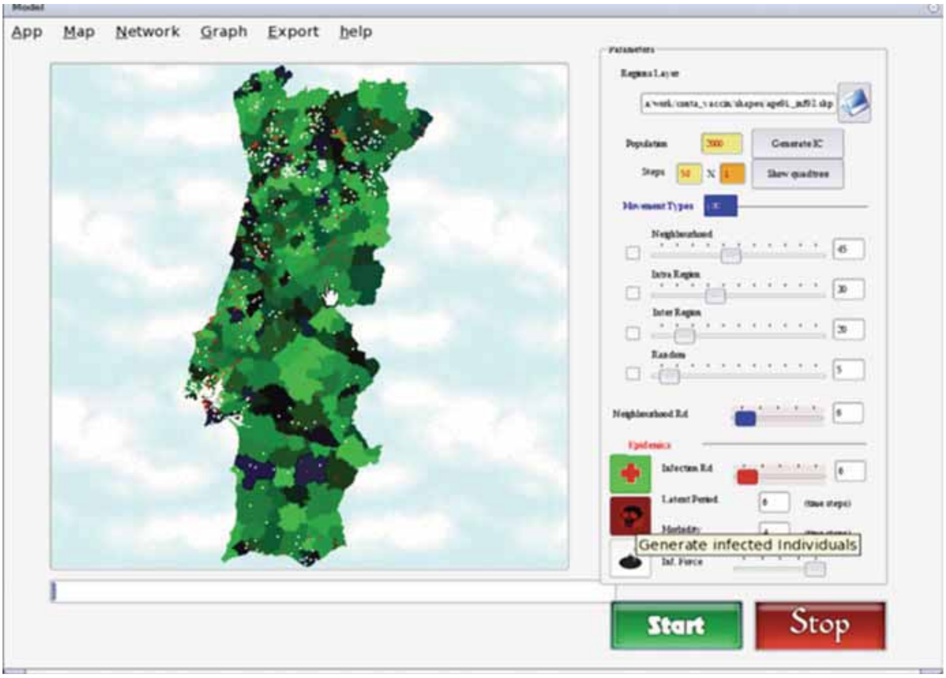


Fig. 29.9 Illustrations of the input of infected individuals through the user interface

In Deijfen (2000), the Bernoulli graph, the Markov graph and the 3-clique model are presented as examples of random graphs used for describing social networks. However, these graphs are too focused on friendship ties to be useful in our representation of movement. On the other hand, there are other social graphs that show that some features can be analogue to the ones expected in a movement network, e.g. the household and the Small-World graphs, which are described in more detail.

The household model divides the population into mutually exclusive groups with a high level of mixing, as is the case with households. This model is described by two separate graphs: one describing the division into households and another one describing the relationships between individuals. Although the household graph itself does not allow for the spreading of the disease, households are linked together by the friendship network. Given a set of N labeled vertices, let $\varsigma 1$ be a graph representing the household structure and $\varsigma 2$ a random graph representing the network of contacts. Formally, we can say that the network ς is formed by the superimposition of the graphs $\varsigma 1$ and $\varsigma 2$ as in (29.1):

$$\varsigma = \varsigma_1 \cap \varsigma_2 \tag{29.1}$$

The movement model developed in this study establishes an analogy between households in a “social space” and regions in a “geographical space”. Instead of households, we use a set of geographical “regions” that are assumed to have a high degree of mixing within them. If we assume that these regions correspond to cities, the movement model establishes a network linking neighboring cities (adjacent vertices) and another network inside the city itself, but contrary to the analogous household model, the network inside this unit **is not** a random graph. The city itself presents an inner structure; each city is subdivided into smaller units called **neighborhoods**. A random graph defines the movement inside each neighborhood and another random graph links these neighborhoods together inside the city. In Fig. 29.2, we can observe the relation between these networks, operating at different scales, which combine together to create the movement network.

Elaborating more on each kind of movement, we can say that neighborhood movement ($\varsigma 1$) represents the movement of an individual in a street or a block, involving activities such as staying at home or going to the pub. Intra-region movement ($\varsigma 2$) is the movement inside the city, e.g. going to work or shopping, while inter-region movement ($\varsigma 3$) means traveling to neighboring cities, e.g. visiting relatives or taking part in leisure activities. Finally, random movement ($\varsigma 4$) links together distant parts of the network and is related to another kind of social graph: the *Small-World* model.

Small-Worlds, also referred popularly as "six degrees of separation", were first introduced by Milgram in 1967, and are now widely used to represent several kinds of social networks (Newman 1999). According to Newman (1999), Small-worlds can be defined as random graphs which possess two properties: short average path lengths and high clustering coefficients. Although at first sight these properties only describe ties between individuals, they are of some interest for the movement model. Following the Watts and Strogatz (1998) model, we introduced small values of probability p on $\varsigma 4$, which produces long range links in the network. Our assumption on the movement network of having a small number of individuals with a high mobility is analogous to the assumption of very well connected individuals in the Kasturirangan (1990) model. On another level, regions with higher populations, for example in big cities, show high coordination numbers, behaving as hubs. It should be noticed that this assumption follows the idea that the accessibility follows the population distribution, and thus can compensate for the lack of accessibility data. An individual's network is also highly clustered inside each region, reproducing in this way another Small-World property.

As a conclusion, we can say that the movement network arises from the super-imposition of the different graphs referred to earlier, which we define in (29.2) as:

$$\varsigma = \varsigma_1 \cap \varsigma_2 \cap \varsigma_3 \cap \varsigma_4 \tag{29.2}$$

Using a bottom-up approach, the global structure of the network emerges from the displacement of each individual a, according to the Eq. 29.3:

$$a_{(i,j)}^{(t+1)} = a_{(i,j)}^{(t)} + d \tag{29.3}$$

The stochastic variable d has a probabilistic distribution as follows:

$$d = (P_1)D_1 + (P_2)D_2 + (P_3)D_3 + (P_4)D_4 = \sum_{x=1}^{4} (P_x)D_x = 1 \tag{29.4}$$

The parameters D_1, D_2, D_3 and D_4 are the different ranges of movement.

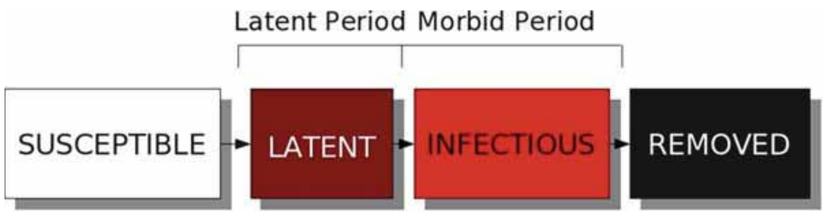


Fig. 29.10 Diagram of the epidemic model state transitions

29.8.2 Infection

In this ABM, we implemented a variation of the SIR model (Epstein 1997) called SEIR (Susceptible-Exposed-Infectious-Removed). The main difference between SEIR (see Eqs. 29.5a–d) and the original SIR model is the introduction of a fourth state very relevant in childhood diseases such as mumps or measles: the Latent or Exposed (*E*) state. In this model, an **infected** individual might not be yet **infectious** and therefore it is important to distinguish between these two states.

$$\frac{dS}{dt} = -\beta SI \tag{29.5a}$$

$$\frac{dE}{dt} = \beta SI - \theta \mathrm{E} \tag{29.5b}$$

$$\frac{dI}{dt} = \theta \mathrm{E} - \alpha \mathrm{I} \tag{29.5c}$$

$$\frac{dR}{dt} = \alpha \mathrm{I} \tag{29.5d}$$

The contact parameter β, also known as infection force, regulates the state transition from susceptible to latent. The transition parameters θ and α regulate the state transition from exposed to infectious and infectious to removed. From the point at which the individual becomes infected until it becomes infectious, there is a latent period (that includes both the symptomatic and asymptomatic period), and from the point that the individual becomes infectious until it is removed, there is an infectious period, or morbid period. Figure 29.10 presents the sequence of states in the SEIR model.

The discrete version of this model is given in Eqs. 29.6a–d.

$$\frac{\Delta S}{\Delta t} = -\beta SI \tag{29.6a}$$

$$\frac{\Delta E}{\Delta t} = \beta SI - \theta E \tag{29.6b}$$

$$\frac{\Delta I}{\Delta t} = \theta E - \alpha \mathrm{I} \tag{29.6c}$$

$$\frac{\Delta R}{\Delta t} = \alpha \mathrm{I} \tag{29.6d}$$

where Δ represents a discrete change over a time period t to $t+1$. Therefore the states at $(t+1)$ can be computed as per Eqs. 29.7a–d as:

$$S^{t+1} = -S^{t}(\beta \mathrm{I}^{t} - 1) \tag{29.7a}$$

$$E^{t+1} = \beta \mathrm{S}^{t} I^{t} - E^{t}(\theta - 1) \tag{29.7b}$$

$$I^{t+1} = \theta \mathrm{E}^{t} - I^{t}(\alpha - 1) \tag{29.7c}$$

$$R^{t+1} = \alpha \mathrm{I}^{t} - R^{t} \tag{29.7d}$$

In this study, we considered a closed population of N individuals, as we show in Eq. 29.8:

$$\frac{dN}{dt} = 0 \tag{29.8}$$

We also assume that we are dealing with micro parasites, in which case a single infection triggers an autonomous process in the host (Diekmann and Heesterbeek 2000). This means that each infection process is independent of other individuals and from the environment, a situation that does not occur with macro parasites. It is important to emphasize once again, that in this model, the contact process is only a function of physical proximity and is not based on social ties as for instance is the case of sexually transmitted diseases. Therefore, the contact is directly determined by the movement model: for infection to occur, it is only necessary to have immediate "proximity" between a susceptible and an infective individual; the exact definition of this "proximity" is controlled by a distance parameter called *infection radius*. Finally, it is important to note the key difference between the model adopted in this study and the compartmental model formulated by Kermack and McKendrick (Epstein 1997), which is not formulated via the previous equations. This difference relies on relaxing the assumption of homogeneous infectivity and homogeneous mixing of susceptibles and infectives. Rather than having such an assumption, their model considered spatially complex scenarios with an irregular distributions of individuals in order to achieve a more realistic representation. As the homogeneous assumptions were abandoned, the final size as well as the growth rate of the epidemic can no longer be computed analytically and these can only be achieved through computer simulation.

29.9 Simulation Experiments

This chapter would not be complete without the application of the model to a real world scenario. Our case study was an epidemic outbreak that took place in Portugal in 1996–1997, an episode usually attributed to vaccination failure (Dias et al. 1996; Gonçalves et al. 1998). As in many ABMs, the input data for the model (as mentioned in Sect. 29.7) is crucial for recreating a certain scenario. Getting the correct data into the model is not a trivial task. Statistical temporal series suffer from widely known problems such as the Modifiable Areal Unit Problem (MAUP) (Lawson 2001), introduction errors and incomplete information, etc. Some assumptions were made in order to use this data but it is important to say that all these things can bias the initial conditions and lead us away from reality. On the other hand, there are limitations in the implementation the model, resulting in some simplifications and consequently to some distortions in the magnitude of the results. In Fig. 29.11, we can see the distribution of infected individuals in the dataset and in the model.

After running one hundred simulations, we calculated the correlation coefficient between the epidemic size in each time step and the observed dataset. This value varied between 0.88 and 0.94, with an average correlation of 0.92. We also compared the spatial distribution of infected individuals in the simulations and the model, and although there are magnitude differences, probably introduced by the

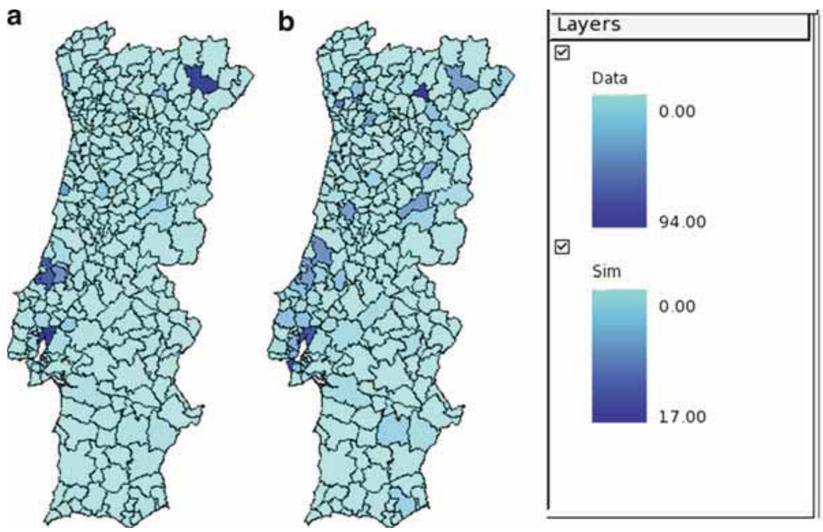


Fig. 29.11 Initial distribution of the infectious individuals: (**a**) in the dataset (1993) and (**b**) in the simulations (t = 0)

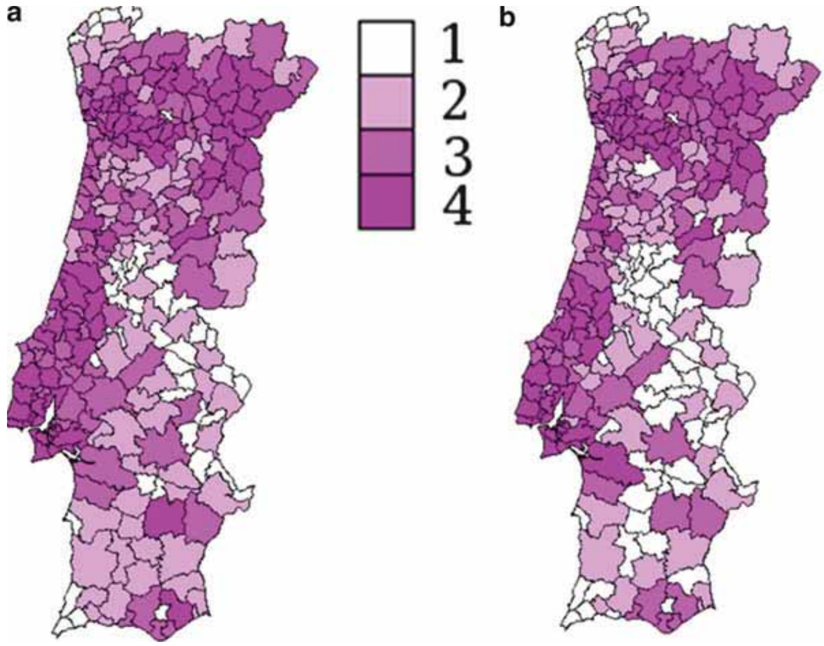


Fig. 29.12 Spatial distribution of the infected individuals, classified in quartiles (**a**) in the dataset and (**b**) in the model

biased initial conditions, we verify the occurrence of the same clusters of infected individuals in both scenarios.

To remove this magnitude difference, the two series were standardized with a classification based on quartiles. Despite some outliers, the comparison between the quartile class in each region in the dataset and in the simulations reveals very close results. This is easily perceived by looking at the normalized distribution of affected individuals in Fig. 29.12.

Finally, it is important to note that a model dealing with many random components usually faces great variability and that the results presented in this section rely on average values of a large number of simulations run over time. Analyzing the distribution of frequencies generated by the model, we came to the conclusion that although they do not follow a standard normal distribution, the levels of skewness and kurtosis are fairly acceptable, and the measures of central tendency indicate a normal behavior, as well as a low dispersion around the mean. These results validate the significance of the simulations and also the use of average values to represent the series.

29.10 Conclusions and Final Remarks

This chapter covers different aspects of the process of implementing and running a bottom-up simulation model applied to human epidemics. We believe that there is much to be gained from this approach of using a hybrid model that combines both networks and multi-agent systems as a way of capturing the complexity of a group of interacting individuals in a non-homogeneous environment. Although the choice of language, C++, makes the programming slightly less trivial than in other languages, the object oriented paradigm is well suited for the development of an individual based model like the one presented.

The exercise with the dataset of mumps made us aware of the high sensitivity of the model to the input data, which compromises its success due to lack of quality and incompleteness of the data. It also made us aware of some weaknesses of the model, namely its incapability of dealing with the full dataset. However, despite all the simplifications and assumptions made, the prediction of spatial patterns of infection was quite successful.

Finally, the most important contribution of this model was to incorporate space into the simulation and to relax the simplistic assumptions of homogeneous space and random movement of individuals. Although it was not covered in this chapter, the use of GIS data and algorithms were central to this model. A realistic spatial simulation model can also be the basis for simulating spatial intervention strategies (Simoes 2007) and therefore provide an important tool for planners and policy makers in the health field.

References

Deijfen, M. (2000). *Epidemics on social network graphs*. Unpublished Masters thesis. Department of Mathematics, Stockholm University, Stockholm. http://www.matematik.su.se/~mia/uppsats.pdf

Dias, J. A., Cordeiro, M., Afzal, M. A., Freitas, M. G., Morgado, M. R., Silva, J. L., Nunes, L. M., Lima, M. G., & Avilez, F. (1996). *Mumps epidemic in Portugal despite high vaccine coverage – preliminary report* (Euro Surveillence Technical Report 4). http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=160. Accessed December 19, 2009.

Diekmann, O., & Heesterbeek, J. (2000). *Mathematical epidemiology of infectious diseases*. New York: Wiley.

Epstein, J. M. (1997). *Nonlinear dynamics, mathematical biology and social science* (Santa Fe Institute studies in the sciences of complexity lecture notes). Reading: Addison-Wesley.

ESRI. (1998). *Shapefile technical description* (An ESRI white paper). http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf

Fuks, H., & Lawniczak, A. T. (2001). Individual-based lattice model for spatial spread of epidemics. *Discrete Dynamics in Nature and Society, 6*, 191–200.

Gonçalves, G., de Araújo, A., & Monteiro Cardoso, M. L. (1998). *Outbreak of mumps associated with poor vaccine efficacy – Oporto Portugal 1996*. Euro Surveill 3 (Technical Report 12, pp. 101). http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=101. Accessed December 19, 2009.

Grimm, V., & Railsback, S. F. (2012). Designing, formulating and communicating agent-based models. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 361–377). Dordrecht: Springer.

Grimm, V., Berger, U., Bastiensen, F., Eliassen, S., Ginot, V., Giske, J., et al. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling, 198*(1–2), 115–126.

Kasturirangan, R. (1990). *Multiple scales in small-world graphs* (Technical Report). http://arxiv.org/abs/cond-mat/9904055. Accessed December 19, 2009.

Lawson, A. B. (2001). *Statistical methods in spatial epidemiology*. Chichester: Wiley.

Mansilla, R., & Gutierrez, J. L. (2000). Deterministic site exchange cellular automata model for the spread of diseases in human settlements. *Complex Systems, 13*, 143–159.

Margolus, T. T. N. (1987). *Cellular automata machines: A new environment for modeling*. Cambridge: MIT Press.

Montgomery, S. L. (1994). *Object oriented information engineering*. New York: Academic.

Newman, M. (1999). Small worlds: The structure of social networks (Technical Report 99-12-080). Santa Fe: Santa Fe Institute. http://ideas.repec.org/p/wop/safiwp/99-12080.html

Nowak, A., & Lewenstein, M. (1996). Modeling social change with cellular automata. In R. Hegselmann, U. Mueller, & K. G. Troitzsch (Eds.), *Modelling and simulation in the social sciences from the philosophy of science point of view* (pp. 249–285). Dordrecht: Kluwer.

Simoes, J. (2007). *An agent-based approach to spatial epidemics through GIS*. Unpublished PhD thesis, Department of Geography, University College of London, London. http://www.casa.ucl.ac.uk/JoanaMargarida/thesis_post_viva.pdf

Watts, D., & Strogatz, S. (1998). Collective dynamics of 'small world' networks. *Nature, 393*, 440–442.

Chapter 30
An Agent-Based Modelling Application of Shifting Cultivation

The An Ngo, Frances Drake, and Linda See

Abstract This paper outlines an agent-based modeling application of shifting cultivation for an upland village in Vietnam, which was developed to improve the management of shifting cultivation and aid forest protection. The model consists of household and land agents situated in a dynamic social, economic and political environment. Adaptation of the agents to changes in policy is incorporated through a trade-off between economic gains and social responsibility, which affect the subsequent decision-making process. The basics of the model are described including the validation process and the results in a business as usual scenario.

30.1 Introduction

Shifting cultivation is commonly practiced in the tropical areas of Africa, Latin America and Southeast Asia (Do 1994; Kerkhoff and Sharma 2006; Tran 2006). Many different practices fall under the term shifting cultivation but it generally involves an alternation of cropping for a few years followed by a relatively long period of fallow (Angelsen 1995; Conklin 1961; Do 1994; Spencer 1966).

The An Ngo (✉)
Department of Environmental Management, Hanoi University of Agriculture, Trau Quy, Gia Lam, Hanoi, Vietnam
e-mail: ntan@hua.edu.vn

F. Drake
School of Geography, University of Leeds, Woodhouse Lane, Leeds, UK
e-mail: f.drake@leeds.ac.uk

L. See
School of Geography, University of Leeds, Woodhouse Lane, Leeds, UK

International Institute of Applied Systems Analysis (IIASA), Laxenburg, Austria
e-mail: l.m.see@leeds.ac.uk; see@iiasa.ac.at

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_30,

Shifting cultivation is often viewed negatively (Gilruth et al. 1995) and in developing countries, it is seen as a primary driver of deforestation, with an estimated contribution ranging from 41% to 60% (Angelsen 1995; Geist and Lambin 2001). In the uplands of Vietnam, shifting cultivation is a common method of agricultural production, with more than 340,000 households practicing this type of agriculture across nearly 74% of the national territory (Institute for Ethnic Minority Affairs 2005). As in other countries, shifting cultivation is viewed as highly destructive to the forest and has become the focus of several government policies in Vietnam. Despite large investments, these policies appear to have had little success as shifting cultivation still remains active outside of the areas where it is legally permitted (Institute for Ethnic Minority Affairs 2005; Tran et al. 2005). Several researchers (Fox et al. 2000; Institute for Ethnic Minority Affairs 2005; Shanks et al. 2003; Vuong 2001) have argued that these polices have been implemented with little consideration for the local socio-economic context or the provision of alternative, culturally acceptable livelihoods. As a result, there has been limited adoption by the local people. To develop more realistic policies for land management and to assist policy makers and stakeholders in evaluating different policy options, it is essential to gain a better understanding of shifting cultivation, in particular the interplay between shifting cultivators, household resources and the constraints associated with production. This requires an approach such as agent-based modeling that takes into account the dynamic properties of the system, including the biophysical, socio-economic and political aspects of shifting agriculture in a larger context. The more generic advantages of this approach are already described in detail in Crooks and Heppenstall (2012) but agent-based modeling is particularly suitable for this application because the decision-making process can be captured at a household level in terms of livelihood, as well as at the landscape scale, in terms of land cover change and natural resource management.

This chapter presents the development of an agent-based model (ABM) to the study of shifting cultivation in Vietnam, which captures the relationships between the political power, the community structure and the need to clear the forest for agriculture. Previous modeling work is reviewed to demonstrate the advances of this application over other related work, which is followed by a description of the model. Implementation, validation and scenario development are also discussed.

30.2 Previous Work on Modelling Shifting Cultivation

Early research into shifting cultivation (e.g. Conklin (1957), Spencer (1966) and Watters (1971)) was largely descriptive in nature and did not explain its dynamic structure or the driving forces. However, simulation modeling in the 1980s allowed further developments, e.g. the forest regeneration model of Wilkie and Finn (1988), which simulated the long term effects of shifting cultivation in north-eastern Zaire. The results showed that population pressure, land tenure systems and fallow length have a strong influence on the spatial patterns of the landscape. In another study, Dvorak (1992) developed an analytical model that established a relationship between shifting cultivation, the labour economy and the fallow cycle in West Africa. This was followed by GEOMOD1 and GEOMOD2 by Hall et al. (1995), which were designed to estimate the total amount of land use change in Southeast Asia and sub-Saharan tropical Africa. Although these models produced results with a high accuracy, ranging from 74% to 96%, they performed less well in predicting the patterns of shifting cultivation when compared to permanent agriculture. Similar to the GEOMOD application, the dynamic spatial model of Gilruth et al. (1995) was based on forest structure, productivity, elevation and the distance to towns and cities. This model predicted the locations of shifting cultivation at different time steps for an area in the Republic of Guinea, West Africa, but the results did not compare well with remotely sensed data.

The reasons for the high uncertainty in the above simulated models are their simple assumptions about the decision-making process, some of which view shifting cultivation simply as a transformation from forest to agriculture. To better capture the decision-making process, Brown (2005) developed a spatiotemporal model of shifting cultivation and forest cover dynamics in the Congo basin. The basic concept for making a decision in this model is based on utility maximisation. The decision making of the farmers is based on personal preferences, which are represented by the relative importance of the various factors influencing the farmers in making land use decisions. The importance of each factor or criterion was quantified by undertaking a household survey in which the respondents were asked to assign scores to reflect the importance of a given factor to the land use decision-making process. The fields with the greatest net benefit would then be selected in a given year subject to labour constraints. This approach to collecting data on personal preferences has been adapted for use in this research. However, the decision function does not include the influence of the socio-economic circumstances of the local area yet it is clear that these external factors (i.e. markets, policies) have an influence on the land use decisions made by a household (Castella et al. 2007; Fresco et al. 1990; Haggith et al. 2003). Therefore, the model of Brown (2005) cannot be used effectively in its current form for policy analysis. Furthermore, field selection operates separately for each household within the individual landholding but it does not clearly represent the interactions between different households in the decision-making process. It is recognised that farmers are often influenced by their neighbours' choices, e.g. sharing fencing works with neighbours (Jepsen et al. 2006) or other kinship relations (Gilruth et al. 1995) Therefore, a model that allows for the simulation of several households simultaneously and interactively is necessary in order to model shifting agriculture at a village level.

Another relevant study in the literature is the ABM of Sulistyawati et al. (2005). It was developed to compare different land use strategies within a subsistence economy of rice and rubber cultivation under scenarios of fluctuating rubber prices and a changing population in Indonesia. The model consisted of modules to calculate: (i) the population dynamics; (ii) land use decision-making; (iii) vegetation dynamics; and (iv) production. However, there are limitations to this model associated with the way in which the decisions were modeled, and the rules were formulated for swidden field selection. To evaluate the attractiveness of potential sites, weights (or personal preferences) were assigned, but these were applied equally to all households, which means that all households effectively use the same decision function. The socio-economic conditions of the households should be taken into account in the decision making as outlined in the research by Brown (2005). Another limitation of this model is that the potential rice yield is only estimated based on fallow age. However, with the same biophysical conditions in a rice field, different management practices could produce quite different yields. It is possible in ABMs to calculate yield separately for each land parcel as a function of the land characteristics and socio-economic conditions of its owner. This modification to the model would have made it more realistic.

Jepsen et al. (2006) were the first to publish on the direct application of agent-based modelling to shifting cultivation in Vietnam. Household agents make decisions in the model regarding which fields they will cultivate based on expected yields and labour investment. Fallow age, which is used to calculate the potential yield and labour required to work on the field, is used to estimate land productivity. The findings of the research showed that local farmers behave as would be expected according to well established shifting cultivation theory, and the spatial output of the model resulted in a good match with the data derived from remote sensing. The ABM of Jepsen et al. (2006) has an advantage compared to the model of Brown (2005) as it is able to capture the interaction between household agents in decision making. However, the model is still overly simplistic in its assumptions such as ignoring all economic factors, e.g. price or household potential capital that could be important in other regions. Adding more constraints to the choice of agents such as soil condition, water availability or distance to the household would make the model more realistic. The rules governing field selection are also too simple and therefore some important decision-making aspects (e.g. the linkage between production goals, preferences and decisions) could be lost. This limitation means that the model is unable to simulate the changing context of agents such as the change in land use strategies due to the application of new policies. A further limitation noted by the authors is the simplistic yield relationship in the model, which is simply derived from the cell age and the labour cost based on the number of adjacent cultivated cells. Other potential factors such as fertilisers, agricultural extension etc. could be added to the model. However, this would require much more field data and further statistical analysis.

The research that is most similar to the application described in this chapter is the Land use Dynamic Simulator (VN-LUDAS) of Le (2005) and Le et al. (2008). The type of study area is the same and the categorisation of households into groups, that are then assigned a specific land use strategy for selecting fields, has been adapted from this research. This categorisation is dynamic so households may change their strategy during a simulation. This is something that is currently not incorporated into other models such as those developed by Jepsen et al. (2006), Sulistyawati et al. (2005) and Wada et al. (2007). However, Le (2005) classified households into groups and then built separate relationships to determine field selection based on the field characteristics. This approach essentially predicts the probability of land use types based on explanatory variables but it does not explain why farmers make these choices.

Instead the method of Brown (2005), which uses household context to predict the probability of selecting land use strategies (i.e. types of personal preferences), was incorporated into this research. Another aspect of the VN-LUDAS model that has been adapted here lies in the land transition module. A large sample was collected for exploring the relationship between stand basal area and forest development. Le (2005) provided ranges of thresholds indicating the transitions between forest land cover types. This calibration is particularly significant for Vietnamese conditions because the forest is quite area specific, and it might not be suitable to use parameters gathered for other regions. Some of these parameters have been used in this ABM application. Furthermore, the VN-LUDAS model incorporates a response mechanism for policy change but the probability that households will violate constraints, including land use policies (e.g. forest protection regulations), is random and treated equally by all household agents. The VN-LUDAS model is therefore limited in capturing the association between individual accountability and the pressure exerted by local authorities to implement policy in the decision-making process of the land users.

The ABM developed as part of this research has incorporated elements from previous research into the development process but it has also attempted to address some of the shortcomings outlined above.

30.3 Study Area and Datasets Used

A model was developed for Binh Son-1 village in Ky Son district, Nghe An province. This village is typical of upland villages in Vietnam, and the ethnicity of the village is almost 100% Kho Mu for whom shifting cultivation is the only agricultural practice. The elevation of the village ranges from 400 to 1,200 m above sea level with a total area of about 7.4 km^2. The village contains 88 households, with a total population of 436 people in 2007. There have been several policies and forest protection programmes implemented in the village. However, a considerable amount of illegal shifting cultivation still occurs in the protected forests.

The data used in this research consists of a survey dataset that was collected during 2007 as well as land cover maps derived from TM satellite images in 2000, 2005 and 2006. The survey data collected include maps that were obtained from participatory mapping exercises, and socio-economic data about the households from questionnaires. Participatory rural appraisal (PRA) was used for the mapping tasks (Chambers 1994), including village territory, soil status and land use maps. The questionnaires were carried out randomly on 63 households, and were specifically focused on gathering information about household conditions related to livelihoods and personal preferences for cultivation. The information on personal preferences was collected based on the method proposed by Brown (2006) in which the survey respondent rated land decision criteria on a zero-to-ten scale according to the importance they placed on each factor. Additional interviews were conducted with key villagers and local authorities about the customary laws and the institutional and other rules that govern the land use activities of the local people.

30.4 Description of the ABM of Shifting Cultivation

An ABM of shifting cultivation consists of household agents, land agents and global parameters. Household agents represent shifting cultivators, who are capable of autonomous actions in the biophysical environment (i.e. on land patches or fields). Land agents are a set of grid cells representing land patches with their characteristics such as soil, land cover and land use. Global parameters are the set of external conditions that include important socio-economic and policy-related parameters. The global environment is considered as the external driving force of land use change. It varies from year to year but applies across the whole grid and influences all agents. Within the shifting cultivation system, household agents interact and make changes to their biophysical environment (i.e. the land patches) while the environment also constrains the activities of the household. The household agents, the policy response mechanism and the land agents are now briefly described below.

30.4.1 The Household Agent

Each household is an autonomous agent that is embedded within a potential decision routine, where field selection and policy adaptation are the two main components:

Field selection routine: this contains the rules that govern the choice of which fields or land patches to cultivate. The rules governing choice are based on the traditional economic behavioural theory of utility maximisation, which is widely applied in the simulation community (Le et al. 2008; Russell and Norvig 1995; van den Bergh et al. 2000; Wooldridge 2002). The main assumption of this theory as applied in this model is that the agent will rationally choose the field that is preferred to or no worse than all the others. The utility represents the ranked personal preferences of a household agent for all possible patches or combination of patches, the set of which is also restricted by labour requirements and accessibility (i.e. physical and tenure accessibility). The utility function (u_i) for field selection is formulated as:

$$Max\sum_{s=0}^{S} u_{st}; \quad u_{st} = [r_{it}][Z_{st}] \tag{30.1}$$

where *i* is a household, *s* is a field or land patch within the accessible area *S* at time *t*; *Zst* are the land specific factors, and *rit* are the preference parameters of the household, which can be derived from the household context (Brown 2005). There is also a stochastic component, which captures the uncertain nature of the perceptions. By multiplying the vector of land variables with the vector of personal preferences of a particular household, the utility function represents the suitability of a given field to the household. Fields with the largest utility values are, therefore, the ones that will be potentially selected by the agent. The general algorithm of the field selection routine is shown in Fig. 30.1.

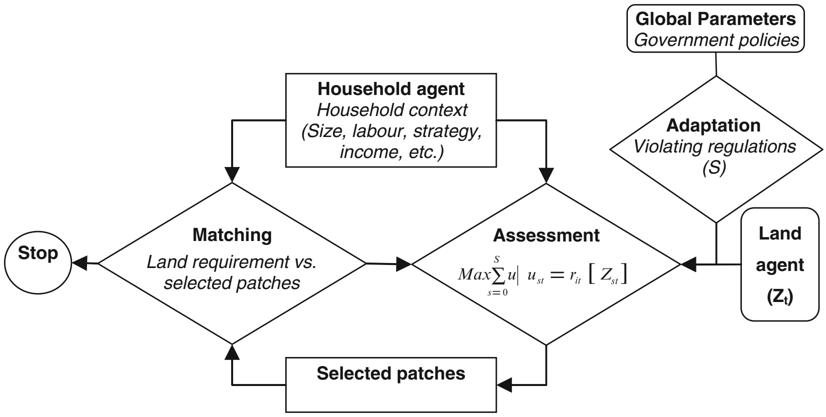


Fig. 30.1 Field selection algorithm

Adaptation routine: this contains the simulation of responses by individual households to policies. The general response mechanism is adopted from the Institutional Analysis and Development (IAD) framework (Ostrom 1999), which has been specifically developed for organising research on institutions and governance structures. The factors that drive an agent's decisions are a combination of incentives (e.g. likely economic benefits) and beliefs (e.g. political accountability) (Clement and Amezaga 2009). The policy response is, therefore, based on the assumption that an individual will behave by trading off their vested interests against their individual accountability or social responsibility (Crano 1995). The general response mechanism is formulated as:

$$Response(R) = f\left\{\left(vestedInterests, accountability\right) * \, pressure\right\} \qquad (30.2)$$

where vested interests comprise the expected economic outcome that the policy implementation will bring to the household. This variable can be expressed by the rate of expected income after and before policy implementation. The accountability is quantified based on a farmer's background and the influence of the body that implements the policy, while the pressure is based on the priority with which the policy is implemented.

The response of an agent to a policy is in the form of a response index that indicates how related policies satisfy each individual. The minimal response index is zero, which indicates no policy is to be implemented or both vested interests and accountability are at the lowest level. Low values indicate potential opposition to the policy while higher values indicate support. The response does not directly change the land use decision of an agent but it contributes to the probability that the agent will undertake activities that support or oppose a given policy and to then act accordingly, e.g. cultivate in the protected forest.

The response affects the *field selection* routine by modifying the definition of total accessible land *S*. For example, land allocation policies are intended to encourage farmers to cultivate only inside the allocated areas. If these policies are implemented, then farmers that respond positively will only choose fields within the total accessible land *S* that fall within the allocated areas. In contrast, farmers that respond negatively to the policy will extend their search to land patches outside the allocated areas and broaden the definition of total accessible land *S*. Similarly, forest protection policies do not allow farmers to cultivate inside any protected forests. Farmers that respond negatively to this policy will not consider the boundary between protected and non-protected areas in their field selection.

30.4.2 The Land Agent

The land agents represent the biophysical environment as a series of grid cells, each of which has associated properties such as land cover, soil, slope, etc. The land patches are updated through direct interaction with the household agents, neighbouring land agents and natural succession processes. The routine that controls the dynamics of the land agents is responsible for regularly updating the dynamic status of the land patches. The land use is simply updated after each field selection since it is a direct driving force of land use change in the uplands (Jakobsen et al. 2007; Jepsen et al. 2006). However, updating land cover is much more complicated as it is driven not only by human intervention but also by natural succession, which is beyond human control. The basic transition model used in this research is expressed as:

$$S_{t+1} = f\left(S_t, P_t, T_t\right) \tag{30.3}$$

where S_{t+1} and S_t are the vegetation state of patch *S* at time $t+1$ and *t*; P_t is the collective gross development at time *t* depending upon both internal and external forces acting on patch *S*; and T_t are the transition rules. P_t is determined from the fallow age and ground basal area, which is estimated from the parameters and empirical models in the literature (Le et al. 2008). The transition rule *T* is estimated using fuzzy sets (see Ngo et al. (2009) for more details).

30.4.3 Model Implementation

The model operates iteratively on an annual production cycle. Each simulation starts with an initialisation stage and continues with cycles of three main phases: categorisation, field selection and update (see Fig. 30.2).

The initial phase: The land patch attributes are imported to the model directly as GIS data layers, which include land cover, fallow age, ground basal area, land use, soil, slope and maps of buffered areas (i.e. distance to settlement areas, roads and streams). Household agents are then created, where the number of agents is equal to the number of households in the study area. Household profiles are assigned based

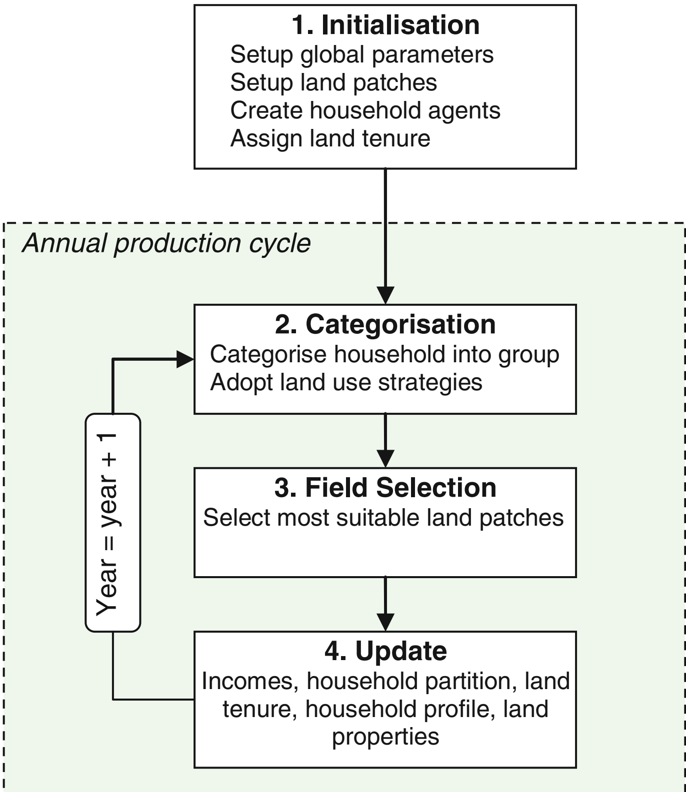


Fig. 30.2 A general simulation cycle of the SCM

on the household survey samples (i.e. mean and standard error). Each household is located randomly in the settlement area. Finally, fields for cultivation are allocated to the households according to the land tenure status that is identified by the global parameters at the time of the simulation.

Household categorisation phase: This phase categorises households into groups based on their characteristics and assigns the field selection strategies (i.e. personal preferences [r_{it}]) to the individual household agents based on the probability that the household will select a certain strategy. During a single iteration of the model, if a household agent changes the group to which it belongs, the field selection strategies will be updated accordingly.

Field selection stage: This phase involves field selection, policy adaptation and prediction of possible outcomes of human activities. The simulation steps are as follows:

- Adopt policies by estimating the policy response index for each household agent, and determine whether or not the agent will violate the land use regulations.
- Search for the most suitable fields based on the perception of the household about land resources and accessibility. The accessibility of land is determined based on the response to the policy in the previous step.

- Decide on the number of patches to be selected based on basic food requirements and the availability of land.
- Record the human interactions (e.g. tenure) for each of the land patches.

Update phase: The phase consists of updating the land tenure, land cover and other household characteristics. Land tenure is regularly updated according to the land use rights in place after each field selection and after changes to the characteristics of the household agents, which are updated using survey information collected at the study site. If the simulation runs long enough, it is possible to produce new household agents based on the demographic attributes of the local population. New households are created when a household immigrates or a newly married couple moves to a new household (i.e. household partitioning). This partitioning can occur when parents are too old or the household size becomes too large. New household profiles are created stochastically based on the mean value and standard errors of the survey data set. The newly created household acquires land tenure based on the global parameters set by the model user.

Each simulation cycle ends when the update phase finishes. However, the simulation can continue to run for other simulation cycles as long as required by the model user. The ABM developed as part of this research operates in NetLogo (Wilensky 1999) with extensions written in Java.

30.4.4 Model Validation

Validation is an essential process to ensure that the model can be applied to examine the patterns of shifting cultivation and produce reliable data for policy analysis. Part of the validation process involves calibrating the model. Some of the parameters in the model were determined from the survey data and a statistical analysis. These include the personal preferences r_{it} and the policy response. The rest of the parameters were determined using a genetic algorithm. Details of how these parameters were obtained can be found in Ngo (2009).

Validation was also applied to the field selection and policy response routines as well as to the final model outputs of land cover change. The Mean Nearest Neighbour distance (MNN) (Campbell 1995) was used to test the hypothesis that the field selection routine behaves differently from a random selection. The test results showed that the selected fields were significantly different from a random selection and the clustering tendency in the simulated maps is quite similar to that of reference maps, which were derived from TM Landsat satellite images.

The policy response was validated by examining the amount of illegal cultivation predicted by the model, which is indicated by the number of land patches cultivated inside the protected forests. The simulated results were comparable with the results from reference maps (i.e. 20.5% compared to 24.6%).

The overall operation of the model was validated by comparing the spatial structure of land cover maps and reference maps derived from satellite images.

Table 30.1 Policy setting for the baseline scenario

Type of policy	Start	Term	Implemented body	Pressure
Land allocation	NA	–	–	0
Techn. support	NA	–	–	0
Credit support	2000	30	Headmen, VCP	4
Extension training	1999	30	Headmen	5
Forest protection	1996	50	Headmen, VCP	8
Population	2000	30	Women's Union	4

Note: *NA* not applicable, *VCP* Vietnam Communist Party cell

Table 30.2 Percentage change in land cover type in the baseline scenario in 2020

Scenarios	Swidden	Bare soil	Grass-shrub	Open forest	Dense forest
Land cover in 2006	10.8	8.8	9.9	53.8	15.9
Baseline (S_0)	13.7	16.7	25.7	31.9	11.6

This evaluation was conducted using a Multiple Resolution Goodness-of-fit index, F_t (Costanza 1989). The result produced an F_t of 80%, indicating a good fit between the model outputs and the reference data, and the model can now be used for further analysis. Details of the validation process of the model can be found in Ngo (2009).

30.4.5 Building a Baseline Scenario

The application of the model in assessing policy impacts is based on formulating a scenario that quantifies the possible transitions and the approximate processes leading to the changes in shifting cultivation in the study site. As an initial start, a baseline scenario was applied to the model which assumes that the development context of the village follows the current trend (i.e. 2007) and remains stable during the simulation. The parameter settings, which are based on the survey in 2007, are shown in Table 30.1. The term of the simulation is 14 years, from 2006 to 2020, which coincides with the time frame of the current forest development strategy (District People Council of Ky Son 2007; MARD 2007). It also approximates a time period suggested in previous research (i.e. 10–20 years) (Huss 1988). The starting point is 2006 because the latest available land cover map (derived from satellite images) is for the year 2006. All scenarios were run ten times and the results were averaged.

The percentage of land cover types and their changes over time are shown in Table 30.2 and Fig. 30.3 respectively. The results show that under a business as usual scenario, the areas under bare land and shrubs increase while open and dense forest decreases by 2020. However, the amount of forest appears to stabilise as the simulation progresses. This finding reflects continued encroachment upon the forest for shifting cultivation and conversion to shrub through fallow and natural succession.

