CASA

MODELING COMPLEXITY: THE LIMITS TO PREDICTION

Michael Batty
Paul M. Torrens

UCL

Centre for Advanced Spatial Analysis
University College London
1-19 Torrington Place
Gower Street
London   WC1E 6BT

[t] +44 (0) 20 7679 1782
[f] +44 (0) 20 7813 2843
[e] casa@ucl.ac.uk
[w] www.casa.ucl.ac.uk

http//www.casa.ucl.ac.uk/paper36.pdf

Date: October 2001

ISSN: 1467-1298

ABSTRACT

A working definition of a complex system is of an entity which is coherent in some recognizable way but whose elements, interactions, and dynamics generate structures admitting surprise and novelty which cannot be defined *a priori*. Complex systems are therefore more than the sum of their parts, and a consequence of this is that any model of their structure is necessarily incomplete and partial. Models represent simplifications of a system in which salient parts and processes are simulated and given this definition, many models will exist of any particular complex system. In this paper, we explore the impact of complexity in validating models of such systems. We begin with definitions of complexity, complex systems, and models thereof. We identify the key issues as being concerned with the characterization of system equilibrium, system environment, and the way systems and their elements extend and scale. As our perspective on these issues changes, then so do our models and this has implications for their testing and validation. We develop these, introducing changes in the meaning of validity posed by the use to which such models are to be put in terms of their users. We draw these ideas together as conclusions about the limits posed to prediction in complex systems. We illustrate our arguments using various examples from the field of urban systems theory and urban science.

ACKNOWLEDGEMENTS

1 Defining Complexity, Modeling Complexity

This inquiry is motivated by a long standing concern for the most appropriate ways of testing and validating large scale models of urban development. These types of model have always been characterized by their extensiveness but since their early development over 40 years ago, there has been a slow but inexorable rise in their complexity. This has been hastened by a sea change in the relation of science to knowledge, and the way we are able to use science in society. In many senses, our growing need for ever more complex models and the increasing difficulties in their validation mirrors the longer term shift from a certain to an uncertain world. The first 50 years of the last century, perhaps even the previous 200 years, was dominated by the notion that science would yield answers of the simplest kind to a wide range of applicable problems but this certainty has gradually dissolved. The reasons for this are diverse. At one level, this may be no more than one of those unfathomable psychological shifts in our awareness of the limits to our knowledge which occur periodically; at another level, it may be due to an increasing body of experiential knowledge of using science in the quest for exact answers to important problems and the growing realization that such certainty is illusory. The recent history of social forecasting in this regard has been salutary; both macro and micro events, from predictions of the stock market and the general performance of the economy to more local issues such as demographic change and traffic movements in cities seem beyond our understanding as well as control in that extraneous events now seem to dominate their behavior. Although this may always have been the case, the models that were fashioned a generation or more ago now seem wholly inadequate.

None of this has daunted our curiosity in using science to explain and predict but it has changed it. 50 years ago, the quest to build useful theories and models was dominated by the view that we could simplify and distill the essence of things so that we might capture enough of the social reality sufficient for comprehension and decision. Despite recognition that the world was complex, it appeared simple enough to produce some absolutism to the theory and models that might be employed in applications. With this growth in uncertainty and the increasing perception that the systems that we deal with are intrinsically complex, simplicity no longer seems the watchword in the development of techniques and models. Prediction is couched in qualification, and our science has become less orientated to prediction but more an aid to understanding, to structure debate. This is seen nowhere more clearly than in the shift to constructing 'what if' scenarios which now dominate all model-building.

The systems approach (Churchman, 1968) that was the foundation on which most operational urban development models were predicated, was strident in its advocacy of three key principles of model-building. The first involves defining the system in its wider environment in such a way that the system has a crisp boundary with the outside world; in short, interactions of interest must be much denser within the system than outside. The second has become more controversial and this revolves around the idea that the system must manifest some equilibrium, that processes of change within it must imply some equilibrium and, if such processes are well behaved, that the equilibrium itself might be the focus of prediction. The third principle involves the elements of the system that must be uniform or homogeneous in some sense, with the focus on explaining the order and regularity that such homogeneity implies. These principles did once appear to be implementable for urban systems but it is now easy to argue that none of these apply to even the simplest system of interest to policy-makers. Systems of any interest are impossible to close, their usual state is far-from-equilibrium; often no such equilibrium ever exists They are composed of heterogeneous agents and objects, indeed their very richness comes from such heterogeneity. The quest of science, it is now argued, should be to grapple with explanation and frameworks that attempt to contain, if not explain, such diversity. None of this bodes well for models in which traditional prediction is the goal.

Complexity theory has gradually overwhelmed systems theory, some would say enriched it, but few are bold enough to define its scope (Gell-Mann, 1994). Most commentators, in fact, define complexity implicitly through its various attributes and dimensions, and this is the path we will follow here. The very best way to characterize a complex system is by the states or conditions it can take on. If there are n elements describing a particular state, and each state is described by the (binary) existence or otherwise of a particular condition for each element, then there are 2^n distinct states. For example, there is a whole class of urban models built around cellular automata where the state of the system might be described by n cells with each cell being developed or not developed. In a system with say 10000 cells or zones, then the number of possible states defies description. Add to this different ways or rules of generating these states, then the problem begins to scale in a manner that cannot be handled by conventional theorizing. Of courses, none of this is very new and this kind of characterization of complexity has been known for a long time. However the change in worldview has transferred attention away from the more restrictive aspects of such models to their existence properties, with the consequence that the systems we deal with appear to have boundless complexity.

Most would agree that complex systems have an extensiveness in their elements or objects that make any fixed description incomplete. This immediately translates into the fact that all possible forms of the system are unrealizable, and their representation is rarely stable. In short, complex systems generate a dynamic which enables their elements to transform in ways that are surprising, through adaptation, mutation, transformation, and so on. This is sometimes described in terms of a system enabling the design of new forms. In any event, the hallmark of this kind of complexity is novelty and surprise which cannot be anticipated through any prior characterization. All that can be said is that such systems have the potential for generating new behaviors. Holland (1995), for example, describes complex (adaptive) systems as being systems that maintain their structure and coherence under all imaginable changes, in short through adaptation. Allen (2001) goes much further and defines complexity in terms of the sources of unexpected change or 'unpredictability'. He says: "The simplest definition of a complex system is one that can respond in more than one way to its environment. The 'choice' in response arises from the fact that non-linear processes within the system can potentially amplify microscopic heterogeneity hidden within it". This, he argues, is the origin of that overworked term emergence, another way of describing behavior that cannot be anticipated.

From this rather casual introduction, we can identify two key elements that define our view of complexity and complex systems models. The first is 'system extensiveness' along any spatial or temporal or topical dimension. Such systems cannot be simplified in the conventional way by reduction or aggregation, for in doing so, the richness of their structure would be lost. This, of course, is directly counter to the usual strategy in science, which involves distilling the essence of a phenomena, the essence being defined in relation to some purpose, and thence using that essence as the basis for theorizing and modeling. The second issue involves process. This can often be portrayed as the system's dynamics in space and time in which unexpected change takes place, new objects emerge, and existing objects transform. There are some logical difficulties in all of this, for once objects have emerged and once one has considered the logical limits to a system and its possible boundaries, to make any progress in terms of crisp and clear representations, then the system has to be bounded in space and time (Epstein, 1999). Much of complexity theory so far has, in fact, been concerned with demonstrating models of systems that were initially deemed inexplicable because they demonstrated surprising behavior. Once understood, this behavior is no longer surprising, but invariably it can only be explained by processes that exist at a micro level giving rise to phenomena at a macro level which, in turn, cannot be explained in traditional macro terms. In short, much of complexity theory and its modeling is rooted in explaining behaviors that have already been observed and in some sense, can thus be said to be no longer complex.

We will put these wider issues aside in this discussion. We are more concerned with demonstrating that a new generation of models alluding to complexity theory often, indeed usually, fall back on traditional strategies which have been conceived for a more certain, simpler world where the kind of ambiguity characteristic of complexity is absent, or at least not to the fore. Our discussion therefore will focus on how we might model complexity in the face of infinite variety; or rather not about how to model it *per se*, but how to face it in terms of traditional ideas based on the validation of some structure against a well defined representation encoded in 'data'. This presupposes that even if a system is infinitely complex, then some simplification must take place. But how much? And more importantly, how do we deal with knowing that our models will always be 'inadequate' in a predictive sense? The issue of parsimony is under fire here, and there is little doubt that our own field has barely broached these issues. *We persist in developing models that are intrinsically complex but which we attempt to validate against some reality which we represent as intrinsically simple.* We do not seek to provide answers to this dilemma although we will identify strategies for dealing with it, which will invariably broaden the context.

We begin by exploring the traditional role of validation in modeling, in confronting models with data, and with replicating the traditional role of experiment in a social context. We will then discuss the problem of system definition – of bounding the system from its wider environment in time and space. This problem has been relaxed in various ways as we have learnt more about models and modeling and this has translated itself into ideas about simulation. Simulation differs from modeling in that simulations are dynamic and open-ended. We will chart this road to simulation which contains the essence of what Epstein (1999) calls 'generative modeling', and we will discuss how this style of modeling has come to replace more traditional parsimonious approaches. We then broaden the context, examining the role of the model, the modeled, and the model user. It is in this context that changes in the emphasis given to the role of validation of a traditional kind, can be legitimized. Finally we discuss some examples from urban science, showing that models built around cellular automata, agents, and network relations need to be validated somewhat differently from the way they have been previously if they are to realize their promise in helping us to understand how cities develop. We conclude by suggesting that this debate should become central to simulation and to the decisions used in constructing, testing, and using any kind of social systems model.

