Environment and Planning A 2005, volume 37, pages 1373 – 1394

DOI:10.1068/a3784

Agents, cells, and cities: new representational models for simulating multiscale urban dynamics

Michael Batty
Centre for Advanced Spatial Analysis, University College London, 1 – 19 Torrington Place, London WC1E 6BT, England; e-mail: m.batty@ucl.ac.uk
Received 22 March 2004; in revised form 23 July 2004

Abstract. New forms of representation at a fine spatial scale, in which units of space are conceived as cells and populations as individual agents, are currently changing the way we are able to simulate the evolution of cities. In this paper I show how these new approaches are consistent with traditional urban models that have gone before, with the emphasis no longer being on spatial interaction but on development dynamics and local movement. I first introduce ideas about urban simulation based on spatial evolution as reaction and diffusion, showing how problems conceived in terms of cells and/or agents enable new implementations of this generic model. I sketch the rudiments of cellular automata which emphasise rules for development transition, and agent-based models which focus on how individuals respond to environmental attributes encoded in cellular landscapes. I illustrate these exemplars through models of residential location. Three applications are then presented at very different spatial scales: pedestrian movement at the building scale, the evolution of systems of cities at a regional scale, and urban growth at the city scale. I conclude with proposals that formal policy analysis in this domain should always be informed by more than one approach.

1 From pattern to process

The earliest mathematical models of cities, developed and implemented in the 1950s and 1960s, were focused entirely on simulating patterns of land use and transport at a single cross-section in time. These models assumed that cities were structures with distinct causes and effects, that what was observed in cities could be factored into different spatial patterns with unique roles as either 'independent causes' or 'dependent effects'. There was some logic to this separation. In the industrial city at least, residential patterns and related services were clustered around export-orientated employment, while transport routes determining the relative pattern of accessibility provided differential competitive advantage which in turn dictated where the most profitable and affordable locations for various activities might lie.

What gave this approach credence was the idea that cities reflected a spatial and structural equilibrium which was relatively unchanging in time. If we could find strong enough correlations within spatial structure to enable robust predictions to be made from one pattern to another, then it mattered little how such a transformation was actually accomplished. In short, the process of moving from one pattern to another could be as simple as possible if the association between them was significant and robust enough. Thus models came to be fashioned around relatively simple associations at an aggregate scale at which variations in spatial patterning were least. Parsimony was the goal, with models being constructed to meet the traditional canons of scientific inquiry: first, the reproduction of existing urban spatial structures, ideally in many different places using the same model, and hence, if their performance was good enough, their subsequent use in prediction.

As we now know, this rather superficial view reflects the fact that, at an aggregative enough scale, all the volatility and dynamism of land-use change in the city are ironed out, smoothed away. When we dig under this surface, this apparent equilibrium is far from being the stable and well-behaved system that we once assumed. In fact, most traditional urban models did not generate extreme predictions in that their users were sufficiently wily to be able to 'massage' their outputs into a form acceptable to their 'client'. But worrying structural features that could lead to untenable forecasts remained. Used sensibly, however, such models still provide good short-term forecasting techniques with many versions in use today (Wegener, 1994).

The main criticism does not revolve around their equilibrating structure and aggregative nature per se. It is more that they do not address the concerns of contemporary planning and policy analysis, now strongly orientated to questions of regeneration, segregation, polarisation, economic development, and environmental quality. Urban sprawl and transportation are still within their ambit, but these new problems exist at a scale these models do not reach. To develop models which simulate finer scale actions requires significant disaggregation, often to the point at which individuals and certainly groups need to be explicitly and formally represented. And it is a consequence of such disaggregation that temporal change comes much more centrally onto the agenda.

In the quest to address different urban issues, we come full circle to the need for simulating urban dynamics; in short, we need to simulate the processes that generate the spatial patterns we observe as if cities are in equilibrium. None of this is very new for there have been many attempts at making static models dynamic (Batty, 1971; Forrester, 1969) as well as attempts to ground these traditional models in the new nonlinear mathematics of chaos, catastrophe, and bifurcation. But all these explorations have failed to yield models that are significantly different from their aggregate predecessors. What has happened is that a new generation of thinking, based not on aggregative, equilibrium-seeking assumptions and without any formal assumptions about dynamics whatsoever has emerged, consistent with models of how activities produce emergent social structures from the bottom up (Epstein and Axtell, 1996). These are the models that I will review in this paper. I will show how new ways of representing change in urban systems through rule-based decision processes can be represented as individual urban elements which I call 'agents' and 'cells'.

I will begin by outlining principles for urban simulation which tie these newer styles of model to the old, defining new ways of representation through three exemplars: first, a traditional urban model in which one spatial pattern (residential location) is immediately predictable from another (accessibility), with this transformation based on an arbitrary allocation or assignment not matched to the reality of how this actually happens; second, a model predicting the same but with the spatial assignment subject to an evolution at the most local level where what has already happened dictates, to an extent, what will happen; and, third, a model in which this evolution is given active form through agents that compare different places in their quest to optimise their residential location. These three exemplars set the scene for a review of new models at three different spatial scales: first, I illustrate movement in buildings and streets at the very small scale at which agents are walkers or pedestrians; second, at the very large scale, I use agents to grow a system of cities, evolving a landscape in which migration generates urban agglomerations; and, third, I generalise agents to cells, simulating the evolution of a metropolitan area. These exemplars and applications provide us with clear conclusions on the problems and opportunities this new perspective offers.

2 Generic structures for urban simulation

All urban models can be written in functional form as some convolution of independent and dependent variables, parameters, and random errors which encapsulate noise or uncertainty in data and behaviour. Assuming K dependent variables Y_i^k, $k = 1, 2, ..., K$, M independent variables X_i^m, $m = 1, 2, ..., M$, $K+M$ parameters

λ^l, $l = 1, 2, ..., K+M$, and a composite random variation term ε_i^k, where the subscript i refers to a location (a census tract, a cell, a parcel), the generic model can be specified as

$$Y_i^k = \mathrm{f}\{\boldsymbol{Y}_i, \boldsymbol{X}_i, \boldsymbol{\lambda}, \varepsilon_i^k\} . \tag{1}$$

The bold symbols define vectors of variables and parameters. This structure contains two features which are often missing in specific models—the positive feedback effect associated with the dependent variables and the error term absent from deterministic models. Equation (1) thus implies a process which is temporal in an implicit sense, for it is inconceivable that such causal effects take place instantaneously.