The visual analysis of the land cover maps between the start of the simulation in 2006 and 2020 shows that the greatest change was an increase in the cultivated land

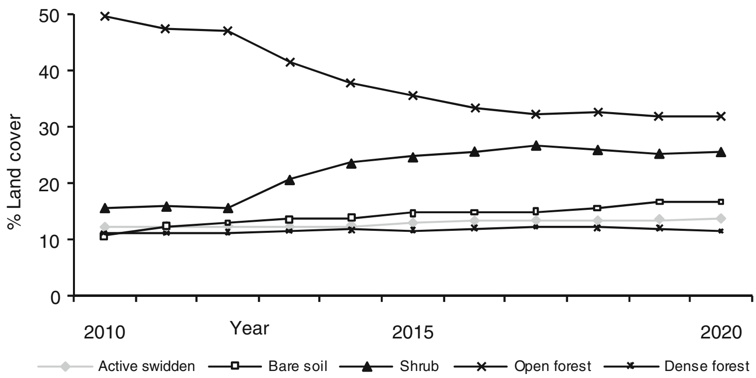


Fig. 30.3 Land cover change over 14 years from the baseline scenario

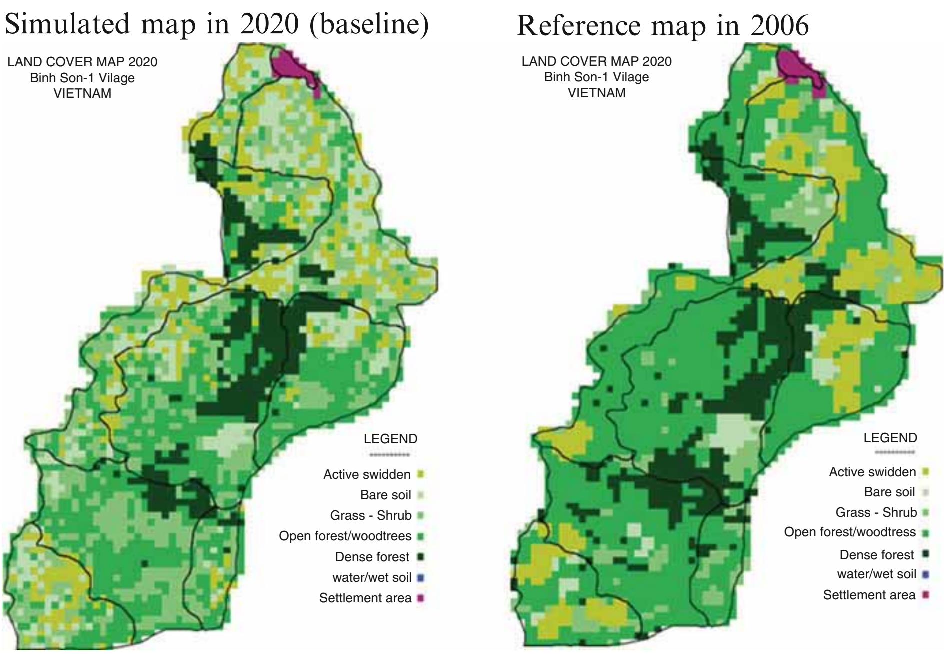


Fig. 30.4 Land cover maps in 2020 compared to 2006

(Fig. 30.4). Although some dense forests remain at the end of the simulation, these areas are either holy forests or forests located in inaccessible locations such as on steep slopes.

Therefore, under a business as usual scenario, the amount of shifting cultivation will increase. The model was also run for a range of different scenarios related to the implementation of potential government policies in the future. The details of these scenarios and their results are described in Ngo (2009).

30.5 Conclusions

This chapter has presented the application of an ABM to study shifting cultivation in an upland village in Vietnam. The application has shown that the relationship between policy implementation and shifting cultivation can be explicitly described using an ABM approach, in particular combining the household context and institutional factors into the land use decision-making process. Given the ability to capture these complex relationships, the model can be used for exploring alternative scenarios and facilitating analysis of policy options towards sustainable forest management.

It is necessary to note, however, that the model has only been validated to date using data from a single village. To develop a really useful and reliable decision support tool for the future, which is the ultimate goal of this research, more data from villages with different types of internal community dynamics and/or policies implemented by the government are required.

Acknowledgements We would like to acknowledge the support of the Vietnamese Government and the School of Geography, University of Leeds, in funding this research project.

References

Angelsen, A. (1995). Shifting cultivation and deforestation: A study from Indonesia. *World Development, 23*, 1713–1729.

Brown, D. R. (2005). *A spatiotemporal model of shifting cultivation and forest cover dynamics*. Ithaca: Cornell University.

Brown, D. R. (2006). Personal preferences and intensification of land use: Their impact on southern Cameroonian slash-and-burn agroforestry systems. *Agroforestry Systems, 68*, 53–67.

Campbell, D. J. (1995). Aggregation and regularity: An inclusive one-tailed nearest-neighbour analysis of small spatially patchy populations. *Oecologia, 106*, 206–211.

Castella, J.-C., Pheng Kam, S., Dinh Quang, D., Verburg, P. H., & Thai Hoanh, C. (2007). Combining top-down and bottom-up modelling approaches of land use/cover change to support public policies: Application to sustainable management of natural resources in northern Vietnam. *Land Use Policy, 24*, 531–545.

Chambers, R. (1994). The origins and practice of participatory rural appraisal. *World Development, 22*, 953–969.

Clement, F., & Amezaga, J. M. (2009). Afforestation and forestry land allocation in northern Vietnam: Analysing the gap between policy intentions and outcomes. *Land Use Policy, 26*, 458–470.

Conklin, H. C. (1957). *Hanunoo agriculture: A report on an integral system of shifting cultivation in the Philippines*. Rome: FAO.

Conklin, H. C. (1961). The study of shifting cultivation. *Current Anthropology, 2*, 27.

Costanza, R. (1989). Model goodness of fit: A multiple resolution procedure. *Ecological Modelling, 47*, 199–215.

Crano, W. D. (1995). Attitude strength and vested interest. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences*. Mahwah: Erlbaum.

Crooks, A. T., & Heppenstall, A. J. (2012). Introduction to agent-based modelling. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 85–105). Dordrecht: Springer.

District People Council of Ky Son. (2007). *The master plan of Ky Son district from 2007 to 2020*. Nge An, Ky Son People Council.
Do, D. S. (1994). *Shifting cultivation: Its social, economic and environmental values relative to alternative land-use* (IIED forestry and land use). London: International Institute for Environment and Development.
Dvorak, K. A. (1992). Resource-management by West-African farmers and the economics of shifting cultivation. *American Journal of Agricultural Economics, 74*, 809–815.
Fox, J., Dao, M. T., Rambo, A. T., Nghiem, P. T., Le, T. C., & Leisz, S. (2000). Shifting cultivation: A new old paradigm for managing tropical forests. *BioScience, 50*, 521–528.
Fresco, L., Huizing, H. G. J., Keulen, H. V., Luning, H., & Shipper, R. (1990). *Land evaluation and farming systems analysis for land use planning – FAO guidelines*. Rome: FAO.
Geist, H. J., & Lambin, E. F. (2001). *What drives tropical deforestation? A meta-analysis of proximate and underlying causes of deforestation based on subnational scale case study evidence* (LUCC Report Series No. 4). Louvain-la-Neuve: University of Louvain.
Gilruth, P. T., Marsh, S. E., & Itami, R. (1995). A dynamic spatial model of shifting cultivation in the highlands of Guinea, West-Africa. *Ecological Modelling, 79*, 179–197.
Haggith, M., Muetzelfeldt, R., & Taylor, J. (2003). Modelling decision-making in rural communities at the forest margin. *Small-Scale Forest Economics, Management and Policy, 2*, 241–258.
Hall, C. A. S., Tian, H., Qi, Y., Pontius, G., Cornell, J. (1995). Modelling spatial and temporal patterns of tropical land use change. *Journal of Biogeography, 22*(4/5), 753–757.
Huss, W. R. (1988). A move toward scenario analysis. *International Journal of Forecasting, 4*, 377–388.
Institute for Ethnic Minority Affairs. (2005). Research on policies for fixed cultivation and sedentarisation in Vietnam: Within the framework of Programme 135 investment. In N. T. Le, H. D. Le, V. T. Nguyen, & V. H. Do (Eds.), *Final report for DFID*. Hanoi: Department for International Development, UK (DFID).
Jakobsen, J., Rasmussen, K., Leisz, S., Folving, R., & Quang, N. V. (2007). The effects of land tenure policy on rural livelihoods and food sufficiency in the upland village of Que, North Central Vietnam. *Agricultural Systems, 94*, 309–319.
Jepsen, M. R., Leisz, S., Rasmussen, K., Jakobsen, J., Moller-Jensen, L., & Christiansen, L. (2006). Agent-based modelling of shifting cultivation field patterns, Vietnam. *International Journal of Geographical Information Science, 20*, 1067–1085.
Kerkhoff, E., & Sharma, E. (Eds.). (2006). *Debating shifting cultivation in the Eastern Himalayas: Farmers' innovation as lessons for policy*. Kathmandu: Quality Printers Pvt. Ltd.
Le, Q. B. (2005). *Multi-agent system for simulation of land-use and land cover change: A theoretical framework and its first implementation for an upland watershed in the Central Coast of Vietnam*. Gottingen: ZEF.
Le, Q. B., Park, S. J., Vlek, P. L. G., & Cremers, A. B. (2008). Land-Use Dynamic Simulator (LUDAS): A multi-agent system model for simulating spatio-temporal dynamics of coupled human-landscape system. I. Structure and theoretical specification. *Ecological Informatics, 3*, 135–153.
MARD (Ministry of Agriculture and Rural Development). (2007). Vietnam forestry development strategy for the period of 2006–2020. Agriculture Publisher, Hanoi, pp. 124.
Ngo, T. A. (2009). *Simulating spatial patterns of shifting cultivation*. Unpublished PhD thesis, School of Geography, University of Leeds, Leeds.
Ngo, T. A., See, L. M., & Drake, F. (2009). An agent-based approach to simulating the dynamics of shifting cultivation in an upland village in Vietnam. *European Journal of GIS and Spatial Analysis, 19*(4), 493–522.
Ostrom, E. (1999). Institutional rational choice: An assessment of the IAD framework. In P. Sabatier (Ed.), *Theories of the policy process*. Boulder: Westview Press.
Russell, S., & Norvig, P. (1995). *Artificial intelligence: A modern approach*. Englewood: Prentice Hall.

Shanks, E., Bui, D. T., Nguyen, T. K. N., Maxwell, O., & Duong, Q. H. (2003). *Community driven development in Vietnam: A review and discussion platform*. Hanoi: The World Bank.

Spencer, J. E. (1966). *Shifting cultivation in Southeastern Asia*. Berkeley: University of California Press.

Sulistyawati, E., Noble, I. R., & Roderick, M. L. (2005). A simulation model to study land use strategies in swidden agriculture systems. *Agricultural Systems, 85*, 271–288.

Tran, D. V. (2006). Indigenous fallow management with Melia azedarach Linn in northern Vietnam. In M. Cairns (Ed.), *Voices from the forest integrating indigenous knowledge into sustainable upland farming*. Washington, DC: RFF Press.

Tran, D. V., Nguyen, V. Q., & Mai, V. T. (2005). *Decentralisation process and its impacts on livelihoods of ethnic minority groups: A study on decentralisation process in forest management in northern and north central uplands of Vietnam*. Hanoi: The Agricultural Publishing House.

van den Bergh, J. C. J. M., Ferrer-i-Carbonell, A., & Munda, G. (2000). Alternative models of individual behaviour and implications for environmental policy. *Ecological Economics, 32*, 43–61.

Vuong, X. T. (2001). Changing land policies and its impacts on land tenure of ethnic minorities in Vietnam. In K. Suryanata, G. Dolcemascolo, R. Fisher, & J. Fox (Eds.), *Policy frameworks for enabling successful community-based resource management initiatives*. Hawaii: East-West Center.

Wada, Y., Rajan, K. S., & Shibasaki, R. (2007). Modelling the spatial distribution of shifting cultivation in Luangprabang, Lao PDR. *Environment and Planning B-Planning & Design. 34*, 261–278.

Watters, R. F. (1971). *Shifting cultivation in Latin America*. Rome: Food and Agriculture Organization.

Wilensky, U. (1999). NetLogo. [online]. Accessed 1 Nov 2007. Available from World Wide Web: http://ccl.northwestern.edu/netlogo/

Wilkie, D. S., & Finn, J. T. (1988). A spatial model of land use and forest regeneration in the Ituri forest of northeastern Zaire. *Ecological Modelling, 41*, 307–323.

Wooldridge, M. (2002). *An introduction to multi-agent systems*. New York: Wiley.

Chapter 31
Towards New Metrics for Urban Road Networks: Some Preliminary Evidence from Agent-Based Simulations

Arnaud Banos and Cyrille Genre-Grandpierre

Abstract Road networks are complex entities, which are arranged hierarchically both in their structure (topology) and by speed. This property has a strong influence on their performance, both at an individual and collective level. Indeed, they intrinsically favour car use, especially for distant trips. In that sense, they may contribute actively to urban sprawl, a non desirable property of urban growth. In this chapter, we propose and explore a strategy aimed at regulating and even reversing such a "speed metric". Using agents, we simulate road traffic on various road network structures and show how limited but well targeted actions can have a strong global impact on the system.

31.1 Introduction

From a planning perspective, road networks do not play fair game: the farther one goes, the more efficient his/her travel is. Indeed, as road networks are hierarchically organised by speed, the proportion of a trip spent at low speeds on minor roads decreases while the proportion of time spent on faster roads increases. This rather evident phenomenon is rarely put forward nor exploited according to its real importance in urban planning. Indeed, the performance of road networks are frequently evaluated using global indices (such as mean speed), without taking into account the distribution of trip lengths.

A. Banos (✉)
Géographie-Cité, and ISC-PIF, CNRS, Paris, France
e-mail: arnaud.banos@parisgeo.cnrs.fr

C. Genre-Grandpierre
ESPACE, Université d'Avignon/CNRS, Avignon, France
e-mail: cyrille.genre-grandpierre@univ-avignon.fr

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_31,

This structurally-based "speed metric" ensures car drivers the possibility to travel farther without necessarily increasing their transportation time in the same proportion. Obviously, public transport modes (bus, tramways) benefits less from such a property, limited as they are by the frequent stops they are supposed to make along their route. Therefore road networks, by their mere structure, intrinsically favour car use, especially for distant trips. As a direct consequence, they also encourage disassociation – within the limits of daily time budgets – the various places of life (home, work) and finally actively contribute to urban sprawl.

Would it be feasible to reduce these negative global scale effects by penalizing road users accordingly to the length of their trips?

In this chapter, we present an agent-based model (S3) aimed at exploring this critical issue. Based on a comprehensive description of the problem at hand, the main components of the model are described and illustrated. Results from first simulations are then presented, suggesting that new metrics could be obtained with rather simple but well targeted actions.

31.2 Road Networks and Trip Efficacy

As Gutiérrez et al. (1998) proposed, efficacy of road networks can be defined in a very simple way. Let's imagine two places, A and B (Fig. 31.1), connected by a road network of length d_r.

It is then very useful to compare the time needed to travel from A to B, that is $t(d_r)$, with the Euclidean distance d_e separating A and B such that:

$$E_{ij} = \frac{d_e(i,j)}{t\left(d_r(i,j)\right)} \tag{31.1}$$

where

$$d_e(i,j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{31.2}$$

Trip efficacy E_{ij} can then be seen as an average speed,[1] defined as the Euclidean distance between origin and destination of a trip divided by its real duration. When describing a road network, efficacy may be privileged compared to speed indicators as it integrates morphological dimensions. Indeed, a given road network may provide high speeds but a poor efficacy with the proportion of detours growing.

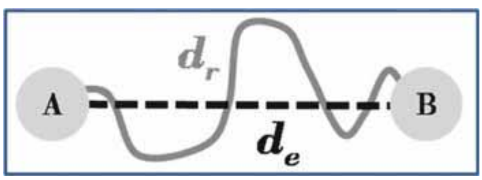


Fig. 31.1 Definition of trip efficacy

[1] In the rest of the chapter it will be expressed in km/h or in m/s, depending on the context.

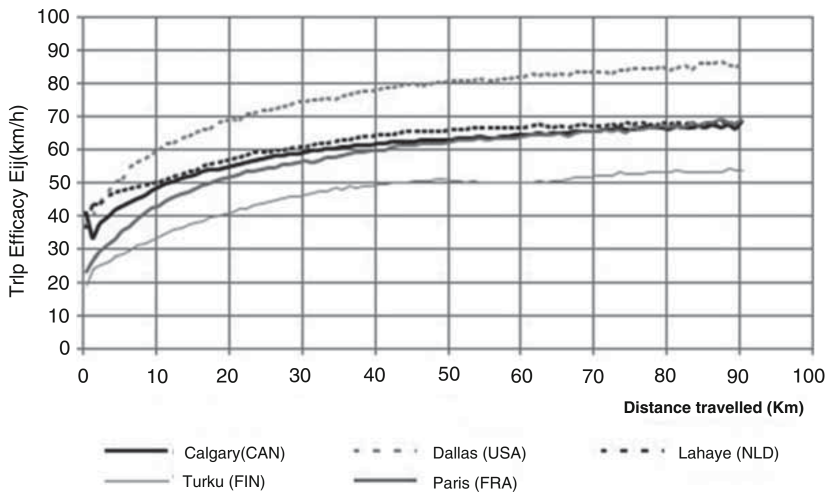


Fig. 31.2 Example results obtained for various cities (average E_{ij} values based on 250,000 pairs of locations selected at random). The five databases used here are directly comparable. However, one may be cautious in interpreting small differences between curves as they are based on average efficacy values

When plotting this index of efficacy against the distance travelled, one can verify that, on average, the level of performance (efficacy) increases non-linearly with the distance travelled (Fig. 31.2).

The main hypothesis we draw here is that by inverting this trend (i.e. short-distance trips being favoured instead of long ones), one could promote higher densities and better functional proximities in urban design, according to the hypothesis of the rational locator (Levinson and Kumar 1994). Therefore, while reinforcing the economic value of physical proximity in a significant way and encouraging short trips, one could actively contribute in reducing urban sprawl.

We call this new kind of metric the "slow metric" (Genre-Grandpierre 2007). However, as road networks are organised hierarchically both by their topology and the speed of their components, it is obvious that one cannot invert this trend simply by reinforcing speed limits or even imposing constant speeds. Coupling this last option with a dramatic alteration of network topology (e.g. imposing a regular structure) would just cancel the trend, but would not invert it (Fig. 31.3).

A war metaphor may then help us overcome this obstacle. Indeed, let us imagine with Stanley and Ostrowsky (1985) a field battle covered with land mines. Moreover, let us assume the distribution of these nasty weapons is fractal, such that one can show an inverse power relation between the number of clusters of size S_i and the size of these clusters:

$$N(S_i) \propto S_i^{-\alpha} \tag{31.3}$$

Let us now imagine a soldier being given a dangerous mission by his officer to deliver a message from place *A* to place *B*. Obviously, this soldier has to find a way to avoid land mines on his way. Let us imagine he has the possibility to find the

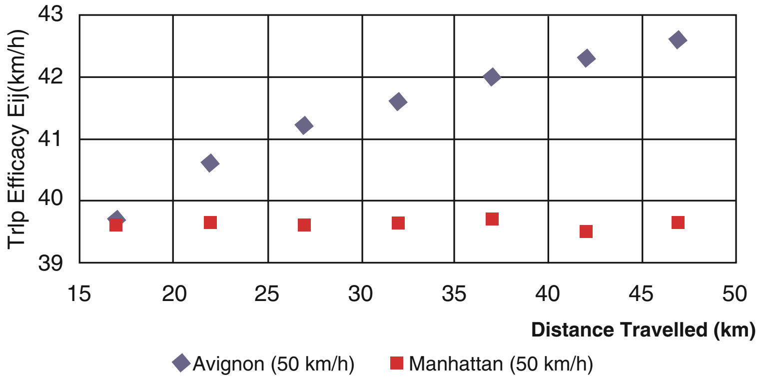


Fig. 31.3 Topology and speed both contribute to structure road networks (average E_{ij} values based on 250,000 pairs of locations selected at random)

shortest safe path from *A* to *B* and that he is able to deliver the message in a given time t_1. As soon as he arrives, his officer asks him to have a 5 min rest and then to deliver a new message from place *B* to *C*. Estimating from his topographic map that the Euclidean distance between *B* and *C* is three times the Euclidean distance between *A* and *B* (i.e. $d_e(B,C) = 3d_e(A,B)$ according to our previous notation), his officer finally asks him to deliver the new (and by the way urgent) message in a given time t_2, proportional to that relation: $t_2 = 3t_1$. However, since the soldier has read the Stanley and Ostrowsky (1985) book, he is able to (respectfully) suggest to his officer that the probability of him delivering the message in this time is very low, if not impossible. Indeed, as the distribution of clusters follows a power law, the time needed to connect increasingly distant places does not increase in a linear fashion, but more likely in a non-linear one:

$$t_n \propto t_1^{\beta} \tag{31.4}$$

Coming back to the issue at hand, what does this war metaphor suggest? Well, a very direct interpretation can be derived: in a road network, traffic lights might be seen as land mines while their duration resembles the cluster's size of the latter. In other words, by correctly positioning a reduced number of traffic lights with their duration following a precise probability law (power law), one could reach the metric inversion goal. Simulations ran in a GIS indeed that specific combinations of location and duration of traffic lights on a network may produce the desired effect (Fig. 31.4).

By modifying the number and duration of traffic lights, we thus obtain various efficacy curves, favouring short-distance trips. These first positive results encouraged us to explore this problem with a more dynamic and microscopic model, in order to address some keys issues identified so far such as: (a) the number, location and duration of stops; (b) the possible structural effect induced by network topology; and (c) the possible impacts on traffic, including congestion and traffic jams.

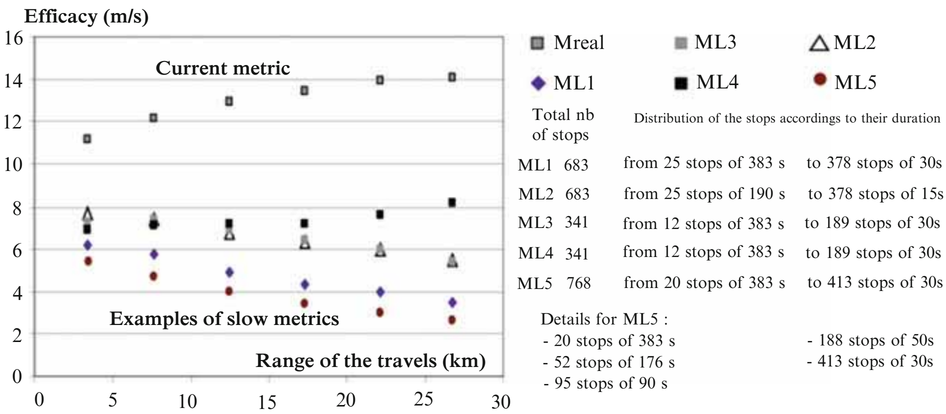


Fig. 31.4 Progressive inversion of current metric in Carpentras (FRA) obtained with various combinations of number, location and duration of traffic lights (average E_{ij} values based on 250,000 pairs of locations selected at random)

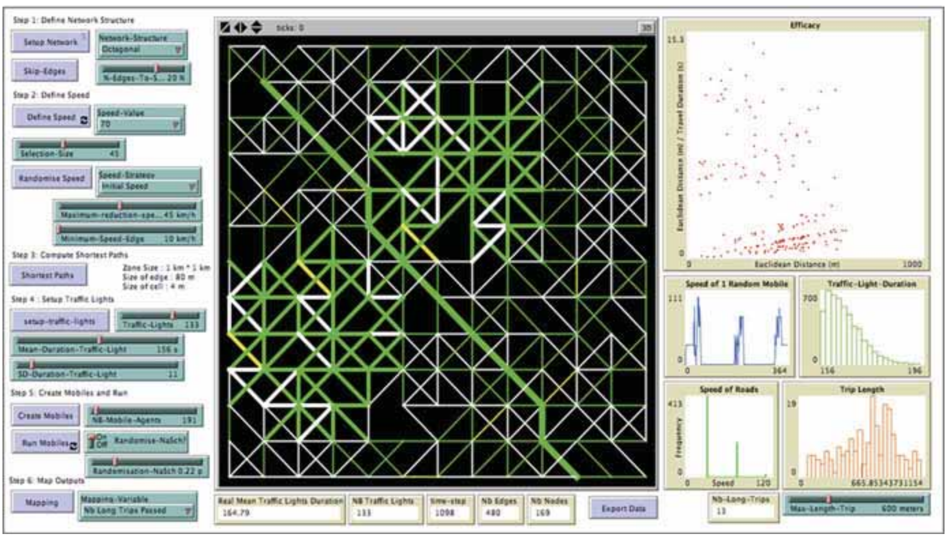


Fig. 31.5 The S3 platform, developed in NetLogo

31.3 Smart Slow Speed (S3): An Agent-Based Simulation Platform

S3 was designed as an interactive platform to explore complex issues with reactive agents as highlighted previously (Fig. 31.5).

S3 is composed of two interacting modules: the first one concerns both the creation of a road network differentiated by speed and the location of traffic lights, while the second one implements an agent-based traffic simulation model.

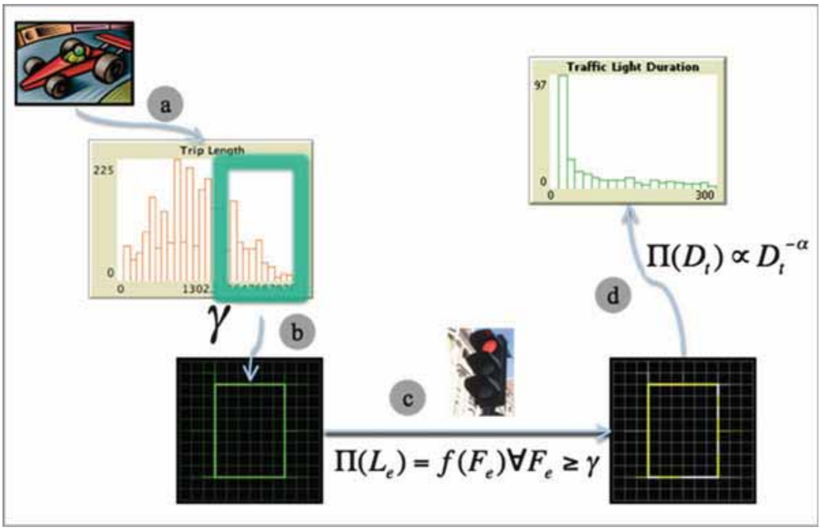


Fig. 31.6 Location and duration of traffic lights: a four step process

31.3.1 Traffic Lights: Choosing Location and Time Duration

The first module (network builder) constructs regular (rectangular or octagonal), networks. The graph generated is non-oriented but weighted by speed. Edges are indeed valued with a given speed *v*, such that $v \in \{30,50,70,90,110\}$, expressed in km/h. On this basis, the shortest paths are computed between any two distinct nodes using the Floyd-Warshall algorithm. As the graph is non-oriented, the total number of pairs is $(n\,(n-1)/2)$ where *n* is the number of nodes.

A population of agents of size *P* is then created, each agent being defined by an origin node, a destination node, and therefore the shortest-path between these two end nodes. The *P* paths created are then used to create the traffic lights, in a four step process (Fig. 31.6).

The main issue concerns the efficient location of a limited number of traffic lights. Let us assume two simplifications. Firstly, let us assume that there is no capacity constraint, which means that the flow on every edge of the network may surpass the edge's capacity: for each edge *e*, $F_e \geq U_e$, where F_e is the flow on *e* and U_e is its capacity. Secondly, let us assume that our agents have no adaptation capacities, i.e. they strictly follow their allocated shortest-path, whatever the context is.

Impacting the maximum number of agents with a limited number of traffic lights then involves identifying target edges, which will be crossed by a large number of agents (Fig. 31.6a). Basically, the probability that a given edge *e* will host a traffic light *L* will be a function of the flow on *e*, $\Pi(L_e) = f(F_e)$. We can think about this process as a preferential attachment one (Newman 2005). Moreover, given the assumptions previously formulated (no capacity constraints and no adaptation behaviour), then it is evident that the flow F_e on a given edge *e* is the number of times that the edge belongs to a shortest path between two given nodes. Therefore,

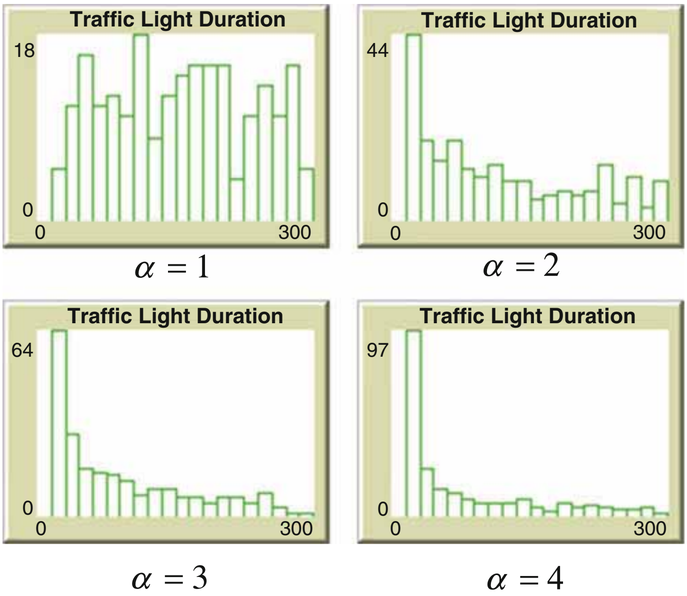


Fig. 31.7 The distributions of the durations of the traffic lights obtained for various values of α

the flow F_e of edge e can also be interpreted as a proxy of its "edge betweenness" (Girvan and Newman 2002), which is a generalisation of Freeman's betweenness centrality (Freeman 1977) to edges. However, introducing slow speed metrics in a network does imply targeting also the longest trips, as they benefit the most from the standard "speed metric" (Fig. 31.6b). Therefore, we reduce the candidates to edges characterised by a flow superior or equal to a given threshold value γ:

$$\Pi(L_e) = f(F_e)\forall F_e \geq \gamma \tag{31.5}$$

Once traffic lights are located (Fig. 31.6c), we have to define their duration. A power law is assumed such that $\Pi(D_t) \propto D_t^{-\alpha}$. Increasing values of parameter α provide various distributions of the durations of the traffic lights, once the minimum and maximum durations are defined (Fig. 31.7).

31.3.2 A Microscopic Traffic Model

The second module handles a microscopic traffic model, aimed at testing the efficiency of the designed network, as well as its impact on traffic fluidity. Before each simulation, m agents are created and located randomly on the n nodes of the network, their destination also being chosen randomly. During a simulation, each agent will have to reach its destination following the shortest route computed as previously defined, and takes into account speed v of the links but also the presence of other agents in front as well as the presence of red lights at intersections.

We use an underlying grid covering the 1 km × 1 km wide area, composed of a large number of small cells (length 4 m). Agents are then located on cells underlying the network. Only one agent can occupy one cell at a time. On this basis and following Banos et al. (2005), we then extend the NaSch model (Nagel and Schreckenberg 1992) in order to introduce traffic lights. According to the prescription of the NaSch model, we allow the speed V of each vehicle to take one of the integer values $V = 0$, 1, 2…, $Vmax$, where $Vmax$ corresponds to the speed v of the current link. At each discrete time step $t \rightarrow t+1$, the arrangement of the m agents is then updated in parallel according to the following driving rules:

Step 1: Acceleration

$$\text{If Vn} < \text{Vmax}, \text{Vn} \rightarrow \min\left(\text{Vn}+1, \text{Vmax}\right) \tag{31.6}$$

i.e. the speed of the *nth* vehicle is increased by one.

Step 2: Deceleration (due to other vehicles/traffic signal).
Suppose Dn is the gap between the *nth* vehicle and the vehicle in front of it, and Dtn is the gap between the car under consideration and the red light in front of it on the road, then:

$$\text{if dn} \leq \text{Vn or dt n} \leq \text{Vn}, \text{then Vn} \rightarrow \min\left(\text{Vn}, \text{Dn}-1, \text{Dtn}-1\right) \tag{31.7}$$

Step 3: Randomisation

$$\text{If Vn} > 0 \text{ and}, \varepsilon \leq p, \text{ then Vn} \rightarrow \text{Vn}-1 \tag{31.8}$$

where ε is a value generated at random from a uniform distribution U [0,1] and $0 \leq p \leq 1$. In other words, the speed of the car under consideration is decreased randomly by unity with probability p. This random deceleration probability p is identical for all the vehicles, and does not change during updating. Three different behavioural patterns are then embedded in that single computational rule: fluctuations at maximum speed, retarded acceleration and over-reaction at braking

Step 4: Movement
Each vehicle moves forward with a certain speed, i.e. $Xn \rightarrow Xn + Vn$, where Xn denotes the position of the *nth* vehicle at any time t.

Figure 31.8 illustrates the kind of traffic patterns generated by this simple but powerful model.

Once a specific hierarchical network is fixed, this traffic model allows exploration of its efficiency as well as the impact of various strategies of speed reduction. More precisely, it allows exploration of the possible conditions under which we may obtain a slow metric and the potential effects of the latter on traffic (fluidity).

Exploring these various issues in a systematic way is a challenge in itself as the possible combination of parameters grows rapidly with their number. In this chapter we will mainly focus on results related to the creation of a slow metric. Its effects on congestion and on global accessibility will be less detailed, even though it remains a fundamental issue.

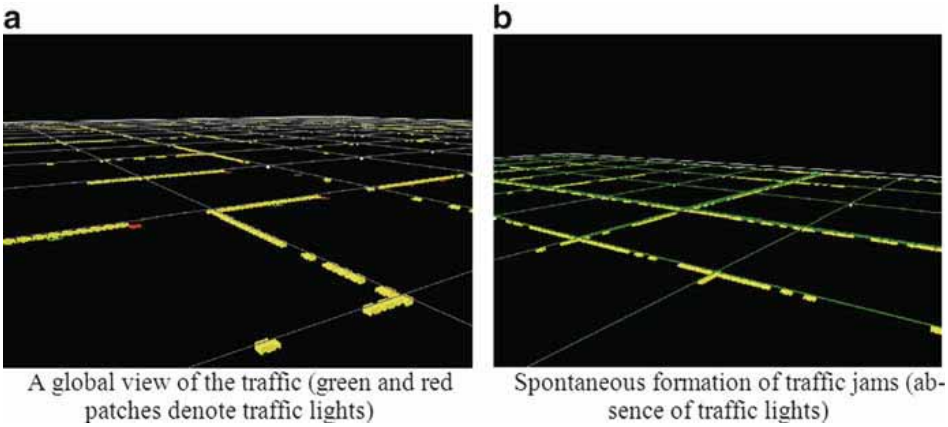


Fig. 31.8 Examples of traffic patterns obtained from the extended NaSch model (**a**) A global view of the traffic (presence of traffic lights) (**b**) Spontaneous formation of traffic jams (absence of traffic lights)

31.4 Preliminary Results

The results we present here may be seen as preliminary as they suggest more than they demonstrate. However, what they suggest is undoubtedly of interest: a small number of traffic lights, with location and duration carefully defined, may be sufficient to obtain a slow speed metric. However, defining the cost of such a measure is trickier than expected at first glance and should be investigated in a much more exhaustive way than we do here.

31.4.1 A Small Number of Traffic Lights May Do the Job

Let us imagine a regular but octagonal network arranged hierarchically by speed in a star-like manner with high speed corridors converging towards the centre (Fig. 31.9). On that basis we can explore the influence of various combinations of traffic lights characteristics on network efficacy, located as described in Eq. 31.5. Importantly, the traffic model (NaSch) is not activated so the pure network component can be identified.

As expected, the speed metric occurs in the absence of traffic lights. Introducing such equipment reduces the discrepancy between short and long trips by penalising the long ones. However, the number of lights seems to play a secondary role compared to their duration, as expressed by the clustering of curves. This graph even suggests that a small number of traffic lights may do the job quite well, if we calibrate their duration and location correctly. It is indeed well known that the pattern of flows depends strongly of the structure of the network, as shown by Penn

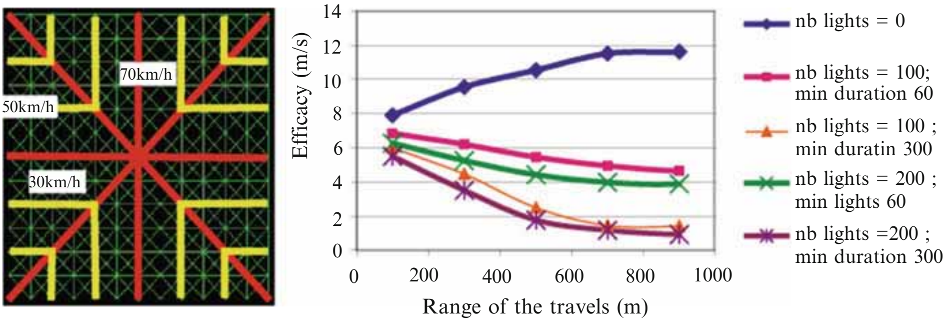


Fig. 31.9 Impact of the number and duration of traffic lights on efficacy for a hierarchical star network under free flow condition (NaSch not activated)

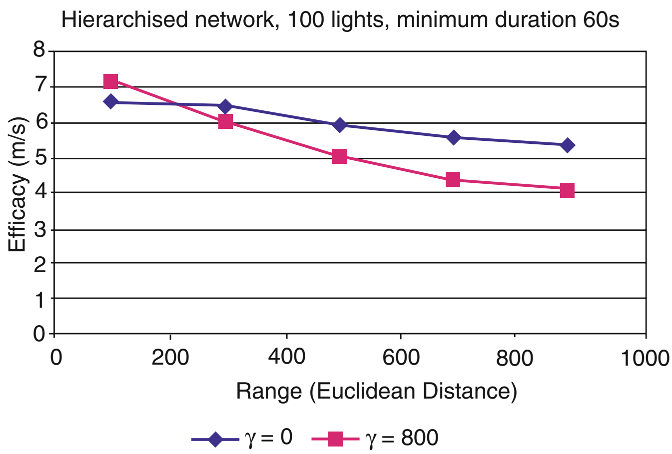


Fig. 31.10 Influence of parameter γ on network efficacy under free flow condition (NaSch not activated)

et al. (1998). Therefore, those edges constituting the backbone of traffic flows should be targeted preferentially in order to setup and reinforce the slow metric process. Moreover, those involved in long distance trips are of specific interest. This crucial point may be illustrated by comparing two values of the threshold γ used in Eq. 31.5. As can be seen in Fig. 31.10, targeting long distance trips amplifies the inversion process in a significant way.

31.4.2 Introduction of Traffic Interactions

Free flow conditions can hardly be defined as representative situations in urban daily life. Introducing traffic interactions between agents, as defined previously, is therefore an important point. A key feature emerging from the simulation may be

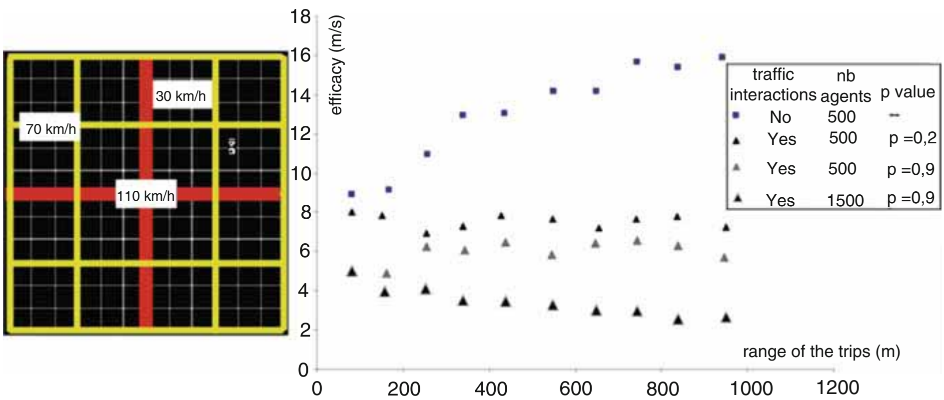


Fig. 31.11 Traffic and congestion as contributing factors to a slow speed metric

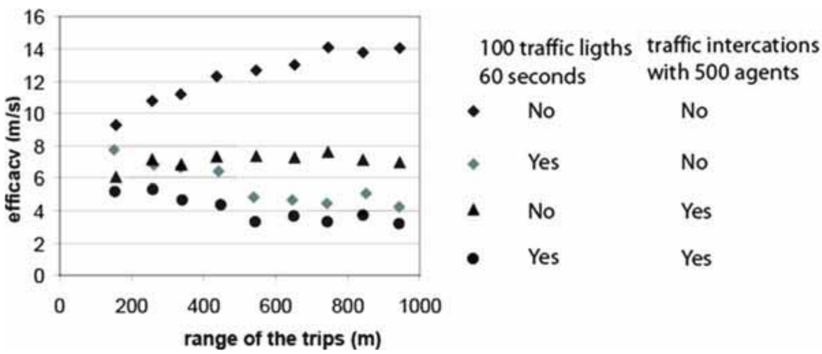


Fig. 31.12 Combining traffic interactions and traffic lights (same road network as Fig. 31.11)

emphasized: traffic conditions and more precisely congestion may contribute actively to the slow speed metric that we are searching for. As Fig. 31.11 reveals, increasing the number of agents and the intensity of the interactions between them (parameter *p* in Eq. 31.8) dramatically alters the efficacy curve in the absence of traffic lights.

Obviously, theoretical performances of road networks are very far from real ones when traffic interactions and traffic lights are taken into account. Interestingly enough, adding traffic lights to this picture amplifies the process, with an additional impact on short distance trips (Fig. 31.12).

However, balancing the efficacy of a given network between long and short trips may not be an end in itself. Indeed, even if the goal of the so-called slow metric is to lower network efficacy, especially for long range trips, the global system should remain efficient. One way to evaluate this loss in efficiency is to compute a "fluidity" index. One basic idea would be to compare a given state, characterised by some constraints (e.g. traffic lights, traffic interactions) with a previous one, free from such constraints. Let us define t_i as the duration of a simulated trip, under given traffic conditions. Let us also define τ_i as the theoretical duration for that same trip under

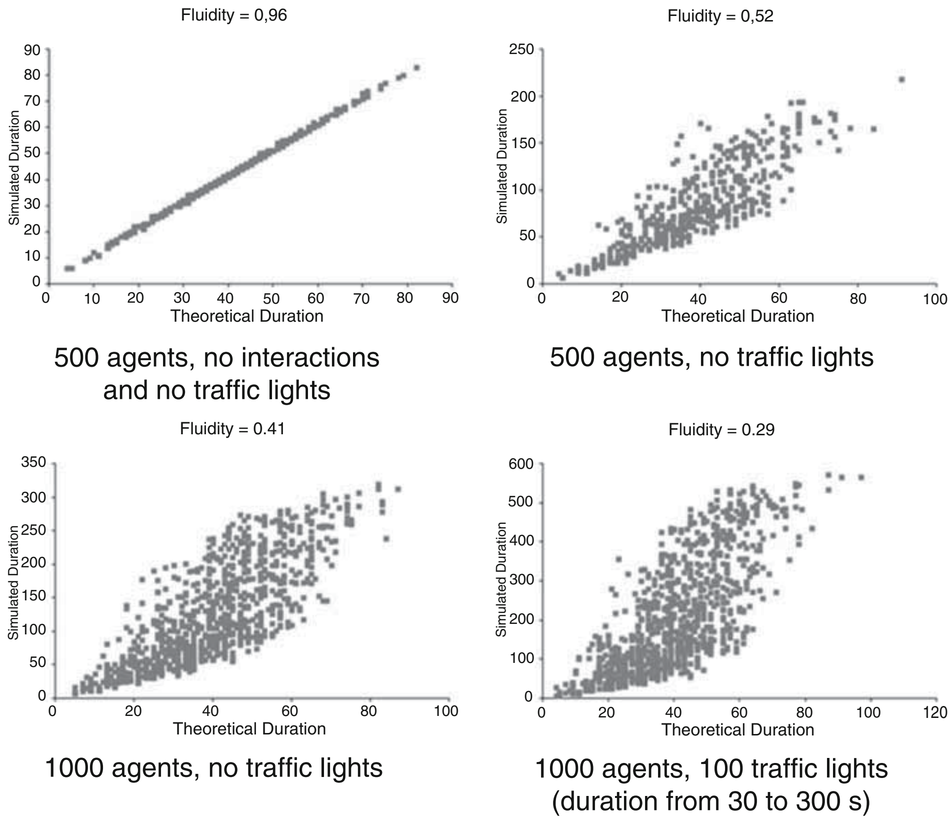


Fig. 31.13 Impact of number of agents, traffic interactions and traffic lights on fluidity (p = 0.25)

free flow conditions (no traffic interactions and no traffic lights). This theoretical indicator may be defined in a simple manner for each trip as:

$$\tau_i = \sum_{k=1}^{n} \frac{l_k}{v_k} \tag{31.9}$$

where l_k is the length of edge k and v_k is its speed. Given step 1 (acceleration) of the NaSch model, it is obvious that $t_i \geq \tau_i$. For each trip (and therefore each agent i), we can then define its loss in fludity f_i as:

$$f_i = 1 - \frac{t_i - \tau_i}{t_i}, (0 \leq f_i \leq 1) \tag{31.10}$$

which evolves between 0 and 1. We can then define an average indicator of fluidity loss F, which will also evolve between 0 (minimum fluidity) and 1 (maximum fluidity) as:

$$F = \frac{1}{n} \sum_{i=1}^{n} 1 - \frac{t_i - \tau_i}{t_i} \tag{31.11}$$

Under free flow conditions, the fluidity loss is null as $t_i = \tau_i$, and it increases when traffic interactions (NaSch rules) and traffic lights are introduced (Fig. 31.13).