2 Traditional Conceptions of Theorizing and Modeling

The conventional process of model-building involves generating and testing theory through a cycle of induction and deduction. In this we follow Popper (1959) who argues that science proceeds through a process of conjecture, thence refutation which is accomplished through a mixed process of induction and deduction. Models are theorized as hypotheses – inductively with respect to data or prior ideas – and are then tested and refined through confronting their predictions with new data both deductively and inductively. The process is confounded. Theories may come out of the blue insofar as they are the product of insights but they can usually be traced to assembling data, deriving relations, constructing hypotheses, testing these on new data, falsifying them, maybe refining or modifying them to make them more or less bolder, as the data and context suggests. In essence, the process is one in which testing and validation involves both theory and data with no privilege given to one or the other.

There are two rules which have been taken as central to the process of developing good models. The first is the rule of parsimony – Occam's razor – which suggests that a better model is one which can explain the same phenomena with a lesser number of intellectual constructs. This is often translated as lesser data or only as much data as is needed, and it is in this sense that theories and models simplify the real world. The second principle relates to independence in verification. A theory which is induced using one set of data needs to be validated against another independent set. In short, if a model is driven by data from one situation, being fine tuned or calibrated to that situation, the only valid test of the model is to then apply it to another situation independent of the first. As we shall see, parsimony and independence in validation are criteria that are rarely fulfilled in traditional modeling. In general, the essential difference between traditional systems and complex systems models is one where it is possible, in principle, to meet the criteria of parsimony and independence for the first set but not for the second.

We will illustrate the principle of parsimony first for its serves to show how data and model structure determine verification. The simplest model is one in which an independent variable y measured over some range of values is explained in terms of some independent variable x measured over the same range. Often more than a single independent variable $x_1, x_2, ... x_n$ is used to explain variation in a single variable y with the implication that each independent variable accounts for some independent component of the variation in y. It could be argued that the more independent variables used in this way, the less parsimonious the model becomes and there is a natural tendency to think of these less parsimonious models as being over-determined. The simple graphical

illustration in Figure 1 makes the point when we compare *(a)* with *(b)*. If we were to try to explain more dependent variables $y_1, y_2, ... y_m$ with less independent ones as we show in diagram *(c)*, then there is a clear problem in that there can be no unique solution. Sometimes it is convenient to think of this kind of problem in terms of balanced equations which to be solvable, simultaneously say, must imply as many unknowns as knowns, as many dependent as independent variables. In fact, econometric models which replicate systems of equations sometimes partitioned into exogenous and endogenous variables invariably invoke such conditions of balance or simultaneity in effecting robust solutions. This is illustrated in model structure *(d)*.

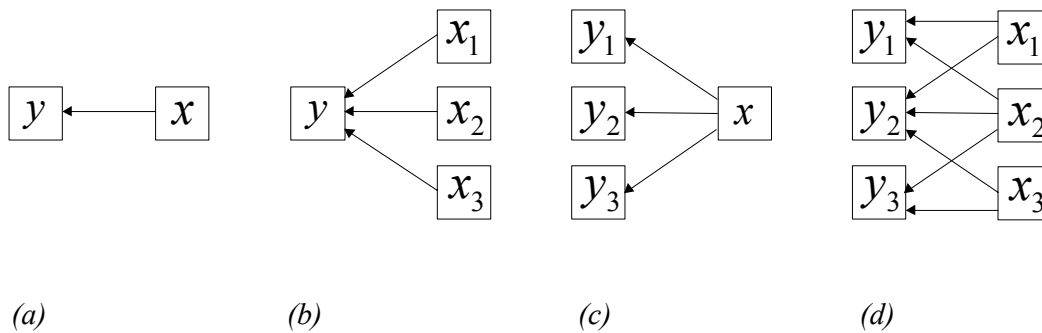|			
(a)	*(b)*	*(c)*	*(d)*

Figure 1: Model Structures Linking Dependent to Independent Variables

All the model structures illustrated above imply some process of linking independent to dependent variables. This process in models which are statistically robust is usually simple: it is additive as in a linear model or at best non-linear but in a tractable mathematical way. Complications begin to arise when the processes involve rule-based systems which in practice cannot be reduced to tractable mathematical operations. Each of these model structures may best be thought of as one in which distinct processes determine the outcomes or predictions. For example in *(a)*, there is one and only one process determining *y*. In *(b)*, there are three processes determining one output. In *(c)*, there are three processes determining three outputs but the concern is that these three processes are determined as variants of only a single input. In *(d)*, seven processes determine three outputs but these seven processes are determined by three inputs and thus the system is in a sense balanced. Strictly where there are more dependent than independent variables, then the processes involved must be further specified with independent data so that there is as much information to determine the outputs as there is input to the model.

Most models of urban development that were constructed for policy purposes from the 1960s on paid some homage to these principles. The best examples were those which mirrored the structures

of linear econometrics, the EMPIRIC model for Boston being the example *par excellence* (see Lowry, 1967). The models built in the spatial interaction tradition (see Wilson, 2000) also tended to meet these conditions, with distance playing a key determining role in predicting trip distributions and enough independent data on location and trip distribution being assembled to provide robust calibrations. There is a sense in these models that you are not getting 'something for nothing', although they suffer enormously from limits posed by the way the systems to which they have been applied were articulated and the inadequacies of theories that were assumed to be at work. Perhaps the clearest model which broke from this tradition and which illustrated distinctly the problems posed by the current generation of models based on complexity was Forrester's (1969) *Urban Dynamics* model. Apart from the fact that the model entirely defined away spatial variation by treating a hypothetical inner city disconnected from its wider environment, the model was not calibrated to data in any way. The model hypothesized countless dynamic relations involving the stocks and flows determining employment and residential activity volumes in the city which were culled from casual knowledge and observation. What validation there was involved superficial observations that the simulation appeared consistent with the characteristic features of US inner city areas at the time. More controversial were the longer term dynamics of the model which mirrored logistic growth and a vicious spiral of decline from which the city area could not break free. It might be, as was argued at the time, that the purpose of this model was to raise the level of debate about the inner city. It was not to provide an operational simulation. It was to foster discussion about possible policy issues. This is an argument we will return to as it is usually used to legitimate complex systems models which cannot be validated in the traditional sense, and it is worth noting that Forrester's model was one of the first to polarize the debate.

The second principle of good model-building involves testing the model in such a way that it can be validated in a context that is independent from that for which it has been initially developed. This is no more or less than the simple requirement in laboratory science that setting up an experiment, then validating the theory once is not a sufficient test; so much fine-tuning goes into setting up the experiment, that it is necessary to see how this transfers – generalizes – to another situation. In terms of urban models, this is a strong requirement; it implies that the first true test of a model is not on the place where it is first developed and fine-tuned but in a second or subsequent place where it performs equally well or badly. Such dual applications have rarely been the case for reasons of happenstance rather than unwitting poor scientific practice. What usually happens is that some model structure is successively refined on different places and at different times, and in this way, some confidence is built up in the model's validity. The problem is often that the model is sufficiently different in each time and place to limit its generality. A formal study of different cross-

sectional land use-transport models in the Lowry vintage which was designed to test the same model variants on different places was mired in data and computer software incompatibilities between these different places and despite the heroic sentiments on which this project was established, the analysis was inconclusive (Webster, et al., 1989).

There is, however, a rather special case where data-rich models containing homogeneous undifferentiated processes linking inputs to outputs does meet this requirement of independence. In situations where the observations are very extensive and homogenous, and where the system can be partitioned into distinct sets or regions without doing gross violence to its structure, then it is possible to develop the model on a sample of the data and validate it on the remaining full data set. This is a little like fitting a model to one part of a city and then validating it on the rest. Invariably this is not possible in cities because spatial variation is such that one would not expect a model of, say, the inner city to apply in quite the same way to a model of the suburbs. Nevertheless, where data sets are extensive and where the relations between inputs and outputs are assumed to be ubiquitous, then model fitting on a sample of the data followed by validation on the full set, less the sample, is quite widely practiced. The best examples involve extracting pattern as in remote sensing data and in neural nets where the assumption is that everything influences everything else. However although this is not usually possible for urban models, it is still possible to build structures which meet the conditions of parsimony and data balance and to validate these types of models in two places rather than one, which the principle of independence suggests. Such tight conditions require the model to be developed in more than one place by the same modeler working under the same conditions, and for practical reasons, this is usually not possible.

So far, we have characterized models as being structures in which inputs are related to outputs in a simple causal manner; that is where the relations between inputs and outputs are not usually the focus of inquiry, with assumptions being made that the way these relations are represented is robust. In the models sketched in the diagrams above, the focus is not on the nature of these relations but on whether or not the outputs are explicable in terms of the input data. If they are not, then the usual strategy is to change the inputs, not the nature of the relations. In so far as the relations embody structure, these are determined from prior theory as to what seems to make sense in common sense terms. It is a straightforward matter to illustrate how these relations become ever more complex by simply adding intermediate outputs and stringing the relations together in the kind of chain that is illustrated in Figure 2 below:
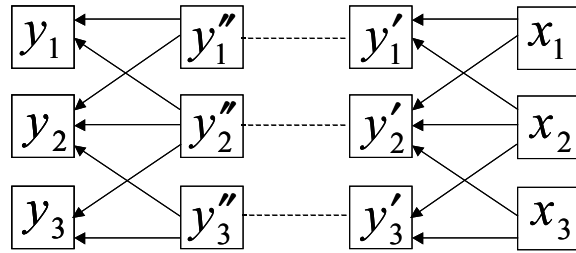
Figure 2: A Model Structure Incorporating Multiple Processes

Here we show successive relations which convolute the original data series in such a way that there is little doubt that the causal processes invoked must be subject to detailed assessment and validation if the structure is to be meaningful. In fact, this strategy is often used in pragmatic model-building where the emphasis is on simply extracting some pattern in the initial data series x_1, x_2, x_3. Neural net models are of this nature but it is unlikely that they have any real meaning for the kinds of models that are now considered to be appropriate frameworks for understanding complex systems.