The simplest model scales one spatial pattern $\{X_i^1\}$ to another $\{Y_i^1\}$ as $Y_i^1 = \lambda X_i^1$, where the causal effect is a proportionate one. In fact, this structure was used quite widely for the first land-use–transportation models in the 1950s. Hansen (1959) proposed his residential location model in these terms as

$$P_i \sim \sum_j \frac{E_j}{d_{ij}^{\alpha}}, \tag{2}$$

where P_i is defined as population in location i, E_j as employment in location j, d_{ij} the travel distance (or travel cost) between locations i and j, and α a parameter reflecting the friction of distance. Equation (2) is the well-known measure of potential or accessibility (Stewart and Warntz, 1958); in this case the accessibility of residential location i to all employment locations j. In applications of this model, equation (2) is usually scaled to ensure that the total population sums to a predetermined total P, $P = \sum_i P_i$ or to an increment of total population ΔP, $\Delta P = \sum_i \Delta P_i$ where ΔP_i, not P_i, is predicted from equation (2).

The largest number of land-use–transportation models developed to date have been based on accessibility equations of this kind. The most widely used structure involves a generalisation of this to two dependent activities, population and employment, as

$$\left.\begin{aligned} P_i(z+1) &\sim \sum_j \frac{E_j(z)}{d_{ij}^{\alpha}}, \\ E_j(z+1) &\sim \sum_i \frac{P_i(z)}{d_{ij}^{\beta}}, \end{aligned}\right\} \tag{3}$$

where the iterator z simply represents the way positive feedback and simultaneity enter the computational process. Iterating in this way until convergence (which is usually guaranteed) provides a unique solution to the two nonlinear equations in equation (3) with respect to the two unknown parameters on distance, α and β.

The first model based on equations (3) is due to Lowry (1964) although it usually involves another independent variable—basic employment—which is added as a driver to the second (employment) equation in the system. Several well-known versions of this model exist (see Batty, 1976) in which the simultaneity between employment and population is conceived of as a multiplier process, thus breaking the positive feedback cycle proposed by Lowry. The structure as stated does not compute spatial interaction per se, although this is implicit in the definition of accessibility. The process used to map the two patterns of population and employment into one another is implicitly dynamic in an artificial way in that the model is started with estimates of population and employment often taken from observed data, and then driven to solution through computer time. This iterative process can be thought of as having a parallel in real time which has been exploited in some temporally dynamic versions, although the usual way of making these models dynamic is simply to compute the increment or decrement of activity through time by using the same structure.

A more appropriate temporal extension involves its specification in reaction–diffusion equations. Here we must introduce scalars directly into the model, for these structures, unlike those on which traditional land-use–transportation models have been built, deal both with growth (and decline) of activities in time and with their distribution across space. I will state only the model for the population equation, for others follow by analogy. Then

$$P_i(t+1) = \lambda P_i(t) + \psi \sum_j \frac{E_j(t)}{d_{ij}^{\alpha}} + \varepsilon_i^P(t), \tag{4}$$

where the postscripts t and $t+1$ refer to time instants, the scaling factors λ and ψ enable appropriate magnitudes to be grown, and the error term $\varepsilon_i^P(t)$ simply provides some noise. In equation (4) the first term on the right-hand side is the action/reaction and the second, the accessibility term, is the diffusion. In this sense, as population changes through time, it is always a function of population at the previous time period (the positive feedback effect) and employment in other locations which acts to diffuse population around its centres. Structures such as these when made operational, are often linked to constraints which enable discontinuities and thresholds to be met. For example, the Lowry model in equations (3) is usually subject to capacity constraints on the density of cells whereas reaction–diffusion equations such as equation (4) are often operationalised by using discrete cellular automata (CA) methods as illustrated below.

Finally we need to show how these structures are consistent with movement through diffusion. If we write the accessibility term in equation (2) as

$$p_{ij}(t) \sim \frac{E_j(t)}{d_{ij}^{\alpha}}, \tag{5}$$

where the term $p_{ij}(t)$ is now the explicit interaction/diffusion between locations i and j, we can state a more complete interaction relation adding the reaction and noise terms as

$$P_{ij}(t+1) = P_i(t) + p_{ij}(t) + \varepsilon_i^P(t). \tag{6}$$

If we now use spatial interaction accounting to examine the population in location i at time $t+1$, we derive the conventional reaction–diffusion equation in equation (4) above as

$$P_i(t+1) = \sum_j P_{ij}(t) \sim P_i(t) + \sum_j \frac{E_j(t)}{d_{ij}^{\alpha}} + \varepsilon_i^P(t). \tag{7}$$

In fact, there is really a rather strong tie to traditional spatial interaction theory and gravitational modelling of city and transport system if we use these forms. Adding the reaction term into the product of the diffusion term and forgetting the noise gives

$$P_{ij}(t) = P_i(t)\frac{E_j(t)}{d_{ij}^{\alpha}}, \tag{8}$$

and with appropriate scaling, total population and total employment can be computed in the same way as constraints are handled in Wilson's (1970) family of spatial interaction models.

I have presented these models in this form so that we can see how disaggregation leads to objects of interest which have a degree of discreteness and wholeness quite different from traditional conceptions of population and employment, hitherto the main working variables of traditional land-use–transportation models. At much finer spatial scales, we reach a level at which the zones or tracts—cells I will call them—become so small that it is appropriate to consider each to have only one state. In other words, imagine that cell i houses only one unit of population, one household. If a cell were to be developed, then $P_i(t) = 1$, if not, then $P_i(t) = 0$. At this point, we might decide that the population might be better represented as something other than a cell; for example, as an independent object or agent. However, the cell would still be of interest in that it would contain land on which the household resides and, as such, would act as a source for attributes of the built or natural environment. This idea clearly changes the nature of the above equations but only in the way they are computed. To make this clear, I must first define the rules needed to work with cities which are represented by cells or agents or both.

3 Representations and aggregations: cells, agents, neighbourhoods, and rules

We will still define a cell as a location i but with the understanding that magnitudes associated with an activity in any cell are usually computed by adding cells within some larger neighbourhood. Most of our traditional models can be defined in terms of cells but, when it comes to the definition of agents, there is no such association with particular locations. An agent k is thus an object $\{w^k\}$ which has attributes and, at any point in time t, is associated with a cell i. In this context, it is defined as $w_i^k(t)$; more than one agent can exist or be associated with a cell and agents can of course move between cells. This changes our conception yet again in that the magnitudes of agents associated with any cell i are computed by adding up the numbers of agents in cell i. Agent-based representation is thus much more general than cellular models in that agents usually exist on a landscape of cells whereas, in cellular models, cells are agents.