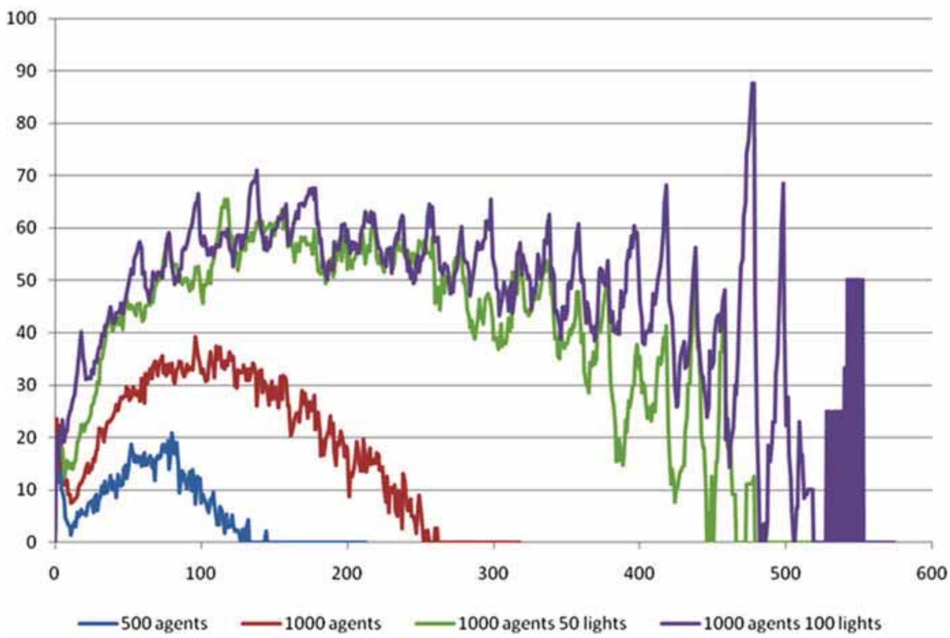


Fig. 31.14 Proportion of cars stopped during the simulation (same parameters as Fig. 31.13). Number of agents and lights increases from left to right

One key issue is therefore to identify socially acceptable ranges of variation for such an index but also to identify winners and losers in a very detailed manner, as very different situations may occur behind the same global value. However, the fluidity index represents the loss of time due to the introduction of the new metric compared to a free flow situation. It is, therefore, closely related to the efficacy index and then provides rather limited insights on traffic jams. Representing the proportion of vehicles being stopped during the simulation then adds some valuable information to this specific issue (Fig. 31.14).

Obviously, adding constraints (number of agents, number of traffic lights) has a strong impact on the proportion of vehicles being stopped (curves tend to 0 as vehicles progressively reach their destination during simulation). Of course, this index merges both stops due to traffic lights and those generated by traffic jams. In that sense, it should be improved in order to display clearer insights on the local and global impacts of the slow speed metric. Comparison with “real world” data may also be of crucial importance here, in order to define the range of what may be seen as acceptable conditions.

31.5 Conclusions

These first results, while being very preliminary, are promising. Indeed they suggest that slow speed metrics could be designed and maintained with limited and thus well-targeted efforts. However, several questions still remain unanswered so far. The first one concerns the calibration of such a metric, which is very complex as it depends on the number, duration and location of traffic lights, the structure of the network, the intensity of the traffic and the nature of its local interactions. In order to optimize this calibration, we must know more about road users' acceptance levels regarding waiting times at junctions, fluidity, and more generally, accessibility.

A second point concerns the strong underlying assumption of non-adapting travelling behaviours. By imposing traffic lights with randomly chosen and ever changing durations, we assume that no driver can predict a better solution to the shortest path (in distance) and therefore has no interest in modifying their route, thus reaching an equilibrium quite naturally. This very strong assumption needs further exploration, even before we imagine possible sources of heterogeneity introduced, for example, by real time traffic information.

Another issue concerns the regulation of traffic jams. In the presence of local traffic jams generated or amplified by a long lasting traffic light, we can imagine self-regulatory processes overriding this targeted traffic light periodicity, especially when edge capacity reaches a critical value. Such self-regulatory systems have already been explored by simulation with promising results in terms of fluidity (Lämmer and Helbing 2008).

The last issue concerns the social acceptance of such a constraining policy. This issue raises a large variety of debates we cannot address here. However, one key issue concerns the capacity of the slow metric to reinforce the economic value of physical proximity in a significant way. This would be mandatory if one's goal is to drive localisation strategies of both households and firms towards more sustainable urban patterns. Moreover, our simulations suggest that the "do nothing" politics that are usually adopted, which consist of letting traffic jams occur and "regulate" the system, do act as non-mastered slow speed metrics. Indeed, it is our belief that long distance trips are still favoured by such a strategy, an intuition we will have to demonstrate in future work. More generally, despite the complexity of the problem and its highly political dimension, our claim is that we can do better than such "do nothing" politics through proactive and well-targeted strategies.

References

Banos, A., Godara, A., & Lassarre, S. (2005). Simulating pedestrians and cars behaviours in a virtual city: An agent-based approach. In *Proceedings of the European Conference on Complex Systems*, Paris, 14–18 Nov 2005.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry, 40*, 35–41.

Genre-Grandpierre, C. (2007). Des réseaux lents contre la dépendance automobile? Concept et implications en milieu urbain. *L'Espace Géographique, 1*, 27–39.

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of National Academic Science USA, 99*, 7821–7826.

Gutiérrez, J., Monzòn, A., & Pinero, J. M. (1998). Accessibility, network efficiency and transport infrastructure planning. *Environment and Planning A, 30*, 1337–1350.

Lämmer, S., & Helbing, D. (2008). Self-control of traffic lights and vehicle flows in urban road networks. *Journal of Statistical Mechanics: Theory and Experiment, 2008*(4), P04019.

Levinson, D., & Kumar, A. (1994). The rational locator: Why travel times have remained stable. *Journal of the American Planning Association, 60*(3), 319–332.

Nagel, K., & Schreckenberg, M. (1992). Cellular automaton models for freeway traffic. *Journal of Physics I, 2*, 2221–2229.

Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics, 46*, 323–351.

Penn, A., Hillier, B., & Xu, J. (1998). Configurational modelling of urban movement networks. *Environment and Planning B, 25*, 59–84.

Stanley, H. E., & Ostrowsky, N. (1985). *On growth and form: Fractal and non fractal patterns in physics*. Nijhoff, Dordrecht.

Chapter 32
A Logistic Based Cellular Automata Model for Continuous Urban Growth Simulation: A Case Study of the Gold Coast City, Australia

Yan Liu and Yongjiu Feng

Abstract This chapter presents a logistic based cellular automata model to simulate the continuous process of urban growth in space and over time. The model is constructed based on an understanding from empirical studies that urban growth is a continuous spatial diffusion process which can be described through the logistic function. It extends from previous research on cellular automata and logistic regression modelling by introducing continuous data to represent the progressive transition of land from rural to urban use. Specifically, the model contributes to urban cellular automata modelling by (1) applying continuous data ranging from 0 to 1 inclusive to represent the none-discrete state of cells from non-urban to urban, with 0 and 1 representing non-urban and urban state respectively, and all other values between 0 and 1 (exclusive) representing a stage where the land use is transiting from non-urban to urban state; (2) extending the typical categorical data based logistic regression model to using continuous data to generate a probability surface which is used in a logistic growth function to simulate the continuous process of urban growth. The proposed model was applied to a fast growing region in Queensland's Gold Coast City, Australia.

32.1 Introduction

This chapter presents a logistic based cellular automata model to simulate the continuous process of urban growth in space and over time. The model is constructed based on an understanding from empirical studies that urban growth is a continuous spatial

Y. Liu (✉)
Centre for Spatial Environmental Research and School of Geography, Planning and Environmental Management, The University of Queensland, Queensland, Australia
e-mail: yan.liu@uq.edu.au

Y. Feng
College of Marine Sciences, Shanghai Ocean University, Shanghai, P. R. China

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_32,

diffusion process over time which can be described through the logistic function (Herbert and Thomas 1997; John et al. 1985; Jakobson and Prakash 1971). It extends from previous research on cellular automata and logistic regression modelling by introducing continuous data to represent the progressive transition of land from rural to urban use. Specifically, the model contributes to urban cellular automata modelling by (1) applying continuous data ranging from 0 to 1 inclusive to represent the none-discrete state of cells from non-urban to urban, with 0 and 1 representing non-urban and urban state respectively, and all other values between 0 and 1 (exclusive) representing a stage where the land use is transiting from non-urban to urban state; (2) extending the typical categorical data based logistic regression model to using continuous data to generate a probability surface which is used in a logistic growth function to simulate the continuous process of urban growth. The proposed model was applied to a fast growing region in Queensland's Gold Coast City, Australia.

Many CA based urban models have been developed over the last two decades (see Iltanen 2012); these models vary considerably based on the configurations of the five basic elements that make up the CA model – the type and size of the cells, the definition of cell states, the type and size of the neighbourhood, the configuration of transition rules, and the way time is modelled. While the sizes of cells and their neighbourhood are fundamentally a scale issue which is common to all geographical problems (Openshaw 1984), it is more challenging for modellers to configure the state of the cells and identify rules that drive the transition of cells from one state to another over time in order to capture the various features and mechanisms of urban growth dynamics.

A common practice in defining cell state is using discrete data to represent a binary state of either non-urban or urban, or using specific land use types (Ward et al. 2003; Li and Yeh 2000; Clarke and Gaydos 1998; Wu 1998a, b, c, 1996; Clarke et al. 1997). Many of these models have made significant contributions to our understanding on urban growth and land use change dynamics both theoretically and empirically, however, a limitation of this modelling practice is in capturing the progressive or continuous change of state geographically and temporally. To overcome this limitation, Liu (2008) and Liu and Phinn (2003) developed models using a fuzzy membership function to represent the fuzzy boundary between the urban core and suburb areas towards the rural land. Other models developed by Mandelas et al. (2007), Dragićević (2004) and Wu (1998b) also introduced the fuzzy set and fuzzy logic concepts in their models following the pioneering attempt by Wu (1996). However, as the membership function and the fuzzy linguistic modifiers are defined in a subjective way, the interpretation of the model's results is largely restricted (Wu 1996).

In terms of identifying and configuring the transition rules, both statistical and non-statistical methods have been developed; the former include the Monte Carlo method (Clarke et al. 1997), the analytical hierarchy process (AHP) (Wu 1998c), multi-criteria evaluation (MCE) (Wu and Webster 1998), principal components analysis (PCA) (Li and Yeh 2002a), multiple regression analysis (Sui and Zeng 2001), and logistic regression (Wu 2002); and the latter include the artificial neural network (Almeida et al. 2008; Li and Yeh 2002b), and the spatial optimization methods such as particle swam optimization (Feng et al. 2011), support vector machines (Yang et al. 2008), kernel method (Liu et al. 2008a), and ant intelligence (Liu et al. 2008b). Amongst these approaches, the logistic regression method has been favoured in a number of studies primarily due to its strong statistical foundation in exploring the association of a dependent variable, measured as presence or absence of a phenomenon (e.g., yes/no, or urban/non-urban), and multiple independent (predictor) variables such as the various environmental, social and institutional factors on urban growth.

For instance, Wu and Yeh (1997) applied logistic regression method for modelling land development patterns based on parcel data extracted from aerial photographs; Gobim et al. (2002) adopted logistic modelling technique to predict the probabilities of local agricultural land use systems in Nigeria; Wu (2002) used logistic regression to compute the probability of a cell experiencing land change and developed a stochastic CA model; Verburg et al. (2002) developed a spatially-explicit land-use change model (i.e. CLUE-S) based on logistic regression to simulate land-use change in relation to socio-economic and biophysical driving factors, which has also been widely applied in other areas (e.g., Zhu et al. 2009); Fang et al. (2005) used logistic regression to analytically weigh the scores of the driving factors of an urban sprawl model for predicting probability maps of land use change.

Indeed, the logistic regression method is effective and efficient in modelling urban growth with cellular automata. However, the use of categorical data representing a binary state of non-urban or urban is limited in modelling areas that are partly developed but are yet to be fully urbanised. Given that the logistic regression model is not restricted to only using categorical data to represent the dependent variable, this method can also be applied to model continuous dependent variable data, so long as the data is within a range from 0 to 1 representing probability values or proportions (Sherrod 2010).

The following section describes the study area where the proposed logistic CA model using continuous data will be tested and calibrated, followed by a section on model description, including the model design concepts and principles, input data and the model configuration details. The results generated from the application of the model in simulating the processes of urban growth in the Gold Coast City from 1996 to 2006 are presented and discussed in the Results section, together with assessments of the model's simulation accuracies using a set of landscape matrix measures. The last section of the chapter summarises findings from the current research and identifies future research directions.

32.2 Study Area

Gold Coast City, the second largest city in Queensland, Australia was selected to apply the logistic regression based CA model to simulate the continuous urban growth from 1996 to 2006. Geographically, the city is situated in the south-east corner of Queensland between the state's capital city of Brisbane at its north and the State of New South Wales at its south. To the west of the Gold Coast are

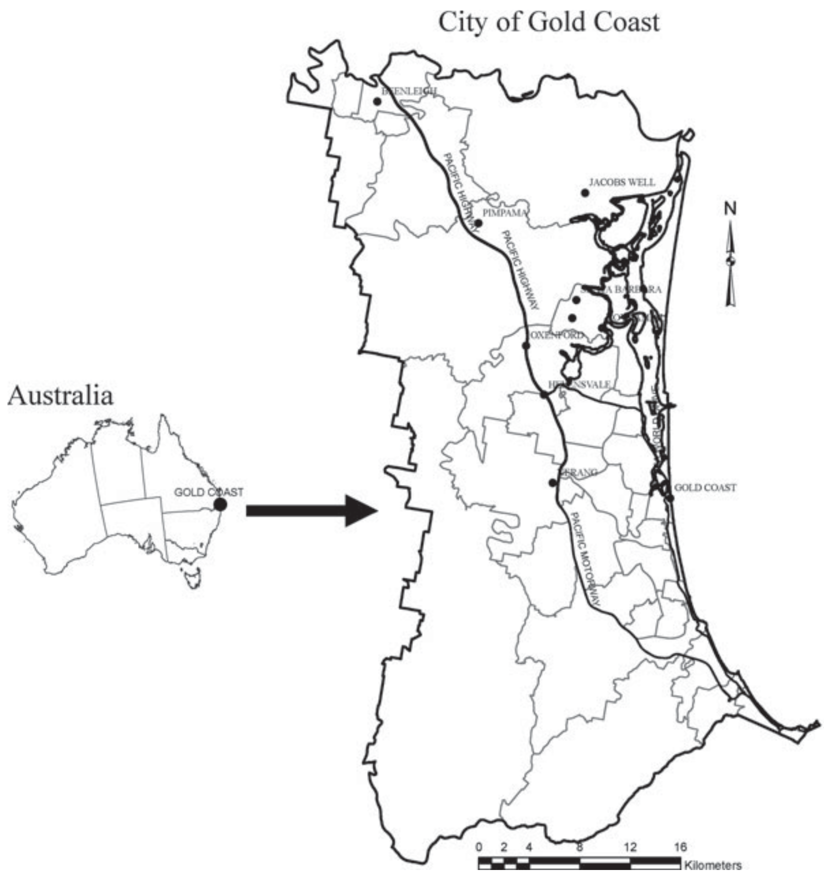


Fig. 32.1 The study area

the Lamington National Park and the foothills of the Great Dividing Range. Spanning across an area of 1,400 km^2 and with 57 km of coastline, this city has some of the most popular surf breaks and beaches in Australia and the world; its west and southwest hinterland also offers the most spectacular scenery in Australia and extremely rich biodiversity which supports rural production, water quality, scenic amenity and outdoor recreation (Department of Infrastructure and Planning 2009).

This city consists of 106 state suburbs prior to the year 2008, with a number of localities, towns and rural districts within the suburbs. Since March 2008, Beenleigh-Eagleby region on Gold Coast's northwest border has been transferred to Logan City. Given that the time frame in this study is from 1996 to 2006, the study area still includes those suburbs that are currently part of Logon City (Fig. 32.1). The total resident population of the city in 2006 was 472,000 (Australia Bureau of Statistics 2006), and the estimated resident population in 2010 was 528,000 (Australian Bureau of Statistics 2011). Over the past 15 years, the urban footprint has been growing rapidly with most of development been concentrated between Yatala at the north and Coolangatta at the south, and continues south beyond the Queensland border into the Tweed Shire in New South Wales (Department of Infrastructure and Planning 2009; Ward et al. 2000).

32.3 Model Description

The model description follows the ODD (Overview, Design concepts, and Details) protocol for describing individual- and agent-based models (Grimm et al. 2006, 2010; Grimm and Railsback 2005, 2012) and consists of seven elements. The first three elements provide an overview of the model; the fourth element explains general concepts underlying the model's design; and the remaining three elements provide details.

32.3.1 Purpose

The purpose of the model is to simulate urban growth as continuous processes in space and over time. It hypothesises that the transition of land from rural to urban use is a continuous process which can be modelled using continuous data and logistic functions in a cellular environment. The proposed cellular model was applied to the rapidly growing Gold Coast City in Southeast Queensland, Australia to test the conceptual framework and effectiveness of the model in simulating the continuous processes of urban growth.

32.3.2 Entities, State Variables and Scales

In cellular automata based urban modelling, the space or area under study is typically tessellated into regular grid cells, each cell representing a piece of land on the ground at a certain scale. For the study area in the Gold Coast City, the spatial scale of cells is at 30 m, with a total of 1,559,283 cells covering the entire study area.

Each spatial entity or cell is represented by a state variable representing the state of a cell undertaking the urban development process. Instead of defining the state variable with a binary value of urban or non-urban, as most other CA based urban models typically do, this model defines the state of cells as continuous values ranging from 0 to 1 (inclusive). For instance, a non-urban cell which has not started its development will have a state value of 0, while a fully developed urban cell will have a state value of 1. Cells that have started their process of urban development but are yet to be fully developed will have a state value between 0 and 1 (exclusive).

Another state variable used in the CA model is the neighbourhood size. In this study, a rectangular 5×5 cells neighbourhood was selected. According to CA

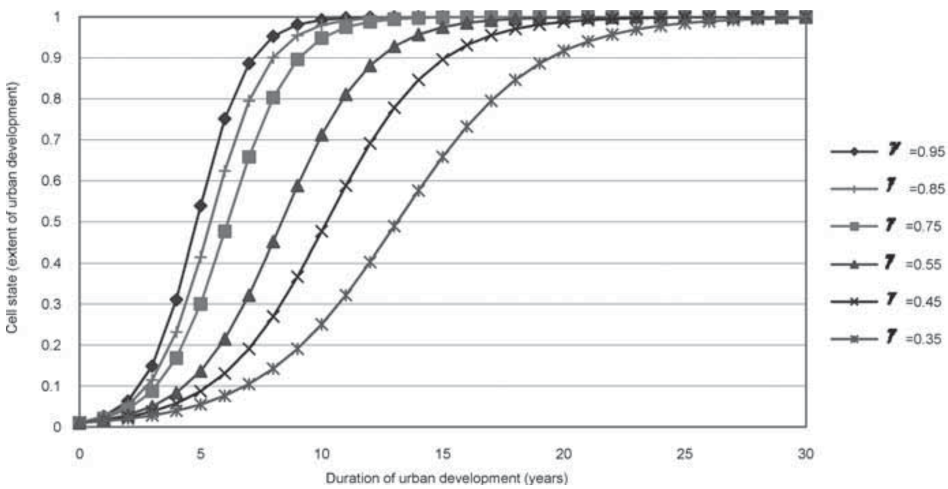


Fig. 32.2 The logistic urban growth process under various growth control speeds

principles, the state of a cell at a certain time is not only dependent on the state of the cell itself in a previous time step, but also the collective effect of all cells within its neighbourhood. This collective neighbourhood effect is modelled in the land use conversion probability sub-model.

32.3.3 Process Overview and Scheduling

Three processes of urban growth have been modelled; these include a continuous urban growth process, a new urban growth process, and a no-growth process. All processes are under the assumption that urban growth progress from a lower state (with a state value close to 0) to a higher state (with a state value close to 1), with the value 1 being the highest state in the urban growth process. Once a cell reaches a state value of 1, it is considered being fully developed and its state will remain as 1 until the simulation is terminated. This excludes both urban re-development and anti-urbanisation processes.

Continuous urban growth
The continuous process of urban growth can be represented by a logistic function (Liu 2008; Herbert and Thomas 1997; John et al. 1985; Jakobson and Prakash 1971).

$$S_i^t = \frac{1}{1 + (1/S_0 - 1) \times \exp(-\gamma \times t)} \tag{32.1}$$

where S_i^t is the state of cell i at time t; S_0 is the initial state of the cell to start its urban growth process. This initial state value can be small but larger than 0 for the cell to start its urban growth process (e.g., $S_0 = 0.01$). γ is a parameter ranging from 0 to 1 which controls the speed of the urban growth process (Fig. 32.2).

This growth control parameter is modelled through a land use conversion probability sub-model, which will be discussed further in the Design Concept and Sub-model sections.

New growth
A new growth process can occur to a non-urban cell if there is a high probability for such growth to occur. For instance, if a non-urban cell is surrounded by cells which are all fully urbanized, this non-urban cell may have the tendency or be influenced by its neighbouring cells to develop towards an urban state. In this case, an initial very low value is assigned as the state of the cell so it can start its urban growth process. Once a non-urban cell is triggered to start its urban growth process, it will then follow the continuous urban growth process; the growth speed is also controlled by the land use conversion probability at the time of development.

No growth
Apart from the above two growth processes, other non-urban cells will remain as non-urban until such conditions where new development can be triggered to proceed.

Model scheduling
The model starts from a historical time where input data is available to represent the initial state of entities in the model. For the City of Gold Coast, this starting date was set to 1996. Another set of data was collected in 2006 which was used to calibrate the model over time. The temporal iteration of the simulation process was determined by comparing the overall urban area generated by the model with the actual urban area size in 2006; if the generated urban area size is within less than 1% difference from the actual urban area size in 2006, then the model is terminated for a simulation accuracy assessment in 2006. This calibration process continues until satisfactory results were achieved which match with the actual urban growth patterns.

32.3.4 Design Concepts

Basic principles: The basic principle imbedded in this urban cellular model is that urban growth is a continuous spatial diffusion process overtime. Spatially, an urban area is featured with high population density and the dominance of non-agricultural land. However, as cities are surrounded by rural or natural land, there is no clear boundary between an urban built-up area and its non-urban hinterland. Between the well-recognized urban land use and the area devoted to agriculture, there exists ‘a zone of transition in land use, social and demographic characteristics, lying between (a) the continuously built-up area and suburban areas of the central city, and (b) the rural hinterland, characterized by the almost complete absence of non-farm dwellings, occupations and land use’ (Pryor 1968:206). This ‘zone of transition’ is a place where both urban and non-urban features occur, which has been broadly termed as ‘fringe’ or ‘rural-urban fringe’ (Bryant et al. 1982; Pryor 1968). The urban fringe area has become the most vigorous part of development in the rural-urban continuum and has attracted much attention in research.

Temporally, the transition of land from rural to urban use is a continuous process which occurs over a period of time. This transition process can be illustrated through a logistic function with five distinctive growth stages which can be illustrated by Rostow's five-stage model of development, Which can be illustrated by Rostow's five-stage model of development, that is, it transits from a traditional rural society to pre-conditional stage for growth. Which can be illustrated by Rostow's five-stage model of development, to a take-off stage for rapid growth, and then to a drive to maturity stage and a final stage of urban society (Potter et al. 2003). Such spatial and temporal processes of continuous growth cannot be captured through a conventional CA model with binary states of non-urban and urban configurations.

While the general process of urban growth can be described through a logistic growth function, the speed of such growth or any possibility of new growth are affected by a number of other factors, namely, the neighbourhood effect, accessibility to infrastructures and services, physical or institutional constraints on the use of land, and other stochastic factors that are uncertain at time of modelling. Such impact factors can be built into the logistic function through a probability index at individual cell scale. This probability index is computed in a sub-model which is elaborated further in the Sub-model section.

Emergence: Urban growth patterns emerge from the transition of states of individual cells, but the transition of individual cells from one state to another is dependent on the state of the cell itself and the states of neighbouring cells at a previous time step according to certain transition rules.

Adaptation: The transition of land use from non-urban to urban is an adaptive process over time; each cell will adapt to change of the land use conversion probability according to its environmental conditions. For instance, a cell with a high probability for growth will develop faster than those with lower probability for growth. A high land use conversion probability is related to advantageous driving factors such as close proximity to transportation network, facilities and services or strong neighbourhood support. On the other hand, a low land use conversion probability is more likely associated with lack of support for growth or other land use constraint, such as steep slope or other planning constraints.

Objectives: The objective of the model is for the simulated patterns of urban growth to match the actual pattern of urban growth at certain point in time. For instance, if the model is initialised with the actual urban growth data in 1996 to generate the urban growth scenario in 2006, the objective of the model is for the simulated urban growth scenario in 2006 to match with the actual urban growth patterns at that point in time.

Learning: Individual cells may change their transition speed over time as a consequence of their experience in matching the expected state of the cell during the model calibration process.

Prediction: Under the basic principle that urban growth will progress continuously from a lower to a higher state over time, it is assumed that individual cells are able to predict their cell state in this growth process according to the logistic function. However, this predictability is inherently associated with their adaptability to change of the land use conversion probabilities in space and over time.

Sensing: Individual cells are assumed to know their own state and the state of their neighbouring cells. They can also sense the environmental conditions such as land slope, spatial proximity to urban centres, and other data regarding the socio-economic and environmental conditions.

Interaction: Interaction amongst cells is modelled through the neighbourhood effect. In essence, the collective state of all cells within the neighbourhood of a processing cell contributes significantly to the land use conversion probability index, which in turn impact on the transition speed of the cell in its urban growth process.

Stochasticity: The land use conversion probability sub-model includes a stochastic disturbance factor on urban development. This stochastic factor is used to represent any unknown or unexpected factors that may impact on land use conversion probability, hence reproduce variability in the simulation processes.

Collectives: The impact of neighbourhood effect on urban growth is modelled collectively. The mean state value of all cells within the neighbourhood of the processing cell will impact on the land use conversion probability which is modelled through the land use conversion probability sub-model.

Observation: Observations of the model's results include graphical display of urban patterns in space and over time. For model calibration, results generated by the model were compared with the actual urban growth patterns through a set of landscape metrics and a spatial confusion matrix analysis.

32.3.5 Initialization

Each cell in the cellular space was initially assigned a state value ranging from 0 to 1 inclusive representing its state in the urban growth process. This initial state value was assigned based on the modified population density map computed from the 1996 Census of Population and Housing data at the Census Collection District (CCD) level; the modification was carried out using the 2006 Census data at the mesh block level (Australian Bureau of Statistics 2005, 2006). As the mesh block data contains higher spatial resolution than the 1996 data at the CCD level, areas identified as non-urban in the 2006 mesh block data were projected back to the 1996 census data to define the area as non-urban in 1996, if the area was not identified as non-urban with the 1996 census data. This modification increases the spatial accuracy of the initial input data for the model to commence.

32.3.6 Input Data

The input data used in the model include two sets of census data in 1996 and 2006. The 1996 census data were processed at CCD level and used (with modification as elaborated above) to define the initial state of the cells in the cellular space. The 2006 census data were collected at the mesh block level to modify the 1996 census data in order to increase the spatial accuracy of the population density map; this dataset was also used to define the state of cells in 2006 which were used subsequently to calibration the simulation accuracy of the model. In addition, a set of spatial proximity or distance variables were also generated to feed the logistic regression model which was used to generate the land use conversion probability due to spatial proximity to facilities and services.

Data defining the continuous state of cells
Two threshold values were used to define the lower and upper limits of population density respectively in delimiting the cell state in the urban growth process. These threshold values were set to 200 and 1,500 persons per square kilometres respectively; this reflects the relatively low population density measure used in Australia to delimit the extent of urban areas (Liu 2008; Linge 1965). A cell with a population density of less than 200 persons per square kilometre is defined as non-urban; its cell state is assigned a value of 0. Similarly, a cell with a population density of 1,500 persons per square kilometre or over is defined as urban; its cells state is assigned a value of 1. All other areas with a population density value of between 200 and 1,500 persons per square kilometre is assigned a state value between 0 and 1, which are linearly interpolated based on the density value of the cell at the specified point in time.

Other input data
The constraint data to urban growth used in the model includes transportation network such as highways, primary and secondary roads, data identifying the major urban and town centres, land use planning schemes, and data illustrating areas where urban development cannot occur, such as large water bodies, forest reserves, nature conservation reserves, large recreational areas, prohibited areas for defence purpose, mine sites, golf courses and major aquaculture areas. These data were collected from various government agencies and processed as raster data in GIS at 30 m spatial scale.

While many spatial factors may impact on urban land use change, in practice, not all factors can be quantified into a simulation model, especially when data reflecting such factors are neither available nor accessible. Historically, Gold Coast's urban growth has been largely driven by existing urban and town centres and the spatial accessibility to transportation and other infrastructure and services; it is also constrained by natural conservation and primary agricultural land. Therefore, eight spatial factors reflecting the spatial proximity of each cell to urban and town centres, to main roads, to residential, industrial, commercial and educational areas, and to natural conservation and primary agricultural land,

Table 32.1 Variables used to compute land conversion probability

Variable	Meaning	Data extraction
d_{ct}	Distance to urban and town centres	Distance variables were extracted from various GIS data layers including urban centres, transportation network, coastal lines and land use data layers
d_{rd}	Distance to main roads	
d_{rs}	Distance to residential areas	
d_{in}	Distance to industrial areas	
d_{co}	Distance to commercial areas	
d_{ed}	Distance to educational areas	
d_{pk}	Distance to parkland	
d_{ag}	Distance to agricultural land	
P_N	Probability of a cell changing from non-urban to urban within a rectangular 5 × 5 cells neighbourhood	Calculated using GIS's focal function
C	Land use constraints including large water bodies, forest reserves, nature conservation reserves, large recreational areas, prohibited areas for defence purpose, mine sites, golf courses and major aquaculture areas as well as land use restrictions through planning regulation	Extracted from urban land use and planning data
R	A stochastic disturbance factor representing any unknown errors	Randomly generated in GIS

together with the neighbourhood effect, physical and land use zoning constraints, and a stochastic disturbance factor representing any unexpected errors on urban growth to simulate the continuous processes of its urban development (Table 32.1).

32.3.7 *Sub-model*

32.3.7.1 Land Use Conversion Probability Sub-model

While urban development follows the logistic function in general, the speed of development varies from one location to another within the urban cellular space over time. This factor is reflected in Eq. 32.1 through the parameter of γ. This parameter can be represented by a land use conversion probability which is determined by the collective effect of its spatial proximity to other facilities and services, the neighbourhood effect, land use suitability constraints as well as a stochastic disturbance factor. Therefore, γ can be expressed as:

$$\gamma = P_i^t = P_{di}^t \times P_{Ni}^t \times C \times R \tag{32.2}$$

Table 32.2 Spatial proximity parameters generated using logistic regression model for the Gold Coast City

Variable	Constant	d_{ct}	d_{rd}	d_{rs}	d_{in}	d_{co}	d_{ed}	d_{pk}	d_{ag}
Coefficient of variable	0.690	−0.102	−0.214	−0.235	0.264	−0.151	−0.207	0.158	0.174

where P_i^t represents the land use conversion probability of cell *i* at time *t*; P_{di}^t is a land use conversion probability due to its spatial proximity to facilities and services; P_{Ni}^t is a land use conversion probability due to neighbourhood support; C represents a (set of) land suitability constraint/s; and R is a stochastic disturbance factor on urban development.

32.3.7.2 Spatial Proximity to Facilities and Services

The land use conversion probability P_{di}^t at location *i* and time *t* is controlled by a set of spatial proximity or distance variables, which can be represented using a logistic regression model.

$$P_{di}^t = \frac{1}{1 + \exp\left(-\left(a_0 + \sum_{j=1}^{k} a_j d_j\right)\right)} \tag{32.3}$$

where $d_j (j = 1, 2, ..., k)$ represents a set of spatial proximity factors on urban land use change. This includes the distances from a cell to urban centres, town centres, main roads, and other service facilities. a_0 is a constant and $a_j (j = 1, 2, ..., k)$ is a set of weighting factors representing the impact of the corresponding distance variables on urban development.

Using the input data from the Gold Coast City, a total of 23,370 sample cells (which accounts for 1.5% of the total cells within the study area) with their state values ranging from 0 to 1 were selected to build the logistic regression model, resulting in a set of spatial proximity parameters (Table 32.2). Table 32.2 shows that Distance to urban and town centres (d_{ct}), Distance to main roads (d_{rd}), Distance to residential areas (d_{rs}), Distance to commercial areas (d_{co}) and Distance to educational areas (d_{ed}) are all having a negative coefficient value, indicating that these factors contribute to the probability for urban growth positively, that is, the closer a cell is to these types of areas, the more likely it is to be developed. On the other hand, the variables of Distance to industrial areas (d_{in}), Distance to parkland (d_{pk}) and Distance to agricultural land (d_{ag}) having a positive coefficient value, indicating that these factors contributes to the probability for urban growth negatively, that is, the closer a cell is to these types of areas, the less likely it is to be developed.

32.3.7.3 Neighbourhood Support

Assume that within the rectangular neighbourhood of $m \times m$ cells, each cell has equal opportunity for development. Thus, the probability a cell develops from one state to another can be defined as:

$$P_{Ni}^{t} = \frac{\sum_{m \times m} S_{Ni}^{t}}{m \times m - 1} \tag{32.4}$$

where P_{Ni}^{t} is the probability that a cell can change from one state to another at time t due to neighbourhood support; S_{Ni}^{t} is the state of a cell within the neighbourhood of cell i at time t; $\sum_{m \times m} S_{Ni}^{t}$ is the accumulative state value of all cells within the $m \times m$ neighbourhood. For the application of the model to the Gold Coast City, the neighbourhood size used in the model is 5×5 cells.

32.3.7.4 Land Use Suitability

In practice, not all cells have equal opportunity for development. For instance, some areas such as large-scale water bodies, national parks, nature conservation reserves, large recreational areas, prohibited areas for defence purpose, mine sites, golf courses and major aquaculture areas and natural reserves cannot be developed into urban land use. Other areas such as areas used primarily for farmland may be prevented from urban development through institutional control, i.e., land use planning regulation. Such constraints can be represented through a constraint function:

$$C = Con(x_i^t = suitable) \tag{32.5}$$

The value of C ranges from 0 to 1, with 0 meaning the cell is constrained from changing its current state, and 1 meaning the cell is able to change its state at the following time step.

32.3.7.5 Stochastic Disturbance Factor

The stochastic disturbance factor R on urban development is defined as (White and Engelen 1993):

$$R = 1 + (-\ln r)^{\alpha} \tag{32.6}$$

where r is a random real number ranging from 0 to 1, and α is a parameter ranging from 0 to 10 which controls the effect of the stochastic factor on urban growth.

With Eqs. 32.3–32.6, the land use conversion probability of a cell i at time t can be rewritten as:

$$
\begin{aligned}
P_i^t &= P_{di}^t \times P_{Ni}^t \times C \times R \\
&= \frac{1}{1+\exp\left[-\left(a_0+\sum_{j=1}^{k} a_j d_j\right)\right]} \times \frac{\sum_{m\times m} S_{Ni}^t}{m\times m-1} \times Con(x_i^t = suitable) \times [1+(-\ln r)^{\alpha}]
\end{aligned}
\quad (32.7)
$$

This land use conversion probability is used in Eq. 32.1 to replace parameter γ which controls the speed of urban growth process.

32.4 Results and Assessment

32.4.1 Simulation Results

Using the initial urban state in 1996 as a starting point, the model was run until it generates an overall urban land size which is within 1% difference from the actual urban size in 2006. Then the simulation accuracy of the model was assessed by comparing the simulated urban pattern with the actual urban pattern in 2006. Through a trial-and-error approach, the best result generated by the model is presented in Fig. 32.3c. As a comparison, the initial state of cells in 1996 and the actual urban pattern in 2006 are presented in Fig. 32.3a, b respectively.

32.4.2 Assessment of Simulation Results

A commonly used method in assessing the simulation accuracy of the model is through a confusion matrix analysis between the simulated results and the observed urban growth patterns (Liu 2008; Li and Yeh 2002b). While the confusion matrix offers a quantitative and locational comparison to measure the 'goodness-of-fit' of the simulated result to the reference map, it lacks the capacity to provide insights on the structural similarity of the map pairs, or the spatial pattern-process relationships between different land patches (Liu 2008; Power et al. 2001). On the other hand, the landscape metrics developed by McGarigal et al. (2002) have been widely applied in landscape ecology studies to quantify specific spatial characteristics of land patches, classes of patches, or entire landscape mosaics. Such landscape metrics, measured directly from fields or maps (e.g., patch size, edge, inter-patch distance, proportion), are more likely to generate meaningful inferences (Li and Wu 2004). Hence, both a confusion matrix and a landscape metrics approaches are used in this study to assess the model's 'goodness-of-fit' as well as the similarities of spatial structure and patterns of the simulated results and the observed landscape patterns.

As both the confusion matrix and the landscape metrics method use categorical data, the continuous cell state data developed in this model were reclassified into

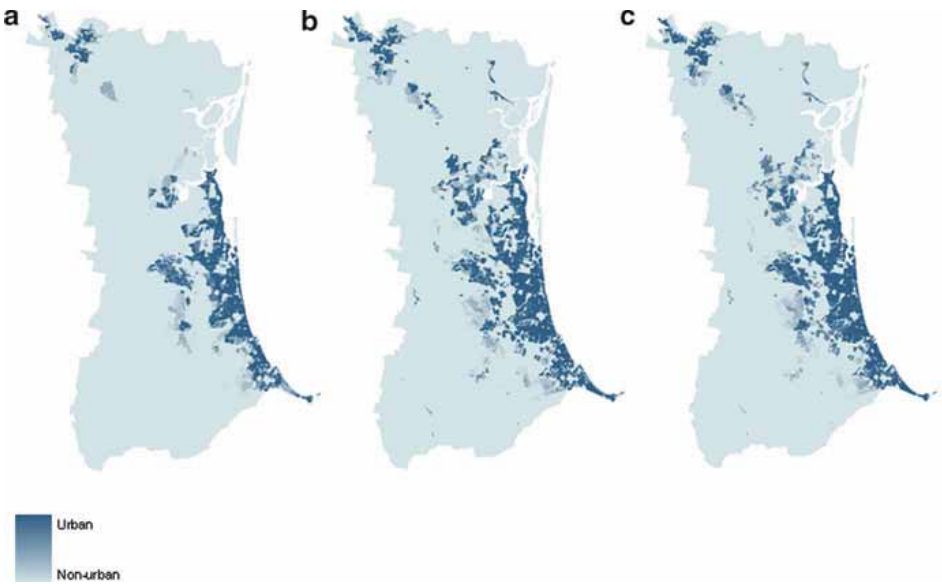


Fig. 32.3 Actual and simulated urban patterns of Gold Coast in 1996 and 2006. (**a**) is the actual urban pattern in 1996 which was used as initial state of cells in the CA model; (**b**) is the actual urban pattern in 2006; and (c) is the simulated result produced by the model which can be compared with (**b**)

Table 32.3 The confusion matrix between the actual and simulated cell states of the Gold Coast City in 2006 (%)

		Actual state of cells in 2006								
		C1	C2	C3	C4	C5	C6	C7	C8	C9
Simulated results for 2006	C1	96.5	53.4	19.2	18.0	13.7	18.8	15.9	17.2	11.0
	C2	0.9	27.2	42.9	26.0	7.2	6.1	5.4	6.0	3.4
	C3	0.3	12.1	14.7	6.7	25.0	22.1	1.8	1.4	2.1
	C4	0.2	1.8	13.8	26.3	4.9	3.3	17.2	22.8	1.4
	C5	0.7	1.4	4.4	8.9	8.7	9.1	10.1	15.1	12.1
	C6	0.2	0.2	1.4	8.2	6.8	4.1	6.8	4.6	3.3
	C7	0.2	0.1	1.0	1.2	12.6	4.2	6.1	4.6	3.3
	C8	0.4	1.5	0.7	1.2	8.2	22.6	22.1	15.2	9.8
	C9	0.9	2.3	2.2	3.5	12.8	9.7	14.6	13.1	53.6

categorical data, with C1 representing non-urban (cell state = 0), and C9 representing urban (cell state = 1), and C2–8 representing states where their state value are linearly interpolated from 0 to 1 (exclusive).

32.4.2.1 The Confusion Matrix

Using a cell-by-cell comparison, Table 32.3 shows the confusion matrix between the actual urban states defined from the census data and the simulated urban states in 2006 using the categorical data reclassified from the continuous data.

Table 32.4 Landscape metrics used to assess the simulated results

Landscape metrics	Meaning of the landscape metric[a]	Unit
NP	Number of patches in the landscape	None
PD	Patch Density. This is the number of patches of the landscape per 100 ha	Number per 100 ha
LPI	Largest Patch Index, which shows the percentage of total landscape area comprised by the largest patch	%
LSI	Landscape Shape Index, which provides a standardized measure of total edge or edge density that adjusts for the size of the landscape	None
PAFRAC	Perimeter-Area Fractal Dimension, which indicates the shape complexity across a range of spatial scales	None
DIVISION	Landscape Division Index, which shows the probability that two randomly chosen cells in the landscape are not situated in the same patch	None
SPLIT	Splitting Index, which indicates the effective mesh number, or number of patches with a constant patch size when the corresponding patch type is subdivided into S patches, where S is the value of the splitting index	None
AI	Aggregation Index, which shows the frequency with which different pairs of patch types appear side-by-side on the map	%

[a]For computation of these metrics, please refer to McGarigal et al. (2002)

Table 32.5 Computed landscape metrics of the Gold Coast City in 1996 and 2006

Metrics	NP	PD	LPI	LSI	PAFRAC	DIVISION	SPLIT	AI
1996 Actual	1,050	0.765	85.270	12.591	1.246	0.272	1.374	98.391
2006 Actual	1,371	0.977	74.950	18.938	1.237	0.435	1.771	97.556
2006 Simulated	5,348	3.895	79.591	21.550	1.345	0.364	1.573	96.949

Based on the confusion matrix, the model generated an overall simulation accuracy of 87%. However, the Kappa coefficient is only 58% indicating a large number of mismatches of cell states due to commission or omission errors. Nevertheless, the high overall simulation accuracy and the higher matching score of 96.5% for non-urban and 53.6% for urban states indicate it is possible to apply the proposed model for urban growth simulation.

32.4.2.2 The Landscape Metrics

Using the FRAGSTATS software program which is freely available from McGarigal et al. (2002), the major landscape metrics used for assessing the model's simulation accuracies are listed in Table 32.4 and such metrics for the actual and simulated urban patterns of the Gold Coast City in 1996 and 2006 are computed and listed in Table 32.5.

Table 32.5 shows that the actual number of patches of the landscape increased from 1,050 in 1996 to 1,371 in 2006; their patch densities also increased from 0.765 to 0.977 patches per 100 ha during the same period. However, the Largest Patch

Index (LPI) decreased from 85.27% to 74.95%, indicating that the size of the largest land patch, that is, the non-urban agricultural land, decreased over this time period. While the shape of land patches measured through PAFRAC was almost stable, the increased value of the Landscape Shape Index (ISI) and the SPLIT index indicate an increased land use complexity and subdivision of the landscape, and the slightly decreased AI value also indicates an increasing tendency of land use disaggregation of the region.

Comparing the landscape metrics generated from the simulated results with those of the 2006 actual cell states, major discrepancies exist in the total number of patches and the patch densities, where the simulated results generated much larger values than those from the actual urban patterns. However, the temporal landscape change patterns of the simulated results match with the actual landscape metrics measurements, indicating the validity of the simulation results.