As we have implied, one focus of complexity theory is on ways in which processes generate patterns. In terms of the structure illustrated above, a typical model in the urban domain might be illustrated as follows. Imagine that each independent variable involves an attribute or attributes of some location – represented by a cell or agent. This cell has some state which might be land use. The process of changing land use which is the dynamic that the model needs to capture, might thus be construed as following a number of stages. This might be the process of land conversion. It will depend of course on other locations or cells in the system and land might go through several processes of change before it becomes fully developed: it might be purchased, assembled, remain idle, be used for temporary facilities and so on. The causal chain above might represent this process. In principle, each element of this process should be explicit and should be capable of being validated with observed data. In practice, this is rarely if ever the case. The data set would be too large, it would be impossible to collect in its entirety, it may be impossible to even observe and measure. Yet the processes are known to be important. Other criteria must thus be used. If the model is broadened and the input elements are no longer confined simply to cells but become heterogeneous with respect to type – x_1 might be cells, x_2 agents, x_3 institutional constraints and so on – then the processes implied by the chain are even more complex and the system diagram above simply shows what might be related to what without any implication that there is a standard way of fitting independent variables to the dependent. Add to this sets of parameters, themselves

11

unknown in strength and value, then the problem begins to explode and very soon, there is no way that all elements of the model can validated. This is even before any consideration that the model structure might contain processes that generate unknown or emergent objects or patterns. In such cases, it is impossible to sweep across the parameter space in an effort to calibrate the structure, even though this time honored method is the usual strategy for confronting the model with data.

In the face of these difficulties, the model-builder often resorts to what we will call here the 'Forrester strategy' – not testing the implied causal structure at all, relying on a simple correspondence between inputs and outputs, but also working up the model from discussions with politicians and decision-makers who do not evaluate their system of interest in the reflective, somewhat detached manner of science. We do not mean to denigrate this strategy because it is often useful and in certain circumstances inevitable. Forrester (1969) himself, in developing his model, said: "I approached these discussions knowing the conceptual nature of the structure being sought, but not the specific details of the structure or the institutional components and behavior to be fitted into it. The others brought the knowledge of the pressures, motivations, relationships, reactions, and historical incidents need to shape the theory …" (page ix). And he continued by saying: "Actually the book comes from a different body of knowledge, from the insights of those who know the urban scene first hand …" (page x).

Moreover in very complex systems, the notion of seeing if a model produces plausible patterns which look right in a superficial way has been lauded as a much more appropriate way than the mindless statistical testing that has taken place with many modeling ventures in the past. It was Mandelbrot (1983, page 21) who said in the context of fractal geometry: "… to see is to believe …". The critical issue in complex systems models is that this is not the only strategy. There are many qualitative tests that are possible with respect to how plausible structures are which generate believable predictions, and these should be mapped out. In fact, there has been hardly any work whatsoever on strategies for validating models which deal with intrinsically complex systems, and one purpose of this paper is to raise awareness and encourage debate in this domain.

To summarize before we begin to sketch out the key elements of intrinsically complex systems and their models, it is clear that the difference between traditional models and the new generation that we are appealing to here, is one which relates to how causal structures are treated. In traditional urban models, the focus is on simple causes. Insofar as these are convoluted in any way, it is through making the system extensive, through repeating these simple causes over many categories but not by elaborating the causal chains that link inputs to outputs. The most extreme variants of

this style simply assume that the causal structure is a homogeneous nexus of additive factors as in multiple regression or in neural nets. The emphasis is largely on validating these kinds of models using data which drives these simple causes. In contrast, complex systems models have multiple causes which display a heterogeneity of processes that are impossible to observe in their entirety. The focus is on more qualitative evaluation of a model's plausibility in ways that relate to prior analysis of the model's structure. In both styles of model, the wider context is important in validating the model too. What the model is to be used for – its purpose – is all important, particularly so where a degree of belief in its predictions may be suspended because of its complexity. Criteria for developing such models are not well-worked out and in urban systems this has become an important challenge.

3 The Problem of Closure

A key concept in general system theory involves the notion of the system and its environment. Systems are usually defined as existing in a wider environment with the system containing all the quantities and qualities of direct interest but with a recognition that for the system to function, it must import and export energy into and from its wider environment. Good system design assumes that the interactions within the system are much denser and connected that those between the system or its environment and in this way, a system is assumed to be relatively independent of its environment. In short, although it may not be possible for a system to function without considering the relations to its environment, the central focus of interest is on the interactions within the system. Whether or not such criteria can be applied to the systems of interest here is a moot question. Much work has assumed that systems can be defined in this way but there is plenty of anecdotal evidence that suggests that only the most trivial systems strictly meet these requirements. Indeed some of the most powerful critiques of contemporary urban modeling have been based on the artificiality of closure as for example in the Forrester (1969) model of the inner city, which entirely ignored the dependence of the inner on the outer city and vice versa, though obvious links such as the journey to work and industrial dependence.

Complexity theory relaxes this criterion somewhat. In particular, the focus is on systems that scale – from the local to the global. Cities and economies are structured in this fashion as all studies of world cities in the modern economy demonstrate. One of the most intriguing features of complex systems is their ability to simulate the way local action generates some global order, and this in itself is often taken as the very definition of complexity. The ability of systems to handle local

action that generates global pattern implies emergence in that there is nothing in the local actions that implies the global pattern. Usually it is the interactions that take place locally that generate the higher order pattern, and it follows that for such systems to be simulated, this kind of link cannot be broken through artificial closure. The archetypal example of local action leading to global order is in the phenomenon of segregation in residential neighborhoods due originally to Schelling (1969), in which a mild preference for living adjacent to one's own group leads to very strict homogeneous spatial segregation. In such cases, it is clearly impossible to simulate neighborhood dynamics without recourse to the entire city as the effect would not be captured without considering all neighborhoods. In a wider context, it is hard to know how to simulate the development of the financial core in a world city, for example, without some dynamics of the more global economy of cities being present within the model, and it is easy to see how this argument might ultimately imply that all cities everywhere need to be modeled simultaneously for the essential features of interest to be captured. This problem also appears to have become more intense as cities have become more interconnected through the development of technologies that aid interaction. Moreover, the very notion of a system of interest being one that is a cluster within a wider system is questionable. Systems function because of local and global connectivity as for example in social networks which are held together by weak ties, rather than strong clusters. In short, the idea of partitioning a system and fixing a level of inquiry at a particular layer in the hierarchy of connections can be problematic. Recent research into the phenomena of 'small worlds' reveals this where the very notion of a small world implies a local-global linkage (Watts, 1999).

Closure is a generic issue in defining systems although we can distinguish between temporal and spatial ways of separating the system from its wider environment. In terms of modeling time, many systems have been closed entirely in that time is defined away and the system studied as though it were in equilibrium. In most systems, equilibrium is an assumption based on convenience. For living systems, there can be no intrinsic equilibrium although there may be steady state activity in which the system renews itself in a balanced manner. For a time in the middle of the 20th century, it did seem as though cities and economies were in some sense stable but with the passing of the industrial era, it is all too clear that cities cannot be explained from their structure at point in time. Equilibrium is a concept that is also inconsistent with interactions between the local and the global. Systems become ever more volatile as we disaggregate to basic units, or rather, systems become ever more homogeneous as we scale them up by averaging activities. In one sense, what is to be explained is how this scaling and averaging takes place – how cities appear stable and equilibrium-like at higher spatial and temporal scales than at the finer scale. Moreover were we to separate their dynamics into coarse and finer spatial scales, then we would miss the fact that policies designed for

one scale often have an impact at a different scale; in dealing with one level of the hierarchy only, true consequences are missed.

In fact, if the focus is broadened a little, cities must be seen as being far-from-equilibrium in that their order is a consequence of continual change. What this implies is that the dynamics of cities is unlikely to be very smooth. Although volatile bubbling change at the low level gets averaged out as we aggregate, this does not appear to lead to radical changes in the trajectories of systems such as those formed through phase transitions and similar discontinuities. Much of course depends on how the system is articulated but the message of complexity theory is that to understand significant urban change, then the system must not be closed in such a way that its dynamics are reduced to only one variety when several different varieties are clearly present. This is not simply an issue of spatial closure for it also relates to the time interval over which the dynamics are captured. Averaging time intervals also reduces variation but an equally significant issue involves the position at which the dynamics are first recorded. Inevitably there has to be some closure in time in that the system must be started at some point in time. This is equivalent to choosing initial conditions and if these are selected in such as way as to destroy critical processes, then the entire dynamics can be confounded.

In urban systems, closure with respect to the range of activities, land uses, agents, or objects which represent the focus of inquiry, is similarly problematic. However it is when classes or attributes interact with dynamic and/or spatial patterns that significant concerns arise. For example, by aggregating two or more activities together, critical dynamics might be collapsed or spatial variation canceled out. In one sense, this kind of issue is related to scaling which in turn relates to how different levels of hierarchy relate to one another. Effects that occur similarly between levels can sometimes act as the basis for partition but in a different vein, might require explicit representation, thus mitigating against any partition. We could continue our discussion of system closure almost indefinitely but the key point that we have made several times is that complex systems are difficult to close from external influences. Indeed the very definition of complexity presupposes that systems have infinite extent and variety and that their unique and novel behavior comes from the interaction of diverse effects that must somehow be accounted for within the system definition and representation. This implies a contradiction in that systems that cannot be bounded and separated from their wider environment must be inherently unpredictable.

4 The Road to Simulation: Artificial Systems as Virtual Laboratories

Models are, by definition, a simplification of some reality which involves distilling the essence of that reality to some lesser representation. Such simplification is usually for some purpose although in science, that purpose may be entirely justified in terms of satisfying our intellectual curiosity. Usually it is more than this although intellectual inquiry is generally a prerequisite. Nevertheless, any model will always contain more assumptions about the reality than are testable in that the very act of defining the system of interest involves contextual assumptions that remain implicit, hence not testable without a radical change in perspective. The difference between complex systems models and those that appeal to the principles of strict parsimony – those that we have been referring to here as traditional models – is one that revolves around the explicitness of assumptions. In essence, traditional models are those in which all relations defining the model are testable while complex systems models have chains of relations that are explicit but untestable in principle and/or untestable because data and observations of their processes are not available.