I will deal with cellular models first. The dynamic which drives such models involves defining rules which enable a cell i to change its state over some interval of time from t to $t+1$. There is usually a limited number of states which any cell can take on, the simplest being developed or nondeveloped, urban or rural, in which the variable $P_i(t)$ is defined appropriately as 1 or 0. Change in any cell is some function of reaction, diffusion, and/or randomness as implied in equation (4), but the critical feature is that the neighbourhood over which diffusion takes place is strictly limited. In traditional CA, this neighbourhood removes any 'action-at-a-distance', with interaction/diffusion confined to the cells immediately adjacent to the cell in question. This is based on the quite obvious notion in physical systems that, when a gas or liquid diffuses, its constituents must move to adjacent locations or cells in order to travel any distance. In urban models, this is appropriate for pollution or walking, for example, but it is problematic for movement involving more manufactured kinds of technology. People can hop over intervening distances between places but particles cannot. In fact, this neighbourhood restriction is often relaxed in many applications and it is more appropriate to refer to these as cell-space (CS) models (Albin, 1975; Couclelis, 1985).

Why locality is so important in CA revolves around the idea that local action leads, in many circumstances, to global order, to emergent structure. Local rules which are applied routinely often lead to structures in the large that look highly ordered but cannot be predicted from any top-down process or model. CA are thus excellent examples of local rules which lead to surprising and possibly unexpected structures. To demonstrate how CA actually work, let us implement the reaction–diffusion structure in equation (4), first without noise and then with. We will structure the solution by examining a conditional reaction first and then consider the diffusion. The cellular array consists of developed, $P_i(t) = 1$, and nondeveloped (empty), $P_i(t) = 0$, cells. We state that a cell can change its state—react—only if it is empty, meaning that already developed cells will not change and remain developed. For reaction

$$\text{if } P_i(t) = 0, \begin{cases} \text{then begin the diffusion test}, \\ \text{otherwise}, \quad P_i(t+1) = 1. \end{cases} \tag{9}$$

If this test is passed, we then see whether or not development can diffuse to an empty cell. That is,

$$\text{if } \Phi_{\min} \leqslant \sum_{j \in \Omega_i} \frac{P_j(t)}{d_{ij}^{\alpha}} \leqslant \Phi_{\max}, \begin{cases} \text{then}, & P_i(t+1) = 1, \\ \text{otherwise}, & P_i(t+1) = 0, \end{cases} \tag{10}$$

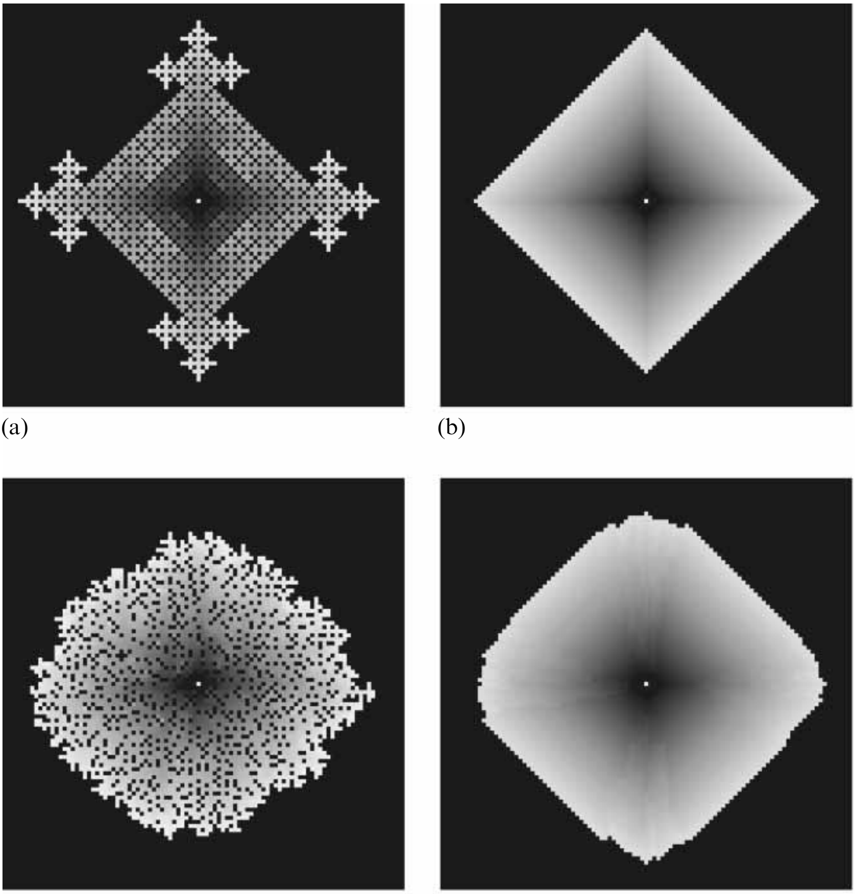
where $\Phi_{\min}$ is the minimum and $\Phi_{\max}$ the maximum access thresholds of the neighbourhood Ω_i that need to be met if the cell is to be developed. In fact, this access threshold is entirely local being simply a count of developed cells. If we define $\Omega_i = \{N, S, E, W\}$, cells which are north, south, west, and east of the cell i—the so-called von Neumann neighbourhood—then each distance in equation (10) is the same, that is, $d_{iN} = d_{iS} = d_{iE} = d_{iW} = 1$, and equation (10) reduces to a count of cells in the neighbourhood.

There are many different rules based on neighbourhood counting. As Wolfram (1994) shows, there is a combinatorially explosive number of rules for even the simplest of two-dimensional CA and it is not usually possible to classify these. There are other rules too, based on voting, for example, as in counting cells associated with preferences (Schelling, 1978) in which states are changed subject to thresholds defined in terms of majorities or minorities. There are different kinds of neighbourhood to consider, such as the eight-cell Moore neighbourhood in contrast to the four-cell von Neumann. All these give rise to many possibilities but to illustrate the essential feature of CA, I will show what happens when we plant a seed in the centre of a square cellular space and then grow the structure—city if you like—around its central business district (CBD). There is one special case worth noting before we continue. We have assumed that a reaction takes place only if the cell is empty as in equation (9) but, if this rule is abandoned, then only the diffusion takes place through equation (10). If we set $\Phi_{min} = 2$ and $\Phi_{\max} = 3$, this generates Conway's Game of Life (Gardner, 1970). When the number of cells is less than 2, a cell which is already developed 'dies' through isolation. When the number is greater then 3, then cell 'dies' through overcrowding. When the number is not between 2 and 3, an already developed cell remains developed but an empty cell is also developed, giving rise to a 'birth'. These simple rules lead to structures which are as self-perpetuating as 'life itself' (Poundstone, 1985).