32.5 Conclusion

Logistic regression is a classical method which has been extensively tested, validated and applied in many research to model the patterns and relationships between the dependent variable and a set of independent variables through a probability surface. The simplicity of its modelling mechanism matches well with cellular automata modelling where complex spatial patterns and relationships can be simulated through a set of simple transition rules. Hence, the integration of logistic regression in urban cellular automata modelling is natural and valid in simulating the spatio-temporal process of urban growth (Dendoncker et al. 2007; Fang et al. 2005; Cheng and Masser 2004; Verburg et al. 2002; Wu 2002; Sui and Zeng 2001).

This chapter contributes to urban growth modelling by developing a logistic regression based cellular automata model using continuous data to simulate the continuous progression of land from non-urban to urban use. The application of the model in Australia's Gold Coast City has generated realistic urban growth scenarios, which matches well for the actual urban growth patterns. Assessment of the landscape metrics shows not only a structural similarity between the simulated and actual landscape patterns, but also consistency in terms of the temporal change of the landscape metrics. This proves the validity of the logistic regression based CA modelling as well as the effectiveness of applying landscape metrics to evaluate the simulation accuracies of the model.

Further research is required to apply the logistic regression based CA model to generate future urban growth scenarios based on projections of population growth and dwelling increases. Population projections made by the Queensland Government Department of Demography and Planning show that by 2031 the expected population of the Gold Coast City will be between 724,000 and 867,000 people (low and high series) (Office of Economic and Statistical Research 2010). This represents 53–83% increase from the resident population in 2006. The application of the cellular automata modelling should provide useful insights by generating different growth scenarios for urban planners and government decision makers to consider in order to managing such substantial urban growth of the city in future.

References

Almeida, C. M., Gleriani, J. M., Castejon, E. F., & Soares-Filho, B. S. (2008). Using neural networks and cellular automata for modelling intra-urban land-use dynamics. *International Journal of Geographical Information Science, 22*, 943–963.

Australian Bureau of Statistics. (2005). Information paper: Draft mesh blocks, Australia, 2005 ABS Cat. No. 1209.0.55.001.

Australian Bureau of Statistics. (2006). CDATA Online. Canberra: Australian Bureau of Statistics. http://www.abs.gov.au/CDATAOnline. Accessed 01 Aug 2011.

Australian Bureau of Statistics. (2011). Regional Population Growth, Australia, 2009–10 ABS Cat. No. 3218.0.

Bryant, C. R., Russwurm, L. H., & McLellan, A. G. (1982). *The city's countryside: Land and its management in the rural-urban fringe*. New York: Longman.

Cheng, J., & Masser, I. (2004). Understanding spatial and temporal processes of urban growth: Cellular automata modelling. *Environment and Planning B: Planning and Design, 31*, 167–194.

Clarke, K. C., & Gaydos, L. J. (1998). Loose-coupling a cellular automaton model and GIS: Long-term urban growth prediction for San Francisco and Washington/Baltimore. *International Journal of Geographical Information Science, 12*, 699–714.

Clarke, K. C., Hoppen, S., & Gaydos, L. J. (1997). A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B: Planning and Design, 24*, 247–261.

Dendoncker, N., Rounsevell, M., & Bogaert, P. (2007). Spatial analysis and modelling of land use distributions in Belgium. *Computers, Environment and Urban Systems, 31*, 188–205.

Department of Infrastructure and Planning. (2009). *South East Queensland Regional Plan* 2009–2031. Brisbane: Queensland Government. http://www.dlgp.qld.gov.au/regional-planning/south-east-queensland-regional-plan-2009-2031.html. Accessed 30 Aug 2011.

Dragićević, S. (2004). Coupling fuzzy sets theory and GIS-based cellular automata for land-use change modelling. In *Fuzzy Information, IEEE Annual Meeting of the Processing NAFIPS'04, Banff*, pp. 203–207.

Fang, S., Gertner, G. Z., Sun, Z., & Anderson, A. A. (2005). The impact of interactions in spatial simulation of the dynamics of urban sprawl. *Landscape and Urban Planning, 73*, 294–306.

Feng, Y. J., Liu, Y., Tong, X. H., Liu, M. L., & Deng, S. (2011). Modeling dynamic urban growth using cellular automata and particle swarm optimization rules. *Landscape and Urban Planning, 102*, 188–196.

Gobim, A., Campling, P., & Feyen, J. (2002). Logistic modeling to derive agricultural land determinants: A case study from southeastern Nigeria. *Agriculture, Ecosystem and Environment, 89*, 213–228.

Grimm, V., & Railsback, S. F. (2005). *Individual-based modeling and ecology*. Princeton: Princeton University Press.

Grimm, V., & Railsback, S. F. (2012). Designing, formulating and communicating agent-based models. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 361–377). Dordrecht: Springer.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S., Huse, G., Huth, A., Jepsen, J. U., Jørgensen, C., Mooij, W. M., Müller, B., Pe'er, G., Piou, C., Railsback, S. F., Robbins, A. M., Robbins, M. M., Rossmanith, E., Rüger, N., Strand, E., Souissi, S., Stillman, R. A., Vabø, R., Visser, U., & DeAngelis, D. L. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling, 198*, 115–126.

Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., & Railsback, S. F. (2010). The ODD protocol: A review and first update. *Ecological Modelling, 221*(23), 2760–2768. doi:10.1016/j.ecolmodel.2010.08.019.

Herbert, D. T., & Thomas, C. J. (1997). *Cities in space: City as place* (3rd ed.). London: David Fulton.

Iltanen, S. (2012). Cellular automata in urban spatial modelling. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 69–84). Dordrecht: Springer.

Jakobson, L., & Prakash, V. (1971). *Urbanization and national development*. Beverley Hills: Sage.

John, F. B., Peter, W. N., Peter, H., & Peter, N. (1985). *The future of urban form: The impact of new technology*. London: Croom Helm.

Li, H., & Wu, J. (2004). Use and misuse of landscape indices. *Landscape Ecology, 19*, 389–399.

Li, X., & Yeh, A. G. O. (2000). Modelling sustainable urban development by the integration of constrained cellular automata and GIS. *International Journal of Geographical Information Science, 14*, 131–152.

Li, X., & Yeh, A. G. O. (2002a). Urban simulation using principal components analysis and cellular automata for land-use planning. *Photogrammetric Engineering and Remote Sensing, 68*(4), 341–351.

Li, X., & Yeh, A. G. O. (2002b). Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographical Information Science, 16*(4), 323–343.

Linge, G. J. R. (1965). *The delimitation of urban boundaries for statistical purposes with special reference to Australia*. Canberra: Australia National University.

Liu, Y. (2008). *Modelling urban development with geographical information systems and cellular automata*. New York: CRC Press.

Liu, Y., & Phinn, S. R. (2003). Modelling urban development with cellular automata incorporating fuzzy-set approaches. *Computers, Environment and Urban Systems, 27*, 637–658.

Liu, X., Li, X., Shi, X., Wu, S., & Liu, T. (2008a). Simulating complex urban development using kernel based non-linear cellular automata. *Ecological Modelling, 211*(1–2), 169–181.

Liu, X., Li, X., Liu, L., He, J., & Ai, B. (2008b). A bottom-up approach to discover transition rules of cellular automata using ant intelligence. *International Journal of Geographical Information Science, 22*(11), 1247–1269.

Mandelas, E. A., Hatzichristos, T., & Prastacos, P. (2007). A fuzzy cellular automata based shell for modeling urban growth – a pilot application in Mesogia area. In *10th AGILE International Conference on Geographic Information Science 2007, Denmark*.

McGarigal, K., Cushman, S. A., Neel, M. C., & Ene, E. (2002). FRAGSTATS: Spatial pattern analysis program for categorical maps. *Computer software program produced by the authors at the University of Massachusetts, Amherst*. http://www.umass.edu/landeco/research/fragstats/fragstats.html. Accessed 03 May 2011.

Office of Economic and Statistical Research. (2010). *Population and housing profile – Gold Coast City council*. Brisbane: Queensland Government. http://www.oesr.qld.gov.au/products/profiles/pop-housing-profiles-lga/pop-housing-profile-gold-coast.pdf. Accessed 30 Aug 2011.

Openshaw, S. (1984). The modifiable areal unit problem. In *Concepts and techniques in modern geography 38*. Norwich: Geo Books.

Potter, R., Binns, T., Elliott, J. A., & Smith, D. (2003). *Geographies of development* (2nd ed.). London: Longman.

Power, C., Simms, A., & White, R. (2001). Hierarchical fuzzy pattern matching for the regional comparison of land use maps. *International Journal of Geographical Information Science, 15*(1), 77–100.

Pryor, R. J. (1968). Defining the rural-urban fringe. *Social Forces*, 202–215.

Sherrod, P. H. (2010). *DTREG: Predictive modeling software*. http://www.dtreg.com/DTREG.pdf. Accessed 01 May 2011.

Sui, D. Z., & Zeng, H. (2001). Modeling the dynamics of landscape structure in Asia's emerging desakota regions: A case study in Shenzhen. *Landscape and Urban Planning, 53*, 37–52.

Verburg, P. H., Soepboer, W., Veldkamp, A., Limpiada, R., Espaldon, V., & Mastura, S. (2002). Modeling the spatial dynamics of regional land use: The CLUE-S model. *Environmental Management, 30*(3), 391–405.

Ward, D. P., Murray, A. T., & Phinn, S. R. (2000). Monitoring growth in rapidly urbanizing areas using remotely sensed data. *The Professional Geographer, 52*, 371–386.

Ward, D. P., Murray, A. T., & Phinn, S. R. (2003). Integrating spatial optimization and cellular automata for evaluating urban change. *The Annuals of Regional Science, 37*, 131–148.

White, R. W., & Engelen, G. (1993). Cellular automata and fractal urban form: A cellular modelling approach to the evolution of urban land use patterns. *Environment and Planning A, 25*, 1175–1193.

Wu, F. (1996). A linguistic cellular automata simulation approach for sustainable land development in a fast growing region. *Computers, Environment, and Urban Systems, 20*, 367–387.

Wu, F. (1998a). An experiment on the generic polycentricity of urban growth in a cellular automatic city. *Environment and Planning B: Planning and Design, 25*, 731–752.

Wu, F. (1998b). Simulating urban encroachment on rural land with fuzzy-logic-controlled cellular automata in a geographical information system. *Journal of Environmental Management, 53*(16), 293–308.

Wu, F. (1998c). SimLand: A prototype to simulate land conversion through the integrated GIS and CA with AHP-derived transition rules. *International Journal of Geographical Information Science, 12*, 63–82.

Wu, F. (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., & Webster, C. J. (1998). Simulation of land development through the integration of cellular automata and multicriteria evaluation. *Environment and Planning B: Planning and Design, 25*, 103–126.

Wu, F., & Yeh, A. G. O. (1997). Changing spatial distribution and determinants of land development in Chinese cities in the transition from a centrally planned economy to a socialist market economy: A case study of Guangzhou. *Urban Studies, 34*, 1851–1879.

Yang, Q., Li, X., & Shi, X. (2008). Cellular automata for simulating land use changes based on support vector machines. *Computers and Geosciences, 34*(6), 592–602.

Zhu, Z., Liu, L., Chen, Z., Zhang, J., & Verburg, P. H. (2009). Land-use change simulation and assessment of driving factors in the loess hilly region – a case study as Pengyang County. *Environmental Monitoring and Assessment, 164*(1–4), 133–142.

Chapter 33
Exploring Demographic and Lot Effects in an ABM/LUCC of Agriculture in the Brazilian Amazon

A. Raymond Cabrera, Peter Deadman, Emilio Moran, Eduardo S. Brondízio, and Leah K. Vanwey

Abstract Building upon previous modelling efforts, an agent-based model of land use change has been developed to model small-scale agriculture in the Brazilian Amazon. The model, LUCITA (Land Use Change in the Amazon), simulates heterogeneous farming household agents during the period of 1970–2000, settling and developing properties west of Altamira, Pará, Brazil. Farming agents, as heuristic agents that satisfy subsistence needs then maximize utility, clear old growth forested areas to pursue agricultural activities. The simulation is utilized here to explore how the development of the household lifecycle is altered when households have access to outside labour resources, and when households settle on previously occupied properties. Simulation results support the assertion that the household lifecycle model alone is insufficient to explain patterns of land use change in these rural frontier-like environments.

33.1 Introduction

Since the early 1970s, a series of initiatives, including government-sponsored economic incentives and the construction of the Transamazon highway, have resulted in rapid land use change within the Brazilian Amazon rainforest. While industrial agriculture, commercial lumber operations, mining, hydroelectric development, and

A.R. Cabrera (✉) • P. Deadman
Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON, Canada
e-mail: raymond.cabrera@uwaterloo.ca

L.K. Vanwey
Department of Sociology, Brown University, Providence, RI, USA

E. Moran • E.S. Brondízio
Anthropological Center for Training and Research on Global Environmental Change, Department of Anthropology, Indiana University, Bloomington, IN, USA

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_33,

highway construction have brought large-scale land use change to the Amazon region, important changes have also happened at smaller scales. In particular, one of the more important local scale actors that have made a significant contribution to land use change in the region are the individual farming households. Researchers focused on understanding land use and land cover change (LUCC) have frequently focused on household-level drivers of change in rural forested frontier environments. There are a number of reasons for this. In many forested frontier environments, smallholders are significant drivers of change. Rural people comprise a very large proportion of the populations of the developing world, providing important foodstuffs and other resources to nearby cities where the majority of people in developing countries are now living. A better understanding of the dynamics of LUCC in these environments can inform policy developments focused on the improvement of rural livelihoods (de Sherbinin et al. 2008).

One forested frontier that has been the focus of much research is the region along the Transamazon highway, west of Altamira, Brazil. For more than a decade, researchers have focused on a variety of questions, including; the rates and patterns of deforestation (Brondízio et al. 2002), the nature and patterns of secondary succession (Lu et al. 2002; Mausel et al. 1993; Moran et al. 1994, 2000), the differences between patterns of land use on neighbouring farms (McCracken et al. 2002), and the influence of natural factors, such as soil quality, on household livelihood (Moran 1995; Moran et al. 2000, 2002). Farming households have frequently been the focus here, as they are the primary unit of production and decision making in this landscape. In particular, neighbouring farms in the Altamira region have been observed to have markedly different patterns of land use (McCracken et al. 2002). This observation has raised research questions related to the relative importance of natural, household, and broader market drivers in shaping land use decisions.

A great deal of debate has focused on the importance of household-level demography in shaping the land use activities of these smallholders. The household lifecycle model, as discussed by Goody (1958) and de Sherbinin et al. (2008), typically starts with a newly married couple arriving on the frontier with limited labour and capital resources. Over time, children are added to the household, grow older, and begin to contribute to the family's labour and economic resources, eventually leaving or staying within an extended family to assume the role of decision maker. Researchers have explored the utility of the lifecycle model for explaining patterns of land use amongst smallholders in the Altamira region over time, where the type of agricultural practices pursued are a function of the labour and capital resources within the household (McCracken et al. 1999; Moran et al. 2002). Here, the lifecycle model indicates that young families with limited resources will initially seek to clear forest on their properties and plant annual crops, such as rice, beans, maize, or manioc. As families age, and labour and capital resources grow, they continue the pattern of deforestation and the planting of annuals, while converting previously cleared land into pasture or perennial crops. Available household resources are seen as a key driver of land use change, as households with greater resources tend to gravitate towards a more diversified land use with an emphasis on perennial crops such as fruit trees, coffee, or black pepper. Households with more limited resources tend to follow land use strategies focused on pasture development and cattle production (McCracken et al. 2002).

More recent studies (de Sherbinin et al. 2008; VanWey et al. 2007) found limited support for the household lifecycle model, noting inconsistent effects of household demography on LUCC. VanWey et al. (2007) found no significant effects of working age males on land use or land cover change, but significant effects for women and children. The study notes the importance of cash, from off-farm employment or government subsidies, as a household resource that facilitates investments in agriculture and subsequent land use and land cover change. These studies call into question the effectiveness of the lifecycle model as a tool to explain patterns of land use. In these complex environments, multiple drivers beyond household demographics appear to be driving land use decision-making. The agent-based simulation outlined here is designed to explore the household lifecycle model.

The challenges posed by the complexity of these environments has led, more recently, to an increasing number of interdisciplinary efforts to explore the complex dynamics of coupled human and natural systems (CHANS) (Liu et al. 2007a, b). These studies have explored the complexity of organizational, spatial, and temporal couplings within CHANS, revealing that these interactions have evolved to become more indirect, distant, and global in scale (Liu et al. 2007b). These couplings also exhibit nonlinear dynamics with multiple thresholds, feedback loops, resilience, and heterogeneity (Liu et al. 2007a). Unravelling and exploring these complex relationships will require a variety of interdisciplinary approaches and tools.

Spatially referenced agent-based models (ABMs) are one tool that has been utilized to explore dynamic interactions within CHANS. In particular, simulations utilizing collections of agents to represent the actions of specific human or natural entities have been explored in many fields, including land use change (Parker et al. 2008b; Rindfuss et al. 2008). These models couple a human system, represented by a collection of agents making land use decisions with an environment system, represented by a raster grid of spatially distributed resources within the landscape, through agent-agent and agent-landscape interactions that feedback and alter the land use or cover change in the area of interest (Parker et al. 2002, 2003). Researchers have also utilized ABM/LUCC to explore the drivers of change in frontier environments (see Parker et al. 2008b for a review). Recent work has focussed on assessing the utility of these tools (Messina et al. 2008) and the need for cross-site comparisons of these models (Hare and Deadman 2008; Parker et al. 2008a, b; Rindfuss et al. 2008).

Here, an ABM, called LUCITA (Land Use Change in the Amazon), is utilized to explore the dynamics of smallholder driven land use change. The study site explored here is the area along the Transamazon highway, west of Altamira, in the state of Pará, Brazil. This is a region characterized by the creation of household farms on 100ha lots located on the highway and a series of side roads running perpendicular to the highway at 5km intervals, creating what has come to be known as the fishbone patterns of settlement in the Amazon. Starting with a baseline simulation that captures some of the dynamics of the household lifecycle model, we observe how the behaviour of the model changes when parameters such as available household capital, available outside labour, and previous land use on a lot are added. These explorations indicate that multiple interacting drivers of change produce different patterns of land use that cannot be explained by the household lifecycle model alone.

33.2 Model Description and Methods

The ODD (Objects, Design concepts, Description) protocol will be used to provide a standardized description of LUCITA (Grimm et al. 2006, 2010; Grimm and Railsback 2012).

33.2.1 Purpose

LUCITA was formulated to measure the effects of household demographics on deforestation in the Brazilian Amazon during the period between 1970, when the area was settled by migrants along the Transamazon Highway, and 2000 (Deadman et al. 2004). Given hypothetical limitations of capital (or lack thereof), as well as limitations on labour based on household size and hired labour (or lack thereof as well), are household demographics (household size and composition) a predictive indicator of deforestation and subsequent land use, or are there more relevant factors to consider?

LUCITA, implemented as a spatial ABM, provides an experimental laboratory for varying household composition, initial capital, immigration cohort size and available land use choices. This allows researchers to explore a series of scenarios from basic behaviours. Such a scenario could be developed to emulate real-world conditions or an alternative hypothetical outcome. In the case of LUCITA, the model is used as a means to explore how variations in household size and composition as well as initial capital and other factors affect macro-level deforestation and land use.

33.2.2 Entities, State Variables and Scales

The ABM is divided into layers: the household layer, wherein agents perform land use and land cover activities, a cadastral layer (a raster grid of properties), and a cellular land cover model. The land cover model is composed of a cellular model of crops, pasture and forest imposed upon a soil grid. One time step in the model represents 1 year. The model is run for a period of 30 years, theoretically covering the period of 1970–2000. The property grid and land cover model have matching extents and resolution: One grid cell represents 1 ha, of which there are 381,000 cells and 3,916 properties comprising the simulation space.

The primary entity of the model is the household agent. The household itself is implemented as an agent that performs farming and harvesting activities, as well as the decision-making involved to pursue such activities. The household resides on a property, which is composed of a set of cells on the cadastral layer and a matching set of cells on the land cover grid. The household's elementary state variables are: property ID and arrival year. Other variables are derived from household members, land cover composition and history (net and total deforestation). The state variables of household members are: age, gender, capital, and contributing labour. The contributing labour variable indicates the amount of labour, in hours per year, which the agent is willing to contribute to farming activities.

Each property has attributes related to its location and use: property ID, a list of cells included in the property, direction to road, and the year the property is first settled. Property grid cells have only a location (x, y) state variable, but also include the distances to the nearest road, highway, and a number of settlements. Other than the year of first settlement, property and property cell attributes are determined *a priori*.

Properties are populated by land cover cells. Land cover cells have a number of attributes related to the lifetime and farming of the occupying land cover class: location (x, y), required labour, initial fixed cost, maintenance cost, age of the land cover (in other words, years since the land was assigned for said land use), death age (total years that the cell can be used for this land use), seed requirements, planting density, expected yield, and years until production. The cell's yield from the previous year is also stored for reference. Concrete land cover classes, which extend the abstract land cover object, include forest, annual crops, perennial crops, pasture and fallow. These crops follow rules established by Fearnside (1986), detailed specifically by Lim (2000) and Robinson (2003).

Soil cells, themselves governed by the Fearnside-based model, include state variables that characterize nutrients and soil conditions: pH, aluminium, nitrogen, carbon, clay, phosphorus, predicted erosion and average precipitation.

Crops may be bought and sold on a market. LUCITA's market is simple, with constant prices throughout the simulation. Similarly, a labour market is implemented with constant buying and selling prices for a year's worth of labour. However, limitations on the number of farming agents in the labour market may be placed to enforce a maximum amount of available labour.

33.2.3 Process Overview and Scheduling

A model year (one discrete time step) begins with an immigration of households, each occupying one property (*immigration*). Households, new and old, convert or maintain land for cultivation, pasture and fallow, according to their needs and resources (*land allocation*). Based on the year's land uses, households can harvest yield and reap revenues (*harvest*). After the harvest stage, *cleanup* is performed. A time-varying environment is not driven from exogenous factors (input data), so the Input Data section of the ODD protocol is omitted from this paper.

Immigration: The number of households immigrating during a given year may be set to correspond to historical cohorts (Brondízio et al. 2002) or other scenarios. A household can migrate into the simulation if at least one property is available. If several properties are available, the incoming household samples 3 properties and selects the one that is closest to the main road leading to Altamira. The new household sets its property ID and arrival year state variables accordingly. Among the 6 available choices of annual crops (bitter manioc, maize, phaseolus beans, rice, sweet manioc and cowpeas), the household chooses 4–6 of these as its preferences upon immigration. (Two uniformly random choices are involved: The household first chooses how many crops to prefer, then selects the crops.) Annual crop choices are restricted to these 4–6 crops for the remainder of the simulation. Perennial crop choices, black pepper and cacao, as well as pasture and fallow are available to all households.

Land Allocation: During the land conversion process, each household iterates through each of its available cells in order from closest to furthest away from the road. As each cell is iterated through, the household will convert or maintain the cell based on its state, subject to available capital and labour. If internal household labour is unavailable and sufficient external labour can be purchased, the household will purchase any labour necessary to process the next cell. If the cell is forested, the household will clear and burn (fallow) the forest. If a cell is fallowed and the soil is moist or last burned more than 3 years ago, a second burn must be performed. If a cell is fallowed, dry and last burned less than 3 years prior to the current year, then a number of options are available. If subsistence needs are met and soil pH is greater than 5.5, then perennials are planted. Otherwise, if pH is too low, then the cell is used for pasture. If pH is greater than 5.5 and the household requires subsistence crops, then a random annual is chosen. An annual crop is chosen from the restricted set of preferred annuals by weighted random selection. Crop choices are weighted by the selling price of the good (per kg) rather than the expected revenue of the cell's yield, as this is considered to reflect local decision-making practices (Moran 1981). During this cell-by-cell traversal, the household is deducted costs and labour. Each land use, other than virgin forest, has an initial fixed cost, an annual maintenance cost and an annual amount of required labour. During the year of conversion (the first year), the initial fixed cost is deducted from the household's capital pool. Subsequent years deduct the maintenance cost. The household continues traversing through cells until it runs out of capital or labour. Resource limitations generally prevent households from clearing their properties in a given year.

Harvest: Similar to the land allocation process, households traverse through their property, cell-by-cell in increasing distance from the road to perform harvesting. The labour and costs required to harvest have been factored into the land allocation step, so this is implemented as a cost-free process. Households calculate their yields by summing the yield from each crop cell. Land cover cells perform natural transitions in this step. For instance, if a fallowed cell is allowed to mature for 20 years, it transitions to a forest cell.

Cleanup: Failed households, those accumulating significant debt, are removed from the simulation and vacate their properties. At the end of the step, household members are aged by 1 year. State variables related to available labour and subsistence needs, derived from the ages and genders of household member agents, are reset for the new year.

33.2.4 Design Concepts

Basic principles: Given utility maximizing households with resource limitations of labour and capital, as well as subsistence needs which must be satisfied, it should be expected that household demographics play a direct, if not linear, role in determining the degree and nature of land use and, by extension, the total amount of deforestation. The development of LUCITA has been an effort to explore this hypothesis by implementing household agents in a virtual laboratory where household sizes and resource constraints could be varied in a series of runs. The environmental model, based on Fearnside's work, has been given sufficient complexity to present households with variable yields based on past actions and environmental conditions. Households have been given behaviours, which while not perfectly rational, reflect local decision-making practices.

Emergence: LUCITA has demonstrated emergent outcomes that have not been modelled explicitly, including fish-bone deforestation patterns representative of historical outcomes. While a relationship between deforestation and demographics can be demonstrated, the relationship between lot life cycles and land use is less obvious without the development of a model.

Adaptation: Household agents are implemented as heuristic decision makers, selecting actions based on conditional if-then decision rules, followed by weighted selection between feasible alternatives. Each of these actions seeks to better the agent's position, first with respect to meeting subsistence needs, then to profitability and future sustainability.

Objectives: The household agents in LUCITA are modelled as utility maximizing decision makers. Households first seek to satisfy subsistence needs, and then perform actions in order to increase profits and/or keep the land suitable for future use (such as fallowing and pasture).

Prediction: Household crop selections are made through a weighted selection based on crop prices (Moran 1981). Within a model where market clearing conditions are easily met regardless of supply and with little risk, purely rational agents would choose far more homogeneous landscapes of fewer crops (Cabrera et al. 2010). The weighted random choice used by households in LUCITA produces heterogeneous landscapes that mitigate risk.

Sensing: In making land allocation decisions, households are assumed to be able to sense (determine or measure, with perfect knowledge) soil conditions, market prices and current household resources. Households do not, however, consider the actions of other households, such as neighbouring agents, nor do they consider actions and consequences occurring in future years.

Interaction: Households generally do not interact. On occasion, a household may hire a member of a failed household for extra labour, but this is done after the latter household has already vacated its property.

Stochasticity: Stochastic processes are used to seed a number of model variables: household size, initial capital. The underlying processes governing these variables are exogenous to the model. Many of the choices made by households are based on uniform or weighted random selection: property selection, crop selection. These random choices are made when a number of feasible alternatives are available to the household, yet the underlying drivers of these choices such as risk, profit maximization and competition are not modelled explicitly, yet heterogeneity between households is desired.

Collectives: Households, as collectives of individual members, are implemented as explicit agents.

Observation: Model outputs are recorded to PostgreSQL, a relational database. All variable input parameters are included, as are household states at the end of each year and property histories. Household state information includes the number of cells assigned to each land use (each type of crop, forest, fallow and pasture), member ages and genders and the year of arrival. Subsistence needs, although derived from ages and genders of household members, are written to the database to simplify later calculation. In addition, land use raster grids are written for each year to separate files outside the database. Property histories are also written to the database, summarized as the year of first settlement. Resettled properties can be determined by comparing the year of first settlement with the residing household's year of arrival, as households do not change properties within the simulation. This provides near-perfect observation of the system for the variables under analysis.

33.2.5 Initialization

The landscape begins in 1970 as virgin forest with no settled households and 3,916 available properties. Migrating households are assigned an initial capital amount of 500–6,000, depending on parameterization (Fearnside 1986). Incoming family sizes may vary from 2 to 12 members, again, depending on parameterization. For the purposes of model analysis, small households consist of smaller than 6 members, whereas large households consist of 6 or more members.

33.3 Analysis

We begin with a discussion of the baseline, or default, version of the model, in which households do not have access to outside sources of labour, and must rely on their own labour resources for agricultural production. Aggregated land use trajectories of these individual household farms verify that household agents follow the generally theorized pattern of initially growing short-term crops (annuals, in particular), followed by more permanent plots of perennials and pasture. Examining Figs. 33.2 and 33.3, we also see that households follow a pattern of deforestation

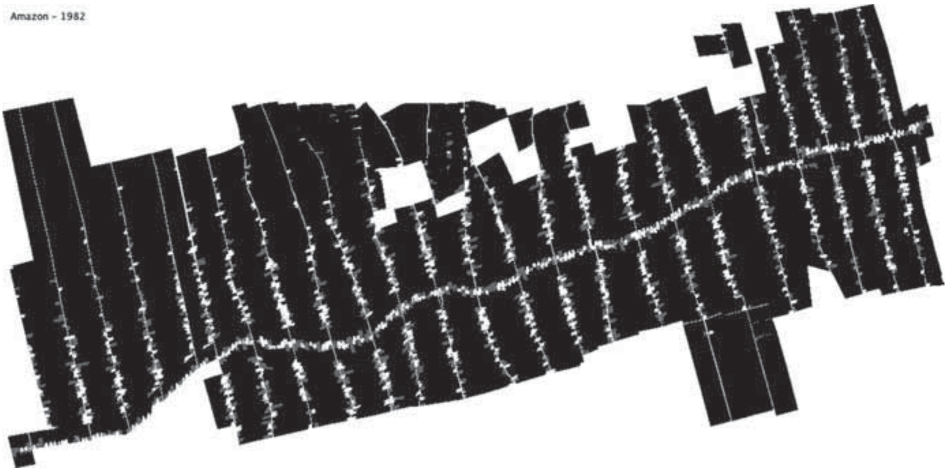


Fig. 33.1 Birds-eye view of land use in LUCITA, illustrating the deforestation pattern

characterized by an initial pulse of higher activity in the first 3 or 4 years, followed by reduced deforestation rates in later years. Spatially, the deforestation follows that of a fish-bone pattern over time, as households settle on a property and begin clearing the forest, starting at the road and moving towards the back of the property (Fig. 33.1). Household agents in the model generally purchase subsistence crops in favour of allocating labour resources toward longer-term and more profitable goods such as perennials and, depending on soil conditions (for cells with low pH, for instance), pasture.

Examining the deforestation trajectories for an aggregated set of households, the effects of household demographics and lot life cycles can be assessed. Three sets of runs are presented, each representing a minimum of 2,000 households. The default run presents a uniform distribution of household family sizes and a broad range of initial capital. The other two runs fix initial capital parameters, and examine the effect of varying household size to a high or low value. Small households consist of 2–6 members, including adults and children, while large households consist of 6–10 members. Within these ranges, a uniform random distribution determines the initial size of the household. An additional three runs are analysed where an outside source of cheap labour is available.

Figure 33.2 illustrates averaged trajectories of total deforestation assumed by households of different sizes, comparing the absence and presence of a labour market. Based on a standard run with no externally available labour (Fig. 33.2a), the model runs exhibit varying levels of total deforestation in response to household size. As would be expected, the additional internal labour resources of larger households allow them to deforest more land, converting it to agricultural production, initially annuals followed by perennials and pasture. In these runs, the level of capital resources in a household makes little difference in terms of long-term deforestation rates, since external labour is not available for purchase.

Larger households, with additional labour, and thus relaxed labour constraints, allow them to cultivate additional plots, each of which carry initial and long-term

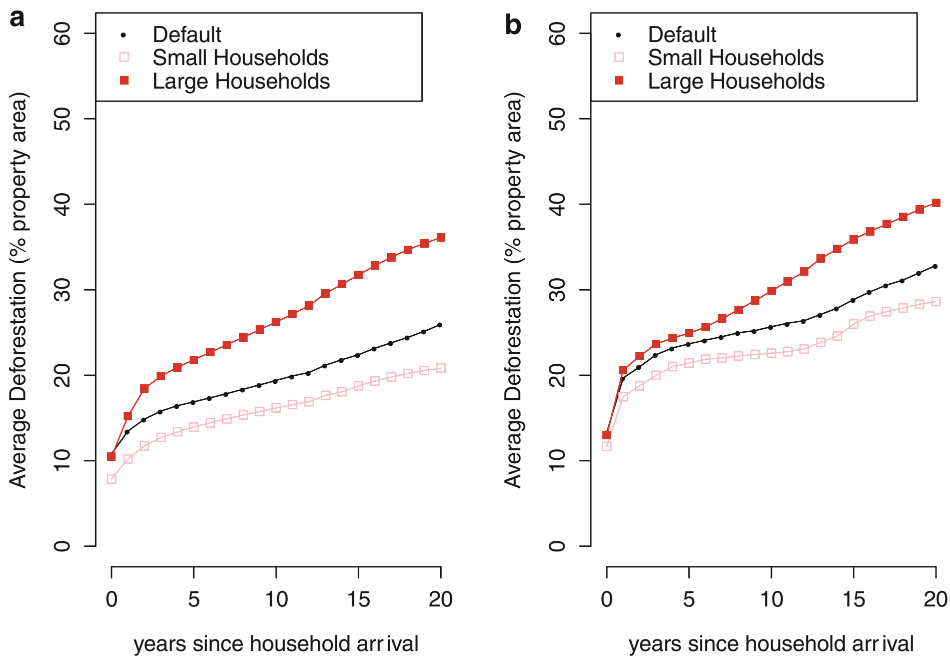


Fig. 33.2 Average property deforestation in the (**a**) absence or (**b**) presence of external labour

maintenance costs. In addition, large households face increased subsistence requirements, in the form of capital expenses and subsistence crops. Households with more available labour clear more forest than those with less labour by an additional 73.4% on average, indicating that labour is the limiting constraint.

We can begin to see how the behaviour of the model diverges from the patterns of the household lifecycle model when we add additional drivers of land use change to the simulation. By introducing a source of cheap outside labour (Fig. 33.2b), households are able to purchase additional labour to cultivate or maintain additional land. This relaxes the labour constraint, especially benefitting smaller households that are quite limited in terms of farming capacity. In this scenario, households accumulate less capital since it is spent on additional labour. These households are not rational revenue-maximizing agents, as their decision-making is governed by a set of heuristic rules. These rules maximize land use, and by extension, household productivity. Here, household demographics play a more limited role in determining overall deforestation rates, and long-term household wealth and productivity, although they are still a factor. Larger households continue to have an advantage in productivity above smaller households, accounting for a 40.4% increase in deforestation when a labour market is present. However, experiments have shown that this difference in productivity can be further mitigated by artificially lowering the cost of labour.

We further explore how patterns of land use change within a household can be altered when we introduce another element of complexity to the model. The deforestation trajectory can be analysed in the context of lot life cycles by comparing the settlement of virgin properties to previously settled properties. The period of

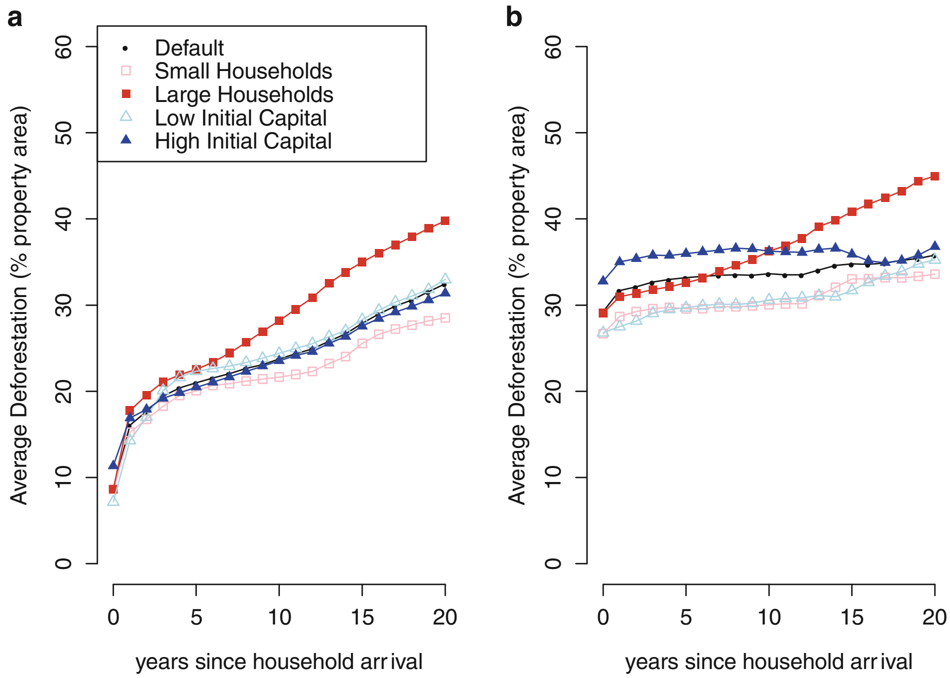


Fig. 33.3 Comparison of deforestation trajectories for (**a**) virgin and (**b**) non-virgin properties

1970–2000 was largely a pioneering period of settlement. Thus, most households in the model settle into virgin properties, especially at the beginning of a simulation run. Most household failures occur within the first few years of settlement. Therefore the typical deforestation trajectory at the time the household vacates its property would likely be, on average, 20%. However, Fig. 33.3, which separates deforestation trajectories by virgin and non-virgin properties (both types of properties in the presence of a labour market), indicates that that the initial clearing of non-virgin properties is significantly more aggressive, as the household settling into the vacated property does not need to expend the effort and costs to clear the property. That initial pulse of deforestation activity is not seen within households that settle on a previously occupied property. Instead, resources are allocated toward maintenance and recovery, with remaining labour allocated to deforestation. Year to year, however, the net rate of deforestation is significantly tempered after the initial clearing. Yet, the differences between deforestation trajectories do not narrow until approximately a 15–20 year period has passed. A similar trend is observed when a labour market is not present.

33.4 Discussion

Overall, exploration of the model reveals how added complexity, interpreted here in terms of additional parameters, can cause patterns of land use to diverge from those seen in a household lifecycle type model. We started with a simple default version of the model, one in which a Chayanovian assumption of a household without access to external sources of labour or capital can be seen (de Sherbinin et al. 2008). This model is capable of producing patterns of land use change in which household size is a fundamental determinant of overall amounts of land use change. Since households may not acquire external labour, the availability of household adolescent and adult males is a key limit to deforestation activity.

By introducing available outside labour to the simulation, the importance of internal household agricultural labour, represented demographically as male members of the household, is diminished, while the importance of capital resources, which may be generated by female off-farm labour or government subsidization programs, becomes more important to overall farm productivity. Available capital, coupled with available outside labour, significantly reduces the difference in deforestation rates between small and large households.

We can also demonstrate a clear change in simulated household behaviour when we examine differences between households settling on virgin properties from those settling on previously occupied properties. This lot effect influences household activities, as seen by the absence of a pulse of initial deforestation for those settling on a previously cleared property. The original pattern of initial land clearing to grow annuals is also less evident in households arriving on previously occupied properties. Household agents are not clearing perennials or pasture to grow annuals.

The simulations discussed here certainly do not provide enough evidence to completely discard the household lifecycle model as a tool to explain patterns of land use change in forested frontiers. The simulations discussed here also do not include many factors that could add a further degree of realism. For example, the simulation does not include multi-sited households, non-farm employment, urban-rural migration, agricultural diseases, the use of sharecroppers, a land market, or government subsidies, However, by demonstrating that the behaviour of a simple household lifecycle model can be altered by introducing additional factors such as access to outside labour and lot effects, we can see that multiple coupled human and natural drivers of land use change are likely operating in these regions – drivers that will require continued investigation.

References

Brondízio, E. S., McCracken, S. D., Moran, E. F., Siqueira, A. D., Nelson, D. R., & Rodriguez-Pedraza, C. (2002). The colonist footprint: Toward a conceptual framework of deforestation trajectories among small farmers in frontier Amazonia. In C. Wood & R. Porro (Eds.), *Deforestation and land use in the Amazon*. Gainesville: University Press of Florida.

Cabrera, A. R., Deadman, P. J., Brondízio, E. S., & Pinedo-Vasquez, M. (2010). Exploring the choice of decision making method in an agent based model of land use change. In D. Swayne, W. Yang, A. A. Voinov, A. Rizzoli, & T. Filatova (Eds.), *Proceedings of the iEMSs fifth biennial meeting: International Congress on Environmental Modelling and Software (iEMSs 2010)*. Ottawa: International Environmental Modelling and Software Society.

de Sherbinin, A., VanWey, L. K., McSweeney, K., et al. (2008). Rural household micro-demographics, livelihoods and the environment. *Global Environmental Change, 18*, 38–53. doi:10.1016/j.gloenvcha.2007.05.005.

Deadman, P., Robinson, D., Moran, E., & Brondízio, E. (2004). Colonist household decision making and land use change in the Amazon Rainforest: An agent-based simulation. *Environment and Planning B, 31*(5), 693–709. doi:10.1068/b3098.

Fearnside, P. M. (1986). *Human carrying capacity of the Brazilian rainforest*. New York: Columbia University Press.

Goody, J. R. (1958). *The development cycle in domestic groups*. Cambridge: Cambridge University Press.

Grimm, V., & Railsback, S. F. (2012). Designing, formulating and communicating agent-based models. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 361–377). Dordrecht: Springer.

Grimm, V., Berger, U., Bastiansen, F., et al. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling, 198*, 115–126. doi:10.1016/j.ecolmodel.2006.04.023.

Grimm, V., Berger, U., DeAngelis, D. L., Polhill, G., Giske, J., & Railsback, S. F. (2010). The ODD protocol: A review and first update. *Ecological Modelling, 221*, 2760–2768. doi:10.1016/j.ecolmodel.2010.08.019.

Hare, M., & Deadman, P. J. (2008). Further towards a taxonomy of agent based simulation models in environmental management. In A. L. Paderes & C. H. Iglesias (Eds.), *Agent-based modelling in natural resource management*. Valladolid: INSISOC.

Lim, K. S. (2000). *Agent-based simulations of land use change in the Amazon*. Unpublished master's thesis, University of Waterloo, Waterloo.

Liu, J., Dietz, T., Carpenter, S. R., et al. (2007a). Complexity of coupled human and natural systems. *Science, 317*, 1513–1516. doi:10.1126/science.1144004.

Liu, J., Dietz, T., Carpenter, S. R., et al. (2007b). Coupled human and natural systems. *Ambio, 36*, 639–649.

Lu, D., Moran, E., & Mausel, P. (2002). Linking Amazonian secondary succession forest growth to soil properties. *Land Degradation and Development, 13*, 331–343. doi:10.1002/ldr.516.

Mausel, P., Wu, Y., Li, Y., Moran, E. F., & Brondízio, E. S. (1993). Spectral identification of successional stages following deforestation in the Amazon. *Geocarto International, 4*, 61–71.

McCracken, S. D., Brondízio, E. S., Nelson, D., Moran, E. F., Siqueira, A. D., & Rodriguez-Pedraza, C. (1999). Remote sensing and GIS at farm property level: Demography and deforestation in the Brazilian Amazon. *Photogrammetric Engineering and Remote Sensing, 65*, 1311–1320.

McCracken, S. D., Siqueira, A. D., Moran, E. F., & Brondízio, E. S. (2002). Land use patterns on an agricultural frontier in Brazil; insights and examples from a demographic perspective. In C. H. Wood & R. Porro (Eds.), *Land use and deforestation in the Amazon*. Gainesville: University Press Florida.

Messina, J. P., Evans, T. P., Manson, S. M., Shortridge, A. M., Deadman, P. J., & Verburg, P. H. (2008). Complex systems models and the management of error and uncertainty. *Journal of Land Use Science, 3*, 11–25. doi:10.1080/17474230802047989.

Moran, E. F. (1981). *Developing the Amazon*. Bloomington: Indiana University Press.

Moran, E. F. (1995). Socio-economic aspects of acid soil management. In R. A. Date (Ed.), *Plant soil interactions at low pH*. Dordrecht: Kluwer Academic.

Moran, E. F., Brondízio, E., Mausel, P., & Wu, Y. (1994). Integrating Amazonian vegetation, land use, and satellite data. *BioScience, 44*, 329–338. doi:10.1016/j.foreco.2005.04.004.

Moran, E. F., Brondízio, E. S., Tucker, J. M., da Silva-Forsberg, M. C., McCracken, S., & Falesi, I. (2000). Effects of soil fertility and land use on forest succession in Amazonia. *Forest Ecology and Management, 139*, 93–108. doi:10.1016/S0378-1127(99)00337-0.

Moran, E. F., Brondízio, E. S., & McCracken, S. (2002). Trajectories of land use: Soils, succession, and crop choice. In C. H. Wood & R. Porro (Eds.), *Land use and deforestation in the Amazon*. Gainesville: University Press Florida.