We appreciate that some might argue with the fact that we are suggesting that complex systems models are not parsimonious. There are clearly examples of models of complex systems, such as the Schelling (1969) models of spatial segregation, which articulate local action that leads to global pattern in the simplest terms. However, even in that case, although the model is simple in its rules, observations of how individuals exercise their preferences to segregate are rarely available and the data to test such models is never complete. A clearer way of signifying the difference between traditional and complex models involves the way they are parameterized. Traditional models are those in which all processes linking their inputs to their outputs can be fine tuned by parameters that enable their outputs to be matched to data. In contrast, although all the processes within a complex system model might be capable of parameterization, many of these are not parameterized; there is no intention of fine tuning these values to match observations for such observation are not likely to be available. A good example of the difference between these two types of model can be elaborated through traffic models of local pedestrian movement. A spatial interaction model based on principles of gravitation would assume that travel densities would vary according to some function of travel time or distance or cost between any origin and destination. That function could be parameterized in such a way that the predicted traffic densities would be matched against observed volumes. It may be that if the functional form could be varied but within the overall bounds of the problem, finding the best fit of some relevant function to data would be possible. However, if the model were conceived of as one in which the actual paths of the pedestrians according to the local geometry of the system were to be modeled – and it might be assumed that for a good prediction of

traffic densities of pedestrians the local geometry were important – then various processes to link pedestrians to local geometries through their cognitive and visual abilities could then be linked to their broader origin and destination behavior. The number of degrees of freedom of the problem thus explodes enormously in that various algorithms for obstacle avoidance and congestion would have to feature. Usually only very general data is available for such obstacle avoidance and it is unlikely in this latter case that the model could be fitted to data in its entirety. Moreover the number of different causal structures which might be equally plausible to enable the agents to proceed through the local environment make testing all model types against data quite impossible.

To summarize, in traditional models we can divide the set of assumptions into those that are explicit, hence testable, and those that are implicit. The parsimony of these models resides in the fact that all the explicit assumptions must be testable. In complex models, explicit assumptions can be divided into those that are testable and those that are not. This immediately presents a dilemma in that these two sets of assumptions often interfere with one another, and it is usually impossible to test one set and not the other. In short, the fact that such models can be only tested partially means that they cannot be validated at all, and even though it is possible to associate some data within some subset of these models' outputs, this is rarely done as such a test is regarded as being arbitrary. Complex systems models are, however, constructed in the full knowledge of these difficulties. Difficulties arise however only when their assumptions are not laid bare and remain hidden. Such models are usually justified on the following premise: that the processes that underpin the model are too important to leave out and that it is preferable to include such processes even though it is not possible to validate them against data. The trouble with this view is that it is difficult to justify and its rational usually depends on intuition. Moreover, the choice of whether to develop a complex systems approach or its simpler antecedent often depends on wider issues involving the purpose of the model, the policy-user context in which it resides, and sometimes the context in which we find ourselves which implies different degrees of acceptance of the problems involved.

Although there is no a one-to-one correspondence between complex systems and simulation models, we will loosely refer to the methods of complex systems models as simulation. Simulation usually implies some form of computational process which in urban systems is often mapped onto a temporal dynamics in some explicit way. Simulation in time involves recursion if only because the same model structures are repeated through time (with one set of outputs becoming the next time period's inputs). Such models often generate more than one outcome. Indeed the possibilities for generating an entire range of scenarios always exists in time for slight changes from time period to

time period might be amplified or damped or both. Moreover, different model structures might converge on the same type of prediction, implying some kind of equifinality that is yet another hallmark of complexity. In talking of simulation however, we must digress slightly and note that microsimulation models of the kind developed originally by Orcutt et al. (1961), and in an urban context by the Leeds group (Clarke, 1996) are not really in the tradition of complexity theory but rather in the spirit of traditional models where simulation refers to the sampling of events from known distributions and thence scaling up to entire populations for predictive purposes.

Prediction involves generating unknown events and in traditional models, such events usually pertain to the future. The ability to calibrate the model by fine-tuning the parameter values in such a way as to replicate a known – present or past – situation provides some confidence in using the model to predict the future. If the model is replicated in a different known situation and performs well in that it survives such a test, there is even greater confidence in its ability to predict the future. Traditional models thus get the present right and are then used to predict the future. In contrast, complex systems model can never predict the present definitively and thus the focus changes on exploring a variety of presents – where the actual present and its variants are just different versions of some unknown future. Simulation enables such models to generate different outcomes, which under some circumstances might appear to be different futures but really define a space of different model outcomes. The way this space is generated is not simply through systematic variations in parameter values, which is the time honored methods of model calibration in the case of traditional models, but through varying the model structures within some limits, that is usually varying the rules that encode different processes into the model, thus simulating different experiments within a kind of virtual laboratory.

This notion of exploring the space of all model outcomes and all model types is central to the simulation of complex systems in that it has become the main way of model testing. This is hardly model validation although it could be regarded as a way to check plausibility and to test the robustness of model structures to changes in causal structures. This space of model structures might be likened to a phase space which defines various model outcomes in terms of model variables. In calibration, the phase space is defined by dimensions associated with the model's parameters and their range of values. In complex systems modeling, the phase space is more qualitative in form, consisting of some mapping of different causal structures onto various dimensions and then some measurement of the model's outcomes under these different structures. Sometimes the various rule sets which mirror these causal structures are parameterized so that the model can be evaluated quantitatively in terms of its performance against standard measures. This is a little like setting up

different experiments within the virtual laboratory where not only the variables defining the experiments are varied but the experimental apparatus is modified from experiment to experiment. In a sense, what constitutes the phase space depends upon how rich the model is. In parsimonious, traditional models such as the spatial interaction type we noted above, it is the space defined by the parameters and their values (associated with the distance functions); in complex models, it might consist of several different spaces relating to different rules and structures, in a wider hierarchy of types. This broadens the problem as it is possible to produce different types and levels of assumption which are capable of being varied although are not capable of being parameterized. In fact, it is even possible to begin to change the very object of study and its representation in this way although in practice such experiments and explorations have rarely been developed.

There are many examples we can review which illustrate these points and we will examine some of these in more detail below. Suffice it to say that the most instructive in the urban domain are those which allude to complexity and the generation of emergence such as those based around ideas of cellular automata (CA). For example, CA models tend to base their mechanism on processes of change in local neighborhoods where action-at-a-distance is espoused. Physical change in cities clearly depends on some local function as is the case of the growth of any structure which must remain connected for its existence, but it is also clear that activities and people do not locate according to local actions restricted to a limited neighborhood space. The rules sets which are used to condition development are also rich and immediately the prospect of testing such models is problematic. For example, there are many neighborhood configurations and many rules sets and the space which is set up by this range of possibilities is enormous. It is not possible to chart such a space and strategies for doing so are quite limited. This problem is best seen in the basic theory of cellular automata as promulgated by Wolfram (1994). Although Wolfram (1994) is able to exhaustively illustrate the possible system outcomes for systems such as the two state (a cell is on or off, developed or not developed), 3 cell regular neighborhood in one-dimensional form where the all possible rules for switching the cells on or off are enumerated, this breaks down for 9 cell neighborhoods in 2 dimensions and all higher orders. Even with these limited possibilities, a bewildering range of behavior is possible and one is forced to conclude that most models that we are working with are arbitrary in this respect, based on a loose consensus of what seems plausible but not on any definitive evaluation of the appropriateness of model structures. Until we are able to move beyond this, then all complex systems model will remain contestable and inconclusive.

Finally it is worth noting that Epstein (1999), amongst others, has argued that complex systems modeling is generative by definition, more a strategy for generating possible model structures and

showing their consequences than a technique for developing fully-fledged definitive models with strong predictive capability. It is arguable whether strong predictive capability is what is required in social systems (because it is probably not attainable anyway) but generative modeling can be equally problematic. In essence, exploring different model types in the absence of data might be useful if equally, but very different, plausible structures are possible, but this is not likely. More likely is the case where there is some agreement about the main elements that can condition or determine some sequence of events but where the operation of these elements is unknown and where different sequences of these can generate very different consequences. In these situations, there is really no alternative but better data and observations so that it is possible to discriminate between these model types. In turn, this pushes the argument back towards traditional modeling where calibration against unique data is the only option.

5 Modelers, the Modeled, and Users: Purposes and Roles

Our entire discussion so far has been without reference to any purpose for studying complex systems or for building appropriate models thereof. Purpose is clearly important, some would say central to what is done, for this conditions how we think about what is modeled and how the system of interest is to be modeled. In a sense, everything follows from this. But in principle, every kind of model can be used for every purpose. Whether this is so will depend on users not on the models per se although there is a tendency for certain types of models to be used for certain groups of users. We can structure models along many different continuums but one which is central to usage depends on the degree to which their predictions, whether or not they pertain to the past or the future, are believable. Many models may be useful even though their outputs might not be believable in that the way they point up and focus events and issues serves a much more important purpose than the generation of hard predictions. We will refer to this spectrum of believability as one which begins with discursive models whose predictions can only be treated in qualitative terms, through to models which generate hard numbers which are as believable as any model might be. There is a strong sense in which complex systems models are less believable than traditional models as we have implicitly argued in the preceding discussion and although such models were rarely used in policy contexts in the past, their use is changing. It is no longer possible to relegate complex systems models to non-policy contexts. In fact as our contemporary view of the usefulness of science has become more uncertain and confused, groups of users have emerged – enlightened one might argue – who are comfortable with engaging in policy discussions using qualitative forms of

modeling, or models whose believability rests on their plausibility and not their ability to replicate known situation.