Figures 1(a) and (b) show two different structures for $\Phi_{\min} = 0$ and $\Phi_{\max} = 1$, and for $\Phi_{\min} = 0$ and $\Phi_{\max} = 8$. The grey tones indicate the order in which the cells are developed. It is quite clear in figure 1(a) how simple rules lead to global patterns; how local rules applied over and over again repeat themselves at different scales generating self-similar, fractal-like structures. In figure 1(b) the filled pattern can in fact be generated when every value of $\Phi_{\max} > 1$, because of the way the von Neumann neighbourhood restricts the counting rule. In figures 1(c) and 1(d), some noise has been added, making the reaction and diffusion rules in equations (9) and (10) subject to a metarule

$$\text{if } P_i(t+1) = 1, \text{ and } \varepsilon_i^P(t+1) < \Lambda, \begin{cases} \text{then}, & P_i(t+1) = 1, \\ \text{otherwise}, & P_i(t+1) = 0. \end{cases} \tag{11}$$

Λ is a threshold above which development is sustained if the random event $\varepsilon_i^P(t+1)$ occurs. This is accomplished through drawing random numbers with appropriate scaling and using these values to modify the diffusion threshold.



(a) (b) (c) (d)

Figure 1. Local rules leading to global patterns: generating concentric structures in von Neumann neighbourhoods with and without noise: (a) with $\Phi_{\min} = 0$ and $\Phi_{\max} = 1$, (b) with $\Phi_{\min} = 0$ and $\Phi_{\max} = 8$, (c) as (a) with noise, (d) as (b) with noise.

Introducing agents into this mix adds an entirely new dimension to such dynamics. An agent defined as $w^k = 1$, where there are K agents in total, $K = \sum_k w^k$, always exists with reference to a location i as $w_i^k(t)$. The number of agents in any cell is given by $w_i(t) = \sum_k w_i^k(t)$. There is a special class of such agent-based models called 'active-walker' models (Kayser et al, 1992) in which there is a strict separation between the landscape of cells $\{P_i(t)\}$ and the walkers who exist on the landscape $\{w_i^k(t)\}$. Agents can change the landscape on which they walk and the landscape changes the agents in that 'it' directs them where to walk. This can be specified in a set of coupled equations

$$\left.\begin{aligned} P_i(t+1) &= \mathrm{f}[w_i^k(t),\, P_i(t)], \\ w_i^k(t+1) &= \mathrm{g}[P_i(t),\, w_i^k(t)], \end{aligned}\right\} \qquad (12)$$

where the two functions f and g represent the landscaping and walker movement functions, respectively.

One feature of 'walking' is that the activity is entirely local. Walkers move only to adjacent cells and, as in CA, the same kinds of local rules and neighbourhoods apply.

Let us now write the joint reaction–diffusion movement function as

$$P_{ij}(t) = \frac{P_i(t)P_j(t)}{d_{ij}^{\alpha}}, \tag{13}$$

where this interaction is possible only if cells i and j are part of the same local neighbourhood Ω. In this case, $d_{ij} = 1$, and $P_{ij}(t) = P_i(t)P_j(t)$. We now define the landscape function as an encoding of the geometry of the system: if $P_i(t+1) = 0$, the space is empty and it is possible to walk on it, whereas if $P_i(t+1) = 1$, it is occupied, with a building say, or is illegal for walking upon. Thus the only possible moves in the system are those in the matrix elements defined by the conjunction $\rho_{ij}(t) = P_i(t) \wedge P_j(t) = 0$. It is now very easy to fashion a simple walking model. An agent at cell i, $w_i^k(t)$, can walk to cell j as $w_j^k(t+1)$ if and only if $\rho_{ij}(t) = 0$. However, this would result in mindless walking because there is nothing else on the landscape to direct motion. A feature of the problems that will be illustrated here is that there is an objective in terms of walking and this can be encoded into the landscape as some form of location-specific resource, $R_i(t)$. This may vary through time by being consumed or replenished, but it serves to direct the walker into available cells in which the resource is optimal in some local way. Thus the walking model might be written as

$$\text{if } \rho_{ij}(t) = 0, \text{ and } R_j(t) = \max_{l \in \Omega_i}\{R_l(t)\}, \begin{cases} \text{then,} & w_i^k(t) \rightarrow w_j^k(t+1), \\ \text{otherwise,} & w_i^k(t) \rightarrow w_i^k(t+1). \end{cases} \tag{14}$$

The reaction–diffusion structure is complete when we add some noise. As with our other models, randomness is essential when we have many walkers so that we can simulate slight deviations from intended direction and other elements of real-world uncertainty (Helbing et al, 2001).

This algorithm can be easily embedded into the generic structures introduced previously. The local optimality which this implies is easily visualised if we imagine an accessibility surface focused on the centre of a city. A walker arrives at the edge with a view to finding the most accessible location. The walking algorithm in equations (13) to (14) not only enables the walker to climb this surface to the optimal point but also to circumnavigate any obstacles. If the access surface is perfectly symmetrical and convex-up, centred on the CBD say, then the algorithm will find the *optimum optimorum*. If there are instabilities in the surface, the algorithm will detect local optima and thus, in real problems, additional mechanisms are required to reduce the possibility of suboptimality. Many examples exist in the simulation of pedestrian behaviour and these are being generalised to other kinds of human motion by means of active particle techniques (Schweitzer, 2003). However, the most impressive attempt to date for socio-economic systems is a model in which agents optimise their consumption by climbing a resource surface, which is called a 'sugarscape' (Epstein and Axtell, 1996).

4 Exemplars: static patterns, cellular growth, and agent-based diffusion

Although I have sketched the rudiments of a general structure for urban simulation, it is far short of a well-worked-out theory. At this point, I will shift tack to illustrate the theories in more practical terms as it is worth emphasising just how close to one another are the different modelling paradigms developed over the last fifty years. I will introduce three hypothetical applications—exemplars—of the way cells and agents can be used to represent the structures of the last section. The focus will be upon explaining residential location in terms of cells becoming occupied by households who seek to optimise their accessibility which reflects the classic trade-off between a consumer's demand to minimise distance travelled and a desire to capture as much living space as possible (Alonso, 1964).