Parker, D. C., Berger, T., Manson, S. M. (2002). *Agent-based models of land-use/land-cover change: Report and review of an international workshop*. Bloomington: LUCC Focus 1 Publication 6. http://www.indiana.edu/~act/focus1/FinalABM11.7.02.pdf

Parker, D. C., Manson, S. M., Janssen, M. A., Hoffmann, M. J., & Deadman, P. (2003). Multi-agent systems for the simulation of land-use and land-cover change: A review. *Annals of the Association of American Geographers, 93*(2), 314–337.

Parker, D. C., Brown, D. G., Polhill, J. G., Deadman, P. J., & Manson, S. M. (2008a). Illustrating a new 'conceptual design pattern' for agent-based models of land use via five case studies – the MR POTATOHEAD framework. In A. L. Paderes & C. H. Iglesias (Eds.), *Agent-based modelling in natural resource management*. Valladolid: INSISOC.

Parker, D. C., Entwisle, B., Rindfuss, R., et al. (2008b). Case studies, cross-site comparisons, and the challenge of generalization: Comparing agent-based models of land use change in frontier regions. *Journal of Land Use Science, 3*, 41–72. doi:10.1080/17474230802048151.

Rindfuss, R. R., Entwisle, B., Walsh, S. J., et al. (2008). Land use change: Complexity and comparisons. *Journal of Land Use Science, 3*, 1–10. doi:10.1080/17474230802047955.

Robinson, D. T. (2003). *Modelling farmer household decision-making and its effects on land use/cover change in the Altamira region, Para, Brazil*. Unpublished master's thesis, University of Waterloo, Waterloo.

VanWey, L. K., D'Antona, Á. O., & Brondízio, E. S. (2007). Household demographic change and land use/land cover change in the Brazilian Amazon. *Population and Environment, 28*, 163–185. doi:10.1007/s11111-007-0040-y.

Chapter 34
Beyond Zipf: An Agent-Based Understanding of City Size Distributions

Timothy R. Gulden and Ross A. Hammond

Abstract George Kingsley Zipf observed in 1949 that the size distribution of cities within nations tends to follow a particular kind of power-law. This distribution is often described as the "rank size rule" or simply as the Zipf distribution. While Zipf convincingly documented this distribution, he was less successful in explaining its emergence. During the ensuing half century, various theories of city formation and development have emerged that have contributed real insights into the geography and economics of cities. For the most part, however, these theories have failed to predict the Zipf distribution of sizes. Another class of theories has been put forward to explain the distribution, but these have tended to rest on unrealistic assumptions, to lack explanatory power, or, at best, to lack the ability to explain the deviations from Zipf that can be observed in many nations. In this paper, we offer a simple agent-based model of city size evolution. This model offers substantial insight into the distribution of city sizes in various countries, while complementing previous work on the economic geography of cities and offering plausible economic interpretations and logic. The model can also account for several important categories of systematic deviation from Zipf that are observed in empirical data, and offers new insights about how such deviations arise.

34.1 Introduction

George Kingsley Zipf observed in 1949 that the size distribution of cities within nations tends to follow a particular kind of power-law (Zipf 1949). This distribution is often described as the "rank size rule" or simply as the Zipf distribution. While Zipf

T.R. Gulden (✉)
Krasnow Institute for Advanced Study, George Mason University, Fairfax, VA, USA
e-mail: tgulden@gmu.edu

R.A. Hammond
Center on Social Dynamics and Policy, The Brookings Institution, Washington, DC, USA
e-mail: rhammond@brookings.edu

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_34,

convincingly documented this rule in cities and many other systems (including the frequency of word usage in most languages), he was less successful in explaining its emergence. During the ensuing half century, various theories of city formation and development have emerged and have contributed real insights into the geography and economics of cities. They have, for the most part, however, failed to predict the Zipf distribution of sizes. Another class of theories has been put forward to explain the distribution, but these have tended to rest on unrealistic assumptions, to lack explanatory power, or, at best, to lack the ability to explain the deviations from Zipf that can be observed in many nations. In this paper, we offer a simple, agent-based model (ABM) of city size evolution. This model offers substantial insight into the distribution of city sizes in various countries while complementing previous work in the economic geography of cities, and offers plausible economic interpretations and logic. The model can also account for several important categories of systematic deviation from Zipf that are observed in empirical data, and offers new insights about how such deviations arise.

In essence, we find that the distribution and its variants arise naturally when people try to optimize their wellbeing by migrating between a given set of cities. We use an agent approach because the dynamics of this process depend on the citizens having imperfect information – they are more likely to move from a congested city to an uncongested one, but may also misread the situation and move to a city that is already overfull. This simulation approach also allows us to model city sizes in dynamic terms, with urban equilibrium sizes responding to population shocks with a lagged adjustment mechanism corresponding to the adaptive provision and decay of infrastructure. We represent these mechanisms in highly simplified, abstract terms beginning with a model that is so abstract that it has little to do with human behavior. We proceed, however, to develop a model that, while still abstract, does present a plausible version of these basic economic phenomena.

34.1.1 The Zipf Distribution

The Zipf distribution is neatly summarized by the expression $S_r = S_0 * r^{-1}$ where S_r is the size of city r, r is the rank of the city (i.e. for the tenth largest city, r = 10) and S_0 is the size of the largest city. This can be restated as the so called "rank size rule" by observing that the second largest city is half the size of the largest city, the third largest 1/3 as large, the fourth 1/4 as large, etc. One property of this distribution is that when it is plotted as an ordered histogram on log-log axes, it results in a straight line with a slope of −1 (the exponent of the power-law) as shown in Fig. 34.1.

34.1.2 Explanations for the Distribution

While the Zipf regularity has been well known for some time, it has resisted attempts at theoretical explanation. Fujita et al. (1999) directly address this fit between theory and observation in their chapter entitled "An Empirical Digression: The Sizes of Cities". They write:

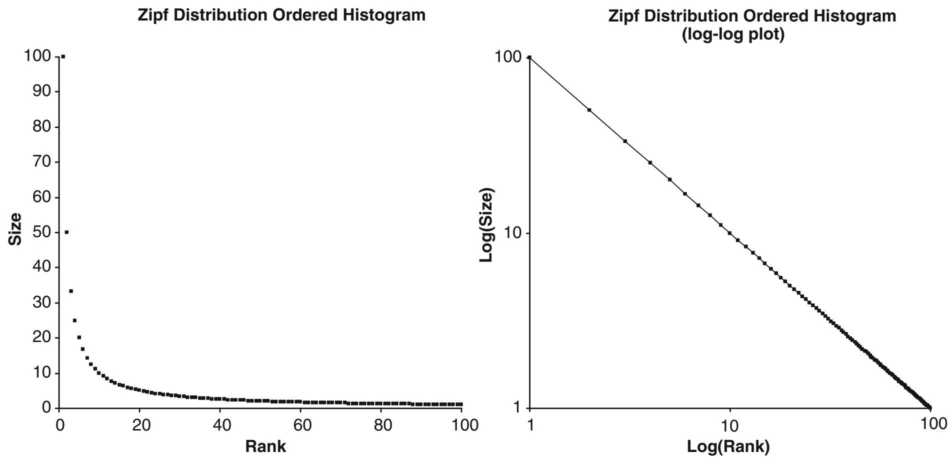


Fig. 34.1 Zipf Distribution ordered histogram on normal and log-log axes

> Attempts to match economic theory with data usually face the problem that the theory is excessively neat, that theory gives simple, sharp-edged predictions, whereas the real world throws up complicated and messy outcomes. When it comes to the size distribution of cities, however, the problem we face is that the data offer a stunningly neat picture, one that is hard to reproduce in any plausible (or even implausible) theoretical model. (p. 215).

The conclusion to this chapter begins: "At this point we have no resolution to the explanation of the striking regularity in city size distributions. We must acknowledge that it poses a real intellectual challenge to our understanding of cities…" (p. 225). Although work in this area has continued in the intervening years, there remain few behaviorally-based candidates to explain the Zipf regularity, and no consensus on how these explanations relate to one another.

Attempts to model the dynamics of city size have largely fallen into one of two categories.[1] Models in the first category extend concepts from standard economic theory to apply to city size dynamics. These include externality models, which apply the "Henry George" theorem from urban economics (Marshall 1890; Jacobs 1984; Henderson 1974; Kanemoto 1980), and models that extend Christaller's (1933) "central place" theory; see Fujita and Mori (1997). Such models are well integrated with the existing body of economic theory, and are often consistent with other economic evidence about city dynamics. Unfortunately, none of these models convincingly produce the empirical regularity of the Zipf distribution.

Models in the second category apply one or more abstract stochastic processes to represent city size dynamics. Early examples included Simon's (1957) proportional growth model and Hill and Woodrofe's (1975) application of the Bose-Einstein process. More recently, the most prominent models in this category have focused on descriptions of city growth as a "Gibrat process" (Gibrat 1931). Papers applying the Gibrat processes include Gabaix (1999a) and Reed (2002). These processes have all

[1] For a more comprehensive review of the literature on urban size distributions, see Gabaix and Ioannides (2003).

been shown mathematically to successfully generate a stable power-law distribution, and in many cases, to closely replicate the Zipf distribution itself. However, such models have little or no economic content. They demonstrate that the Zipf regularity can be generated using a variety of statistical mechanisms, but they do not offer a set of comparable behavioral principles or realistic economic mechanisms that are sufficient to produce Zipf. As one recent paper put it: "this collection of models is essentially statistical – they seek to *generate rather than to explain* the regularity" (Overman and Ioannides 2001). It is often unclear how the abstract mechanisms represented in many of these models can be useful metaphors for real-world social or economic processes. Indeed, in some cases, closer examination has found strong empirical evidence that mechanisms such as the Gibrat process are *not* good descriptions of real city-size dynamics; see Cuberes (2004). Abstract stochastic models have also tended to be "brittle" – they can generate the Zipf distribution, but they are "one-process-fits-all" and cannot generally account for the exceptions to or variations in Zipf that are observed in the data.

Duranton (2002) presents a model based on Grossman and Helpman's (1991) quality-ladder model of growth that produces both Zipf and certain observed variations from Zipf. This model is similar to the model presented here in that it treats urban population as a largely conserved quantity that is redistributed among interconnected cities. In this respect, these models differ sharply from other models that produce good fits, e.g. Gabaix (1999a). Most notably, this property allows these model to produce Zipf-like distributions without assuming that shocks are uncorrelated with city size (Gibrat's law). The model presented here differs from Duranton's in that it is considerably more general, while still having a strong behavioral basis. Indeed, the goal of this paper is to establish a general behavioral framework within which successful economic city size models can be built.

34.1.3 Deviations from Zipf

While the Zipf distribution offers a remarkably good fit for many nations, the fit is imperfect in many cases. In this paper, we will examine three countries that are particularly interesting with regard to their adherence to and deviations from Zipf. These three countries are: the United States, Russia, and France. All three countries provide excellent data on urban agglomerations. The United States represents a relatively good (though significantly imperfect) fit for Zipf, while France and Russia deviate in different ways that may offer lessons applicable to broad classes of countries.

Before attempting to analyze the extent to which cities in different countries do or do not deviate from Zipf, we need to address the definition of a city. In this paper, we are interested in the city as a social and economic phenomenon, rather than as a legal entity. Our unit of analysis is thus not the population within the official city limits, but rather the population of the urban agglomeration of which the legally incorporated city is often only a part.

Consistently defining an urban agglomeration is challenging (Le Gleau et al. 1996), but in the cases we have chosen, it is possible to derive reasonably satisfying definitions of urban agglomerations. The statistical agencies of both the United States and France have addressed this problem directly by developing various functional definitions of urban agglomerations, while Soviet central planning produced Russian cities that are clearly separated, compact and well defined. We will discuss the specifics of each of these cases in turn.

34.1.3.1 USA

The cities of the United States have generally been regarded as being very nearly Zipf distributed. Because of the sprawling nature of many US cities, and the high daily mobility of the US population, the definition of an urban agglomeration for the US has proven particularly difficult. Over the past several decades, the set of Metropolitan Statistical Areas (MSAs) developed by the US Office of Management and Budget (OMB) and the Census Bureau were the standard measure of urban agglomerations. The MSA definitions had significant limitations, however, and were often influenced by local electoral politics.

In 2003 the OMB released a new series of consistently and objectively defined agglomeration data that it terms Core Based Statistical Areas (CBSAs) or "Metropolitan and Micropolitan" areas (Federal Register 2000). This definition attempts to capture spatial and economic integration with a rigor that had not previously been attempted. The result is a consistently defined set of 922 cities. These cities follow the Zipf distribution fairly closely over a tremendous range: from greater New York City with 18.3 million people down to about the 800th city (Jennings, Louisiana) with a population of about 30,000. Although several of the largest cities are significantly smaller than Zipf would predict, the distribution generally fits a power-law exponent very close to −1.

For convenience in the analysis that follows, we will restrict our data to a subset of the 250 cities with populations over 150,000 (Fig. 34.2). This reduced set of cities looks very much like the full set, displaying a power-law exponent of 1.005.

Comparing the US city size distribution to a pure Zipf distribution for a comparable number of cities and citizens, we find the Zipf assumption misplaces about 15% of the population overall, with an error of 9.7% at the median city.[2]

[2] We can produce an objective measure of how well a "constructed" Zipf distribution fits the observed data by dividing the number of people which the Zipf rule misplaces relative to the data (the cumulative error) by the total population of the cities. The cumulative error is calculated as the sum of the absolute values of the errors for each city divided by two (because each citizen which is in the wrong place is also missing from the right place). We will refer to this measure as the total error.

While the overall error is well reflected by this measure, it does not give a sense of how the error is distributed. A sense of this distribution is given by the error at the median city. This is to say that we measure the error for each individual city ((abs($Data_i - Model_i$)/2)/$Data_i$) and report the median of these values. This indicates whether the error is concentrated in a few large cities which fit poorly or is distributed throughout the range of the cities. We will refer to this measure as the median error.

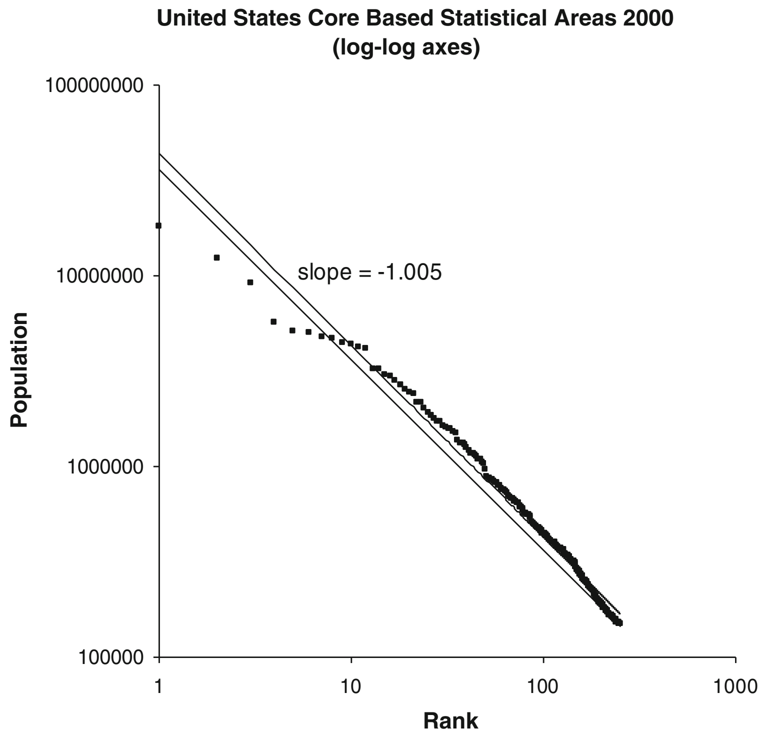


Fig. 34.2 United States core based statistical areas, 2000

34.1.3.2 France

The French National Institute of Statistics and Economic Studies (INSEE) produces a variety of excellent data on French cities using various definitions. These include the municipality (*commune*); the urban pole (*pôle urbain* or *unité urbain*); and the urban area (*Aire urbain*).

One of the most salient features of the French city size distribution is the dominance of the Paris metropolitan area. Of the three ways of defining a city offered by the French statistical agency (INSEE), the "urban pole" definition is the most appropriate for our analysis, but under-represents the size of the Paris metro area. We will use a modified definition of "urban pole" which, following Le Gleau et al. (1996), we will call an "urban center". This revised definition better captures the dominance of Paris in the French urban system.

The urban center data conforms fairly closely to Zipf, displaying an overall power-law exponent of –0.98 (see Fig. 34.3). The primary deviations from Zipf are that Paris is about two and half times the size that the rest of the distribution would predict while the second agglomeration, Marseille-Aix-en-Provence, is about two thirds the size that the distribution would predict. The combination of these two factors makes Paris about seven times as large as France's second city – whereas in a

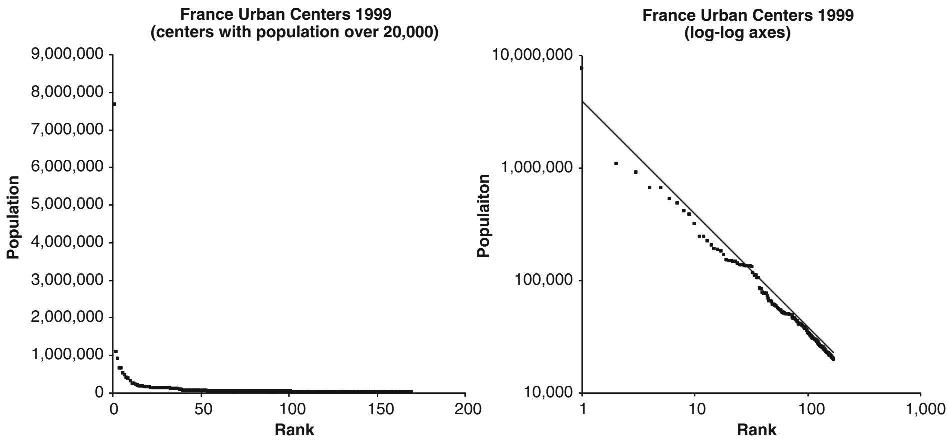


Fig. 34.3 France urban centers, 1999

strict Zipf distribution it would be twice as large. Overall, the Zipf distribution displaces 17% of the French population, but this is largely due to the very poor fit of Paris – the error at the median city is only 7%.

Because France is much less populous than the United States, its urban structure is also much smaller. Whereas the United States has about 900 cities with populations greater than 20,000, France (following the 1999 urban center definition) has only 170 cities above this size.

34.1.3.3 Russia

Unlike the United States and France, which both adhere closely to the Zipf regularity for all but their largest cities, the Russian city size distribution displays a distinct curvature on log-log axes over the entire range of its urban structure (see Fig. 34.4).

The substantially different Russian urban structure is not surprising given the radically different physical, social, and economic environment in which it developed. Much of Russia's urbanization took place during the Soviet period when internal migration was intensely managed by the central government, which pursued objectives such as the extraction of natural resources, the occupation of territory that might be claimed by China, and the movement of industrial production away from the potential front with Western Europe. Policies of forced and incentivized migration, costly investments in infrastructure, and intensive subsidies to far-flung cities in inhospitable locations increased both the number and the size of cities in remote parts of the Soviet Union (Hill and Gaddy 2003). A basic reality of this system, which we will make use of in our model, was that moving down the urban hierarchy was generally easier than moving up it. A person living in Moscow might be assigned a job in a minor industrial center in Siberia, but a person living in that Siberian city would be unlikely to be assigned to Moscow. This system insured that the smaller (and often colder and

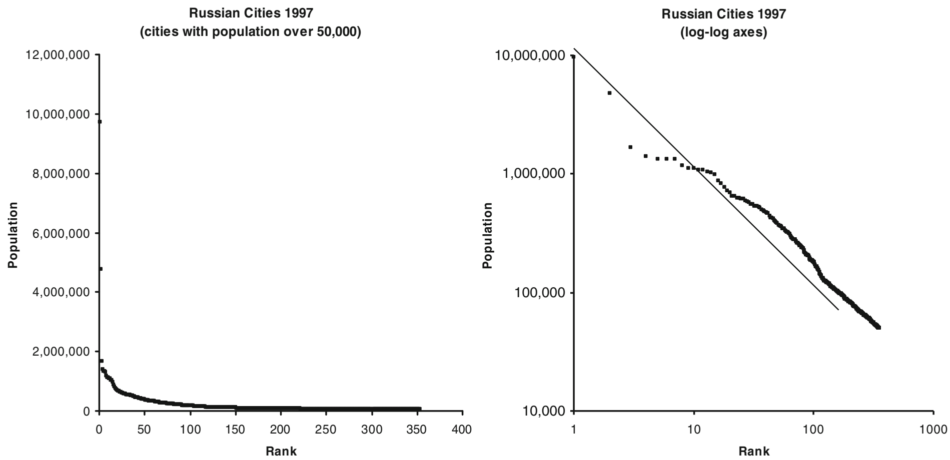


Fig. 34.4 Distribution of Russian city sizes

generally less hospitable) industrial cities of Siberia remained populated in spite of Russian citizens' inclinations to move elsewhere (Hill and Gaddy 2003; Iyer 2003).

Russian urban agglomerations are easier to define than their US and French counterparts because of the way that Soviet planners designed the Russian urban structure (Hill and Gaddy 2003). The desire to spread population over the vast territory of the Russian empire created large distances between cities, while the planned nature of these cities reduced or eliminated urban sprawl in most cases. Because Russian cities tend to be distinct and compact, Russian city population numbers and urban agglomeration numbers tend to coincide, requiring the aggregation of suburbs with central cities only for Moscow and St. Petersburg. The data generated by the Russian census are therefore appropriate for our purpose without adjustment beyond the agglomeration of these suburbs.

The overall best fit power-law for this data has an exponent of −0.92 – a number close enough to unity that some authors have failed to remark on it. Our measure of total error indicates that the fit between the Russian distribution and the Zipf distribution is similar to that for the US and France, misplacing 16% of the population (as compared to 15% and 17% respectively), but this apparent similarity is misleading. This shows up in a median error figure of 17% (as compared to about 10% for the US and 7% for France). While distributions for the US and France are generally Zipf like, with departures in the largest cities, the Russian distribution is distinctly curved as shown in Fig. 34.5.

We can demonstrate this curvature by dividing the Russian city distribution into two parts and examining the exponents of the best-fit power-law that describes each part, measuring the power law exponent for cities larger than 500,000 separately from those between 500,000 and 100,000. These sets of cities display two distinct exponents. The upper part of the curve has a slope of −0.68 while the lower part has a slope of −1.19. These slopes are significantly different with $p \ll 0.001$. Similar tests on data from the US and France yield slopes that are not significantly different.

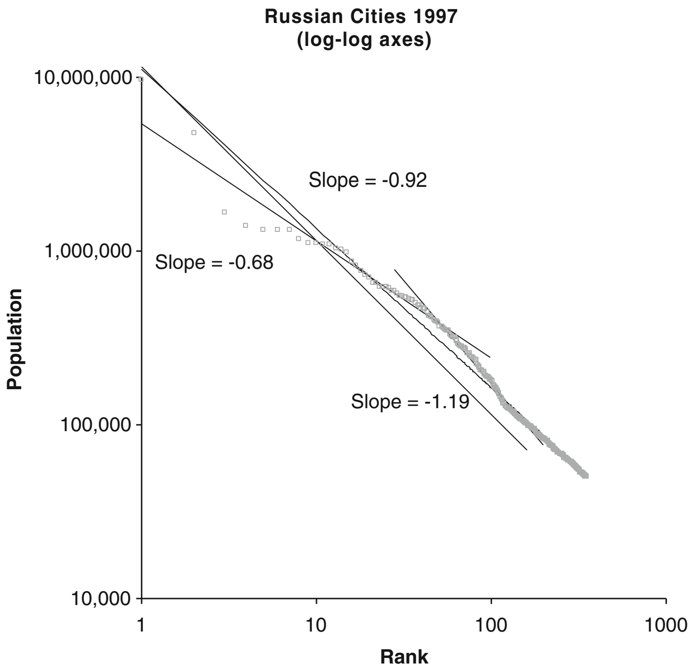


Fig. 34.5 Curvature of the Russian city size distribution

By inspecting the graph (Fig. 34.5) we can see that our cut off of 500,000 between the two groups is an arbitrary one, and that the distribution of cities larger than 100,000 is better described by a curve that is concave toward the origin. In this sense, the Russian distribution departs from the Zipf distribution for all of the 161 cities in this range.

34.2 A Simple, Abstract Model: Jars and Beans

34.2.1 Model Description

In the sections that follow, we will attempt to explain both the tendency of urban systems to approximate Zipf, and the reasons why various countries depart from it, by constructing a model that is as simple as possible while capturing the essential features of the systems in question.

We begin with an abstract model that produces remarkably good agreement with real city size distributions. This model is designed to explore the way in which power-law distributions can emerge from systems involving stochastic exchange. Because the abstract model does not itself contain sufficient detail to capture plausible urban dynamics, we describe it in terms of "jars" (rather than cities) exchanging "beans" (rather than citizens). In the next section, we will extend the model in such a way that it demonstrates a plausible relationship to social and economic realities.

The rules of the abstract model are simple. The model begins with some number of jars, each of which contains some number of beans. The jars interact in random pairings. In each interaction, the jars exchange some number of beans ("the bet") equal to half of the beans in the smaller jar. In the base case, both jars have an equal probability of winning the bet. Once the winner is determined, the beans are exchanged and a new random pairing of two different jars is made. There is a floor size of 1 bean. If a jar of size 1 loses a bet, nothing happens and it remains at size 1. If it wins a bet, it wins a whole bean (rather than half a bean).

One important feature of this model is that it assumes that urban population is conserved, unlike other models (e.g. Gabaix 1999a; Fujita et al. 1999) that assume people freely enter and leave the urban system. Our assumption of conservation fits with empirical evidence that once people have migrated to a city and have traded their rural skills for urban ones, they tend to remain in the urban system – migrating from one city to another in search of opportunities, but seldom returning to live in the hinterlands. In our simple abstract model, this is reflected in a strict conservation law: beans are neither created nor destroyed, they simply move from jar to jar.

The model also differs from other stochastic models (typified by Gabaix (1999a)) by not needing to assume that the growth rates of cities are independent. These previous models generally depend on a Gibrat process for their results, in which cities grow (or shrink) by random amounts that are uncorrelated but share a common mean. In our model, growth rates *are* correlated (one city's gain is another city's loss). We believe that this is a more plausible assumption for modeling city size dynamics. We also assume that growth rates depend on city size. When a small city faces a larger city, it faces a gain or loss of half its size, whereas the larger city faces a gain or loss that comprises a smaller fraction of its population. Small cities, therefore, face greater size volatility than large ones, a fact that also coincides with real world observation (Gabaix 1999b).

34.2.2 Results from the Abstract Model

Although the model is very simple, it can produce statistically robust Zipf distributions as well as some interesting variations on the distribution. If the model is run with the appropriate number of beans[3] for the given number of jars, it will approach the Zipf distribution regardless of the initial distribution of the beans between jars. Initializing the model with "too many" beans – more beans than would be required

[3] From the definition of the distribution, it follows that a certain number of jars requires a certain number of beans to fill the distribution. When the floor size (the size of the smallest jar) is one bean, the largest jar should contain a number of beans equal to the number of jars. The sizes of all the jars between the largest and the smallest are then given by the rank/size rule, rounding to the nearest whole bean. For example, for 100 jars, 516 beans are required to fill the distribution.

to fill a Zipf distribution for the given number of jars – produces instability in the top of the distribution with large fluctuations in the sizes of the largest jars, with the excess beans tending to float among the top few jars. Radical overfilling of the distribution tends to produce "jamming" at the top, where the largest jar ends up with the majority of the excess beans. Initializing the model instead with "too few" beans produces a curvature of the distribution, maintaining the power-law exponent in the lower tail and progressively lowering it in the upper tail. As we will see below, these two related results have important parallels to the real world deviations from Zipf's law observed in the cases of France and Russia.

The model is reasonably robust to changes in key parameters. For example, while it is important that the "bet" be related to the size of the smaller jar in a given interaction, the exact proportion used generally affects only the speed with which the system approaches equilibrium, not the nature of that equilibrium.[4]

We can make a first analogy from this abstract model to urban dynamics by thinking of the jars as cities and the beans as groups of citizens. Each bean represents the number of citizens in the smallest city (size = 1) in the sample. Actual population data can therefore by translated for use in the jars and beans model by dividing the total population of the urban system by the size of the smallest city in the system. This translation means that the units of exchange in the model are the size of the smallest city. This coarse assumption leads to discontinuities in the lower tail of our graphs, but it produces some interesting initial results and we will subsequently refine them.

Population figures for the United States, inserted into this simple model, produce a distribution that bears a noticeable resemblance to real data. In the year 2000, according to the Census Bureau data discussed above, the US had 250 cities with a population larger than 150,000 and these cities were home to a total of 220,227,293 people. We translate this for use in the jars and beans model by dividing the total population by the size of the smallest city (150,000), giving 1,468 beans in total. We can then get a first approximation of the US urban distribution by initializing the model with 250 jars and 1,468 beans (initially distributed randomly). Running the model with these parameters gives a fit that is quite suggestive.

Figure 34.6 shows the discretized version of the US data compared to output from 100 runs of the simple model using 250 jars and 1,468 beans. The heavier, central line on the graph indicates the median size for the city of each rank across all model runs; the lighter lines represent a 90% confidence interval around this median. The US data do not fit precisely within this envelope, but it is not far off. The gray circles in the figure represent one of the hundred sample runs that is particularly suggestive. We will return for a more careful analysis with a more complex model in the next section of the paper.

[4] Extremely small bet sizes can begin to cause the lower tail of the distribution to collapse. This does not occur with a bet sizes close to 50% of the smaller jar (the setting used throughout this section of the paper). We will discuss the sensitivity of the model to bet sizes in more detail in the next section of the paper.

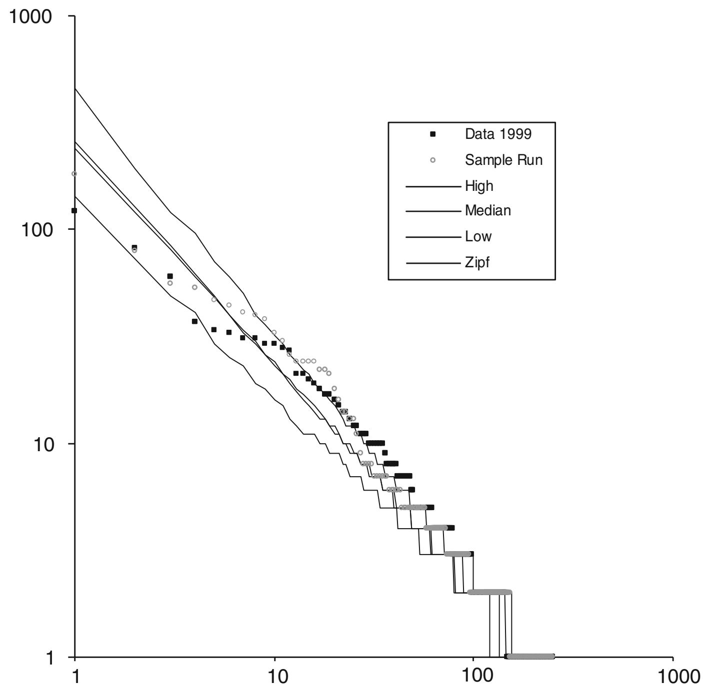


Fig. 34.6 Simple model output compared with discretized US data

Conducting the same exercise for France produces similarly provocative (although again not entirely realistic) results. Using our definition of an urban center, France has 170 cities with populations larger than 20,000 that collectively contain 22,386,598 people. We thus initialize the model with 170 jars and 1,119 beans (again distributed randomly).

Again, we see that the real data generally fit within the range of model results (Fig. 34.7). We can see from the representative sample run (grey dots) that in a case where the first two cities are of the proper size, the fit of the rest of the distribution is also very close. Although the simple model does not fully predict the primacy of Paris in the French urban system, the median model run does reflect an increase in slope in the top three or four positions. This is consistent with the notion that a small urban system with a relatively large population will tend to see disproportionately large cities at the top of its range.

Finally, we can obtain intriguing results for Russia by applying the model with a slight variation. In 1997 Russia had 161 cities with populations over 100,000 that collectively contained 70,282,100 people. This yields 703 beans in 161 jars.

Initializing the model with these values gives us a distribution that is concave toward the origin on log-log axes, but which has a somewhat different shape than we see in the data from Russia. If, however, we approximate Soviet era restrictions on internal migration by introducing a slight bias into the process (simulating the asymmetry in difficulty between moving up and moving down the urban

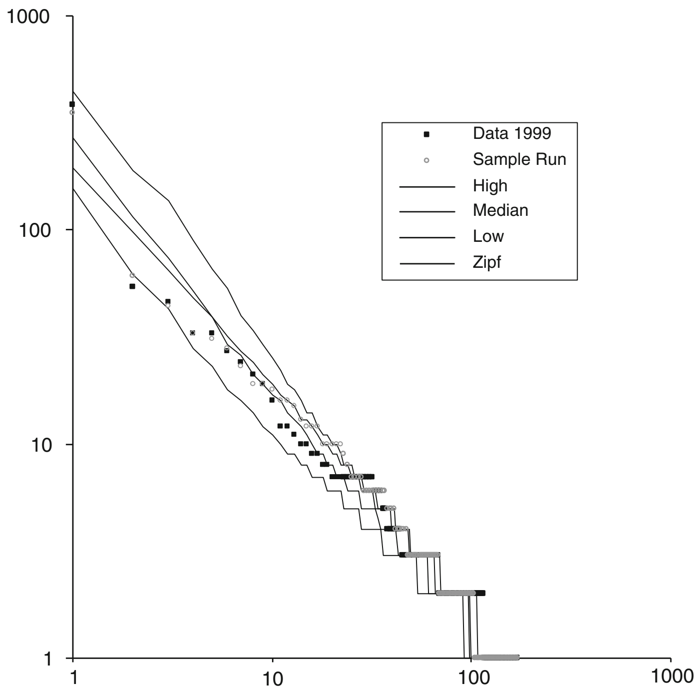


Fig. 34.7 Simple model output compared with discretized France data

hierarchy by giving the smaller city in each pairwise interaction a small advantage) the shape of the distribution comes to match the Russian case much more closely, as shown in Fig. 34.8.

34.2.3 Limitations of the Abstract Model

While the abstract model offers a simple mechanism that is capable of generating distributions resembling real city size distributions, it suffers from several serious limitations in interpretation. Although this model incorporates more realistic assumptions (such as correlated growth rates) than other stochastic models have employed, the dynamics of the model still bear little resemblance to those of real cities: cities do not engage in "tournaments" of flipping coins for half of their citizens. In addition, the floor assumption of the abstract model provides a subsidy to the smallest jars – in each interaction they stand to either remain unchanged or to double their number of beans. This mechanism tends to move beans from the upper parts of the distribution into the lower tail in a way that has no clear analog in the dynamics of urban migration.

Also, the dynamics of the simple model involve a high churn rate, with cities changing rapidly changing their rank within the distribution and the largest individual cities varying tremendously in size over time. In the time scale that is required

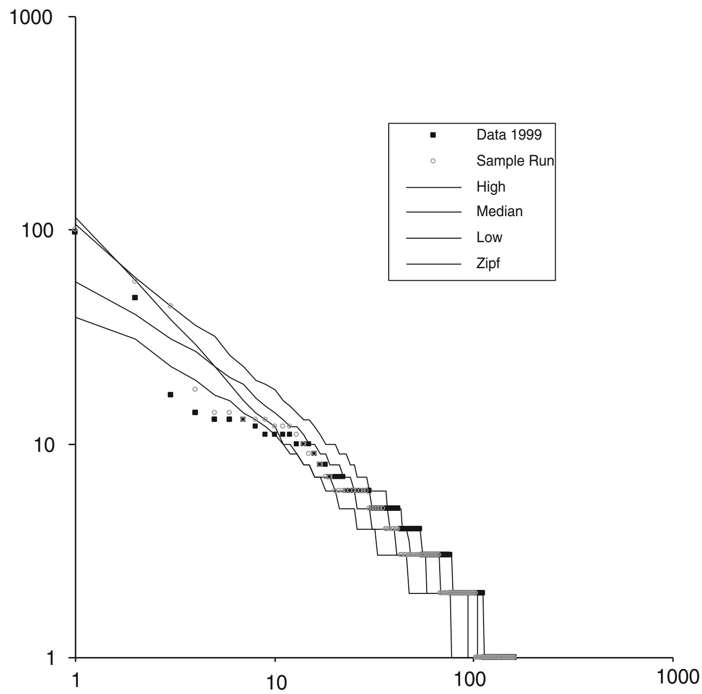


Fig. 34.8 Simple model output compared with discretized Russia data

to achieve the power-law distribution, Chicago might change places with Peoria several times. This unrealistic dynamic highlights the fact that the abstract model has no place in it for differences in site suitability. Some sites (natural ports, for example) are inherently better than others for large cities, and any plausible model of urban dynamics should be able to reflect this fact.

34.3 A Richer Model: Cities and Citizens

34.3.1 Model Overview

To address these deficiencies, we will now introduce a richer model that comes closer to representing real urban dynamics. This model preserves and improves upon many of the desirable qualities of the abstract model while remedying some of its shortcomings. The richer model relies on the notion that a city has a short-term equilibrium size that balances economies of agglomeration (reasons to move into the city) with diseconomies of congestion (reasons to move out). A city can be thought of as being oversized if it moves above this equilibrium value and undersized if it moves below it. This short-term equilibrium is subject to shocks that result from the bounded rationality of citizens. The equilibrium reacts to these shocks over the longer term according to a lagged adjustment mechanism. Finally, the model introduces a conception of "core size" – a size below which it is not economically feasible for a given city to shrink.

34.3.2 Bounded Rationality

The concept of bounded rationality underlies the exchange mechanism in the abstract model and provides us with guidance in refining it in terms of both the size of and bias in exchanges. We can see the central role of imperfect information in the model by assuming (temporarily) that all cities are at their equilibrium sizes. In this case, with each city at its optimal size, perfectly informed and rational agents would have no incentive to move from one city to another because any move would leave their home city underfilled and their new city overfilled – making the mover worse off. Under the assumption of perfect information, any distribution of city sizes in which all cities are in local equilibrium would be stable indefinitely.

The citizens in our model, however, have imperfect information and bounded rationality. Some citizens, therefore, will move from city to city *even at* an "equilibrium" distribution of sizes. People are more likely to move from a more crowded city to a less crowded city, but the reverse is also possible. The size of the exchange between cities, therefore, is a parameter of the model. It represents the degree to which the rationality of the citizens is bounded – the percentage of the citizenry that will move between two equally attractive cities because of imperfect information (which we are modeling only in the abstract). With full information and no bounds on rationality, the exchange between two cities at their equilibrium sizes would always be zero. In the extended model we present below, the expected value of the exchange is zero, but the actual exchange amount varies symmetrically around zero. In this sense, the exchange mechanism is "unbiased".[5] This principle of unbiased exchange differs from the abstract (jars and beans) model discussed above. In that model, the floor mechanism provides a significant subsidy to small jars. With 100 Zipf distributed jars, an exchange size of 50% of the smaller jar, and a floor of one, about 1/3 of the jars face positive expected returns – and the rest face negative expected returns.

As with the abstract model, the primary effect of changing the size of the bet in this richer model is simply to change the speed with which the system moves. However, when the bet is small enough, very few small cities face positive expected returns. Over the long run, this leads the lower tail of the distribution generated by the model to sag (i.e. to bend toward the origin) and produces long oscillations in the extent of this sagging. These features are not observed in real data. A closer

[5] When the exchange amount is decreased to 1% of the smaller jar (as it is in the runs of the model that follows), only the single smallest jar can be expected to be within 1% of its floor, and the bias that it introduces into the system is vanishingly small.

match to empirical reality can be achieved by introducing a small amount of growth into the system. When all cities grow by a tiny amount each round, the lower tail restabilizes near a slope of −1. The amount of growth does not need to be carefully tuned to achieve this result. The growth rate needs only to be large enough to keep the tail from sagging, and small enough that the system can "digest" the new citizens. Within that range, the growth rate can vary by an order of magnitude without significant impact on model output.

This assumption of growth is consistent with the real world, where all of the world's major urban systems are still growing. This is most apparent in developing countries, which are experiencing both population growth and urbanization. It is also true, however, of OECD countries such as France, where urbanization continues even as population has stabilized (Julien 2001a).

34.3.3 Lagged Adjustment

Cities' equilibrium sizes adapt to the shocks imposed by the bounded rationality of their citizens through a lagged adjustment mechanism. If a city grows above its equilibrium size, it will become congested in the short run. If it remains congested for long enough, however, the city will adapt. Firms will move in to hire idle workers. New housing, roads and facilities will be built. Once these things happen, the city can comfortably accommodate more people than it did before – its equilibrium size has increased. Similarly, if citizens move out and stay out for long enough, firms will leave and infrastructure will deteriorate, leaving the city able to comfortably accommodate fewer people than it once could.

Adding an adjustment lag does not change the dynamics of the model, but does impact the rate at which individual cities change size over time and therefore the rate at which the distribution changes. Because the parameters of this mechanism only influence the speed with which the model changes (and we are not attempting to calibrate the model to real time), we will not dwell on the lagged adjustment mechanism here. Any mechanism that retains the unbiased quality of the exchange system from the simple model, and that does not introduce excessive noise into the model will produce similar results.

34.3.4 Inherent Suitability

A further requirement for the richer model is to account for the influence of geographic suitability and the persistence of great cities. We accomplish this by positing "core size", determined according to more conventional economic logic, which is one component of observed size.

We begin with the assumption that only some fraction of the population of a city is tied to the city's specific geographic location. Chicago, for instance, is in a unique location to serve as a port for a huge section of the American Midwest. Many of the jobs in Chicago need to be located exactly where they are geographically – at the base of Lake Michigan. Many other jobs in Chicago, however, do not have to be in that location. But they do have to be somewhere. We thus divide the population of a city into a "core" population, which is dependent on the city's geographic location (and is subject to more or less standard microeconomic rules for its size), and a floating population, which is subject to the mechanisms of the model.

A recurring problem for theorists of city sizes has been that models containing appropriate economic content (e.g. Fujita et al. 1999) tend to predict distributions that look quite different from those that are actually observed. The model presented here solves this problem and dovetails nicely with such models by freeing them from the need to predict a Zipf like distribution. A model like that of Fujita et al. (1999) is probably well suited for predicting the core sizes of cites. These core sizes should be much more readily subject to "rational" analysis. The core sizes, however, are not the only component of the observed distribution. The sizes we observe are based on the sum of the core size and the size of the floating population that can potentially live elsewhere. We will equate the "core" size of a city to its "floor" (i.e. minimum size) in the model.

Remarkably, the presence of some cities with higher floors (larger core sizes) does not change the basic dynamics of the model. It still produces Zipf distributions and the aforementioned characteristic departures from Zipf. However, the cities with higher floors tend to stay in the upper part of the distribution, thus reflecting much more realistically the persistence of major cities that we observe in the real world.

An analogy to a cake with icing is a useful way to visualize the relationship between the core and observed distributions of city sizes. The core distribution is the cake, while the floating population is the icing. All that we observe in city size data is the height of the top of the icing. While the cake of the core size distribution might be rather lumpy and vary depending on economic and geographic structure, the icing of the floating population flows smoothly over the cake and finds its level. In the case of cities, the attractor is not flat – as it is in the case of a physical cake – but rather follows the shape of the Zipf distribution and its related departures as outlined above.

Because this study is concerned with the overall shape of the various city size distributions, it is sufficient to note that adding heterogeneous core sizes does not change the distributions that emerge from the model. The simulations that follow will use uniform core sizes unless otherwise stated, with the core size being equal to the size of the smallest city in the system. The results would not be changed if a more complex or dynamic core distribution were used.

34.3.5 Results from the Richer Model

34.3.5.1 USA

This richer model produces a fit for United States core-based statistical area data that is significantly *better* than the Zipf approximation. The only significant

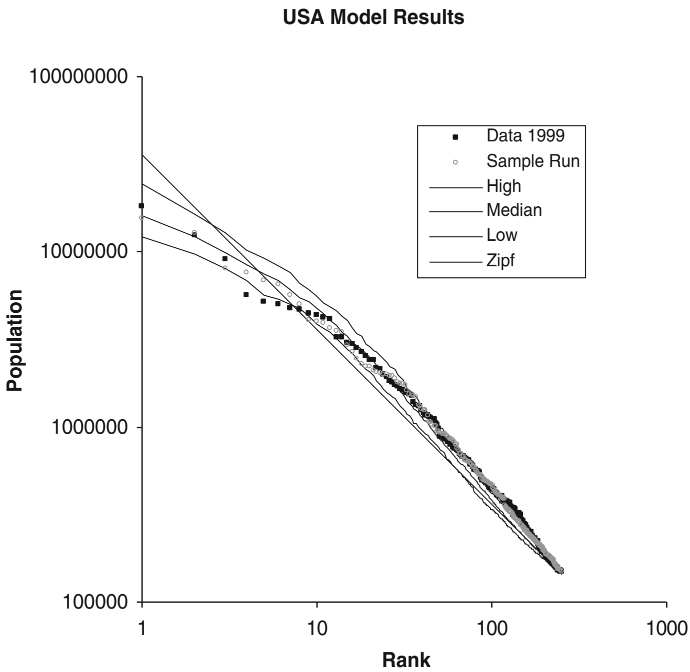


Fig. 34.9 USA model results

parameters in this model are the number of cities with populations over 150,000 (250 of them), the number of people in these cities (220,227,293 in total), the rate by which each city grows at the end of every round, and the fraction of the smaller city which will serve as the exchange amount in interactions. The first two (cities and citizens) are given by the data. The results are insensitive to the growth rate over a broad range of values (roughly and order of magnitude). The size of exchanges alters the degree of variance between runs, but does not have a noticeable impact on their median outcome. The model has no other free parameters.