We need to unpack this spectrum in much more detail but it is usually only fruitful with particular applications in mind and these we will introduce briefly in the next section. A particularly useful way of defusing the role of modeling is to consider the process of use as one of 'story telling'. The extent to which the story told is believable of course is always at issue but no matter what the model says about the past or the future, it tells a story (Guhathakurta, 2001). The story may not be very good, just as the model may not be good either but the notion that models provide just another way of examining a situation, is a good starting point in any application. In fact, this is always the case anyway as the notion of learning about any problem involves different and contrasting viewpoints, some of which may be dismissed, others which will gain ground, and in this, quantitative systems model play a central role. In short, there is always an educational role for modeling and if this role is construed as broadly as possible, then it is clear that any models can in principle be used for any purpose. Pedagogy is important in all contexts for in solving problems directly or indirectly, individually or in group discussion, models which polarize and focus on the key issues are essential. Most models are designed to do this through their role in simplification.

Perhaps the most obvious use of complex systems models which generate unexpected change is for learning, education, and in the broadest sense for entertainment. Models for these purposes do not have to meet strict requirements of validation, unless the purpose is to educate and learn about those specific types of models. More usually such models are designed to stress specific issues, to highlight and to focus, rather than to predict for purposes of problem solving and policy. In fact, models with emergent properties based on evolutionary principles such as those which have been developed in artificial intelligence and artificial life, are increasingly being adopted in game simulations, in web site design, and in digital transactions processing. These kinds of system are strongly influenced by the design of new methods for automated reasoning, rather than any concern for testing how such models might replicate the past or the present. However it is now clear that the considerable effort which has already gone into the development of computer games at all levels from pure entertainment to formalized education, has already had a major impact on simulation. There is evidence that what is state-of-the-art game design today is often incorporated into the e-science of tomorrow. Computer graphics interfaces are a classic example but so is algorithm design and perhaps more importantly, the ability of game designers to think 'out of the box', reveals possibilities for scientific modeling that ordinarily would not be attempted in normal science (Johnson, 2001).

In one sense, thinking about complexity in the way we have been sketching is so new in terms of a science of cities, that it is not surprising that the traditional norms of theory development and hypothesis testing have been relegated to the background. For example, most urban models based on analogies to cellular automata have been more concerned with simply getting such models constructed and demonstrating that a rich dynamics can be generated, rather than with any strict methods for their validation. This was perhaps the case thirty years or more ago with Forrester's (1969) Urban Dynamics model which was one of the first attempts to demonstrate the kind of digital richness that was possible with modern computer systems. Although we clearly need much more explicit principles for complex systems model development which broach directly the question of validation from all perspectives, there is a parallel problem which has become significant. The systems of interest – in our case cities – have themselves changed during this period when digital science has become possible. The traditional attempts to classify and describe cities in coherent terms, which have dominated urban science for the last 100 years or more, have increasingly come under scrutiny. The very systems that we have been concerned with have become more complex as much through the development of digital technologies as through changing life styles and economic conditions of the urban population. In other words, there is now a strong debate about how we should classify cities – how we should describe them – that takes us back to an earlier stage of science. The difficulties in validating complex systems models may be as much to do with the fact that the categories and classes, the objects and elements that we consider significant, have also changed. These new models are as much for engendering the debate about classification as for developing new robust theory which can be validated in the traditional way.

Let it be clear that in this paper we are not arguing at all that complex systems models should be abandoned and that we should return to more traditional strategies of developing parsimonious models. Nor are we arguing that parsimony is not relevant to complex systems models for some of the best models are parsimonious in a way that illustrates the principles of emergence and surprise. What we are calling for is a new strategy for dealing with complex systems models. We need to be explicit about the purpose for modeling and we need to consider the extent to which a complex systems model contains hypotheses that should be validated numerically against observable data. We need to be clear about the line between explicit and implicit assumptions, about the role of prediction and exploration. In fact, a tentative suggestion would be that all models – traditional or complex – should mix calibration with exploration. In the last analysis, it is hard to see the value of a model that does not touch reality at some point in which that reality can be replicated 'unambiguously'. In this way, we consider exploration of model structures as well as more detailed

methods for calibration to be essential in any process of model validation. It is in those models – and there are currently many of them – where such validation is not invoked whatsoever except perhaps at the most casual level, that we feel that a much more explicit process of modeling should be invoked. We also feel that all modeling should be paralleled with extensive debate, with the construction of alternative models – through counter modeling as in the debate over national economic futures in some western countries – and with alternative conceptions of data and observations which are drawn from the same basic set. Where these ambiguities remain, there should be extensive questioning of model structures and purposes.

The contrasting focus for urban modeling is policy-making. Traditional models were largely built from the middle of the last century with urban problems and policy response in mind. It is worth noting that this genre of models were often called 'operational' models in contrast to theoretical models, although in urban studies there was a strong correspondence between operationality and theory in that operational modelers invariably invoked the macro theory of the city that had been crudely fashioned in urban geography and economics during the previous half century. In short, the notion that models for policy making must replicate in some measure the past was widely assumed. The idea that systems theory and analysis which appeared so useful for military and logistical problems at that time could be used for solving social problems also reinforced this assumption. At the same time, this suggested that a good model for policy making would be one in which if the key variables governing the city could be identified and used to predict the present, that same model could be used for managing and controlling the city through some optimization of its structure. The concept of a system and its control was central to this consensus.

Furthermore, the idea that the systems of interest were stable in some way was also essential to this quest. The concept that static models could be developed which replicated the situation at a point in time – assumed to be an equilibrium – and that these models could be used to predict a future point in time which in turn would be in equilibrium, was to the fore. This, as we have seen, has become an untenable assumption and has lead to the unraveling of the basic idea that we are dealing with systems that are a simple enough to describe and predict in this way. There has been a loss of faith in this style of modeling although the view still persists that for models to be operationally useful they should replicate the past and the present in a sufficiently robust way to give confidence in any use they may have for prediction. Models which cannot be validated are thus no different from qualitative reasoning, from intuition, or even dictat which were the usual schemes used to develop policy prior to the computer era.

There is a sense in some policy making that models which cannot predict the present and are unlikely to be able to predict any kind of future, might still be useful. This forces the argument back to education and learning, to ' modeling as story telling', and to the use of models to engender and structure discussion and debate. There is a limit to how far this perspective can be justified. Much depends upon the specifics of the situation where such models are found useful, on the nature of the problem and the relative values and disposition of the users and decision-makers. In these cases, then the use of models to generate 'what if?' scenarios is the main basis for application where such scenarios define bounds to the solution space within which possibilities for the future might be discussed and debated.

Before we conclude this section, we should note two related issues. First models are increasingly being used to communicate other kinds of idea – as vehicles on which other less controversial issues might be conveyed. The use of models to help in visualization is a classic example where visualization often requires a more specific focus which only predictions from models can enable. Scenario building is the classic case in point, as are all kinds of urban design which require visualization. Our second digression is no less significant. The lines between the modeler, the modeled, and the user are increasingly blurred. This is a question of cognition in that the kinds of actions and interaction which form the substance of the new generation of complex systems models mirror processes of decision which in turn are those that are employed by model users in policy making. This is no more or less than the idea that the user is part of the system to be modeled and is often no different in behavior from the rest of the system that is being modeled. In less charitable terms, complex problems have been described as those in which the solution is part of the problem – the plan or planner is part of the problem – and this is an issue that is taxing us in how we might represent decision processes in the city while using those same models to engage in similar decision-making. These are problems for future research but they are nonetheless important.

6 Applications from Urban Science: Cellular Automata Models of Urban Development

We have developed the limits to predictive modeling with general reference to urban models which are disaggregate. They treat activities as individual agents or objects at a micro-spatial level, and are temporally dynamic, embodying processes which do not imply any equilibrium end state to the simulation. In particular, we have alluded to cell space models which have some analogy with the

principles of cellular automata in that development is conceived as being based on local rules which generate development, one cell at a time, over many time periods. Here we will illustrate the problems with this genus of models by presenting a plausible, although hypothetical simulation, which suffers from all the limitations we have noted but is constructed with all the advantages that complex systems theory brings to urban modeling.

First however, let us consider four significant applications of CA models which have recently been developed to simulate the growth (and decline) of urban systems. These models are typical of those which are currently being developed with a loose appeal to ideas from complexity theory, namely the notion that local 'uncoordinated' or individualistic actions lead to global patterns with order at higher scales. Each of these models is based on the idea that what happens in a restricted neighborhood based on cells which are physically contiguous to each other, acts as a generator of development. The changes in development which take place in each neighborhood are invariably a function of the size, number, and shape of the already developed cells in the neighborhood as well as physical constraints on what can and cannot be developed. These models do not consider specific interactions or movements between cells, and as such they do not connect land use to the transport system. They do not treat the demand for activity (in the form of land use) as being a function of supply which is the main prediction from such models and as such they do not invoke any market clearing mechanisms which address questions of balance or equilibrium between demand and supply. This poses a critical problem in that it is hard in these models to control for total land coming onto the market. As cells are developed in response to local rules and constraints, to keep the total activities generated in line with what is feasible and observable in terms of growth rates, then some arbitrary mechanisms have to be imposed so that realistic totals emerge.

Another critical problem relates to the fact that the idea of activities influencing development locally, with no action-at-a-distance taking place, is at best a gross simplification of how cities work. Only in one very restrictive sense is this acceptable and this relates to the physical growth of the city which, for it to remain connected in the widest sense of the word, means that development takes place closer to existing development than farther from it. Although development can be fragmented, if cells are considered large enough, as local spheres of influence for example, the growth of cities can be seen as embodying contiguous development. However, new development cannot be traced exclusively to single local sources for agents who determine new land uses make decisions with information about many spatial scales in mind. In an effort to meet this criticism, many modelers have relaxed the local neighborhood criterion, simulating decisions within wider

fields where action-at-a-distance can take place, thus moving these models from the strict premise of CA to what Albin (1975) has called 'cell-space' models.