The first exemplar operationalises Hansen's (1959) residential model in which accessibility in equation (2) is measured on a cellular landscape around a CBD. This is measured linearly by using 'negative' distance $D - d_{i0}$, whereas space available is measured as a 'positive' nonlinear function of distance ξd_{i0}^{η}. D is some limiting distance at the edge of the city, ξ and η are parameters, and the CBD is defined at $j = 0$. In fact, it makes more sense presentationally to describe locations i, j in terms of coordinates x, y, where accessibility d_{xy} and space available s_{xy} are given by

$$\left.\begin{aligned} d_{xy} &= D - (|x - x_o| + |y - y_o|), \\ s_{xy} &= \xi\{[(x - x_o)^2 + (y - y_o)^2]^{1/2}\}^{\eta}. \end{aligned}\right\} \qquad (15)$$

These values associated with cell x, y are defined respect to the CBD at x_0, y_0, and accessibility T_{xy} is a simple sum of these two components

$$T_{xy} = d_{xy} + s_{xy} \,. \qquad (16)$$

To differentiate the two equations in expression (15), a Manhattan-like distance is defined for travel to the CBD, and a direct crow-fly distance from the CBD with respect to the space available.

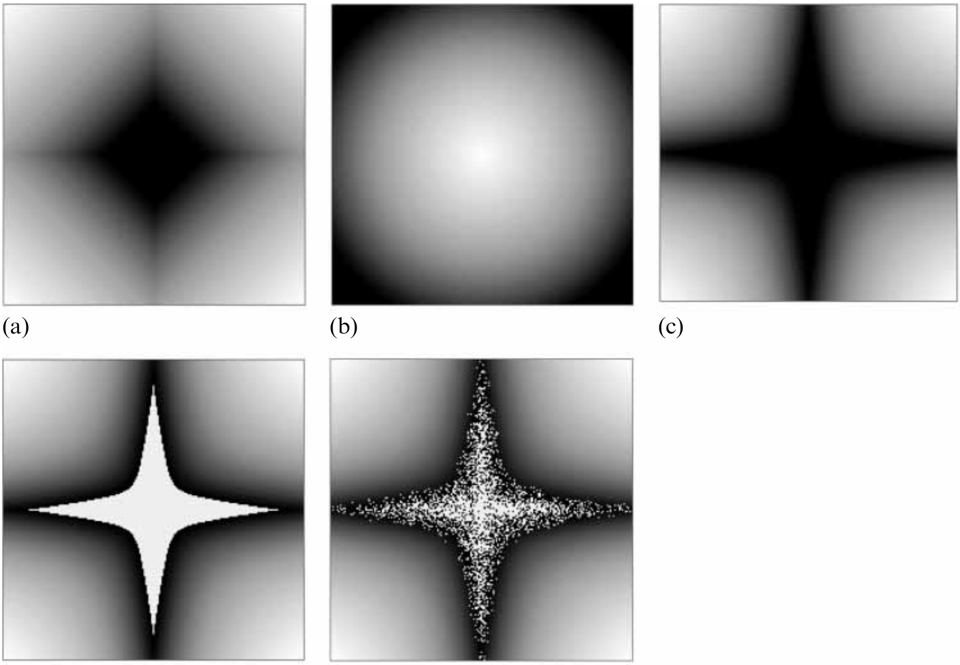
The model is very easy to state. We define one unit of population (a household, say) associated with a cell as $P_{xy} = 1$, the total units of population to be allocated as $P = \sum_{xy} P_{xy}$, and we allocate population to cells so that

$$P_{xy} = 1, \text{ and } P_{x'y'} = 0, \text{ if } T_{xy} > T_{x'y'} \,. \qquad (17)$$

This is a simple assignment that ensures that every household is in a cell which has higher accessibility than any empty cell and that this allocation is exhaustive. To illustrate its application, a cellular space of 201×201 pixels is used in which the central cell 101, 101 is defined as the CBD. Figure 2(a) (over) shows the linear travel accessibility, figure 2(b) the space available, and figure 2(c), the aggregate accessibility, all defined by means of equations (15) and (16) where $D = 101$, $\xi = 0.4$, and $\eta = 0.1$. In figure 2(d), 4000 households units are allocated where it is clear that the symmetric pattern generated simply reflects the form of the aggregate accessibility in figure 2(c). This model is based entirely on an implicit diffusion process with no growth or decline (reaction) and no error or noise. It is easy enough to add noise as ε_{xy} and, with an appropriate scaling, allocating 2000 households leads to the pattern shown in figure 2(e). It is clear from these images that this kind of model is simply a mapping of one pattern onto another. It has all the hallmarks of a system generated from the top down without the kinds of dynamics and bottom-up processes which are so necessary to understand how cities change and evolve.

The second example is even simpler in that the conception of travel accessibility and space available is entirely local. This is a model in which we seek to locate households around a CBD which is the first active location initiating the development process. We will now mix coordinate with index notation. For development to occur at time $t+1$ in cell i, the cell must be linked to the growing city—that is, it must be adjacent to some already developed cell $P_j(t) = 1$, any $j \in \Omega_i$, and the amount of space around this cell i must be a maximum for the system. We can easily implement these rules by using the following conditional:

$$\left.\begin{aligned} &\text{if } P_i(t) = 0, \text{ and } \sum_{j \in \Omega_i} P_j(t) > 0, \text{ and } N_i(t) = \min_i \sum_{j \in \Omega_i} P_j(t), \\ &\qquad \text{then,} \qquad P_i(t+1) = 1, \\ &\qquad \text{otherwise,} \quad P_i(t+1) = 0, \end{aligned}\right\} \qquad (18)$$



(a) (b) (c)

(d) (e)

Figure 2. Hansen's (1959) accessibility model implemented as a cellular automata with and without noise: (a) travel accessibility, (b) space available, (c) composite accessibility, (d) allocation without noise, (e) allocation with noise.

where $N_i(t)$ is the number of neighbours around cell i which needs to be a minimum for the number of empty spaces to be a maximum and the cell to be developed.

We begin this process with the central cell as the seed, that is, $P_{i=x_0y_0} = 1$ and this leads to the structure shown in figure 3(a) where 4750 cells have been occupied. It is clear that, as the structure grows, it becomes compact as the average number of neighbours for the entire space begins to converge. The pattern of neighbours is shown in figure 3(b), and at $t = 1600$ when the simulation is stopped, the average number is 2.78. Although this structure can be grown only from the bottom up, there is still a uniformity about its morphology which is unrealistic.