We begin the simulation of the United States city size distribution with 250 cities and a reduced population of 50 million citizens (about 1/5 of the actual population) distributed evenly between the cities. The initial population size is not significant so long as it is small enough to allow the model to approach equilibrium before the full population is reached. We run the simulation forward with each city growing by a small amount (1/20,000th) at the end of each round, stopping when the population reaches the year 2000 total urban population of 220,227,293 (Fig. 34.9).

For the sake of simplicity, we begin these simulations with a uniform distribution and with a fixed number of cities, although in the real world the urban system is always in the neighborhood of the Zipf distribution (with the number of cities increasing along with their populations). Such a growth pattern is supported by history (Zipf 1949; Pumain 2004) and emerges from certain theoretical formulations (Simon 1957; Gabaix 1999a; Axtell and Florida 2001). When the initial state is close to Zipf, the growth rate becomes much less critical. It needs to be great enough

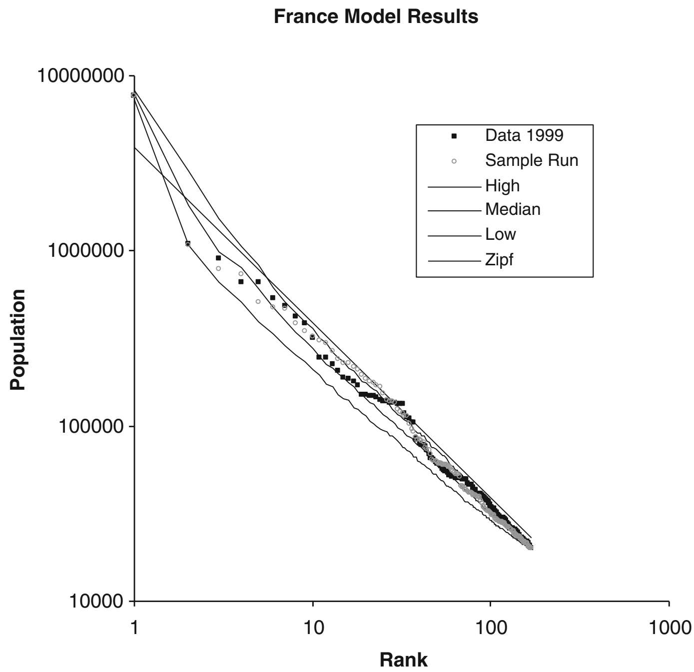


Fig. 34.10 France model results

to prevent the collapse of the lower tail, but more rapid growth is not a problem because the system does not need to produce major structural changes.

34.3.5.2 France

As discussed above, France is generally characterized by a Zipf distribution with Paris considerably larger than the distribution would predict. Although the abstract model was capable of producing results that were consistent with French data, this occurred in only a small fraction of model runs. The richer model performs considerably better in this respect, although it requires a somewhat more complex assumption about growth[6] (see Appendix). The distributions are shown in Fig. 34.10.

[6] The previously discussed issue with collapse of the lower tail in the absence of growth is particularly problematic in this case because of the very large size of Paris. Starting from a uniform distribution of city sizes, any growth rate large enough to prevent the collapse of the lower causes the population to reach its target size before the model has had time to grow Paris to its full size. We therefore begin the model with approximately 90% of the total population of France and run it forward until Paris has reached 90% of its actual population. We then introduce growth at the same rate of 1/20,000th per iteration used for the US simulation and run it until the model population is equal to the French population.

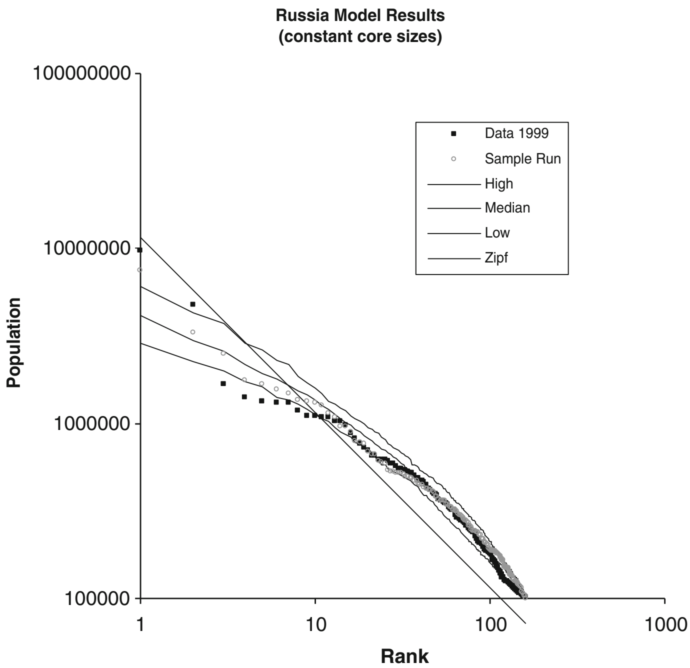


Fig. 34.11 Russia model results with constant core sizes

Whereas Zipf produced a total error of 17% and a median error of 7%, the model produces a total error of 5.3% and a median error of 3.3%.

34.3.5.3 Russia

To simulate Russia, we initialize the model with 161 cities, a population of 70,282,100 in these cities and a floor of 100,000 (the size of the 161st city). As with the simple model, we introduce a bias into the migration probability to simulate the effects of internal movement restrictions. The degree of this bias is a free parameter of the model, which we calibrate to 0.0025 in favor of the smaller city in each pair-wise interaction.[7] The slight bias toward smaller cities eliminates the tendency of the lower tail of the distribution to collapse and makes the model behavior invariant in the presence or absence of population growth (Fig. 34.11).

[7] Given that the model does not attempt to represent the urban system in actual space and time, it is not possible to calculate this movement bias using actual data. Because it is the only free parameter in the model, however, we can calibrate it by comparing model results to the observed data. We obtain a good fit by assuming a bias of 0.25% in favor of the smaller city in each pair-wise interaction. That is to say that, in each interaction, the probability of the larger city receiving the migration is 49.75% while the probability of the smaller city receiving the migration is 50.25%.

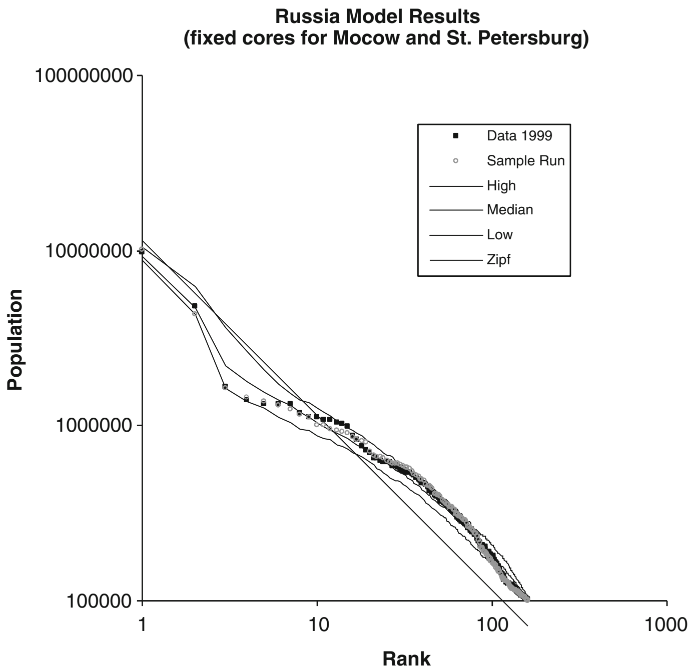


Fig. 34.12 Russia model results with larger cores for Moscow and St. Petersburg

When run with these parameters, the model captures the basic shape of the Russian city size distribution, but misses the primacy of Moscow and St. Petersburg. These cities have each played unique roles in Russia's economic and political history, serving as capitals of highly centralized political systems under both the Czars and the Soviet system. St. Petersburg is also unique in serving as European Russia's only ice-free port. The continuing pressure of internal immigration on these cities – even in the face of falling population in Russia generally (Iyer 2003), indicates that these cities remain at or below their equilibrium size in the collective mind of the Russian people. We incorporate the unique economic and geographic appeal of these two cities by assigning them core sizes that are 90% of their observed sizes, while leaving the cores of the remaining cities uniform at 100,000 people.

We observed earlier that introducing heterogeneous floor sizes alters the stability of individual cities but does not change the shape of the overall distribution *unless* floors are set so high as to make a city "protrude" from the distribution. In this case, we are conjecturing that political and geographic forces have caused the core sizes of Moscow and St. Petersburg to protrude from the Russian city size distribution.

When we incorporate these larger core sizes for Moscow and St. Petersburg into the model, it produces an excellent fit for the data (Fig. 34.12). Overall, the median model run misplaces only 3.25% of the population. This is much better than Zipf, which displaces 12.5%. The error at the median city similarly drops yet further to 2.5% as compared to 16.5% for Zipf.

34.3.6 Limitations

The richer model presented above displays a good deal of success in reproducing the distribution of city sizes in the United States, France, and Russia, but does have some limitations.

While the model can predict the overall shape of the urban distribution for various countries, it does not predict the movements of particular cities within that distribution. In order to control volatility of city size in the model, we have employed the "core size" concept – but we do not model explicitly how such core sizes evolve. The model produces similar distributions over an extremely broad range of possible core size including, we believe (but do not show here), core sizes that are compatible with observed levels of volatility.

A second, related, limitation of the model is that its current formulation does not lend itself to calibration to real time scales. Real urban systems generally expand simultaneously in both population and number of cities, whereas we hold the number of cities fixed. We believe that this assumption, although unrealistic in the long term, can yield insights in the shorter term by keeping the model simple enough for ready analysis and insight.

A third limitation is that the model uses a simple but highly unrealistic interaction network. Cities in the model interact randomly, regardless of their size or location – indeed, location is not represented in the model at all. We do not explore here the sensitivity of the model to different interaction regimes.

34.4 Discussion

34.4.1 Implications for Developing Nation Megacities

One of the more interesting and policy-relevant insights generated by the model is that the primacy of Paris (and, by extension, other disproportionately large capitals) might have more to do with the number of small cities than it does with the nature of the large city. Previous efforts to explain urban primacy (e.g. Ades and Glaeser 1995) have tended to focus on the political economy of the capital as the reason that it grows disproportionately large. These theories would attribute the massive size of Paris to the highly centralized nature of the French political system and the fact that it is "the capital of everything" including politics, finance and culture, for the nation. This contrasts with the United States where the political capital (Washington) is different from the finance capital (New York and to some extent Chicago) and the cultural capital (which one might argue is split between New York and Los Angeles). Our model allows for such theories – we invoke this kind of reasoning to explain the size of Moscow and St. Petersburg in Russia – but the model suggests that this kind of explanation may not be required to explain the size of Paris. While the central role that Paris plays in French political, economic, and cultural life undoubtedly does endow it with a substantial core size, it is not clear that this role requires the city to be as large as it actually is.

The stylized result from the model is that a country with a large population and relatively few cities will tend to produce a Zipf distributed population in all but the largest city (or few cities) with the "overflow" population collecting at the top of the distribution (here, in Paris). Our framework suggests more generally that there is a relationship between number of cities and number of people in an urban system. This relationship has important implications for urban planning in the developing world. Our analysis presents a reason to expect the emergence of megacities such as Sao Paulo in Brazil, Dhaka in Bangladesh, and Jakarta in Indonesia. These countries generally have highly centralized governments and severely constrained capital availability. These factors make it very difficult for their urban systems to expand in terms of number of cities at a rate that bears any resemblance to their rates of population growth and urbanization. Developing nations are therefore left with a small number of cities and a large urban population. While a person's first move from rural to urban life may be from the countryside to a nearby city (a tendency that would tend toward balanced urban growth), our model suggests that the next step of inter-urban migration will tend to concentrate the urban population.

Megacities create numerous policy challenges, involving growth management and the provision of adequate infrastructure for a rapidly growing population. Failure to meet these challenges can create disastrous situations in the areas of environmental protection, public health, and human development and can lead to social unrest, political instability and violence (Bugliarello 1999).

The model further suggests that efforts to encourage migration from the first tier cities to middle sized cities are not likely to succeed over the long term. A government hoping to stem the growth of a primate city would do better to focus limited resources on providing the infrastructure and economic base that would allow large towns to become full participants in the urban system – thus expanding the number of cities and thereby reducing pressure on the capital.

34.4.2 Implications for Russian Urban Structure

The model suggests that two factors have played a role in creating the odd distribution currently observed in the Russian urban structure: a large urban system relative to its population and movement restrictions that have historically biased movements toward smaller cities. Unlike the urban structures of the US and France, the Russian urban structure was not created by free mobility and free markets. Soviet central planning created, instead "a structure of production – location, capital, employment, materials, energy use, etc. without any regard for economic opportunity costs, in an environment free of economic valuation" (Ericson 1999).

The result of this non-market resource allocation was an extensive urban structure that post-Soviet leaders have continued to work hard to preserve through subsidies and other measures. For a host of ideological and security related reasons, Soviet central planners aimed for relatively even dispersal of cities of fairly uniform size while at the same time creating a highly centralized system of power (Demko and Fuchs 1984). These factors contributed heavily to the creation of the odd urban structure that we see today.

One of the major Soviet era policies used to maintain this sprawling urban structure was a system of permits that were required for one to move from the hinterlands into an industrial center, and from a smaller industrial center to a larger one. This policy may be likened to biasing migration toward the smaller city in our model. While these policies are officially no longer in place since the fall of the Soviet Union, traces of them remain – particularly with regard to migration into Moscow and St. Petersburg. President Putin remained committed to avoiding Siberian "ghost towns" at almost any cost and many subsidies to these towns are in place long after the end of the Soviet system (Gaddy and Ickes 2002).

While the current form of our model is not useful for estimating the speed with which the size distribution might change with the relaxation of these restrictions and subsidies, we can use it to speculate about their general nature. We expect that unmanaged movement would lead to continued growth pressure on Moscow and St. Petersburg. We would further expect strong growth in a small handful of second tier industrial cities with current populations between 1 and 1.5 million (Novosibirsk is a typical example). However, we would expect this growth to extend to only three or four such cities, with the vast majority of cities with population between 100,000 and 1.5 million experiencing a prolonged period of population decline.

34.5 Conclusions

Our adaptive agent framework has allowed us to design and explore a framework for understanding city size distributions which, in spite of its extreme simplicity, is able to generate close approximations of the actual city size distributions for the US, France, and Russia. Although simple, this model is hard to examine analytically because of the high degree of interaction among its parts. Previous attempts to explain the Zipf distribution have, in general, gained analytical tractability by assuming independence of the growth rates of cities. While it is possible to generate the Zipf distribution using such assumptions (Gabaix 1999a) it is hard to imagine how the departures that we have reproduced could be derived in such a setting, and independent growth rates seem implausible for real city interactions. Our agent-based simulation methodology allows us to drop this restrictive assumption.

The use of this approach has made it possible for us to make real progress in understanding a phenomenon that has puzzled economists, geographers and others for over 50 years. Our model establishes a basis for moving beyond the assignment of mystical significance to the Zipf distribution of city sizes and allows us to see city size distributions as the result of straightforward behavioral rules. We can further understand Zipf as only a special case of city size distributions and see deviations from Zipf not as noise or error of some sort, but as the products of differing policies and situations.

Appendix: City Definition for France

The French National Institute of Statistics and Economic Studies (INSEE) produces a variety of excellent data on French cities using various definitions. These include the municipality (*commune*); the urban pole (*pôle urbain* or *unité urbain*); and the urban area (*Aire urbain*).

Of these three ways of defining a city, the first and third are inappropriate for use in this analysis. The municipality definition is not useful because most major cities are composed of many municipalities. The municipality of Paris, for example, had a population of only about 2.1 million people in 1999. The urban pole of Paris, in contrast, was composed of 396 such municipalities and was home to over 9.6 million people (Chavouet and Fanouillet 2000). While the legal definition of a municipality reflects historical and administrative realities, it tells us little about the urban agglomerations that we are studying.

Where the city as municipality definition is too restrictive, the city as urban area definition seems to be too broad. French urban areas are defined as those areas where at least 40% of the workers commute into an urban center that employs at least 5,000 people (INSEE 2004). These areas can be very large, often many times the area of the urban pole. A major problem with this definition for our purposes is that this surrounding area mixes people who commute into the city center with people whose social and economic lives are not integrated with the city. This commuting based definition also creates the impression of rapid growth for many cities, not because the cities have changed significantly, but because French commuting patterns have been changing, with workers traveling increasing distances to work (Julien 2001b). French cities have therefore been expanding their areas of influence more rapidly than they have been growing in terms of employment, built area, or other measures of city size (Julien 2001a).

The French definition of an urban pole strikes something of a balance between these two definitions. An urban pole is defined as a collection of contiguous communes in which more than half of the population lives in an area where buildings are separated by no more than 200 m. This definition is thus a reasonably close approximation of the built up area of the city. However, because this definition includes whole communes that are only partly urbanized, it tends to over count the urban population at the edges of cities. Because the circumference of a circle increases more slowly than its area, this bias tends to inflate the size of smaller cities.

In an effort to avoid this problem, we adopt a slightly more restrictive definition of a French city, that we will call an "urban center". Our definition follows the spirit of the one described by Le Gleau et al. (1996) while adapting it to better

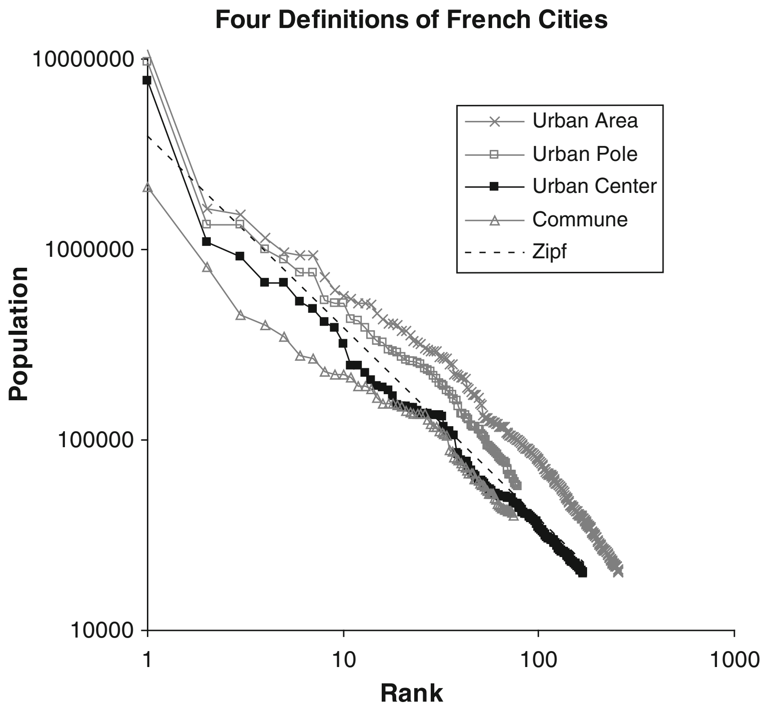


Fig. 34.13 Four definitions of French city sizes

capture the dominance of Paris in the French urban system. Le Gleau defines an urban center such that, if a single commune within an urban pole contains more than half of the pole's population, then this commune is the urban center. If the central commune contains less than half of the population of the pole, then it is agglomerated with the other communes of the pole that have at least half of the population of the largest commune. This definition has the effect of making the urban centers of France appear very nearly Zipf distributed (Fig. 34.13) – but it makes little sense as a definition of a city. Most notably, the central commune of Paris is much larger than any of the other 395 communes that make up the Parisian urban pole. This means that, by Le Gleau's definition, the urban center of Paris is represented by only this one commune, putting its size at 2.1 million people (as compared to 9.6 million in the urban pole).

We retain Le Gleau's concept of omitting the fringe areas by changing the criteria for agglomerating secondary communes, but refine it to avoid distorting large cities (particularly Paris). Under our definition, we agglomerate all of the communes in the pole that have a population greater than 20,000 people. Because communes tend to be of roughly uniform size, this is a reasonable proxy for density. We choose the number 20,000 because it is also the minimum size of a city in our dataset. Thus, any commune within an urban pole that would qualify as a city in its own right by virtue of its population of 20,000 is agglomerated into the urban center. This definition eliminates the inflation of the urban periphery that is present in the urban pole definition while retaining the basic idea of a city as a contiguous built-up area. Our analysis uses this definition of a French urban center.

References

Ades, A. F., & Glaeser, E. L. (1995). Trade and circuses: Explaining urban giants. *Quarterly Journal of Economics, 110*(1), 195–227.

Axtell, R. L., & Florida, R. (2001). *Zipf's law of city sizes: a microeconomic explanation* (Working Paper). Washington, DC: The Brookings Institution. www.brook.edu/dynamics/papers/cities

Bugliarello, G. (1999). Megacities and the developing world. *The Bridge, 29*(4), 19–26.

Chavouet, J. M., & Fanouillet, J. C. (2000). *Forte extensions des villes entre 1990 et 1999* (INSEE Paper #707).

Christaller, W. (1933). *Die zentralen Orte in Süddeutschland*. Jena: Gustav Fischer (Charlisle W. Baskin, Trans. (in part)). (1966). *Central places in Southern Germany* . Englewood Cliffs: Prentice Hall.

Cuberes, D. (2004). *The rise and decline of cities*. Unpublished manuscript. http://economics.uchicago.edu/download/cities3.pdf

Demko, G. J., & Fuchs, R. J. (1984). Urban policy and settlement system change in the USSR, 1897–1979. In D. Harris, G. J. Demko, & R. J. Fuchs (Eds.), *Geographical studies on the Soviet Union: Essays in honor of Chauncy*. Chicago: Department of Geography, University of Chicago.

Duranton, G. (2002). *City size distributions as a consequence of the growth process* (CEPR Discussion Paper 3577).

Ericson, R. (1999). The structural barrier to transition hidden in input-output tables of centrally planned economies. *Economic Systems, 23*(3), 199–244.

Federal Register. (2000). Vol. 65, No. 249, 27 Dec 2000.

Fujita, M., & Mori, T. (1997). Structural stability and evolution of urban systems. *Regional Science and Urban Economics, 24*(4–5), 399–442.

Fujita, M., Krugman, P., & Venables, A. J. (1999). *The spatial economy: Cities, regions and international trade*. Cambridge, MA: The MIT Press.

Gabaix, X. (1999a). Zipf's law for cities: An explanation. *Quarterly Journal of Economics, 114*, 739–767.

Gabaix, X. (1999b). Zipf's law and the growth of cities. *American Economic Review Papers and Proceedings, 89*(2), 129–132.

Gabaix, X., & Ioannides, Y. M. (2003). The evolution of city size distributions. In J. V. Henderson & J. F. Thisse (Eds.), *Handbook of urban and regional economics* (Cities and geography, Vol. IV). Amsterdam: North-Holland.

Gaddy, C. G., & Ickes, B. W. (2002). *Russia's virtual economy*. Washington, DC: Brookings Institution Press.

Gibrat, R. (1931). *Les inégalités économiques; applications: aux inégalités des richesses, à la concentration des entreprises, aux populations des villes, aux statistiques des familles, etc., d'une loi nouvelle, la loi de l'effet proportionnel.* Paris: Librairie du Recueil Sirey.

Grossman, G. M., & Helpman, E. (1991). Quality ladders in the theory of growth. *The Review of Economic Studies, 58*, 43–61.

Henderson, J. V. (1974). The sizes and types of cities. *The American Economic Review, 64*, 640–656.

Hill, F., & Gaddy, C. (2003). *The Siberian curse: How communist planners left Russia out in the cold*. Washington, DC: Brookings Institution Press.

Hill, B., & Woodrofe, M. (1975). Stronger forms of Zipf's law. *Journal of the American Statistical Association, 70*(349), 212.

INSEE. (2004). http://www.recensement.insee.fr/EN/RUB_MOT/ACC_MOT.htm

Iyer, S. D. (2003). Increasing unevenness in the distribution of city sizes in post-Soviet Russia. *Eurasian Geography and Economics, 44*(5), 348–367.

Jacobs, J. (1984). *Cities and the wealth of nations: Principles of economic life*. New York: Random House.

Julien, P. (2001a). *Les Grandes Villes Francaises Etendent Leur Influence* (ISEE Paper No. 766).

Julien, P. (2001b). *Les Deplacements Domicile-Travail: de plus en plus d'actifs travaillent loin de chez eux*. ISEE Paper No. 767. (Trans. (part) More and more workers are working far from home).

Kanemoto, Y. (1980). *Theories of urban externalities*. Amsterdam: North-Holland.

Le Gléau, J. -P., Pumain, D., & Saint-Julien T. (1996). Villes d'Europe: à chaque pays sa definition. *Économie et Statistique*, no. 294-295,1996-4/5 [in English as "Towns of Europe: To each country its definition"]. http://www.tu-chemnitz.de/phil/geographie/material/WS_03_04/BK/villeseuro.pdf

Marshall, A. (1890). Principles of Economics. London: Macmillan.

Overman, H. G., & Ioannides, Y. M. (2001). The cross-sectional evolution of the US city size distribution. *Journal of Urban Economics, 49*, 543–566.

Pumain, D. (2004). *Scaling laws and urban systems.* Santa Fe Institute (Working Paper No. 04-02-002).

Reed, B. (2002). On the rank-size distribution for human settlements. *Journal of Regional Science, 41*, 1–17.

Simon, H. A. (1957). *Models of man: Social and rational*. New York: Wiley.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge: Addison-Wesley.

Chapter 35
The Relationship of Dynamic Entropy Maximising and Agent-Based Approaches in Urban Modelling

Joel Dearden and Alan Wilson

Abstract Entropy maximising models are well established within the field of urban modelling as a method for predicting flows of people or material within an urban system. The dynamic urban retail model (Harris and Wilson, Environ Plan A 10:371–388, 1978) is one of the most well known applications of this technique and is an example of a BLV (Boltzmann-Lotka-Volterra) model. We define an agent-based model (ABM) of urban retail and explore whether it can be made equivalent to a BLV model. Application of both models to the metropolitan county of South Yorkshire in the UK indicates that both models produce similar outputs. This direct comparison provides some insights into the differences and similarities of each approach, as well as highlighting the relative strengths and weaknesses. The ABM has the potential to be easier to disaggregate, while the entropy maximising model is more computationally efficient.

35.1 Introduction

In this chapter, we explore the relationship between dynamic entropy maximising models of spatial interaction and agent-based models (ABMs). The dynamics are added to spatial interaction models through a form of Lotka-Volterra equations, which has led to these models being designated as BLV (Boltzmann-Lotka-Volterra) models. These models have a long history. ABMs are based on agents which have 'development rules'. There is sometimes a confusion with CA (cellular automata) models in which the cells are given development rules and so in some ways can seem like agents. This is resolved by Epstein and Axtell (1996) by distinguishing agents from an 'environment' (which can be a grid of cells).

J. Dearden (✉) • A. Wilson
Centre for Applied Spatial Analysis (CASA), University College, London, UK
e-mail: j.dearden@ucl.ac.uk; a.g.wilson@ucl.ac.uk

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_35,

Here, we want to explore whether it is possible to define a set of agents, an environment and associated rules in such a way that there is an ABM model that is equivalent to any BLV model. If this can be established, since BLV models are typically more highly developed and realistic, then this should point the way to building more realistic ABMs. We proceed as follows. We take the retail model as an archetypal BLV model and then seek to formulate an equivalent - or near-equivalent - ABM.

35.2 The Retail Model as an Archetypal BLV Model

Define S_{ij} as the flow of spending power from residents of i to shops in j; let e_i be spending per head and P_i the population of i. W_j is a measure of the attractiveness of shops in j which, for these illustrative purposes, we take as the logarithm of 'size' – reflecting the range of choice and the lower prices through economies of scale. The vector $\{W_j\}$ can then be taken as a representation of urban structure – the configuration of W_js. If many W_js are non-zero, then this represents a dispersed system. At the other extreme, if only one is non-zero, then it is a very centralised system. There is clearly, potentially, a measure of order in this specification of structure. An obvious order parameter would be the number of zones greater than some size $N(W_j > M)$ for some constant M.

A spatial interaction model can be built by maximising an entropy function in the usual way (Wilson 1967, 1970) to give:

$$S_{ij} = A_i e_i P_i W_j^{\alpha} e^{-\beta c_{ij}} \tag{35.1}$$

where

$$A_i = 1 \Big/ \sum_k W_k^{\alpha} e^{-\beta c_{ik}} \tag{35.2}$$

to ensure that

$$\sum_j S_{ij} = e_i P_i \tag{35.3}$$

and

$$\sum_{ij} S_{ij} \log W_j = X \tag{35.4}$$

where $\log W_j$ is taken as the measure of consumer benefits and X is an estimate of the total benefits achieved. We also have

$$\sum_{ij} S_{ij} c_{ij} = C \tag{35.5}$$

α and β are parameters (which are the Lagrangian multipliers associated with Eqs. 35.4 and 35.5). Because the matrix is only constrained at the origin end, we can calculate the total flows into destinations as

$$D_j = \sum_i S_{ij} = \sum_i \left\{ \frac{e_i P_i W_j^{\alpha} e^{-\beta c_{ij}}}{\sum_k W_k^{\alpha} e^{-\beta c_{ik}}} \right\} \tag{35.6}$$

The model is essentially based on a microcanonical ensemble with double labels (i, j) for 'energy' states instead of the single i-labels in the classical gas model, hence the 'Boltzmann' element of the model. A_i is the inverse of the statistical mechanics partition function – but at a zonal level. C can be taken to represent total 'energy' in some sense and the c_{ij} are individual energy states. c_{ij} is a measure of impedance, i.e. some kind of transport cost. c_{ij} is usually taken as a generalised cost, i.e. a weighted sum of elements like travel time and money cost.

Note that W_j^{α} can be written as

$$W_j^{\alpha} = e^{\alpha \log W_j} \tag{35.7}$$

so that

$$u_{ij} = \alpha \log W_j - c_{ij} \tag{35.8}$$

can be taken as a measure of utility.

A suitable hypothesis for representing the dynamics is (Harris and Wilson 1978):

$$\frac{dW_j}{dt} = \varepsilon (D_j - KW_j) W_j \tag{35.9}$$

where K is a constant such that KW_j can be taken as the (notional) cost of running the shopping centre in j. This equation then says that if the centre is profitable, it grows; if not, it declines. The parameter ε determines the speed of response to these signals. Equation 35.9 is a form of the Lotka-Volterra equation and hence the characterisation of these models as BLV models (Wilson 2008). The equilibrium position is given by

$$D_j = KW_j \tag{35.10}$$

which can be written out in full as

$$\sum_i \left\{ \frac{e_i P_i W_j^{\alpha} e^{-\beta c_{ij}}}{\sum_k W_k^{\alpha} e^{-\alpha c_{ik}}} \right\} = KW_j \tag{35.11}$$

and these are clearly nonlinear simultaneous equations in the $\{W_j\}$.

It is possible to characterise the kinds of configurations that can arise for different regions of α and β space: for larger α and lower β, there are a smaller number of larger centres; and vice versa - as characterised in broad terms by the order parameter, $N(W_j > M)$, defined earlier (Wilson and Oulton 1983; Clarke and Wilson 1985, 1986; Clarke et al. 1986; Lombardo 1986).

35.3 An Agent-Based Retail Model

We define two kinds of agents:

- Consumers (C), with retail expenditure
- Retailers (R), each running a single shop[1]

Each consumer and retail agent can be located at a unique point in our region of interest, rather than being aggregated into zones. We then need to define an associated set of running costs for each R-agent and a matrix of interaction costs.

The C-agents will each be given a residential location, i – possibilities of changing these would involve a model extension. Each R-agent will seek a possible shop location, j. The i's and j's are now nodes in the environment. In one time period, the consumers will each deploy a utility maximisation rule – on a probabilistic basis - to find a shop – the utility being given in Eq. 35.8. Note that this involves 'looking' further than neighbouring cells in the environment and hence mimics – in fact generalises – Epstein's and Axtell's 'vision' mechanisms in their sugarscape model (Epstein and Axtell 1996). (This also has a relationship to Potts' models in the statistical mechanics of crystal lattices in which interactions extend beyond nearest neighbours.) At each j, the retailer will be able to sum the in-flows and decide whether the revenue exceeds the costs or not – cf. Eq. 35.9. In our trial ABM, there will be a probability that a loss-making retailer will seek to relocate. In the terminology of Holland (1995), each agent has a stimulus and a response – consumers having income as the stimulus and spending it as a response; each retailer having the total revenue inflow as a stimulus and the possibility of relocation as a response. The retailer totalling inflows is equivalent to Holland's 'tagging' of (C) agents as a means of aggregating. The decision to relocate is an (R) agent interaction with the environment. We expect that running a model of this type would generate emergent behaviour as in the BLV model and we put this to the test below.

The model to be tested works as follows. Each retailer calculates a *range of choice* factor R_j for its current shop location j by counting the number of other shops within n metres of itself, where n is, say, easy walking distance to other shops nearby – here we use n = 200 m. We will refer to this range as the *range of choice distance*. R_j then represents the number of other shops a consumer would also be able to visit if they travelled to shop j. The value of R_j is recalculated every iteration of the model.

Each retailer calculates the net income f_j for their shop j every iteration:

$$f_j = D_j - K \tag{35.12}$$

where D_j is the total income of shop j and K its operating costs. The total operating costs of all shops in the region are set equal to the total spending money of all

[1] For simplicity we are only modeling independent retailers. An interesting extension of the model might be to include chain stores with one retailer owning multiple shops.

consumers in the region. As a result some shops will always be unprofitable and as a result their owners will be looking to relocate.

At each iteration a proportion of randomly chosen retailers are allowed to relocate. Only those retailers making no money ($f_j \leq 0$) will consider relocating when prompted. A retailer that decides to relocate has an equal chance of either moving to a random position in the region or moving near to one of its competitors. If it decides to move near a competitor, it will evaluate the profit made by every other shop k and use this to calculate a probability m_{jk} of moving near that shop:

$$m_{jk} = \frac{t_k}{\sum_l t_l} \tag{35.13}$$

where t_k is the profit made by shop k. If a shop k is making a loss, then $t_k = 0$ and there will be no chance of another retailer relocating near it. It is possible for a retailer to decide to relocate near itself in which case it remains where it is. Here we define "moving near to a competitor" as: moving to a random position that is at most n metres away from it, where n = *range of choice distance*.

A consumer's position in the region is fixed and represents the position of its home. Each consumer calculates the probability p_{ij} that it will travel from its home i to visit shop j:

$$p_{ij} = \frac{R_j^{\alpha} e^{-\beta c_{ij}}}{\sum_k R_k^{\alpha} e^{-\beta c_{ik}}} \tag{35.14}$$

As with the entropy maximising model, α represents the impact of range of choice on shopping decisions, β the impact of travel cost and c_{ij} is the travel cost between house i and shop j. The set of all p_{ij} for one consumer agent i makes up a probability distribution that represents the likelihood of that consumer shopping at each shop in the region. Each time a consumer is prompted to go shopping it generates a uniform random number to choose a shop based on these probabilities. Each consumer agent has a fixed amount of money to spend when it goes shopping and it always spends the full amount in its chosen shop. This amount can obviously vary across agents depending on the data used to initialise the model.

During the course of a model run, several retailers may locate within each other's *range of choice distance* and so mutually benefit each other by increasing each other's range of choice factor, R_j. We can think of a group of retailers that does this as making up an "emergent" retail zone, which could, in practice, represent a row of shops, a high street or a shopping mall. Some pairs of retailers in the retail zone might be outside of each other's *range of choice distance* and so not mutually benefit each other but still be part of the same group because they are linked indirectly via other retailers. We can identify these groups as they form in the model using a recursive algorithm to identify closed groups of retailers that are all directly or indirectly connected to each other. The full algorithm is given in Appendix 1. The number of retailers in each "emergent" retail zone can be thought of as equivalent to the

W_j term in the entropy maximising model, and the value of R_j for each retailer in the centre may be less than or equal to this term depending on its layout. An alternative measure of attractiveness for a retailer could use the W_j term for the retail zone it belongs to rather than the R_j term. The formula for p_{ij} would then be:

$$p_{ij} = \frac{W_j^{\alpha} e^{-\beta c_{ij}}}{\sum_k W_k^{\alpha} e^{-\beta c_{ik}}} \quad (35.15)$$

One iteration of the model comprises the following steps:

1. Each consumer chooses a shop and spends all their money there.
2. All retailers calculate their profit level.
3. ε% of retailers are given the option of relocating.
4. All retailers recalculate their range of choice factor, R_j (because one retailer moving can affect multiple neighbours).
5. The size of each emergent retail zone is calculated.
6. All consumers recalculate their set of probabilities p_{ij}.

The percentage ε represents the rate at which retailers respond to profit levels and so performs a similar role to ε in the entropy maximising model.

35.4 Results

35.4.1 System Overview

In order to properly test the ABM, we choose to model the metropolitan county of South Yorkshire in the UK. By modelling a real system, we are better able to judge whether the outputs are realistic or not. Full details of data sources and how we use them to initialise each model are given in Appendix 2. Figure 35.1 shows the raw data plotted on a map and gives an idea of the distribution of retail outlets within the county.

35.4.2 System Equilibrium

The BLV model generally tends towards an equilibrium solution. We can detect this by examining the change in size of each retail zone, W_j. Here we define equilibrium as less than 0.001% change in the size of each retail zone for at least 250 iterations. We abandon a model run after 10,000 iterations in case the model does not converge.

For the ABM to detect equilibrium is more difficult because whatever structure emerges is not fixed. For this purpose we use the Fuzzy Numerical technique

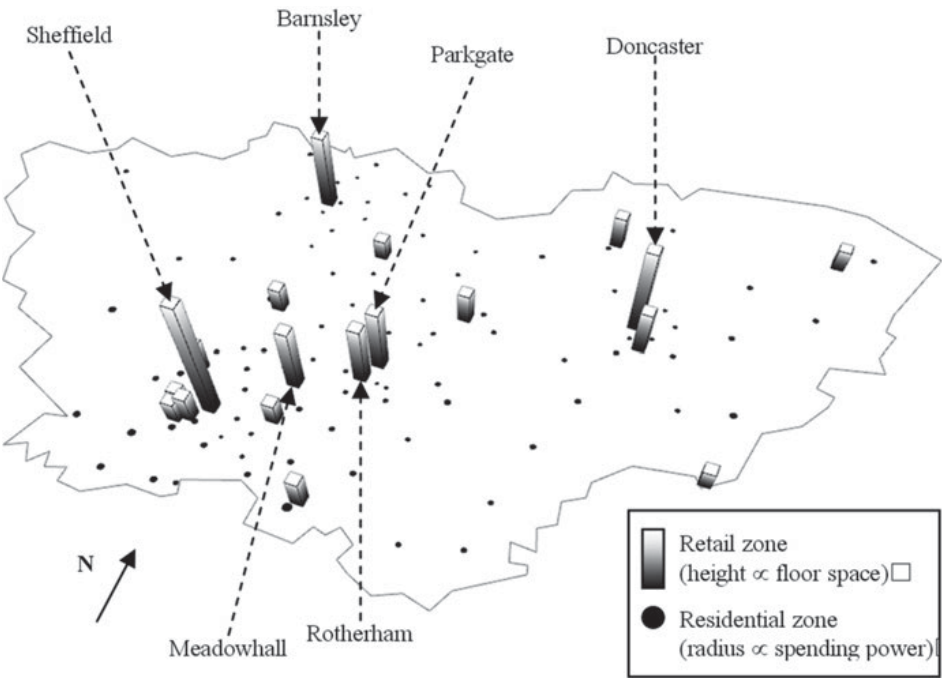


Fig. 35.1 Visualisation of the South Yorkshire data

developed by Hagen-Zanker et al. (2006) for comparing continuous raster maps. At 100 iteration intervals, we calculate the number of retailers in each cell of a 40×40 grid that covers the region. We use the fuzzy numerical similarity metric to compare consecutive grids,[2] and if this metric exceeds 0.996, we assume that the model has reached a stable state.

The rate of response parameters in each model were set as follows: for the BLV model we use $\varepsilon = 0.003$, and for the ABM we set $\varepsilon = 1\%$.

35.4.3 Emergent Structures

Our first job is to see what kind of structures and behaviour, if any, emerge from the ABM. For this we use population data for South Yorkshire to generate consumer agents (Fig. 35.2) but start with a uniform random distribution of retailers across the region. For this first exploration, we choose the model parameters $\alpha = 1.0$, $\beta = 0.5$. The output using Eq. 35.14 is shown in Fig. 35.3, and using Eq. 35.15, is shown in Fig. 35.4.

---

[2] We use a neighborhood size of 10.

Fig. 35.2 Distribution of the 50,000 consumer agents in the South Yorkshire model

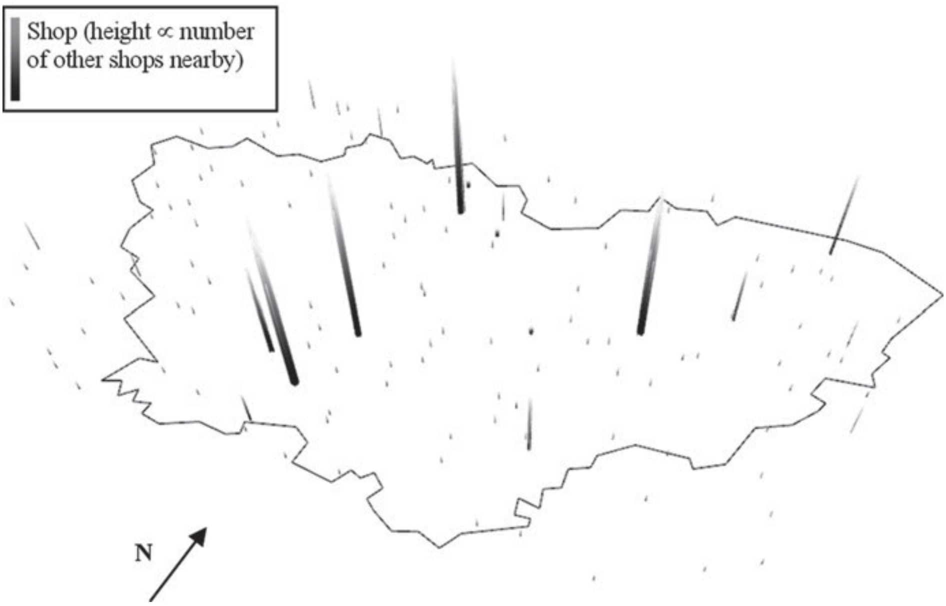


Fig 35.3 Structure in the model using Eq. 35.14

In both cases, realistic structures emerge with large clusters of retailers appearing at the major cities and towns in the region: Sheffield, Barnsley, Rotherham and Doncaster. Equation 35.14 appears to produce more compact retail centres than Eq. 35.15, presumably because there is more benefit in locating close to as many other retailers as possible.