Anas (1986) refers to these kinds of models as 'physicalist'. In fact, what Anas really had in mind was the previous generation of land use-transport models which only superficially appealed to economic mechanisms, with the implied criticism that such models should incorporate markets and their clearing much more explicitly. This has not really happened; the kinds of complexity that have come to dominate our thinking, has largely broken with this previous generation of models on the premise that static physicalist (or geographical) and economic models could never embrace the kinds of dynamics that generate urban complexity. However the new models are even more physicalist. Although an argument can be made that, in some situations, such physical-based models might be useful in modeling the development process for example, the CA or cell-space type models that have been developed to date are considerably less applicable than the operational models originally developed a generation or more ago which are still being applied in an urban policy-making context. It is in this sense then that CA or CS models applied to urban development still largely remain pedagogic vehicles and it is important to note that most of these have not been motivated by policy-makers faced with practical problems.

The four models that illustrate these dilemmas are significant in that they are all serious applications which attempt to match model predictions against actual observations. One of the first CA models developed by White and Engelen (1993) simulated the development of land use in a hypothetical city with the same kinds of dimensions as the medium sized US cities of Cincinnati, Houston, Milwaukee and Atlanta in 1966. This model broke directly with the notion that neighborhoods should be local. Somewhat arbitrary decisions about the size of neighborhoods were made, arbitrary in that there was no sensitivity testing of the model's predictions with respect to varying the size of these neighborhoods. The standard neighborhood chosen was based on a circular neighborhood of radius six cells containing some 113 cells in total and being divided in 19 distinct distance bands. In fact, probabilities for development of each cell in this neighborhood varied inversely with respect to distance from the central cell, and as such the entire model could be considered as being a cell space equivalent of a simple population density model but with as many poles of development as cells in the system. A simple hierarchy of land uses was imposed on the simulation and the growth in each time period was constrained to match measured values, with the cells for development being chosen until the growth limit was reached.

In this application, it is hard to see how anything that is not already built into the data of the model and its structure can explain the resultant city growth, which is what one might expect. The seeds of growth are all important and the model can be regarded as simulating a relatively simple diffusion of an existing structure whose future form is strongly conditioned by the shape of the city embodied in its initial conditions. Although CA is invoked as a mechanism for growth and representation, this model is similar to a land use model of the Hansen (1959) variety in which land use is located according to relative accessibility of cells but with the spatial system being rigorously laid out on a grid with limited accessibility fields of no more than six cells in radius

A second model – that developed by Clarke and Gaydos (1998) for the simulation of urban growth over long historical time periods since the late 18th century for San Francisco and the Washington-Baltimore corridor relaxes the neighborhood assumption entirely in that cells are chosen for development across the entire space without any constraints imposed by local neighborhoods. In fact, this kind of cell space model is based on computing a development potential for each cell based on nearness to roads, topographic slope, and constraints on land to be urbanized with already developed cells being the seeds for growth which occurs through 'breeding' and local diffusion which appears to be based on the fact that cells near to those that breed are more likely to be favored for new locations. In this sense then, there is a neighborhood effect but the size of such neighborhoods does not seem to be fixed in extent. In fact, the model probably works by considering the factors such as nearness to roads, slopes and constraints not as any composite index of potential but as stages in the simulation of new development. This model, which is hardly a CA in the traditional sense of the term and is more like a land development model, has in fact been calibrated to data over the long historical times periods considered. The various mechanisms and weights of the factors influencing development – roads, slopes, other constraints etc. – are fixed by an exploration of the solution space within which the model is operated. It is not entirely clear how large this space is, nor if the weights produced are time dependent or unique in their convergence against some best fitting statistics although the model does produce plausible and well-fitting patterns. Again the structure of the model is simplistic and there is no sense that anything surprising emerges from its structure although this is not what was intended.
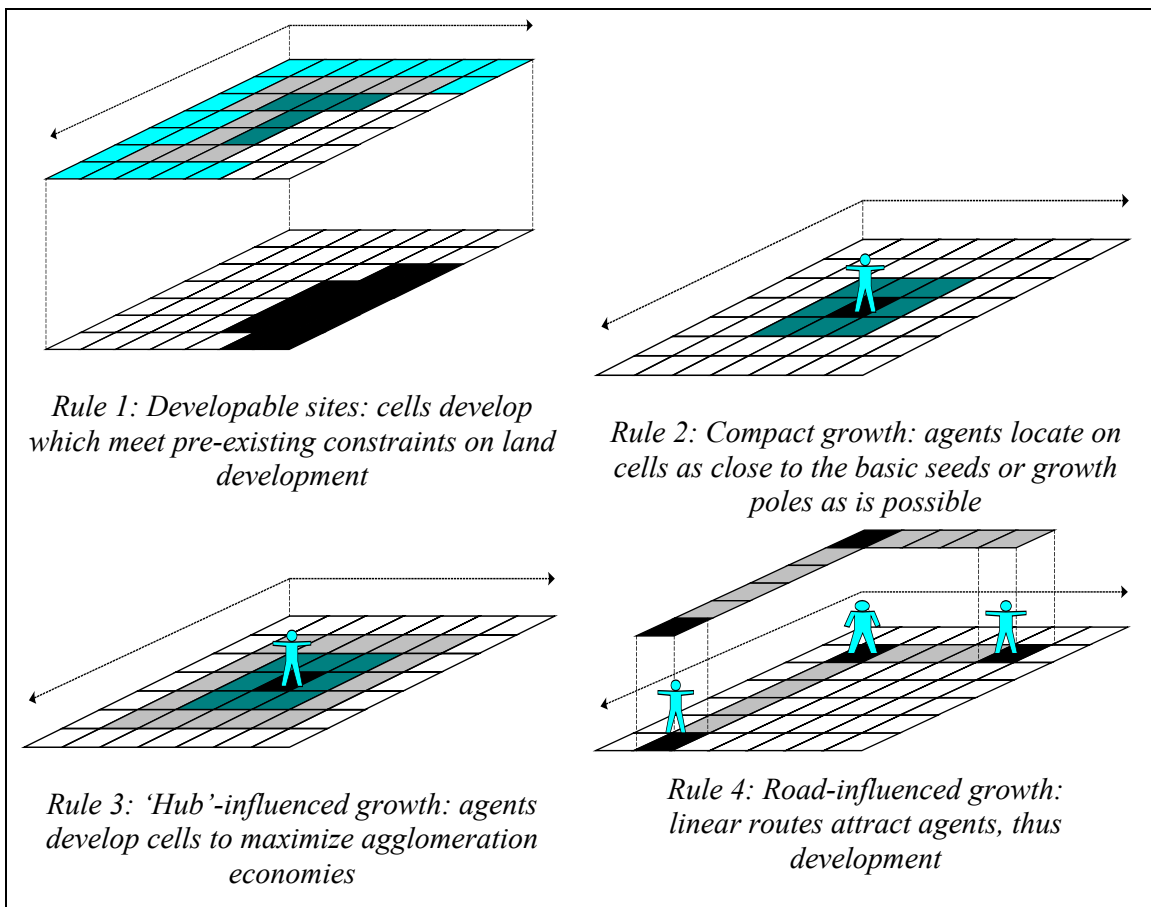
Wu (1998) has developed a number of idealized CA-like models for land development where the focus is on choosing the best transition rules for development within neighborhoods where development takes place. In fact, this is what Clarke and Gaydos (1998) do, although they focus on simple methods for determining weights empirically through conventional methods of calibration using trial and error fitting within the solution space. Wu (1998) invokes more structured methods

which are based on considering the importance of different factors, one to another, in influencing decisions to develop zones. He suggests that a method, called the 'analytic hierarchy process', be used where a consistent weight set is extracted through pairwise comparisons made by a decision-makers in their consideration of each factor against one another. This method is used to develop the transition rules for land development in the city of Guanzhou in southern China where the neighborhood is based on the nine cell Moore neighborhood. The interest here is that this is an example where consistent weights are derived for the transition rules using substantive comparisons which can be related to what we know about how land uses are developed. There are also methods which we will not review here for generating weights on land development which can be incorporated in the simulation through cell transition rules. These take large sets of known transitions and extract weights of factors embodying independent data which might explain these rules using techniques such as neural nets, and related methods of AI (Papini et al., 1998).

Our final example also involves a model for regional urban development in southern China in Guandong province which has been subject to rapid urbanization in the last 20 years. Xia and Yeh (2000) have built several CA-like models of this region and have also recently attempted to fit their models using neural net methods as alluded to in the previous paragraph. Here however we will focus upon the way in which they handle land constraints in the model. In essence, they define each cell as having a land development potential which takes account of the extent to which that land is constrained from development. Apart from binding constraints which restrict any land from development, the cell size at which they are operating is such that only a proportion of the cell might be developed and this needs to be reflected in the way constraints are handled. In fact, this is a feature which is not discussed in any of the previous models although it must be important given that the cell sizes tend to be rather large in all cases. Only when the cells reach the level of resolution of the plot is it possible to design a model in which state transitions are discrete in the development sense. Xia and Yeh's model handles this issue rather well but the size of the neighborhood is fixed at the entirely local level and the problem of action-at-a-distance is defined away in their model. For issues of sprawl, this model may be well-suited although it is not very different from the kinds of models developed for handling urban growth at the cell level in the 1950s and 1960s.