The third model adds the notion of agency to CA. Essentially the model is the same as the previous two in that all the action begins at the CBD. Agents w^k are launched at the edge of the city space and then engage in a random walk. If they walk outside the city space, they are moved back to the edge, but if they walk to a cell adjacent to one already developed, given by $P_j(t) = 1$, any $j \in \Omega_i$, they decide to locate there and the cell i is developed, that is, $P_i(t+1) = 1$. The simulation is begun with the central (CBD) cell being seeded, and with all the walkers being located randomly at the edge of the city space, that is, $w_i^k(0)$, $\forall k$, where $d_{i=xy} \geqslant D$. Random walking is simply a method of exploring the space and, in this model, walkers move randomly to adjacent cells in their von Neumann neighbourhood $j \in \Omega_i = N, S, E, W$. Essentially the structure grows when walkers make first contact with developed cells and, as in the previous model, the fact that cells are developed one by one ensures that the structure remains connected.

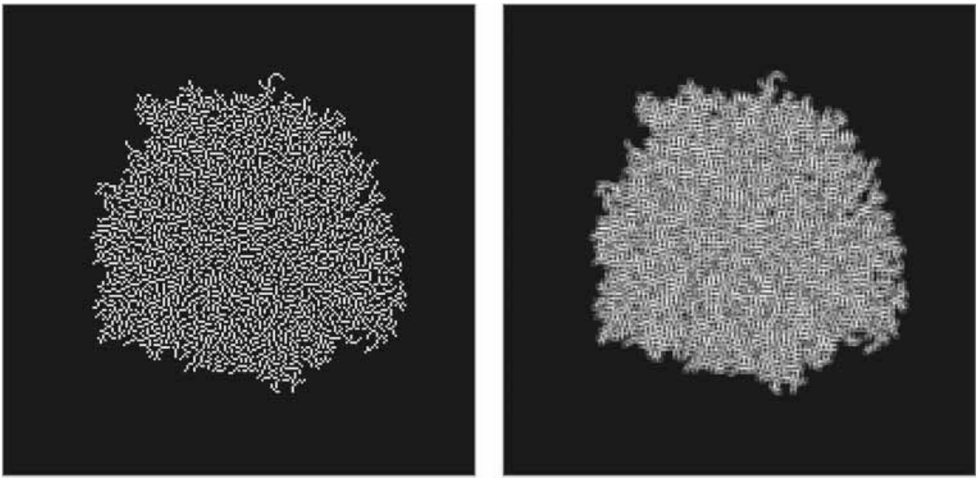

(a) (b)

Figure 3. Connected cellular growth around a central seed with local maximisation of residential space: (a) developed cells, (b) numbers of neighbours.

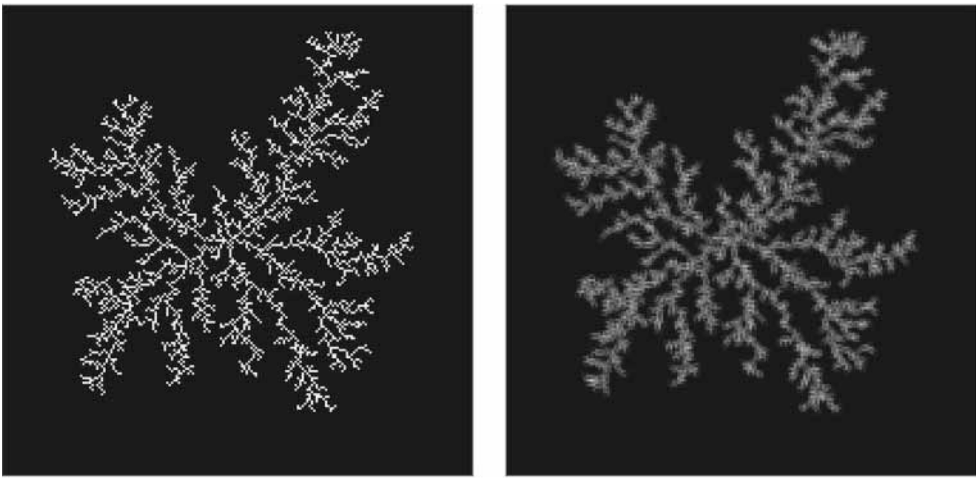
Formally this model can be posed as follows. At each time t, we execute the test

$$\left.\begin{array}{llll} \text{if } w_i^k(t) \;=\; 1, \text{ and} \displaystyle\sum_{j \in \Omega_i} P_j(t) \;>\; 0, \text{ and } P_i(t) \;=\; 0, & & & \\ \quad \text{then,} & P_i(t+1) \;=\; 1, & \text{and } w_i^k(t) \;\rightarrow\; w_z^k(t+1), & \\ \quad \text{otherwise,} & P_i(t+1) \;=\; 0, & \text{and } w_i^k(t) \;\rightarrow\; w_j^k(t+1). & \end{array}\right\} \tag{19}$$

In the first line of equation (19), we test to see whether the cell location in which the walker is located is empty and whether there is development in its neighbourhood. In the second line if this is so, the cell is developed and the walker returns to its initiating point on the edge of the city space where the index z = random x, y with $d_{xy} \geqslant D$. The third line of the test is associated with failing the test in the first line and then the walker simply continues walking to a cell in its neighbourhood, chosen randomly as random $j \in \Omega_i = N, S, E, W$.

This is the very well-known model of diffusion-limited aggregation (DLA). It was first introduced by Witten and Sander (1981) and has been used to grow many kinds of structures which have a dendritic form. Essentially what is generated is surprising. Unlike the previous model in which the mass is compact, this structure is much more tree like, with branches reaching out to capture as much space as possible, not unlike the way cities grow into their surrounding hinterland. It is essentially a fractal structure and its morphology can be tuned to produce dendrites of differing compaction with varying fractal dimensions. The dielectric breakdown model represents an equivalent form. Rather than being conceived as a bombardment of a growing mass with particles from the edge, growth takes place from the centre where tips of the evolving structure have the greatest probability of growth (Niemeyer et al, 1984; Stanley and Ostrowksy, 1986).

The structure is illustrated in our 201×201 cellular space in figure 4(a) (over) which shows its morphology and in figure 4(b) which is a plot of the number of neighbours associated with the developed cells. As the structure grows, the average number of neighbours declines inexorably as the development reaches out into greater and greater regions of empty space, although the average number of those cells which have at least one developed neighbour is about 2.3, not so different from the



(a) (b)

Figure 4. Diffusion-limited aggregation: cellular growth from agent-based random walks: (a) developed cells, (b) numbers of neighbours.

previous model. This kind of irregularity can be generated only from the bottom up. It is a product of randomness combined with locational principles based on keeping the structure connected, or agglomerated, combined with the search for greater and greater space in which to grow.