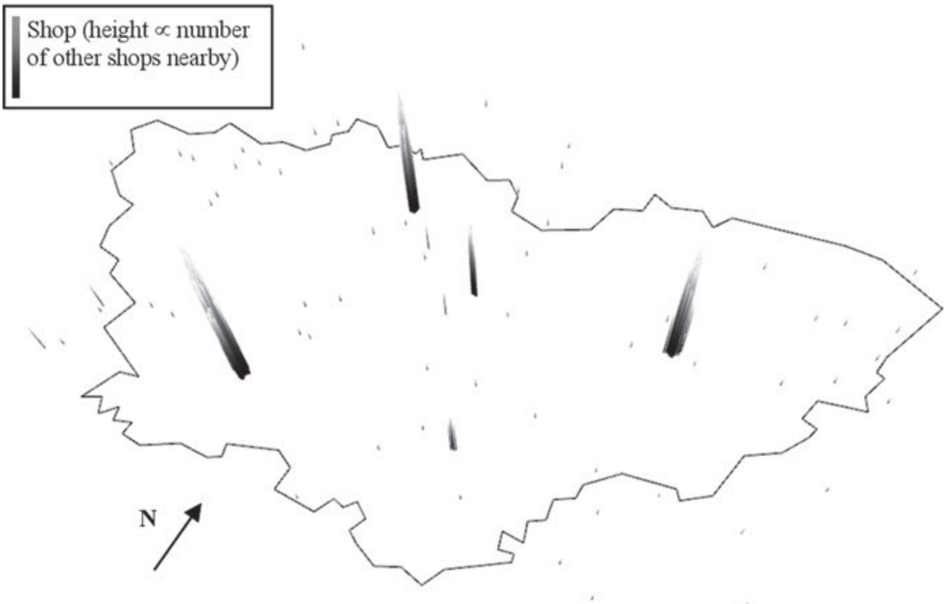


Fig. 35.4 Structure in the model using Eq. 35.15

35.4.4 Results Grids

For the BLV model, a results grid can be plotted (Wilson and Dearden 2011) and this shows the possibilities of emergent behaviour and of phase changes through plots of the order parameter. The results grid in Fig. 35.5 represents the parameter space formed by varying the parameters $\alpha = 0.1–2$ and $\beta = 0.1–2$, using a step size in both cases of 0.2.

Figure 35.6 shows a surface plot of the order parameter $N(W_j > 300{,}000)$ for the same parameter space as Fig. 35.5.

For comparison, we produce a results grid for both variants of the ABM and show the results in Figs. 35.7 and 35.8. A visual comparison of the grids suggests that both models are producing similar outputs with the BLV model grid in Fig. 35.5.

Plots of the order parameter $N(W_j > 300{,}000)$ in Figs. 35.9 and 35.10 indicate that the behaviour of the ABM is largely similar to the BLV model across the parameter space; however, there some clear differences, especially when Eq. 35.14 is used. This may be, in part, because the ABM is a much more noisy system than the BLV model.

35.4.5 Model Calibration

We can calibrate the model runs in the results grids above to find a best fit for each model. The goodness of fit of the output from the BLV model is easily checked against data using the coefficient of determination, R-squared, because we have a

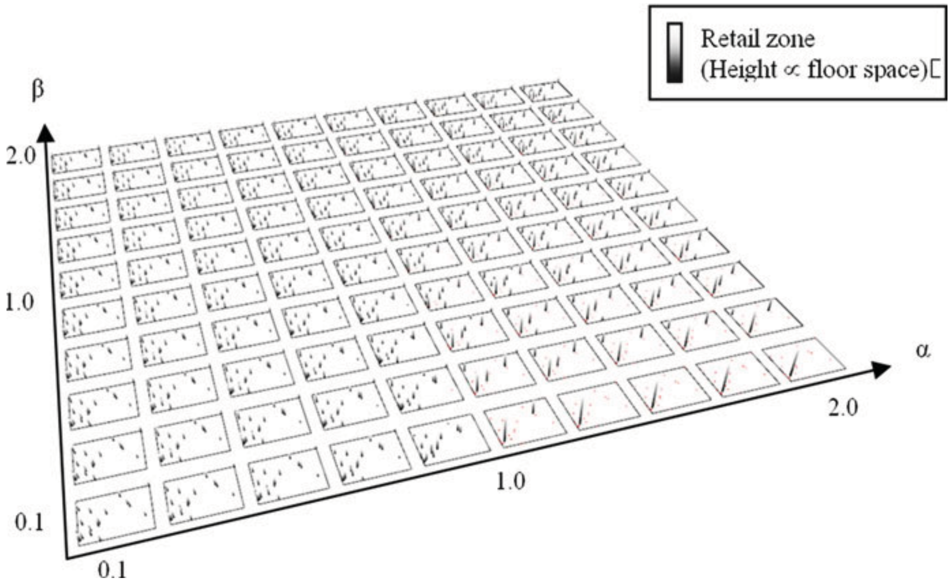


Fig. 35.5 Results grid for the BLV model in (α, β) space

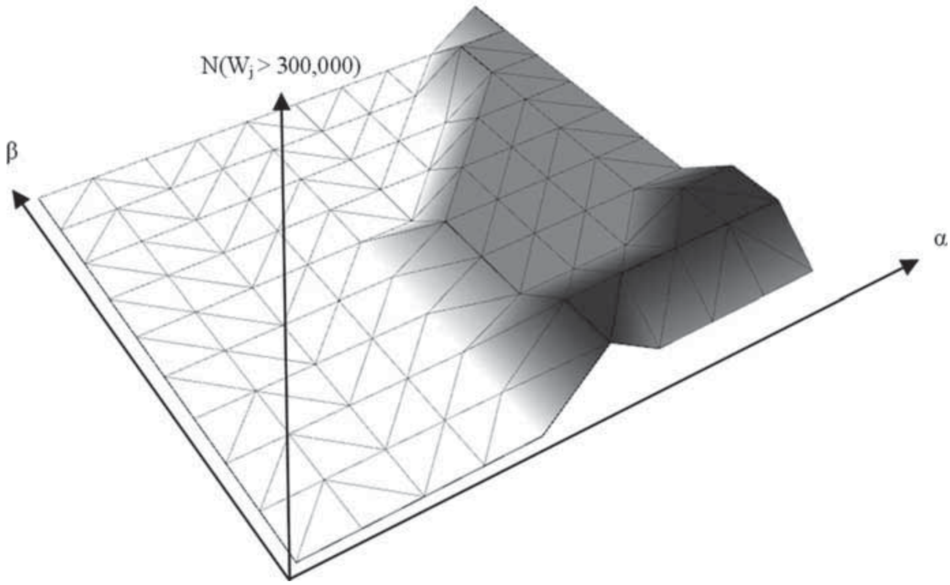


Fig. 35.6 Order parameter surface plot for the BLV model in (α, β) space

fixed zone system – each retail zone in the model is at the same position as its equivalent zone in the data. We find the best fit at α =1.57 and β=0.85, which produces an R-squared value of 0.74. The output (Fig. 35.11) matches the real system reasonably well. However, because we are not taking into account the complexities

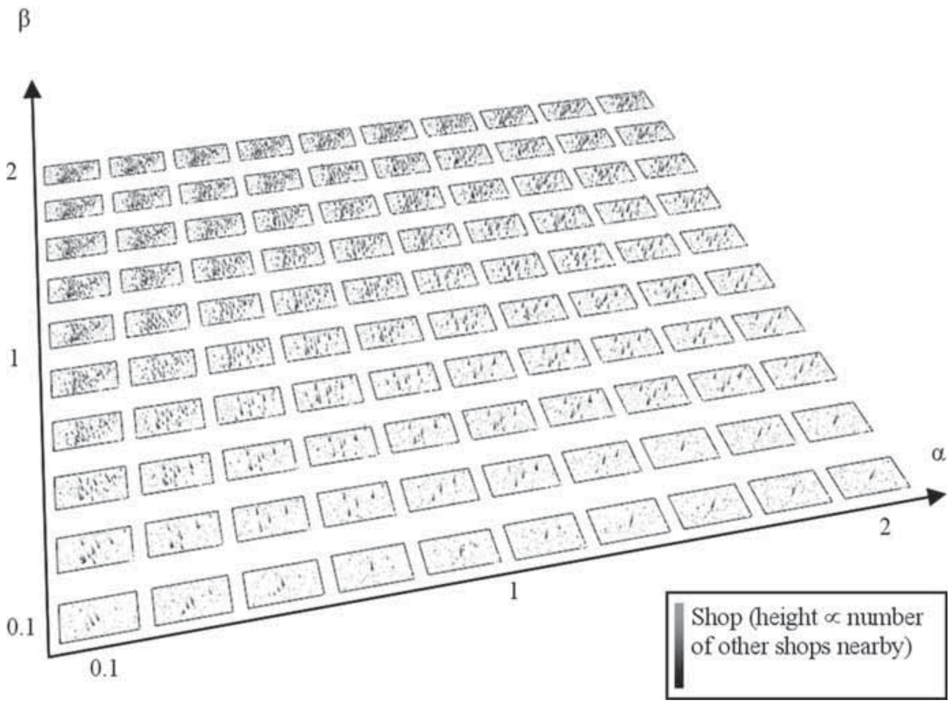


Fig. 35.7 Results grid in (α, β) space for the agent based model using Eq. 35.14

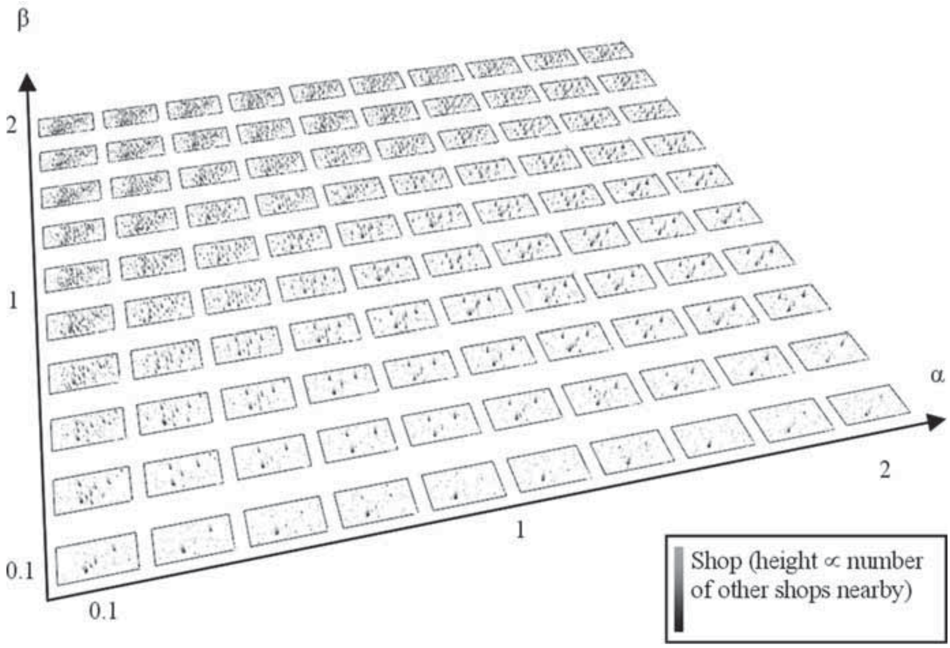


Fig. 35.8 Results grid in (α, β) space for the agent based model using Eq. 35.15

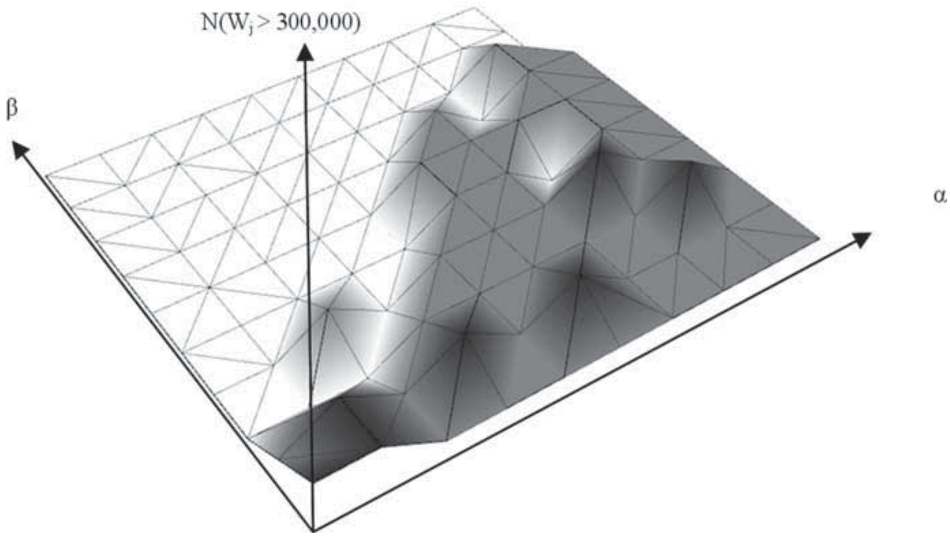


Fig. 35.9 Order parameter surface plot for the agent based model using Eq. 35.14 in (α, β) space

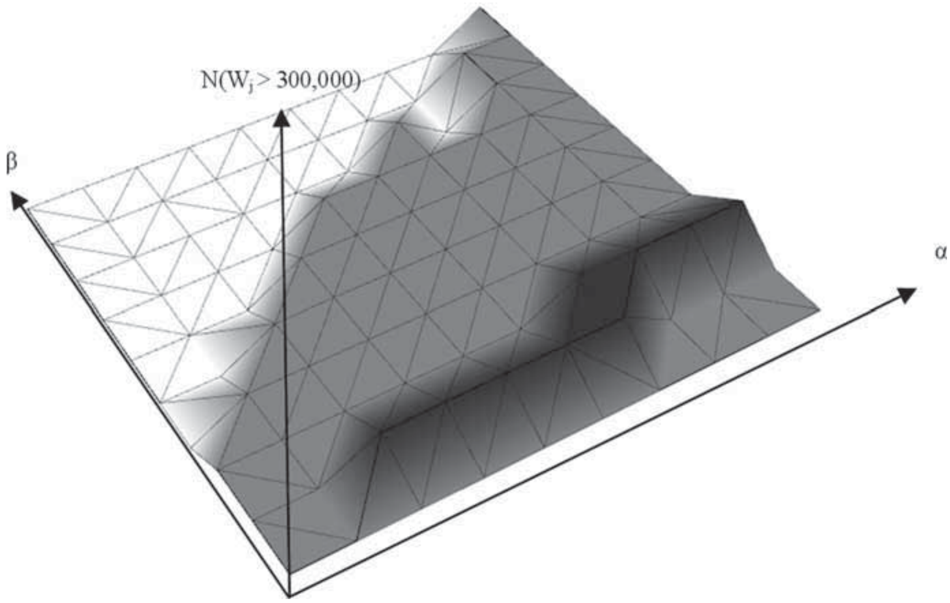


Fig. 35.10 Order parameter surface plot for the agent based model using Eq. (35.15) in (α, β) space

of the transport network in our measure of distance, we are missing some retail zones, most notably the large Meadowhall shopping centre near Sheffield.

Measuring the output from the ABM is more difficult because retail zones can emerge anywhere on the map. We again make use of the Fuzzy Numerical comparison technique because it allows us to compare two retail systems that differ in both

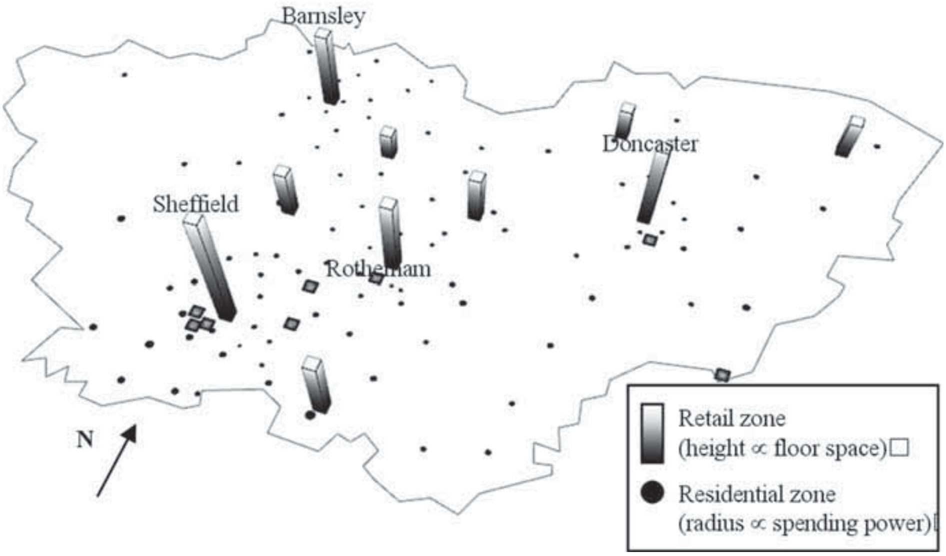


Fig. 35.11 Best fit BLV model run for South Yorkshire

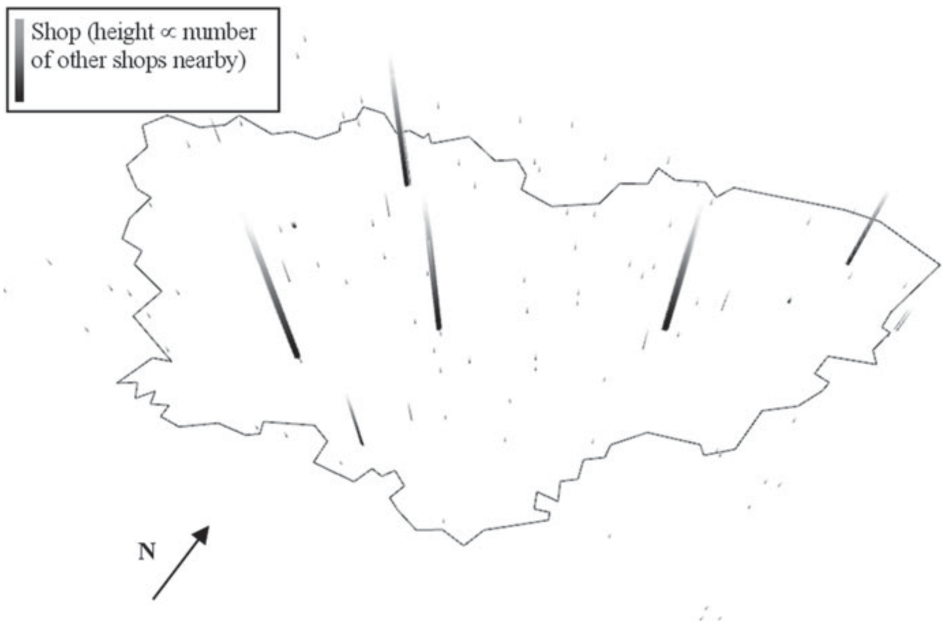


Fig. 35.12 Best fit model run using Eq. 35.14

position and number of centres. To do this we convert both the town centres data (see Appendix 2) and the ABM output to a raster grid. We find the best fit parameter set for Eq. 35.14 at $\alpha = 1.57$ and $\beta = 1.15$ (35.2) and for Eq. 35.15 at $\alpha = 1.57$ and $\beta = 0.73$ (35.3) (Figs. 35.12 and 35.13).

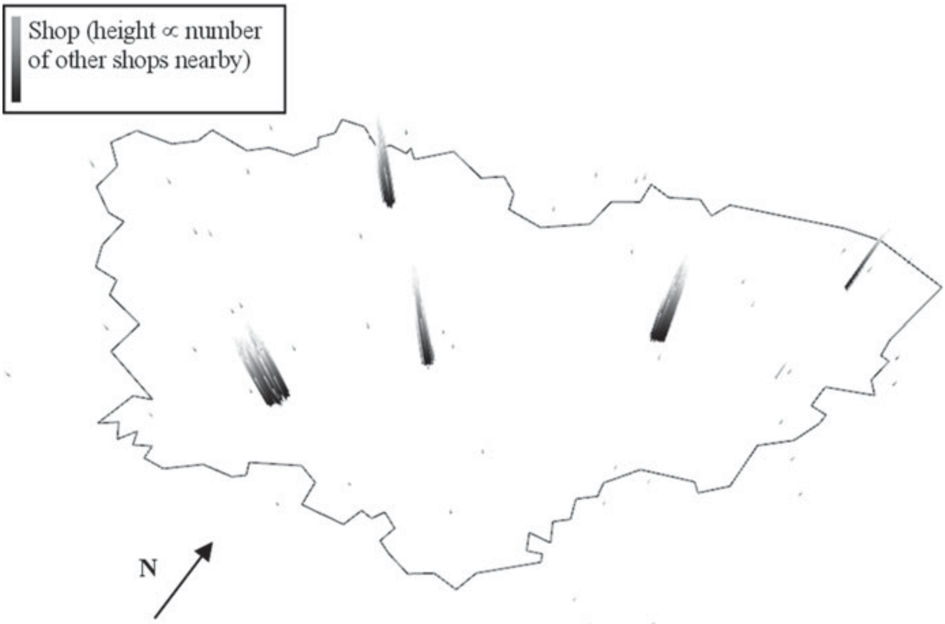


Fig. 35.13 Best fit model run using Eq. 35.15

35.5 Conclusion

We have demonstrated a zoneless ABM of urban retail in which retail centres emerge due to individual retailers locating near each other. By comparing the outputs across a portion of (α, β) parameter space, we have demonstrated that it produces similar results to the well established BLV urban retail model. This work is a first step towards defining an ABM that is equivalent to any BLV model.

In comparing the models side by side, it becomes clear that the BLV model is far less computationally intensive to run when dealing with very large systems. For example, the real population of South Yorkshire is approximately 1.2 million people. However, in order to produce fast run times, we modelled the region using ~50,000 consumer agents and ~500 retailer agents (see Appendix 2 for more details). Given more time and/or computing power the number of retailer and consumer agents could be increased closer to the real number in South Yorkshire.

The ABM approach does provide benefits when built using an object oriented programming language because the system can be constructed in a modular fashion. This may mean that the ABM is easier to disaggregate because we can quickly introduce multiple agent types.

Here we used macro level to data to calibrate our ABM output. It would be preferable to use micro-level data to generate and calibrate the retailer and consumer agents, although locating appropriate sources of data is difficult.

The next steps would involve more disaggregation, i.e. more agent types, and then to extend the model in the direction of a comprehensive model that draws in a wider range of urban sub-models - possibly using the Lowry (1964) model as an archetype, but also moving towards realism (cf. Wilson 2006).

Appendix 1: An Algorithm for Calculating the Boundary and Membership of Emergent Retail Zones

Pseudo code is given here for calculating the emergent retail zones that appear in the model.

```
Setup an empty list of shops called processedList
While there are still shops not in processedList
  Choose a shop s that is not in processedList
  Setup an empty list called shopList
  Call function findClosedGroup with s, shopList
  and processedList as parameters
  shopList now contains all the shops in one retail
  zone
```

The recursive function **findClosedGroup** does the following:

```
For each shop t nearby
  If t is not already in shopList
    Add t to shopList
    Add t to processedList
    Call function findClosedGroup with t,
    shopList and processedList as parameters
```

Appendix 2: Data Sources

The retail data comes from the **Town Centres project** 2004. We use the *total retail floor space* attribute from each town centre area to:

- **For the BLV model**: set the floor space of each retail zone.
- **For the ABM**: determine the number of retailer agents we need to generate inside the town centre area. We do this by dividing the *total retail floor space* by an average shop size of ~2,800 m^2, which produced about 500 retailer agents for the region.

The average shop size was chosen to reduce the computational load but could obviously be reduced given more time.

The population data are from the 2001 UK Census. We use the *All people* field from the KS001 Usual resident population table for the CAS Ward boundaries. Then:

- **For the BLV model**: the centroid of each CAS Ward is the location of each residential zone and the P_i value is set to the *All people* value.

- **For the ABM**: the *All people* value, divided by an aggregation factor, decides the number of consumer agents we generate at random positions inside the CAS Ward boundary. In this case the aggregation factor was set to 24 (meaning that each consumer agent represents 24 people) and produced ~50,000 consumer agents in the model. Again this was done to allow for reasonable computation times.

The average income data are from the CACI Paycheck data for 1999 at postcode area level. We aggregate these data up to CAS Ward level and then use the aggregate value for each CAS Ward to:

- **For the BLV model**: set the e_i value for the corresponding residential zone.
- **For the ABM**: set the spending money available to each agent living inside the CAS Ward.

For simplicity, all travel costs were calculated from the Euclidean distance between two points.

References

Clarke, M., & Wilson, A. G. (1985). The dynamics of urban spatial structure: The progress of a research programme. *Transactions of the Institute of British Geographers, 10*, 427–451.
Clarke, G.P. & Wilson, A.G. (1986). *Combining theoretical and empirical research in retail location analysis* (Working Paper 468). Leeds: School of Geography, University of Leeds.
Clarke, G. P., Clarke, M., & Wilson, A. G. (1986). Multiple bifurcation effects with a logistic attractiveness function in the supply side of a service system. *Systemi Urbani, 7*, 43–76.
Epstein, J. M., & Axtell, R. (1996). *Growing artificial societies*. Cambridge: MIT Press.
Hagen-Zanker, A., Engelen, G., Hurkens, J., Vanhout, R., & Uljee, I. (2006). *Map comparison kit 3: User manual*. Maastricht: Research Institute for Knowledge Systems.
Harris, B., & Wilson, A. G. (1978). Equilibrium values and dynamics of attractiveness terms in production-constrained spatial-interaction models. *Environment and Planning A, 10*, 371–388.
Holland, J. H. (1995). *Hidden order: How adaptation builds complexity*. Reading: Addison Wesley.
Lombardo, S. R. (1986). New developments of a dynamic urban retail model with reference to consumers' mobility and costs for developers. In D. A. Griffith & R. J. Haining (Eds.), *Transformations through space and time*. Dordrecht: Martinus Nijhoff.
Lowry, I. S. (1964). *A model of metropolis* (Memorandum RM.4035-RC). Santa Monica: Rand Corporation.
Wilson, A. G. (1967). A statistical theory of spatial distribution models. *Transportation Research, 1*, 253–269.
Wilson, A. G. (1970). *Entropy in urban and regional modelling*. London: Pion.
Wilson, A. G. (2006). A generalised representation for a comprehensive urban and regional model. *Computers, Environment and Urban Systems, 31*, 148–161.
Wilson, A. G. (2008). Boltzmann, Lotka and Volterra and spatial structural evolution: An integrated methodology for some dynamical systems. *Journal of the Royal Society, Interface, 5*, 865–871.
Wilson, A. G., & Dearden, J. (2011). Phase transitions and path dependence in urban evolution. *Journal of Geographical Systems, 13*(1), 1–16.
Wilson, A. G., & Oulton, M. (1983). The corner shop to supermarket transition in retailing: The beginnings of empirical evidence. *Environment and Planning A, 15*(2), 265–274.

Chapter 36
Multi-agent System Modelling for Urban Systems: The Series of SIMPOP Models

Denise Pumain

Abstract The SIMPOP simulation model was the first application of a multi-agent system (MAS) in geography and has been followed by a series of related applications. This chapter summarizes a few specific features of this series of models and the main insights that have been gained from this experiment. As the first objective of these models was to reconstruct the evolution of urban settlements at broad scales in geographical space and time (i.e. at national or continental levels and for decades or centuries), we explain the selection of the stylised facts making up knowledge about the dynamics of complex urban systems. They are used in the simulation models so as to reconstruct the interaction networks structuring the systems of cities. The originality of SIMPOP is thus to simulate the emergence and further hierarchical and functional differentiation of interdependent cities from interactions between them, so that the agents in this model are immobile entities, representing complex aggregation of individuals at meso-level. The quality of MAS is underlined for its flexibility in modelling spatial interactions with varied geographical configurations, and for its ability to deal with objects occurring on different scales, between geospatial entities that have expanding range of activities. We explain how we set about restricting the impact of the difficulties inherent in this type of modelling, which have been the subject of frequent criticism, in particular for their excessive complexity, thought to make any validation procedure unfeasible. In particular, we endeavour to describe in detail the various stages of model construction, on the basis of stylised facts obtained via numerous observations and comparisons, and set out to perform a multi-scale validation by testing the plausibility of the results delivered by the model at different aggregation levels. However, despite these promising methods of validation, further improvements are necessary for fully exploiting the capacities of simulation by using more powerful computing devices and validation methods. In this direction, the generic model SIMPOP will be completed and

D. Pumain (✉)
Institut Universitaire de France, UMR Géographie-cités, Université Paris I, Paris, France
e-mail: pumain@parisgeo.cnrs.fr

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_36,

transferred to an open and scalable simulation platform, and specific versions will be developed and tested for the main regions of the world.

36.1 Introduction

There is considerable interest in multi-agent system (MAS) modelling in the human and social sciences. This is because the analytic methods that are usually developed for dealing with complexity in physical or biological systems meet two major limitations when applied to this domain. First, societal complex systems share a large path dependence regarding their historical and geographical context, due to a functional directedness in social innovation and institutions, and resulting in a very high speed of their evolutionary processes, when compared to biological or geological evolutions (Lane et al. 2009). Second, they rarely admit reductionism in explanation, because the structuring of societies through reproduction and innovation always involves a variety of processes of a different nature that are analysed and formalised through specific axioms and concepts by each discipline. For instance, in order to describe urban systems dynamics, we can model separately population growth in demography, markets in economics, conflicts in political science and spatial configurations in geography in terms of complex systems analytical tools, but they have to be rearticulated for making a concrete situation understandable in a satisfying way (Pumain 2011a). Thus it is not only the intentionality and liberty of human behaviour that imposes limits to these models for predicting their future evolution, but probably as well a complexity of a different kind from natural phenomena.

The development of simulation models as "artificial laboratories" for exploring the behaviour and evolution of complex systems in social sciences is still in its infancy, especially regarding the principles of parsimony in the construction of, and the reliability in validation methods during experimentation that were identified by Batty and Torrens (2001) as major issues for their possible use in prediction (the issue of validation is further discussed by Ngo and See 2012). However, we think that MAS models do have a potential for progress in these two directions (Bura et al. 1996); this has been a focus over the last 15 years in the series of the SIMPOP models.[1] The SIMPOP models simulate how a structured system of cities emerges over a long time period in a given territory from interactions between individual cities. This is achieved by reconstructing interaction networks from previous information and knowledge in urban theory (in the form of "stylised facts"). In our view, it is not only the capacity of modelling the behaviour of reacting, adaptive or learning individual

[1] We acknowledge here the participation of computer scientists, Jacques Ferber, Stéphane Bura, Alexis Drogoul, Benoît Glisse, Jean-Louis Giavitto, Guillaume Hutzler, and Thomas Louail, as well as statisticians and geographers, Lena Sanders, Hélène Mathian, France Guérin-Pace, Anne Bretagnolle, Céline Vacchiani-Marcuzzo, Arnaud Banos, Clara Schmitt and Sébastien Rey to the conception and use of these models. The European programme ISCOM (2002–2006) directed by David Lane supported the development of the SIMPOP2 model.

agents that represents the major epistemological change in urban dynamics, labelled as "the shift from macro-static to micro dynamics" that has occurred during the last 50 years (Batty 2008). Indeed, it is the possibility of merging entities from different levels (of space and time) in the same model, together with the parallel development of multi-level validation methods that may as well be a fruitful way for MAS model design in geography. Such tools already can be used for testing theoretical hypothesis about the historical evolution of urban systems and could perhaps help in exploring their possible future evolution, and reducing the uncertainty about the future by delineating plausible dynamic trajectories.

36.2 Some Stylised Facts for Explaining How Urban Hierarchies Emerge in a Geographical Theory of Urban Systems

The series of SIMPOP models was first conceived as a tool for theory testing. The main objective was to model stylised facts that would retrace the generic dynamics of urban systems over long periods of time. Urban systems are complex adaptive systems whose evolving geographical structure can be explained in the framework of an evolutionary theory (Pumain 2000). As with any geospatial system, they are largely self-organised (Allen 1997), even when constrained by strong forms of political or economic control. This is because urban systems involve very large numbers and a wide variety of interactions of different kinds at different scales that change over time. Similar to territorial systems like countries or regions, urban systems share common dynamic properties which produce regularities in the organisation of their geographical diversity. When considered as self-organised systems, systems of cities in any part of the world can be described at two basic levels of observation, that each are emerging from the interactions occurring mainly at lower levels: indeed one city can be described as a spatial organisation that is produced by the interactions between different urban actors, in majority from inside and partly outside the urban field generated by the "daily urban system" (Berry 1964). In contrast, is the second upper level systems of cities that are mainly shaped by the exchanges of persons, goods, capital and information from one city to the next, under the constraint of evolving ecological and overall political, technological, cultural and economic rules. These networks of cities are characterised by emerging properties that remain relatively stable over time and space, as: the increasing *hierarchical differentiation* in their sizes (as abstracted by Zipf's law or lognormal distribution, and measured by the accumulation of population or wealth as well as in qualitative way by the complexity of their society (see Gulden and Hammond 2012 for a further discussion of Zipf's law)); their *functional differentiation* (economic and social specialisation) that can maintain over much longer periods of time than the corresponding characteristics of the individuals that are composing them (change in profession, residential migration, replacement of generation); their rather regular *spatial organisation* (originally summarized by central place theory, today more frequently conceived as complex networks, organised according scale free small worlds networks (Rozenblat 2010; Rozenblat and Pumain 2007)).

The evolutionary modes of such systems are similar to those observed for other types of complex systems (Pumain and Saint-Julien 1979; Allen 1997; Pumain 2006). The *competition between cities* for attracting people and activities in order to maintain and increase the local value (in terms of economic, human, symbolic, patrimonial… capitals) leads to many *local fluctuations* (growth, decline) that in general do not modify the *global hierarchical structure of the system* that remains much more stable (Robson 1973; Pumain 1982). However, that structure is progressively transformed over time, especially through a process of hierarchical diffusion of innovations (concerning for instance in the last decades the business services, that are still concentrated in the largest metropolitan areas), and through cycles of functional specialisation (examples of manufacturing, tourism…) (Pred 1977). Such processes are essentially non linear: urban growth induces a concentration in the largest cities, partly due to positive feedbacks between size and accessibility (early adoption of modern rapid transportation and communication means, such as paved roads, then rapid trains, airlines…) and partly to the capture of the benefits of the innovations (concepts of initial advantage, associated to a site or a situation). During the course of time, the intermediary steps are short-circuited, because of the more limited number of necessary stops within the rapid networks that hampers the chances of development of smaller towns and systematically weakens them (Bretagnolle 1999). Moreover, the urban dynamics, when expressed in terms of demographic or economic growth, are likely to show reversals: unlike in the economic product cycle, there is never a total substitutability in the case of cities, but on the contrary a possible reuse of some old urban locations, that may have been momentarily abandoned but become attractive again in the course of a new economic cycle (Fig. 36.1). A well known example among many others is the city of Montpellier in France that was a brilliant university town in the eighteenth century, but completely ignored by the industrial revolutions of the nineteenth and twentieth, and had a spectacular revival after 1950.

Zelinski (1971) suggested to call "urban transition", an expression analogous to "demographic transition", that can be interpreted as a phase transition in physics, the transformation of the world pattern of settlement, from a mainly rural habitat, constituted of relatively homogeneous and dispersed small population clusters, into an urban form of habitat, made of much larger, concentrated and differentiated nodes. According to the region of the world, the transition occurred more or less early in time (from beginning of nineteenth century until about 1950) and at different paces. But everywhere it was accompanied by an intensification of the communication between towns and cities, by multiplying the networks that were connecting them, especially through the gains in speed and capacity provided by new means of transportation and communication. Meanwhile, the transition everywhere lead to huge increases in city sizes (while in 1800 there were very few urban agglomerations approaching one million inhabitants in the world, there are today about 400 of them and about 40 are above 10 million inhabitants) and to a wider concentration of population within cities (half of the world population is located in urban areas in the

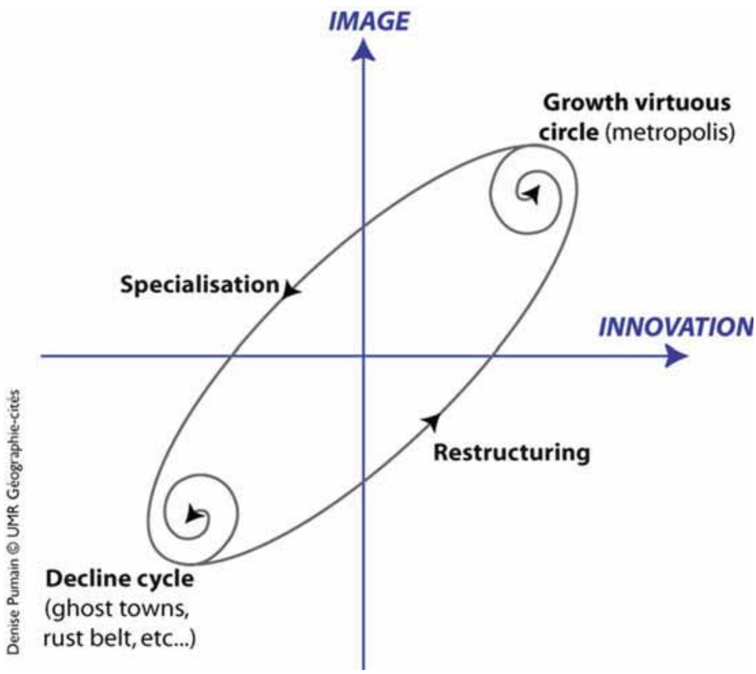


Fig. 36.1 Urban trajectories within innovation flow and relative positions in the system of cities

2000s, 80% in developed countries – it was only 20% two centuries ago). Moreover, the societal and economic evolution of these towns and cities is so coherent that, first within regional then national frames (and even later at world scale), it has been suggested to consider them as forming *systems of cities*, as early as in 1841 (Jean Reynaud, quoted by Robic 1982). The emerging properties of such systems can be interpreted as resulting from the interactions between cities (Pumain 1997).

These stylised facts were tested in the SIMPOP1 model, which was the first application of MAS in geography (Bura et al. 1996; Sanders et al. 1997), developed with the help of Jacques Ferber's (1995) research group. The objective was to identify the conditions of emergence of a functionally differentiated system of towns and cities from an initial much more homogeneous rural settlement system, over duration of some 2,000 years. This model differs from most of MAS in two aspects: its agents are immobile, as they represent places in geographical space, even if their interactions are affected by the technical innovations in the communication systems and the increasing level of urban resources which modify the relative situation of elementary cities in the network of their relational space. Second, as SIMPOP is a model of interactions between territories (aggregated geographical objects), the "behaviour" of such agents is not reducible to the behaviour of individual persons, for instance as in models of cognitive economy. SIMPOP simulates the emergence, structuring and evolution of a system of cities, starting from an initial spatial distribution of settlements in a large region, state or set of states, a repartition of resources which can be assessed randomly or exogenously, and rules defining how cities interact, grow and specialise in urban functions that appear exogenously. In this model, the environment is represented by cells having different kinds of resources, for agricultural production or industrial purposes (those last ones can be exploited from year 1800 only), and various facilities or obstacles for circulation. They can be allocated randomly or according to a specific pattern. Towns emerge as centres of accumulation of population and wealth, through first the trading of agricultural surplus with their surrounding cells, then from their competition for the acquisition of other urban functions, as other types of trades, or administrative roles, or manufacturing activities, within broader networks. Interurban competition is simulated by relating profits (from trade or taxes) and growth rates, weighted by a random factor. Meanwhile, the spatial range of interactions is increased when cities acquire new functions and with technological progress going. As a result, different patterns of towns and cities in terms of spatial and hierarchical distribution emerge.

Three main results were established from this first experiment: (1) interactions among settlements can indeed produce the emergence of a system of cities, even when starting from a pattern where settlements are all of same size; (2) if there are no interactions in the system, an urban hierarchy cannot emerge; (3) the urban hierarchy requires a recurrent appearance (creation) of new urban functions for being maintained over time as a major structural feature of the system.

The computing power available at the time restricted the number of interacting geographical entities to 400 cities. With SIMPOP2 (prepared in a collaboration with Alexis Drogoul's research group (Glisse 2007)) we have been able to extend the capacity of the model to about 5,000 cities, enabling simulation of the European cities from the Middle Ages to the year 2000 (Pumain et al. 2009); these cities are differentiated by way of some ten main functions corresponding to the main waves of socio-economic innovation during the period. This generic model was then used in an adapted form to simulate the evolution of settlement in the USA, thus enabling identification of the particular features of the urban dynamics of this New World country (Bretagnolle and Pumain 2010) by confronting results of the simulations with the historical database reconstructed by Anne Bretagnolle from census data (Bretagnolle et al. 2008).

36.3 Innovation and Interurban Competition as Driving Processes

At the heart of urban dynamics is the very strong linkage between innovation and urban development (Lane et al. 2009). Innovation has here to be understood in a very broad sense, including not only production processes or technological novelty but also societal evolutions that may be of cultural or symbolic nature and therefore influence urban change. Contrary to the modern theories of local development (that are rightly criticized for instance by Shearmur 2010), we believe that this developmental process is not so much embedded inside localities – although deeply rooted in local societal systems once installed- but on the contrary highly dependent for its appearance and evolution on interactions between urban places. We have investigated this process by observing the quantities of economic activities in urban systems and using scaling laws as a tool for detecting their behaviour regarding urban population size at different stages of their evolution within urban hierarchies (Paulus 2004; Pumain et al. 2006b; Pumain 2011b). As scaling laws are a tool for identifying constraints on the development of complex systems, there is a rather strong and general evidence that the already well documented process of hierarchical diffusion of innovation (as identified by Hägerstrand 1952, modelled by Morrill (1965) as early as 1965 and revisited by Pred including more complex network processes in 1977) is a major generic feature in the dynamics of urban systems.

Innovations are propagated in the urban system by a variety of exchanges including information. Interactions between cities keep over time some permanent features, among them the most important is their competition for adopting social change and capturing the benefits from innovation. In the SIMPOP models, a city participates to this interurban competition through the functions (mainly defined by economic specialisation) that it successively acquires over time. A function enables a city to supply a type of product or service to other cities, which provide more or less returns in terms of economic growth and attractiveness on population, according to the level of productivity of that function. The criteria for establishing a list of relevant specialisations for defining urban functions are related to an evolutionary perspective, under the main hypothesis that the relative dynamics of an urban entity in the system of cities is depending on the innovation cycles that the city has adopted (or to which it has better adapted). The question is to identify, for the entire system of cities, which innovation cycles have produced noticeable urban specialisations, affecting in a durable way the relative evolution of the specialised cities, by determining specific urban growth trajectories. For the SIMPOP1 and SIMPOP2 models, a limited number of urban functions (about 10) were selected as representative of the major economic cycles which gave rise to differential urban growth and cities specialisation over the past four centuries (Bretagnolle and Pumain 2010). A few more functions were added in the EUROSIM version of SIMPOP designed for modelling the evolution of European cities between 1950 and 2050 (Sanders et al. 2007).[2] Cities as agents have a total or partial (as constrained by the network of their partner cities) information about the emergence of new functions (that remain exogenous to the model). Cities also have a power to decide whether or not to invest in a given innovation, according to the wealth they have previously accumulated and to their line of urban strategy, that can be more or less risk-oriented. This decision process is represented by a "cognitive" attribute named "urban governance"[3] The urban governance also may represent in the model the possible intervention of the individual actors, which represent a third level in the modelling of urban systems.

[2] This project was supported by the EU Programme called TiGreSS directed by N. Winter.

[3] Although implemented in the SIMPOP2 and EUROSIM models, the role of this political entity in the dynamics has not yet been fully explored. It may reveal essential for modelling the recent dynamics of urban systems in countries such as Russia or China.

This level can be lower than the city level (for instance, an investor choosing a specific location for a firm, or a mayor defining a type of urban policy) or above the system of cities (for instance a multinational firm choosing a place for investment, or a political system imposing a centralised administration that can lead to the emergence of a prominent capital in an urban system (for example, that of France), whereas a more decentralised government may lead to a more regular urban hierarchy (for example, that of Germany)).

As we simulate the development of urban systems which include a large number of towns and cities, it would be unrealistic to think of implementing their interactions in the model as "real" flows of exchanged goods, people or information. The interactions which are simulated in SIMPOP models are not these "first order" interurban exchanges, but more abstract, "second order" interactions, which represent an interpretation of the effect of these concrete flows on the relative dynamics of cities (Pumain et al. 2006a). The urban functions that are essential attributes of the cities do not give an exhaustive description of their economic profiles, since they are attributed only to the cities having developed a major specialisation in a particular sector of activity during the corresponding innovation cycle. In a similar way, the exchange of products and resources among cities on the "market place" (cities selling and buying according to their level of supply and demand) does not reflect the totality of the urban economy but only the specialised part of the interurban market, the one that is likely to give rise to urban growth differentials. The rules which define the ability of a city to adopt innovations (i.e. new functions) are partly deterministic, in order to reproduce the powerful trend to hierarchical diffusion of urban innovation (this is the case for most of central functions, a given level cannot be acquired if the other are not yet there), and partly random: when new urban specialisations appear, they can select locations (or become acquired according to some decision of urban governance) which do not necessarily correspond to the largest cities. There are sometimes necessary "seeds" for such location of specialised activities, as mineral resources for manufacturing industries of the nineteenth century, or knowledge (human capital) in the case of technopoles of the twentieth. Such specification about the "environment" of the urban systems was implemented when calibrating SIMPOP2 on the European and USA cases (Bretagnolle et al. 2010).