To illustrate the difficulties facing empirical testing of these types of model which we also face in our own work, we will extend the standard CA model to incorporate 'mobile cells' – agents – which interact with each other through a layer of cells which represent fixed locations defining the city. This model is designed to show how a large metropolitan area with the dimensions of Chicago can

be simulated over the last 200 years from the time when it was a small trading post prior to its growth as the key market for the mid-Western United States. Here we will assume the local nine cell von-Neumann neighborhood and place seven seed sites across the region at Chicago, Madison, Milwaukee, Gary, South Bend, Lansing, and Grand Rapids. Each cell in the region can be occupied by any number of agents – its population – and the model is driven by agents reproducing and spreading in the search for developable sites around their initial seeds. Cells have five different attributes: developed or not, developable or not, and a measure of size based on the number of agents located there at any one time. We have defined seven rules which we consider agents 'can' invoke when deciding to locate. These rules are 'plausible' and are culled from the literature rather than from any considered analysis of the land development process over the last 200 years, although there is a good, general literature that has informed our specifications (Cronon, 1991).

Rule 1: Developable sites: cells develop which meet pre-existing constraints on land development

Rule 2: Compact growth: agents locate on cells as close to the basic seeds or growth poles as is possible

Rule 3: 'Hub'-influenced growth: agents develop cells to maximize agglomeration economies

Rule 4: Road-influenced growth: linear routes attract agents, thus development

Rule 5: Scattered growth: agents develop cells by leapfrogging the growing cluster to optimize access

Rule 6: Bifurcated growth: agents develop new cells which connect and compete with the existing cells

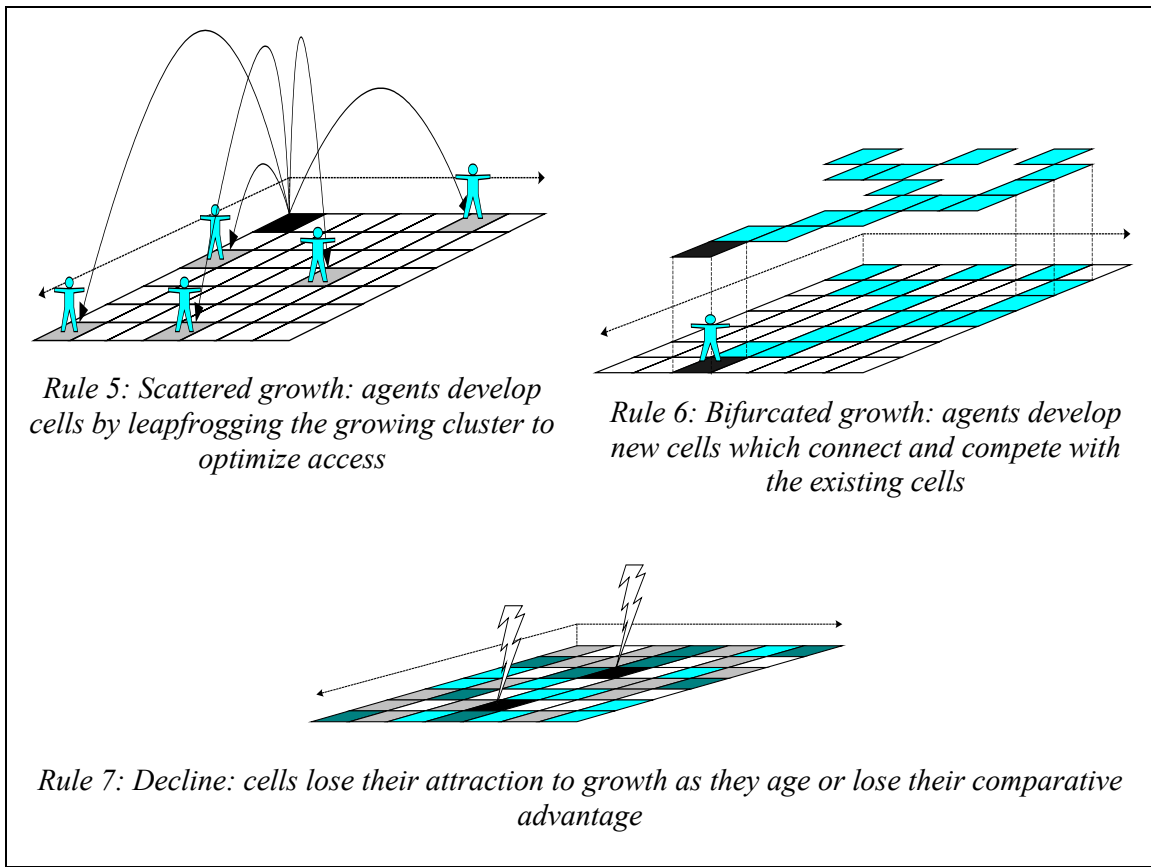*Rule 7: Decline: cells lose their attraction to growth as they age or lose their comparative advantage*

Figure 3: The Seven Rules Used by Agents in their Search for Cells to Develop

The seven rules are illustrated graphically in Figure 3. First cells are only developed if there are no constraints on their development. This rule has a major influence on structure and in a sense, represents the exclusion of sites which we know we cannot simulate. Of course, if we were to exclude all sites that have not been developed at the final simulation date we are aiming towards, then we might simulate the development exactly, thus restricting the model entirely to a simulation of only what had been developed already. Clearly the use of constraints in this way must be a matter for good judgment. The second rule assumes that growth is compact, thus optimizing the economies of local agglomeration but this is contrasted by the third rule that suggests that local centers might attract development disproportionately if they get a head start through random decisions. Rules 2 and 3 thus balance the forces of centralization against those of decentralization. Rule 4 implies that agents locate to maximize their accessibility and thus have a tendency to locate along roads, while the sixth rule allows the possibility of growth bifurcating or deviating from the existing pattern in a small number of cases which are determined randomly. Finally, cells can lose their attraction as a function of how long ago they have been developed as well as the shifting pattern of accessibility to all other cells within the growing metropolis.
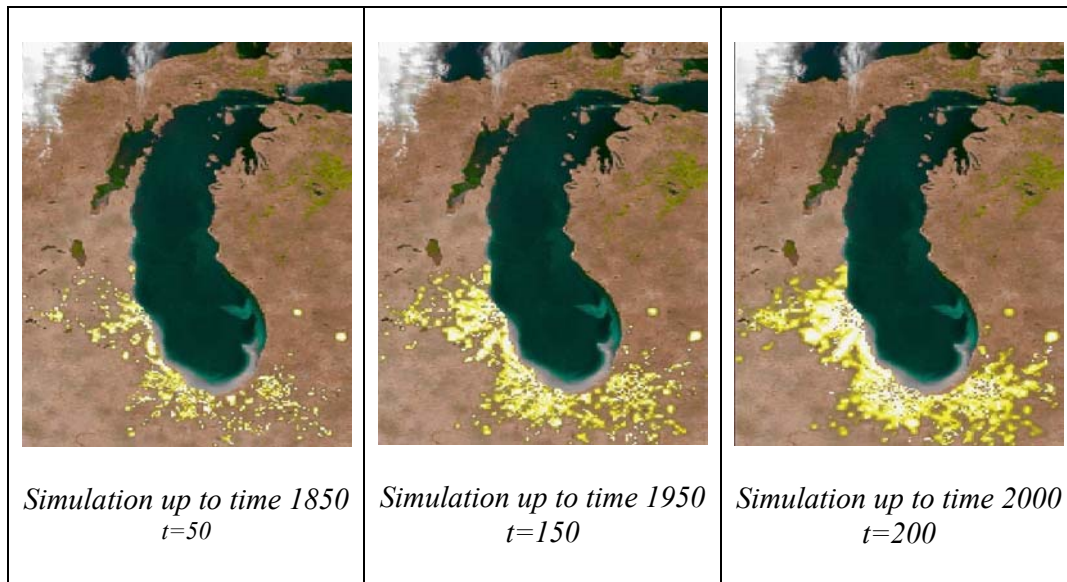
| Simulation up to time 1850
t=50 | Simulation up to time 1950
t=150 | Simulation up to time 2000
t=200 |

Figure 4: The Development of the Chicago Region from 1800 to 2000

Growth rates which guide this process by only allowing a given number of cells to be developed in each iteration/time period have been determined from population census data. The simulation at times 50, 150 and 200, the last time, are shown in Figure 4 where the development is grey toned to show the age of development, and thence overlaid on an RS image of the region (disregarding the cloud cover to the north of Lake Michigan). The development is clearly plausible in that from six seed sites, Chicago grows fastest and biggest but it is also clear that the sprawl is greater than that which currently exists and that there is less fragmentation of urban land use than actually occurs. The model is able to generate some edge city phenomena and there is some evidence, although perhaps less significant than in reality, that growth is directed along the radial routes which dominate the region. It is hard however to say that the phenomena generated is emergent in any fundamental sense although as we argued much earlier, once an emergent phenomena is understood, then it is no longer surprising. This kind of model does have some potential for simulating urban sprawl although the fact that cells and agent locations are chosen randomly within the predetermined rules does limit the use of model outputs for specific forecasting in the Chicago region. The model, even in this more elaborate form than the simplest CA equivalents, is still a pedagogic tool first and foremost.

The issue of using the model for policy purposes turns on the extent to which it can be fitted to real data. As we argued above, the output data is crude – simply what is developed and not developed in contrast to the elaborate nature of the mechanisms and the assumptions that determine the model's

structure. There are seven explicit rules and these are not really in a form where they are testable against data. Although we do not show this here because we have not yet attempted an exhaustive analysis of the model's sensitivity to these rules, it is already clear that very different outputs can be obtained with quite minor changes in the rules themselves. In fact, dramatic changes take place by missing out sets of these rules. With 7 rules, there are 7! possible models. Combine this with the intrinsic randomness of allocation and even with this simple model, the solution space quickly becomes intractable. The problems that we noted above for generic CA and CS models – difficulties of closing the system, the artificiality of the local neighborhood which inhibits action-at-a-distance effects, and the inability of the model to handle interactions across space and between demand and supply sectors, are all present in our Chicago model. From this limited foray into such models, it seems that apart from their ability to show that surprising patterns, or rather patterns that we judge in real cities difficult to explain, hence surprising, such as edge cities, gentrification etc., such models are really demon-strations of systems principles rather than vehicles for operational analysis and policy-making. Models which might be built along these lines to demonstrate emergence and which are capable of being calibrated to real data are likely to be much more specific than those that we have illustrated here.