5 Simulation at the very smallest scales: pedestrian movement in buildings and streets

The three applications represent a classification of problems at different scales which also reflect different varieties of dynamics and different assumptions about the extent to which cells and agents engage in goal-seeking activities. At the smallest scale in built environments, routine, repetitive movement based on 'fast' dynamics is the focus where the frequency of interaction is measured in terms of seconds and minutes, sometimes hours and days but never any longer. Such activities usually respond to the environment through agents 'using' what has been already created rather than recreating it over much longer time periods. In contrast, at the very largest scales, at which we are dealing with systems of cities (Berry, 1964), the dynamics is 'slow'. These are based on decisions which take place much more infrequently through migration, decisions to establish new settlements that evolve over decades, if not centuries. Somewhere in between, we will deal with the city, the mesoscale—the focus of the exemplars so far—at which the temporal scale is over years and decades.

We will begin with models that simulate small-scale movements over short time intervals in buildings and streets where the focus is on visiting places. We have already mapped out a general structure for such agent models in which movement—walking—was articulated as the intersection of geometry defining where one might walk in response to resources defining the landscape on which movement takes place. The previous notation will be retained: walking takes place on and between cells i and j adjacent to one another in appropriately defined neighbourhoods $j \in \Omega_i$, which meet the requirement that both are empty of any other activity, that is, $\rho_{ij} = 0$. For movement in streets and buildings, we assume that the matrix $\boldsymbol{\rho}$ does not vary through simulation time, and thus it defines the 'container' within which walkers respond to the resource landscape $\{R_i\}$, also unchanging in time. In fact, these models at this scale are not active-walker models at all, but passive-walker models. The landscape never changes although walkers do respond to each other.

There are many variants ranging from those in which geometry is all important to those in which the accessibility of resources takes precedence. When geometry is important, these models apply to very fine scales at the level of corridors and rooms and tend to be used to predict panic situations and evacuation events in hazardous environments. Very detailed issues involving the physics of acceleration characterise these models (Helbing et al, 2001). At the other extreme when geometry is unimportant but patronage of different locations is (as in shopping activities, for example), the attraction surface is all important (Borgers and Timmermans, 1986). In the example here, both are necessary as we will simulate movement in a complex building where geometry does dictate where people go but the attraction of different exhibits is the prime reason why people move within the building in the first place. Modifying equation (14), a walker will move

$$\text{if } \rho_{ij} = 0 \begin{cases} \text{then,} \quad w_i^k(t) \rightarrow w_j^k(t+1), \text{ where } j \leftarrow \max\limits_j \{\nabla R_{ij} + \varepsilon_j^k(t)\}, \\ \text{otherwise the walker engages in obstacle avoidance.} \end{cases} \tag{20}$$

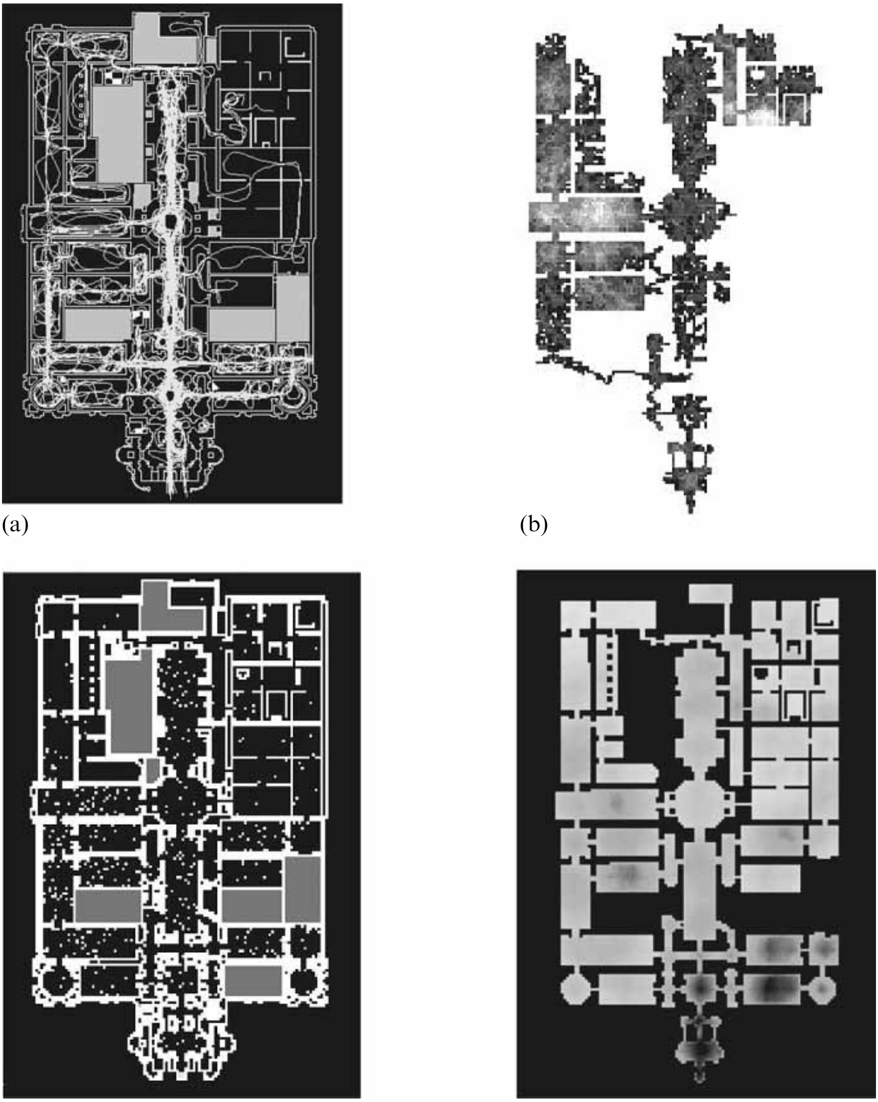
We define ∇R_{ij} as the local gradient in the resource surface in the direction from cell i to cell j which is a maximum but add to this some random noise $\varepsilon_j^k(t)$. All this does is push the walker in the direction of greater resources. In fact, obstacle avoidance is probably the more frequent occurrence in complex geometries, and routines to effect this consist of moving walkers in different directions, dependent upon their previous reactions to obstacles, how far they are able to see, and so on. The other feature involves interactions between different walkers. There are limits on congestion which involve dispersing walkers if too many attempt to reach the same location. This is simply a matter of ensuring that $\sum_k w_i^k(t)$ is within a certain threshold and initiating local movement if it is not. Panic can set in if congestion occurs across a wider region of cells and dispersion is not possible. In contrast, flocking or herding based on walkers 'following the crowd' is considered by assessing how the number of walkers $\sum_{j \in \Omega} \sum_k w_j^k(t)$ in a wider neighbourhood Ω attracts even more walkers.