36.4 Urban Dynamics, History and Evolution

The series of the SIMPOP models is called "evolutionary" because a major underlying hypothesis is that the pervasive structural features of urban systems that we can observe are produced both by a generic dynamics (reproducible, or repeated processes) and an historical evolution (specific trends), over long periods of time. This evolution involves systematic, time-oriented changes in major circumstances of the system over time, including the demographic and urban transitions, the increase in gross and per capita economic wealth, the trendy increase in the speed of transportation means, as well as the recurrent appearance of technical, economic and cultural innovation. Thus, it is this social, historical evolution which supports the dynamics of urban systems, even if in a concrete way the dynamics is made through the mechanism of interurban interactions. Those are the "bottom-up" processes leading to the emergence of the structure of the system, whereas the evolutionary trends can be thought of as emerging trends, which are produced as feed-back effects by the system of cities itself, and become new constraints on the dynamics of individual cities. Actually, we do not know yet how to make these large evolutionary trends emerge (as sets of qualitative and quantitative changes), and they are represented in an exogenous way within our models, whereas it is possible to represent the endogenous process of building an urban hierarchy from the interactions between cities.

Another peculiarity of our simulation experiments is thus that they are not exactly "historically driven", but at least widely guided by the general historical trends that we do not intend to generate through the model. These historical trends are useful in two ways during the course of modelling: first, they help to rightly specify the values of some of the parameters and the syntax of rules during the stage of conception of the model (these parameters and rules help to maintain the simulations in a range of plausibility, they help in reducing the number of parameters whose values are unknown or not measurable and have to be estimated from the simulations). Secondly, a few historical series are used as benchmarks for checking the ability of the model to reproduce the temporal trajectories of different variables as urban population and wealth, share of employment in different types of urban functions, or distribution of city sizes. This calibration exercise is conducted as parsimoniously as possible in order to keep a large generic dynamic core in each application. Only modifications that are necessary for retracing the general evolution of the urban system (urban population and distribution of city sizes) are allowed when transferring the model from one system to the next (Bretagnolle et al. 2010).

These experiments are used for testing major hypothesis about the constraints and evolutionary paths that have differentiated the geographical structure of urban settlement systems all over the world. Conceiving generic dynamics and specific histories for reconstructing a variety of observed evolutions is thought of as a way of testing our urban theories through comparative analysis. This method relies not only on the design of simulation models, but as well on large historical urban data bases that are built according to precise and harmonised rules (Bretagnolle et al. 2008). The main problem is to define urban areas that despite their expanding and often fuzzy spatial limits keep a coherent geographical significance over long periods of time and for a large variety of political, economic and societal contexts.

36.5 MAS as Adapted Simulation Tools for Geographical Dynamics

When compared to our previous attempts at modelling the dynamics of urban systems using systems of differential equations (Pumain et al. 1989; Sanders 1992), MAS models have introduced a broader flexibility in the model conception, both for describing a variety of spatial interactions and including interactions from different levels. However, in order to fully exploit their potential from what we discovered in the nineties, we had to wait for the recent increases in technical computing capacities for developing applications of large systems of cities, at national or continental scales.

36.5.1 Simulating a Variety of Spatial Interactions

Our model is a geographic model, in the sense that spatial interaction are supposed to reflect the power of cities in terms of range of influence of their activities and support for new developments from their access to more or less extended markets. Three types of spatial interactions are distinguished for reflecting the most frequent types of interurban exchanges, linked to different constraints: (1) proximity constrained interactions are representative of many activities for which the distance between supply and demand is an essential constraint, they are the rule for all central place functions, whatever their level and range, and even if that spatial range is increasing over time; under that rule, the probability of exchanges are distributed according to an interaction model of gravity type. (2) territorially constrained interactions are limiting a city's influence within boundaries, regional of national, they correspond to all types of administrative or political activities; the interaction rule is modulated according to the nature of the activity, for instance, a capital can levy taxes in an exhaustive way on all cities belonging to its region or state, whereas in the case of other activities this rule can attribute only a preference for a territorial market. (3) interactions within specialised networks are free from distance constraints, even if exploring them for developing new markets along this line may have a differential cost according to the distance. Long distance trade, maritime transport, part of tourism and manufacturing industry follow this type of spatial interaction rule. From these modulated rules of spatial interaction, a variety of patterns of exchange networks are generated during the course of simulation by the model.

36.5.2 Mixing Entities of Different Levels

The paradigm of contemporary complex systems theory focuses on the emergence of properties at a macro level as resulting from micro-level behaviours. Most of the applications, for instance in cognitive economics or social networks, refer to a "social ontology" that includes only two levels: the micro level of individuals whose interactions define an emerging structure or configuration at a macro level. These models do not include entities of different levels that could also have an important role in the structuring of social systems, as for instance institutions. Most of the time their intervention remains as a black hole for instance in economic modelling. There are few exceptions, such as the EURACE model developed for simulating divergences or convergences in the economic development among European regions (Deissenberg et al. 2008). Actually there is no reason for limiting the use of MAS to simulating micro-macro interactions, as this type of modelling enables including entities of different levels and scope that are characterised in the same way as the "agents", with specific attributes, behavioural rules and information protocols. For instance, in the more recent versions of the SIMPOP model, the urban functions are designed that way, and in EUROSIM cities were related not only to their country of location but to entities of higher levels defining "blocks" with specific parameters and rules for demographic and economic evolutions. These implementations can represent steps toward multilevel modelling, that was experimented in the conception of another model named SIMPOPNano that was designed by Thomas Louail (2010).[4] In this model we made explicit the possibility of interactions between the location of urban functions inside the space of one city and the role played by the city in the system of cities. The model was intended to test the hypothesis above, by comparing the connexions between intra-urban spatial structures and urban systems dynamics in stylised large metropolitan areas of European or North-American type.

36.5.3 Developing Multi-scale Validation Methods

Different granularity of space and time in the analysis of simulation results can be of great help when trying to assess the validity of such simulation models. Each simulation produces large quantities of data and figures that can be analysed in order to check if the simulated dynamics are plausible when compared to observations made at different geographical levels and for different historical periods of time. Helène Mathian (Mathian and Sanders 2009) contributed in a decisive way to improve our ability to evaluate the quality of different simulations by implementing analytical tools for exploring and comparing the results. The principle is to observe the ability of the simulation to reconstruct the urbanisation process at several levels of observation. For instance, at the most aggregated level, curves of the evolution of total population and wealth of cities (for the system under consideration) are compared to observed curves of their development over the same duration. At that level, rank-size distributions of the system of cities are established and the evolution of their slopes over time is compared with observations. At a meso-geographical level, comparisons are made of the mean size and evolution of number of cities by classes of size of by types of urban functions. In the case of Europe, the plausibility of urban patterns obtained within large groups of countries (in three blocks, North-western, Southern and Eastern Europe) was also examined for assessing the quality of simulation at this meso-level. Eventually, the individual trajectories of cities were classified and analysed to check if their relative frequencies, type of curve and dispersion indices were matching the stylised observations, at this micro level of individual cities evolution.

Although constructed under the principle of parsimony, the SIMPOP models as most of MAS models can be blamed for being "complex" and perhaps "complicated"

[4] We acknowledge the support of Thomas Louail 's PhD by the R2D2 DIM of Region Ile-de-France.

models (Pumain 1998). To avoid classifying them as purely "ad hoc" models that would replicate specific urban evolutions by excessively reducing the degree of freedom in the variation of city sizes and functions we adopted several compromises between designing a generic model and simulating realistic ones. For instance, we can start from initial conditions were the locations of cities are totally random, or regular, and compare the simulation results obtained with the same rules and parameters when introducing the map of geographical locations. This method could be a way of measuring the impact of previous historical interactions on the spatial structure of urban systems (Bretagnolle and Pumain 2010). But in our view the best way for validating our approach of urban dynamics is to design a series of models that can be applied to different periods in time and parts of the world and check which rules and parameters have to be adapted for satisfying the change in the context and environment of the model. This could be named a rule of conceptual consistency, if the necessary changes actually reflect well what is already known about the historical and geographical conditions prevailing in a particular stage of evolution of each settlement system. Of course, we also have to develop in the future new methods that are now under experiment for improving the model testing.

36.6 A New Series of SIMPOP Models as Virtual Laboratory

Over the last 15 years, a series of models were elaborated in the laboratory Géographie-cités (SIMPOP 1 and 2, EUROSIM, SIMPOPNano, SIMPOPlocal, SimpopNet etc.…), to compose one overall view of systems of cities. This series will grow with the elaboration of several new models designed for testing complementary hypotheses about urban dynamics. To be preserved, it must be integrated into a formal framework of exploitation, which will help to re-use and share models: standardization of in/out of models, traceability of realized experiences and evaluations, coupling with communitarian platforms.

36.6.1 Enlarging Urban Dynamics Towards Urban Evolution

Urbanisation dynamics used to be strongly dependent on local environmental constraints at the time of the emergence of the first cities and during the following centuries. At that time, natural disasters or even smaller recurrent difficulties as climatic variations or epidemics could completely destroy the population and resources of human settlements as well as predations from other groups. A simplified SIMPOP model called SIMPOPlocal is now designed using Netlogo language by Clara Schmitt[5] and Sébastien Rey with the help of Arnaud Banos for simulating

[5] We acknowledge the supports of ADEME for Clara Schmitt's PhD and of R2D2 DIM Région Ile-de-France for Sébastien Rey.

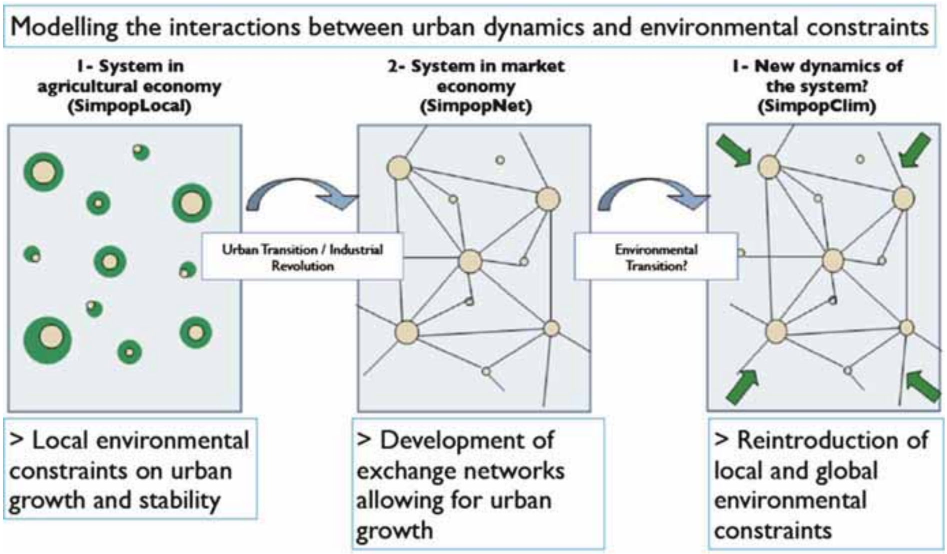


Fig. 36.2 Three versions of SIMPOP for modelling the changing relationship of systems of cities to the environmental constraints (Source: Clara Schmitt 2011, PhD thesis (University Paris I and ADEME))

this type of dynamics. Once developed, this model could afterwards be tested by archaeologists on specific environments. It is intended to be generic for the period of emergence of urban systems when they were still highly dependent on the local amenities and uncertainties of their environment (Diamond 1997) at a time where spatial interactions had a much less important impact on settlement systems than the local ecological constraints. These local constraints were then overcome through the development of long range exchanges between cities. Cities became interdependent and were co-evolving because all-connected by communication networks. The impact of local resource limitation was considerably reduced during the industrial revolution and the contemporary urban transition, during which the development of exchange networks between cities allowed them to import what they lacked and export surplus thus initiating a seemingly unlimited urban growth. Another simplified model called SIMPOPNet will be designed for representing the evolution of urban systems during that period of quasi-autonomy in their dynamics (Fig. 36.2). Today, environmental constraints on city development do reappear but in a top-down way, through the spiralling global awareness of climate change, energetic resources limits, biodiversity loses, air and water quality, etc. Cities could now find their growth trajectories re-affected by environmental limits. Because the adaptation to the new constraint is similar to an innovation diffusion process following urban hierarchy and specific networks, this evolution can be modelled in our agent-based simulation model SIMPOPClim (Fig. 36.2).

36.6.2 Endogeneisation of the Appearance of Innovation in the SIMPOPlocal Model

There are decades and centuries of exchanges related to socio-economic innovations. In the first SIMPOP models, two processes of emergence are represented: one is bottom-up and the other top-down. Indeed, the level of the system of cities influence the local level of each city as a constraint (by reducing the space of its possible trajectories) while it also acts as an activator of urban development (through interurban emulation). Only a few transformations of the generic model are necessary for its application to different systems of cities. They reveal for instance how the historical and geographical context influence the emergence of specific properties in structuring urban systems in ancient or more recent settlement systems or in developing countries (application of SIMPOP2 to South Africa by Céline Vacchiani-Marcuzzo with the help of Thomas Louail). Nevertheless, the actual societal innovation remains exogenous to the model as represented by new urban functions that are introduced over time during the simulation. Thus, only weak emergences can be simulated through this type of modelling, that is, for processes and properties that are already identified. Strong emergence, producing social innovation, as new artefacts or technologies, ideas or social practices is not yet well understood enough for any predictive or prospective simulation – unless a high level of abstraction is accepted, but this operation implies the emptiness of the applications in terms of societal significance.

The appearance of innovation is made endogenous in the SIMPOPlocal model. This is possible because very little information is available about effective interaction processes due to innovation during the early times of emergence of urban settlement systems. A very general process is thus imagined for producing innovation from probability of encounter between populations, inside and outside each settlement. The impact of innovation cannot be detailed according to differentiated urban functions but is converted in a more productive rate of exploitation of local resources allowing for settlement growth, whatever the concrete type and content of innovations (that can be related to any aspect of social and economic life of the time). The frequency of appearance of this abstracted innovation is thus estimated in this model from a generic historical pattern of the evolution of urban population. This model is an attempt for quantifying the pace of emergence of innovation and their impact on the development of urban systems, in an evolutionary mode where the availability of local resources impose a major constraint on the growth of each settlement (Schmitt 2011).

36.6.3 Scenarios for Exploring the Future

Studying the dynamical properties of urban systems in a variety of historical and geographical contexts may bring an important contribution to the issue of the future or urbanisation: on the long run, one may wonder if, after the end of the urban transition, settlements will continue to concentrate and to differentiate, or if a bifurcation toward a different form of organising human habitat will occur. On a shorter period of time, a better knowledge of the dynamics of urban systems can help to predict the effects of the increasing connections linking towns and cities, either from political decisions (as following the European integration process) or, more generally, due to the diversity of linkages created through the globalisation process. The main question then is about how cities are going to redefine their relative positions within enlarged networks and hierarchies. On the very short term, many questions remain debated, as the possible effects of the new means of communication and trade on urban systems: are the new information and communications technologies (NICTs) going to overturn the existing urban hierarchies, or will they be progressively integrated into these systems without changing much their relative configuration, as this happened in the past for previous innovations (telegraph, railway, telephone, automobile…). Our hypothesis is that a generic model that would succeed in simulating on the long term the transformation of urban systems, respective to their capacity of reaction and adaptation to the changing conditions that allow for and constrain communication and trade, would be a great help in understanding the past co-evolution of cities and predicting the future.

36.6.4 A Generic Simulation Platform

A key issue in relying with confidence upon MAS models for sustaining scientific predictions is to improve our capacity of understanding the complex dynamics they simulate by screening the massive data bases that they generate during a simulation (Banos 2009). As complex systems more generally include many non linear relationships, MAS models have differentiated properties of sensitivity and robustness regarding the parameters and rules that they activate. The classical method for running simulations estimating parameters value by trial and error is no longer valid when trying to calibrate models that have huge numbers of possible evolutionary paths. New calibration methods automatically generating simulations and varying parameters in a systematic way within a given range and under specific protocols of control have to be experimented. Such scientific workflow management systems are now available, and will inspire the development of our MAS modelling platform. It integrates the model as a generic module, loosely coupled but easily exchangeable, with a series of module-tools, organized around it, and resulting from the community or to the intern development of the laboratory.[6]

This platform includes at least four types of components: one for describing the type of experiment by choosing a set of parameters values and planning their variation for running series of simulation; another one is dedicated to the distribution of modelling tasks on a computer grid for reducing the duration of computation and

[6] We acknowledge the support of ANR Transmondyn and ERC grant GeodiverCity for various developments of this platform.

entailing the necessity of replication of simulations in large quantities to obtain good evaluation of mechanisms; a third one consists in a variety of exploitation tools for analysing the results of each simulation and comparing them (it is here that the multi-scale validation methods are implemented). Eventually, a communication tool for sharing the information about the simulations among the pluri-disciplinary research group (in which each one intervenes with own skills, lexical fields, formalisms and methodologies) is essential for establishing a memory of the experiments and contributing to the scientific value of the exercise (Rey 2011).

36.7 Conclusion

As with all pioneer works, the fascinating adventure of the SIMPOP series of models required a lot of efforts, from a multi-disciplinary community of researchers. When trying to solve theoretical questions through MAS modelling, a major challenge is to practice a right selection for defining a tractable granularity in the multiple dimensions of space, time, attributes and processes that specify the model. This adjustment between modelling capabilities and pending urban mysteries or gaps in theoretical construction is made through negotiation (conceptual as well as procedural) between geographical scientists and computer scientists. The exercise is thus highly constrained by the state of the art in both disciplines, including as well purely technical considerations like computing power or data availability. It is thus not surprising if at that stage we prefer not to conclude, but mentioning our intuition that the efforts are worthwhile!

References

Allen, P. (1997). *Cities and regions as self-organising systems*. Amsterdam: Gordon and Breach.

Banos, A. (2009). Simulating pedestrian behavior in complex and dynamic environments: An agent-based perspective. In F. Bavaud & C. Mager (Eds.), *European handbook of theoretical and quantitative geography* (pp. 1–27). Lausanne: FGSE.

Batty, M. (2008). Fifty years of urban modelling: Macrostatics to micro dynamics. In S. Albeverio, D. Andrey, P. Giordano, & A. Vancheri (Eds.), *The dynamics of complex urban systems* (pp. 1–20). Heidelberg: Physica Verlag.

Batty, M., & Torrens, P. (2001). Modelling complexity. The limits to prediction. *Cybergeo, 201*, 1–24.

Berry, B. J. L. (1964). Cities as systems within systems of cities. *Papers of the Regional Science Association*, *13*, 147–163.

Bretagnolle, A. (1999). Les systèmes de villes dans l'espace-temps : effets de l'accroissement de la vitesse des déplacements sur la taille et l'espacement des villes. Thèse de doctorat de l'Université Paris I.

Bretagnolle, A., & Pumain, D. (2010). Simulating urban networks through multiscalar space-time dynamics (Europe and United States, 17th -20th centuries). *Urban Studies, 47*(13), 2819–2839.

Bretagnolle, A., Giraud T., & Mathian H. (2008). La mesure de l'urbanisation des États-Unis, des premiers compto.irs coloniaux aux *Metropolitan Statistical Areas* (1790–2000). *Cybergeo, 427*, 1–36.

Bura, S., Guérin-Pace, F., Mathian, H., Pumain, D., & Sanders, L. (1996). Multi-agent systems and the dynamics of a settlement system. *Geographical Analysis, 28*(2), 161–178.

Deissenberg, C., van der Hoog, S., & Dawid, H. (2008). Eurace: A massively parallel agent-based model of the European economy. *Applied Mathematics and Computation, 204*(2), 541–552.

Diamond, J. M. (1997). *Guns, germs and steel: The fates of human societies*. W.W. Norton & Co.

Ferber, J. (1995). Les systèmes multi-agents. Vers une intelligence collective. Paris: Inter Editions.

Glisse, B. (2007). Exploration scientifique par la conception collaborative de systèmes multi-agents. Thèse d'informatique de l'Université Paris 6.

Gulden, T. R., & Hammond, R. A. (2012). Beyond Zipf: An agent-based understanding of city size distributions. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 677–704). Dordrecht: Springer.

Hägerstrand, T. (1952). *The propagation of innovation waves*. Lund: The Royal University of Lund.

Lane, D., Pumain, D., van der Leeuw, S., & West, G. (Eds.). (2009). *Complexity perspectives on innovation and social change* (Methodos series, Vol. 7). Berlin: ISCOM, Springer.

Louail, T. (2010). Comparer les morphogenèses urbaines en Europe et aux Etats-Unis par la simulation à base d'agents – Approches multi-niveaux et environnements de simulation spatiale. Université Paris 6, thèse de doctorat.

Mathian, H., & Sanders, L. (2009). "Jeux d'échelles dans les pratiques de calibrage et de validation des modèles de simulation en géographie ", atelier scientifique du CNRS, "Modélisation des phénomènes sociaux et spatiaux: évaluation et validation des modèles de simulation individu-centrés", Berder 12–14 mai. http://s4.csregistry.org/tiki-index.php?page=Les+contributions

Morrill, R. L. (1965). *Migration and the spread and growth of urban settlement* (Lund studies in geography, series B, Vol. 26). Lund: Royal University of Lund, Department of Geography.

Ngo, T. A., & See, L. M. (2012). Calibration and validation of agent-based models of land cover change. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 181–196). Dordrecht: Springer.

Paulus, F. (2004). Coévolution dans les systèmes de villes : croissance et spécialisation des aires urbaines françaises de 1950 à 2000. Université Paris I, thèse de doctorat.

Pred, A. (1977). *City systems in advanced societies*. London: Hutchison.

Pumain, D. (1982). *La dynamique des villes*. Paris: Economica.

Pumain, D. (1997). Vers une théorie évolutive des villes. *L'Espace Géographique, 2*, 119–134.

Pumain, D. (1998). Urban research and complexity. In C. S. Bertuglia, G. Bianchi, & A. Mela (Eds.), *The city and its sciences* (pp. 323–362). Heidelberg: Physica Verlag.

Pumain, D. (2000). Settlement systems in the evolution. *Geografiska Annaler (Series B), 82*(2), 73–87.

Pumain, D. (2006). Alternative explanations of hierarchical differentiation in urban systems. In D. Pumain (Ed.), *Hierarchy in natural and social sciences* (Methodos series, Vol. 3, pp. 169–222). Dordrecht: Springer.

Pumain, D. (2011a). Le contexte en sciences sociales et dans les modèles de simulation en géographie. In J. P. Muller (Ed.), *Le contexte dans les modèles de systèmes complexes*. Paris: Chemins de Traverse. forthcoming.

Pumain, D. (2011b). Urban systems dynamics, urban growth and scaling laws: The question of ergodicity. In J. Portugali, H. Meyer, E. Stolk, & E. Tan (Eds.), *Complexity theories of cities have come of age: An overview with implications to urban planning and design*. Heidelberg/Berlin: Springer.

Pumain, D., & Saint-Julien, T. (1979). Les transformations récentes du système urbain français. *L'Espace Géographique, 3*, 203–211.

Pumain, D., Sanders, L., & Saint-Julien, T. (1989). *Villes et auto-organisation*. Paris: Economica.

Pumain, D., Bretagnolle, A., & Glisse B. (2006a). Modelling the future of cities. In *ECCS'06, Proceedings of the European Conference of Complex systems*, University of Oxford, 25–29 Sept 2006. http://halshs.archives-ouvertes.fr/halshs-00145925/en/

Pumain, D., Paulus, F., Vacchiani, C., & Lobo, J. (2006b). An evolutionary theory for interpreting urban scaling laws. *Cybergeo, 343*, 1–20.

Pumain, D., Sanders, L., Bretagnolle, A., Glisse, B., & Mathian, H. (2009). The future of urban systems. In D. Lane, D. Pumain, S. Van der Leeuw, & G. West (Eds.), *Complexity perspectives on innovation and social change* (Methodos series). Berlin: ISCOM, Springer. chapter 12.

Rey, S. (2011). *A grid and GIS-based simulation platform for computational geography*. PhD thesis, University Paris I.

Robic, M. C. (1982). Cent ans avant Christaller, une théorie des lieux centraux. *L'Espace Géographique*, 1, 5–12.

Robson, B. (1973). *Urban growth, an approach*. London: Methuen.

Rozenblat, C. (2010). Opening the black box of agglomeration economies for measuring cities competitiveness through international firm networks. *Urban Studies, 47*(13), 2841–2865.

Rozenblat, C., & Pumain, D. (2007). Firm linkages, innovation and the evolution of urban systems. In P. J. Taylor, B. Derruder, P. Saey, & F. Witlox (Eds.), *Cities in globalisation* (pp. 130–156). London: Routledge.

Sanders, L. (1992). *Systèmes de villes et synergétique*. Paris: Anthropos.

Sanders, L., Pumain, D., Mathian, H., Guérin-Pace, F., & Bura, S. (1997). SIMPOP, a multi-agents system for the study of urbanism. *Environment and Planning B, 24*, 287–305.

Sanders, L., Favaro, J. M., Glisse, B., Mathian, H., & Pumain, D. (2007). Intelligence artificielle et agents collectives: le modèle Eurosim. *Cybergeo, European Journal of Geography, 392,* 1–16.

Schmitt, C. (2011). *SIMPOPClim: From local to global environmental constraint on urban evolution*. PhD thesis, University Paris I.

Shearmur, R. (2010). Innovation et développement territorial? L'innovation comme processus (presque) a-territorial. *Bulletin de la Société Géographique de Liège, 55*, 17–27.

Zelinski W. (1971). The hypothesis of the mobility transition. *Geographical review, 61*, 219–249.

Chapter 37
Reflections and Conclusions: Geographical Models to Address Grand Challenges

Alison J. Heppenstall, Andrew T. Crooks, Michael Batty, and Linda M. See

Abstract This chapter provides some general reflections on the development of ABM in terms of the applications presented in this book. We focus on the dilemma of building rich models that tend to move the field from strong to weaker styles of prediction, raising issues of validation in environments of high diversity and variability. We argue that we need to make progress on these issues while at the same time extending our models to deal with cross-cutting issues that define societal grand challenges such as climate change, energy depletion, aging, migration, security, and a host of other global issues. We pick up various pointers to how we might best use models in a policy context that have been introduced in many of the applications presented within this book and we argue that in the future, we need to develop a more robust approach to how we might use such models in policy making and planning.

A.J. Heppenstall (✉)
School of Geography, University of Leeds, Leeds, UK
e-mail: A.J.Heppenstall@leeds.ac.uk

A.T. Crooks
Krasnow Institute for Advanced Study, George Mason University, Fairfax, VA, USA
e-mail: acrooks2@gmu.edu

M. Batty
Centre for Advanced Spatial Analysis (CASA), University College, London, UK
e-mail: m.batty@ucl.ac.uk

L.M. See
International Institute of Applied Systems Analysis (IIASA), Laxenburg, Austria
e-mail: see@iiasa.ac.at

A.J. Heppenstall et al. (eds.), *Agent-Based Models of Geographical Systems*,
DOI 10.1007/978-90-481-8927-4_37,

37.1 Where Do We Stand in Modelling Geographical Systems?

The models dealt with in this book represent our most recent thinking as to how we might best represent geographical systems to describe the variety and complexity that confronts us in both our abstractions and our casual perceptions of the world around us. Agent-based modelling (ABM)[1] represents a softening of the rather parsimonious, aggregative, largely static models that we began with over 50 years ago. As we have argued throughout this book, this is because of what we have learnt about such systems but also because these systems have become more complex, and our models have been reduced in their severity. They are now richer and more descriptive, thus probably more informative but less predictive. Yet what we have been at pains to portray is not that ABM or cellular automata (CA) or microsimulation models substitute for any of those that have gone before, for they complement our knowledge. The array of modelling approaches now available provides us with a much greater menu of ideas, which we might apply to critical spatial problems. Of course, some of the applications in the past can now be done much more effectively with newer styles of theory and model, but all the approaches that have been developed build on this past, and in the kind of pluralistic world that we have implicitly adopted as anchoring our field, we consider all these approaches to have value.

One of the central issues in all of science is dramatically illustrated by our experience of ABM. It has long been regarded that good science can only be generated if theory is tested under controlled conditions, in laboratory contexts where extraneous events are excluded or at least accounted for in some definite way. As we have learnt more about the world, it has become increasingly clear that science has two faces: one where strong theory can be generated and tested in the classical tradition and one where such strong theory breaks down in more open applications. It is the latter that dominates our quest to apply scientific principles to more open problems and we thus face a dilemma. Some rather good and obvious examples are widely known. For example, weather forecasting, which is based on strong classical theory, is generally of weak predictability. When such theory and their models are put into the real world, such extensive variability based on extraneous unpredictable forces is simply unknown and this destroys strong predictions. Although we may be able to test and confirm or falsify the sorts of strong theory that underpin the hydrodynamics of controlled atmospheres, when it comes to making strong predictions in the wider context of the world's weather, this is simply not possible. In fact in our own field for geographical systems such as cities and regions, strong theory is even less likely to be possible because it is almost impossible to set up controlled human experimentation, and once what theory there is, is put to the test, the whole basis of any such knowledge that is culled from experiment, changes.

It is still an open question as to whether or not there are intrinsic differences between science and social science. Traditionally these have been assumed to relate to animate and inanimate matter – that physical systems are not able to manufacture

[1] ABM is also taken to mean Agent-Based Model (s) as well as Modelling.

their own destinies whereas human systems have that potential. But notwithstanding this point, which may not be resolvable in any case in the context of the variability that models of geographical systems need to work with, we must accept that strong prediction is not likely, nor is it appropriate. Thus the role of models is intrinsically different from that which was assumed 50 years ago. Models must be to inform rather than predict or if to predict, then these are entirely conditional on the context in which such prediction is made. To an extent, this is encapsulated in the notion of 'what if' experiments and scenarios, i.e. conditionals that are context dependent.

So far we have not explored this dilemma in detail for it is compounded by the fact that our systems of study are getting more complex as they evolve, and thus our theories and models tend to lag behind the systems that we observe. Moreover, our models also tend to change the very system that they attempt to simulate as we learn more about the world and apply more considered actions to its problems. In some respects, as soon as models came to be applied to geographical systems, their designers and users became quickly aware how problematic was the issue of validating them and using them to generate robust predictions. The very act of constructing the model soon became a focus of such efforts, imposing on the discipline what were usually inchoate and ill-defined problem contexts. The ideology of modelling quickly moved from notions about prediction to information, from making forecasts to providing informed advice concerning the problem in context. Despite the longevity of these notions, little has been done on pursuing the logic of using models in this way, for science in general and policy analysis in particular has been stymied by the dilemma of getting to grips with the uncertainty of scientific knowledge in the human domain. What has been done, and this is very clear from the various contributions in this book, is that as models have become richer to deal more effectively with the systems which they are designed to represent, the process of calibration, estimation, or validation against external, purportedly independent data has become more involved. For example, Ngo and See (2012) develop considerable detail with respect to model validation which is considerably more elaborate in its testing than that developed for more aggregate parsimonious models. Verification has come onto the agenda largely because as models have got bigger, it has become clear that simply defining a model and then assuming it runs as specified, generates considerable uncertainty. Indeed visualization of model data and outputs, even the processes of simulation, is essential because of errors that creep in during the process of assembling data, encoding algebra, and operating such models in complex computational environments.

One of our conclusions is that we must engage in a more vibrant and wide ranging discussion of formal theory and models of geographical and other social systems in terms of their validation. We need to continue to explore the limits of how far we can expect our models to replicate reality, to examine the conditions that we observe in the past and expect in the future with respect to what our models are able to say about different circumstances. We need to think much more out of the box about complexity, validation and the world that we wish to influence, following the exemplar of van der Leeuw (2004) who in his discussion of modelling the ancient past raises a series of dilemmas that are clearly different from attempts to simulate the contemporary past and the present. In the social sciences, there has always been an uncomfortable tension between developing theory for its own sake – curiosity driven knowledge – in contrast to theory that is practically inspired, particularly for circumstances where we ourselves are involved in making proposals and plans for designing the future, for resolving social problems and for engendering a better quality of life for all. We need to grasp this nettle more directly, and only by doing so, will we be able to reconcile the key issues involving the use of the new generation of models introduced in this book. It is highly likely that the future will be dominated by a plurality of model types and styles, and to confront this world, we will need a much clearer sense of where and when to use what particular model (Epstein 2008). This plurality has been anticipated and there is already some rudimentary research into how two or more models might be compared (Axtell et al. 1996). The idea that we build more than one model for any and every situation is growing. It was suggested many years ago by Greenberger et al. (1976) as counter-modelling but it has taken a long time in coming. It is a challenge that will underpin all others.

37.2 The Need to Address the Grand Challenges

Most of the applications of ABM presented in this book involve city systems in the mainstream traditions of urban geography, regional science, urban economics, and transportation modelling. Much of the field however, has been influenced by a more rural focus, particularly with respect to land cover modelling and applications to development in developing countries. Although the concept of an agent does differ between rural and urban, and across spatial scales, common applications deal with spatial structure, form, mobility and demographics. Most ABMs to date have focused on specific sectors rather than on comprehensive representations, except where, the tradition in urban modelling, large models such as TRANSIMS have attempted to model a wider range of sectors. As we have implied, many models still lie at the level of 'proof-of-concept' with a strong pedagogic focus, and the field is ripe for more focused applications.

Because ABM tends to be a generic style of modelling, it is often used as more of a toolkit to develop models, rather than as an exclusive framework for large scale applications. This however, is changing as we become more familiar with its potential, as evidenced by the large scale epidemiological, transportation, financial markets and the human immune system models currently being attempted (Castiglione 2003). However, the grand challenges which have dominated the social and policy sciences of late are not well represented so far in terms of these new styles of model. Currently most applications to geographical systems have focused on different activity or land use sectors, dividing the city system into housing markets, education and schools, the health sector, transportation, with specific spatial behaviours relating to crime, pedestrian flow and movement, segregation, traffic flow and related processes. In a developing countries context, there has been a focus on aid and development, but the bigger issues of energy, climate change, and security have rarely been tackled. As we have noted, there has been a focus in ABM on diffusion, particularly disease but it is timely to stand back and inquire into how such modelling styles can be best used to inform the really big, pressing questions that now dominate both local and global policy.

Perhaps the most focused of any of the models presented here are those involving microsimulation where policy questions are to the fore, specifically demographics. Aging, for example, is of central concern certainly in western nations but also in China, and such secular change can have an enormous impact on the structure of geographical systems that define populations and their life cycles. The models developed here by Wu and Birkin (2012) show how some really big challenges might be addressed in terms of changing life cycles but these have not yet been linked to other sectors where the impact of changing demographic profiles is likely to have an enormous impact on the spatial consequences of such change. It is only when these types of model are integrated in some way, coupled, that the level of comprehensiveness of the simulation is much increased, and that the real effectiveness in discussing the impact of these challenges on modern urban society and its cities can be thoroughly appreciated. To an extent, this is achieved via a loose coupling of various models, which tend to be different perspectives on how a variety of tools such as microsimulation, visualization, online mapping, and database organization for city simulation are being developed in the NeISS project (see http://www.neiss.org.uk/) but the extent to which the integration is based on coupling ABMs is not strong. Integrated modelling insofar as it exists for geographical systems has so far been based much more on aggregative models as, for example, in the climate change impact assessment of flooding in the Greater London region, which was based on an integration of various macro models such as input-output, land use transportation interaction (LUTI) models, land development using CA, and hydrological flood models (Walsh et al. 2011). The domain is thus wide open for more ambitious ABM simulation frameworks which integrate different models as well as interfacing them with appropriate policy analysis and stakeholder participation (Batty 2010).

In the next section, we will sketch some of the grand challenges that are key to social and public policy at the present time and suggest how ABM and related techniques might relate to some of these applications. Our focus is more on coupling, on how we use ABM to enrich and inform the dialogue, rather than on producing highly integrated structures that attempt complete comprehension. Moreover, we also see these grand challenges being informed by many different kinds of models, thus reinforcing our sentiments that in the future, many models, models at different levels of aggregation, both static and dynamic, will define the portfolio that decision-makers and scientist alike consider appropriate to the kinds of advice that science can bring to bear on these critical questions.

37.3 Climate Change, Energy Issues, Poverty, Aging and Migration: Can ABM Make a Difference?

Models of geographical systems tend to cut across the kinds of problems that are now widely regarded to be of major societal significance. The grand challenges that have emerged over the last 20 years involve themes that cross many sectors for which models are individually built and thus integrated modelling is usually required if these challenges are to be informed by the tools and methods presented in this book. In particular, the key challenges at present involve aging and demography, urbanization and migration, energy depletion, climate change, poverty (which is a recurrent issue), health and disease, and security and conflict. These are by no means all those that are identifiable, for at more modest scales, issues involving housing, financial markets and global trade all raise profound issues that cannot be tackled sector by sector. Moreover, technology change is compounding, complicating and continually changing the nature of these problems. This poses a dilemma, for our models tend to be focused on quite manageable and identifiable sectors and activities, whereas the grand challenges are cross-cutting. Indeed, these challenges tend to be important because they cannot be handled in traditional ways, by traditional models.

The quest therefore is to develop our models so that they can address pieces of these major challenges. ABMs in fact are rather focused on behaviours central to the way many of these big problems need to be resolved. Behavioural change is often a clear and obvious solution to some of these issues and ABMs have the potential to simulate such behaviours. However, one key paradox is that the grand challenges appear to involve changes in behaviour, which represent not a continuous evolution of current patterns of behaviour but often radical shifts. How we use models which simulate current patterns of behaviour which need to change if the key issues are to be resolved poses enormous difficulties for implementing and using models that are based on current and past patterns of behaviour. For example, shifts in responses to climate change such as adaptation involves changing behaviours to reduce their impact while mitigation policies will give rise to changed behaviours as a result of new policies designed to reduce the drivers of change such as carbon emissions.

There are some areas of ABM that are well suited to developing insights into some of these global challenges. In fact the intersection between ABM and micro-simulation modelling involves representing life cycle effects that incorporate issues of aging related to geodemographics and health. There are strong policy issues involved here with respect to the provision of facilities, particularly housing and health services, for different age groups. Wu and Birkin (2012) address these issues directly in one of their contributions here but the contributions from Leeds more generally, particularly from Malleson (2012) on crime, Harland and Heppenstall (2012) on education, and Smith (2012) on health, all inform these key issues. The various contributions on city size distributions from Gulden and Hammond (2012) and Pumain (2012) do not quite address migration although the elements of such modelling are implicit in their models, but as migration is likely to be a key challenge at every spatial scale as we become more mobile, then the seeds of how we might explore this through ABM are reflected in various of the contributions which deal with spatial interaction and movement.

Issues of security and conflict have been handled extensively at the micro-spatial scale in pedestrian models useful in testing evacuation scenarios. There is a lot of expertise in dealing with crowding in confined spaces as illustrated in the articles on crowd movements by Johansson and Kretz (2012) and on traffic congestion by Banos and Genre-Grandpierre (2012). Many of these challenges are global in scope and to address these will require generalizing many of these approaches to deal with large spatial scales. There is also considerable potential in many of the models reviewed here to simulate diffusion in many media. Simoes (2012) provides a direct example of such spatial simulations by modelling the spread of a childhood disease in Portugal, but it is easy to see how these kinds of models might be generalized to the global scale in the manner proposed by Epstein in the PACER (2011) project.

Mobility is central to these grand challenges and the models contained herein address these issues directly. There is the scale change of course in the way we have posed these challenges, for many applications here address these issues at a more local scale. For example, questions of compactness and sprawl intimately affect energy issues. Here we have reviewed various CA models that are being used to generate insights into such development (see Iltanen 2012; Barros 2012) but what we sorely need is to extend such models to explicitly deal with such activities. In fact, one problem in ABM is to represent actions and interactions in a sufficiently robust way to begin to generate predictions that deal explicitly with mobility and movement that can be matched against more aggregate observations. Indeed the need to aggregate from the individual level to a level where patterns are more meaningful is a technical challenge that is only just being addressed. Last but not least, development and aid are key to poverty and in some of the contributions here, particularly those dealing with development in developing countries at the finer scales of land cover, there are useful suggestions for showing how ABM can begin to address questions of equity and distribution.

37.4 Anticipating Future Forms of Modelling and Prediction

Our focus on modelling and simulation here has been pluralistic, focusing on ABM, the newest variant of geographical models but arguing that good models of geographical systems inevitably mix and match the best from many different simulation frameworks. Our best guess is that there will be many more examples in the future of a hybrid variety, which take the best tools and methods from different types, and produce model structures that combine the best of many worlds. In the future, integrated models of a hybrid type may well become the norm but perhaps the greatest changes will not come from new insights into how to model, but from new sources of data. A change in focus from what in the past has been meso-level approaches to the truly micro – local – and the macro – global – is possible. Dealing with routine fine scale spatial and temporal behaviours is more closely matched to ABM than location predictions of the more aggregate kind, while at the global level, the policy context is ripe to be informed by insights for a new class of aggregate ABM. These can combine the individualistic behaviours in such models with aggregates or groups treated as individuals at the highest levels.

It is tempting to second-guess new styles of model that might emerge which build on this evolution of the last 50 years. We are not able to do this but we can identify responses that appear promising: integrated and hybrid models, a sustained and direct approach to simulating spatial behaviour, which is something more than the tokenism that currently besets most behavioural simulations, a new sense of how models might be developed and used all the way from model specification through to implementation and validation with new ways of figuring out how such models can be used to inform the future, how new data sources might change what we consider important in simulation, and how many different and competing models of the same phenomena might be reconciled in the quest to enrich our understanding of key problems: these are all pointers to a future that will take this field further and improve its relevance to ongoing social challenges.

References

Axtell, R., Axelrod, R., Epstein, J. M., & Cohen, M. D. (1996). Aligning simulation models: A case study and results. *Computational and Mathematical Organization Theory, 1*, 123–141.

Banos, A., & Genre-Grandpierre, C. (2012). Towards new metrics for urban road networks: Some preliminary evidence from agent-based simulations. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 627–642). Dordrecht: Springer.

Barros, J. (2012). Exploring urban dynamics in Latin American cities using an agent-based simulation approach. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 571–589). Dordrecht: Springer.

Batty, M. (2010). Integrated models and grand challenges. *ArcNews,* Winter 2010/2011. Available at http://www.esri.com/news/arcnews/winter1011articles/integrated-models.html. Accessed 16 June 2011.

Castiglione, F. (2003, November 3–6). *Large-scale agent-based models: Perspectives and requirements, hot topics*. Workshop: Agent Based Modeling and Simulation, The Institute for Mathematics and its Applications, University of Minnesota. Available at http://www.ima.umn.edu/talks/workshops/11-3-6.2003/castiglione/IMA-WrkshpABM2003.pdf. Accessed 16 June 2011.

Epstein, J. M. (2008). Why model? *Journal of Artificial Societies and Social Simulation*, *11*(4), 12. Available at http://jasss.soc.surrey.ac.uk/11/4/12.html. Accessed 16 June 2011.

Greenberger, M., Crenson, M., & Crissey, B. L. (1976). *Models in the policy process: Public decision making in the computer era*. New York: Russell Sage Foundation/Basic Books.

Gulden, T. R., & Hammond, R. A. (2012). Beyond Zipf: An agent-based understanding of city size distributions. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 677–704). Dordrecht: Springer.

Harland, K., & Heppenstall, A. J. (2012). Using agent-based models for education planning: Is the UK education system agent based? In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 481–497). Dordrecht: Springer.

Iltanen, S. (2012). Cellular automata in urban spatial modelling. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 69–84). Dordrecht: Springer.

Johansson, A., & Kretz, T. (2012). Applied pedestrian modeling. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 451–462). Dordrecht: Springer.

Malleson, N. (2012). Using agent-based models to simulate crime. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 411–434). Dordrecht: Springer.

Ngo, T. N., & See, L. M. (2012). Calibration and validation of agent-based models of land cover change. In A. J. Heppenstall, A. T. Crooks, L. M. See & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 181–196). Springer.

PACER (2011). National Center for the Study of Preparedness and Catastrophic Event Response (PACER). Available at http://www.orau.gov/dhssummit/factsheet/PACER2011.pdf. Accessed 16 June 2011.

Pumain, D. (2012). Multi-agent system modelling for urban systems: The series of SIMPOP models. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 721–738). Dordrecht: Springer.

Simoes, J. A. (2012). An agent-based/network approach to spatial epidemics. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 591–610). Dordrecht: Springer.

Smith, D. M. (2012). Simulating spatial health inequalities. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 499–510). Dordrecht: Springer.

van der Leeuw, S. E. (2004). Why model? *Cybernetics and Systems, 35*, 117–128.

Walsh, C. L., Dawson, R. J., Hall, J. W., Barr, S. L., Batty, M., Bristow, A. L., Carney, S., Dagoumas, A. S., Ford, A. C., Harpham, C., Tight, M. R., Watters, H., & Zanni, A. M. (2011). Assessment of climate change mitigation and adaptation in cities. *Proceedings of the Institution of Civil Engineers: Urban Design and Planning, 164*(2), 75–84.

Wu, B. M., & Birkin, M. H. (2012). Agent-based extensions to a spatial microsimulation model of demographic change. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 347–360). Dordrecht: Springer.

Index

DOI 10.1007/978-90-481-8927-4,

S

W

X

Y

Z