7 The Limits to Prediction

As we have explained, the time honored principle used in testing theory against data involves specifying causal structures or chains in which independent variables are used to explain an equivalent number of dependent variables; this occurs in such a way that the predictions are useful and interesting, thus adding to our knowledge of how the world works. This is the principle of parsimony, or requisite variety. Invariably theories contain more than can be tested in that there are basic, often implicit, assumptions which might be testable against data but are not, while many theories imply relationships which cannot be tested against data for observational reasons or simply for lack of data. Such theories may well be plausible but they cannot be regarded as parsimonious in the sense we have portrayed here. However there remains a basic problem with theories and models that meet our criterion of parsimony. We may well be able to rigorously test a theory by setting up a model which can be uniquely calibrated, hence testable against data but the causal explanation implied by the theory may not be unique in itself. For example, by changing the independent data driving the explanation, we may be able to build equally good models which predict the same outputs but we are still left with choosing between one model and the other. One might argue that if there are equally good reasons for choosing either set of independent data, then we are still unable

to choose. One can push the argument back and say that if there are two or more different sets of independent data which explain the same phenomena equally well, then there must be correlation between these sets of data and thus they are not in and of themselves independent from one another. This, however, may not be the case or at least we may not be able to decide that it is.

A good example relates to city growth. It is often argued that cities grow around their central cores and that more wealthy income groups can afford to commute longer distances to work in these cores, thus exercising their preference for more space at the edge of the city. This is a crude paraphrasing of the way urban economic theory explains the organization of land used around the core of the industrial city. In short, land use organization is based on the tradeoff between travel cost, rent, and space required which is a function of income. However, an equally good explanation of urban growth is that cities have grown outwards around their cores because this is the only way they could do and still function with their work at the center. As technology has progressed during this period, densities of development get lower and the size of sites get larger as the city grows in that we can generally afford more. Higher income groups are thus forced to live in these more distant areas even if they do not want to due to the fact that there is no place else they can reside and consume the space even if they could afford it. Of course, one might argue that these two competing explanations might be merged in some way but it is entirely possible that each may explain actual development in the monocentric city to the same (numerically) high level.

Thus parsimonious models are not all that they might seem and there is a strong case for building models which contain plausible mechanisms even if we cannot test these mechanisms. The problem is that short of statistical or numerical criterion, good rules for choosing models based on a combination of discursive and reflective analysis as well as standard quantitative evidence are not well-developed. In the case of the CA-like models which we reviewed in the last section, there are so many assumptions about the representation of space and the nature of the transition rules that are used to determine development that it is not possible to definitively use such a model to make predictions that we can act upon. In such instances, we are always forced back to the argument that such models are pedagogic, that they are demonstrations of what is possible, and in the last analysis, provide vehicles for discussion, for counter modeling, and for argumentative discourse. This may, of course, be said for all science and its application to human affairs.

There are two issues that we should draw together which further illustrate the limits to prediction posed by models of complex systems. The first is the issue of emergence which strongly relates to scale and space in urban systems in that models of local action can be demonstrated to give rise to

global order. These we would submit are invariably pedagogic in that once understood, then the phenomena of global order explicable in terms of local action is no longer surprising or mysterious. What is more to the point is that models that can give rise to specific objects or places that emerge from local and other actions are still worth exploring. For example, models that generate growth poles where none existed before are important in that although we might know the generic reasons for such growth, the precise conditions for their emergence may not be known. Such models can be used to explore these conditions, again in a pedagogic way. Suffice it to say that none of the models with the possible exception of our own agent-based CA in the last section contain the mechanisms that give rise to such structures.

This leads to the second issue which limits our predictive powers. Most of the models that we are alluding to here contain mechanisms which involve choosing the drivers of growth using some sort of random processes. For example, choosing cells to be developed is often a process of determining the probability that they might be developed and then choosing actual allocations to these cells, based on these values but by Monte Carlo methods. It is entirely possible to structure these types of model in a deterministic frame but most would agree that the certainty implied in this is problematic. If the usual course involving random simulation is adopted, then there is the problem of knowing what actual simulations mean when they can vary from run to run. The notion of taking some central limiting simulation is problematic too when decisions within these structures are invariably determined by applying discrete thresholding. In short, when one is able to generate let us say 200 different simulations of the development of Chicago using the same model but with different random number seeds, then what is the single limiting case ? How do we take an average of different cells developed in different runs of the same model ?

There are important issues too that, alongside questions of developing rules for building plausible models of complex systems, are relevant to knowing how to deal with predictions that emanate from such models. One feature that we have not addressed here is that because all models are now digital and can be communicated across networks, it is possible to set up effective interfaces to their use in many different situations. Such digital dissemination is now possible and this we urgently need ways of illustrating how these models work so that we can assemble teams to improve such simulations and to learn about their limits in the widest possible domains. Where we are dealing with systems that are intrisincally uncertain and infinitely complex, then the only way forward is to learn the limits to such systems and in this way, to fashion our models to account for such limits. In this way, their use in predictive contexts would become more explicable.

8 References

Albin, P. S. (1975) *The Analysis of Complex Socio-Economic Systems*, D. C. Heath, Lexington, MA.

Allen, P. M. (2001) Models, Knowledge Creation and their Limits, a paper presented at the NEXSUS Workshop on *The Limits to Knowledge*, Central European University, Budapest, June 23-24, 2001.

Anas, A. (1986) From Physical to Economic Urban Models: The Lowry Framework Revisited, in B. G. Hutchinson and M. Batty (Editors) *Advances in Urban Systems Modelling*, North-Holland, Amsterdam, 163-172.

Churchman, C. W. (1968) *The Systems Approach*, Dell Publishing Company, New York.

Clarke, G. (Editor)(1996) *Microsimulation for Urban and Regional Analysis*, Pion Press, London.

Clarke, K. C., and Gaydos, L. J. (1998) Loose-Coupling a Cellular Automaton Model and GIS: Long-Term Urban Growth Predictions for San Francisco and Baltimore, *International Journal of Geographic Information Science*, *12*, 699-714.

Cronon, W. (1991) *Nature's Metropolis: Chicago and the Great West*, W. W. Norton, New York.

Epstein, J. M. (1999) Agent-Based Computational Model and Generative Social Science, *Complexity*, *4*, 41-60.

Forrester, J. W. (1969) *Urban Dynamics*, MIT Press, Cambridge, MA.

Gell-Mann, M. (1994) Complex Adaptive Systems, in G. A. Cowan, D. Pines, and D. Metzer (Editors) *Complexity: Metaphors, Models, and Reality*, Addison-Wesley Publishing Company, Reading, MA, 17-45 (with discussion).

Guhathakurta, S. (2001) Urban Modeling as Storytelling: Using Simulation Models as a Narrative, a paper presented to the World Planning Schools Congress, July 1-15, 2001, Shanghai, PRC.

Hansen, W. G. (1959) How Accessibility Shapes Land Use, *Journal of the American Institute of Planners*, *25*, 73-76.

Holland, J. H. (1995) *Hidden Order: How Adaptation Builds Complexity*, Addison-Wesley Publishing Company, Reading, MA.

Johnson, S. (2001) *Emergence: The Connected Lives of Ants, Brains, Cities, and Software*, Scribner, New York.

Lowry, I. S. (1967) Seven Models of Urban Development: A Structural Comparison, in G. C. Hemmens (Editor) *Urban Development Models*, Special Report 97, Highway Research Board, Washington DC, 121-163.

Mandelbrot, B. B. (1983) *The Fractal Geometry of Nature*, W. H. Freeman and Company, San Francisco, CA.

Orcutt, G. H., Greenberger, M., Korbel, J., and Rivlin, A. M. et al. (1961) *Microanalysis of Socioeconomic Systems: A Simulation Study*, Harper and Row, New York.

Papini, L. Rabino, G. A., Colonna, A., Di Stefano, V., and Lombardo, S. (1998) Learning Cellular Automata in a Real World: The Case Study of the Rome Metropolitan Area, in S. Bandini, R. Serra, and F. Suggi Liverani (Editors) *Cellular Automata: Research Towards Industry: ACRI'96: Proceedings of the Third Conference on Cellular Automata for Research and Industry*, Springer-Verlag, London, pp. 165-183.

Popper, K. R. (1959) *The Logic of Scientific Discovery*, Hutchinson, London.

Schelling, T. C. (1969) Models of Segregation*, American Economic Review, Papers and Proceedings*, *58* (2), 488-493.

Watts, D. (1999) *Small Worlds: Networks between Order and Randomness*, Princeton University Press, Princeton, NJ.

Webster, F. J., Bly, P. H., and Paulley, N. J. (Editors) (1989) *Urban Land-use and Transport Interaction: Policies and Models*, Report of the International Study Group on Land-use/Transport Interaction (ISGLUTI), Avebury, Aldershot, Hampshire, UK.

White, R., and Engelen, G. (1993) Cellular Automata and Fractal Urban Form: A Cellular Modelling Approach to the Evolution of Urban Land Use Patterns, *Environment and Planning A*, *25*, 1175-1199.

Wilson, A. G. (2000) *Complex Spatial Systems: The Modelling Foundations of Urban and Regional Analysis*, Prentice-Hall, Harlow, UK.

Wolfram, S. (1994) *Cellular Automata and Complexity*, Addison Wesley Publishing Company, Reading, MA.

Wu, F. (1998) Simland: A Prototype to Simulate Land Conversion through the Integrated GIS and CA with AHP-Derived Transition Rules, *International Journal of Geographic Information Science*, *2*, 63-82.

Xia, L. and Yeh, A. G. (2000) Modeling Sustainable Urban Development by the Integration of Constrained Cellular Automata and GIS, *International Journal of Geographic Information Science*, *14*, 131-152.