To illustrate this model, my colleagues and I applied it to the movement of visitors in the Tate Gallery on London's Millbank. We have good data on the circulation patterns of 550 visitors observed over a short time period in August 1995 which are shown in figure 5(a) (over) (UAS, 1996). Paintings were then on display in forty-nine rooms of the building and the problem was simplified to consider only those visitors—some 97% of those visiting in fact—who entered the gallery through the main entrance. What we were interested in was the pattern of visitation to the various rooms. To measure this, we introduced walkers into the gallery over a short period of time and then examined the pattern of room occupancy in the steady state which emerges when the model is run through many time periods. We assess the average number of walkers visiting each room Ω_r over T time periods

$$\sum_{\tau=1}^{T} \sum_{j \in \Omega_r} \sum_k \frac{w_j^k(\tau)}{T},$$

and once this quantity stabilises, we are then in a position to assess the 'fit' of the model to the actual pattern of room visitation.

One of the greatest advantages of agent-based models is that, as we run the model with different numbers of agents, we can derive different kinds of information about the problem. In our Tate model, when we launch just one agent, we can consider this as a 'probe' to explore a complex space and in doing so, assess how well it is dealing with obstacle avoidance which is shown in figure 5(b). As we run the model with more, then



(a) (b) (c) (d)

Figure 5. Movements of real and simulated walkers in the Tate Gallery, Milbank, London: (a) walker paths in the gallery, August 1995: 12 noon to 1pm, (b) the progress of a single agent exploring the gallery, (c) a snapshot of agents moving in the gallery, (d) the steady-state pattern: visitors to cells and rooms.

many agents, we can also assess the role of randomness on the pattern of visitation, enabling the 'right' level of variation in overall spatial behaviour to be defined. Figure 5(c) illustrates a typical snapshot of agents within the gallery. Although this does tend to show these rooms more frequently patronised, we cannot say anything about the long-term pattern of visitation. In short, this is a good example in which cross-sectional patterns mean very little in terms of the longer term dynamics. Figure 5(d) shows the average patronage in the steady state in each cell, not in each room although it is possible to get a sense of the frequency of room visitation from this. The use of this model for 'what if' analysis is fairly obvious. In this case, closing or opening rooms or changing their configuration for various types of exhibition as well as showing how different kinds of attractions in rooms affect movement is what the application is all about.

I have but touched the surface of this kind of modelling as the field is currently exploding. In terms of using this particular model, the main applications have been to simulate shopping patronage in town centres (Haklay et al, 2001), to predict shortest routes in pedestrian networks, and to model street parades, in particular the Notting Hill Carnival where public safety was the main focus (Batty et al, 2003). Models at much finer scales involving panic and evacuation possibilities tend to include much more basic physics and there are strong links to CA models of traffic movement (Helbing et al, 2000). Useful summaries are provided by Schreckenberg and Sharma (2002). There are also active-walker versions of these models in which the landscape is altered by the act of walking (Helbing et al, 2001). It is to these kinds of problem that I now turn, but at a very different scale.

6 Simulation at the very largest scales: the emergence of systems of cities

We will evolve a system of cities where walkers are 'migrants' and resources are their 'jobs'. These are active-walker models in which walkers respond to jobs which define an economic landscape which in turn directs where migrants search. The landscapes have no geometry, being featureless plains in the grand sense, although it is entirely possible that geometric obstacles could be introduced and thus the mechanisms of the very small scale (in the previous section) might feature in directing walkers. The ultimate model is in fact composed of two such landscapes, the first linking people to jobs through physical networks based on origins and destinations, the second being defined in terms of 'resource' potential which enables new walkers—new migrants to the system—to be located which reflect an appropriate growth rate. It is quite possible to specify models in which these landscapes are compatible but separate, their networks simply being the consequence of where jobs and people are located. However, in the model I will present, these landscapes interact with one another through time, thus adding a metalevel coupling, defining not simply 'active-walker' models but 'active-landscape' models. In this sense, I consider this to be an extension to the state of the art in agent-based modelling.

I will first introduce a model with fixed origins for walkers and fixed destinations for resources, and this will enable us to predict the paths that walkers take between them. I will then show how the capacity of the network channels which emerge can be used to define the potential for locating new walkers. In this way, the network model is limited to a location model, thus tying together walkers and landscapes in several different ways, each reflecting various positive feedbacks and diffusive effects that drive the evolution of the system. Each origin cell for a walker will be indexed as I and each destination cell for a resource as J, where $w_I^k(0)$ and $R_J(0)$ are the initial distributions of these quantities. In fact in the first model these will not change in time as the focus is entirely on generating the networks that connect these two distributions.

The way the model works is by letting walkers move randomly through space, starting from their origins in search of resources at the given destinations. When a walker discovers a resource, essentially it tells other walkers of its discovery. But most of these walkers will not be in the same vicinity and thus the walker has to have some means of communicating this information. The walker does this by returning to its origin with the resource—back home to consume the food, if you like—and in making this trip, it lays down a path that other walkers can observe. This path will mark the straight-line distance from the origin to the destination, subject to any noise that interferes with the process. As the process continues, more and more walkers will discover resources and those walkers who have not yet discovered resources will detect the paths that lead to these resources. Ultimately, everybody will be travelling on a route that takes them directly from their origin to a resource destination.

This is quite similar to the Helbing et al (2001) model of trail formation and it figures widely in the way insect populations such as ants forage for food (Camazine et al, 2001). The model is sometimes called a swarm algorithm because, when all movement is random, this is akin to a swarm moving out from some source. It is used to predict shortest paths in the Notting Hill model in which such paths within the street network were unknown (Batty et al, 2003) although it will be used it here to predict straight-line distances in the featureless plain. In essence, if there are enough walkers swarming out from known origins $\{w_I^k(0)\}$ to known destinations $\{R_J(0)\}$, then, once such a destination has been discovered, the agent in question heads back directly from J to I, impressing a track $s_{ij}(t+1)$ on every i, j cell pair which defines this line. This is added to the existing track, if there is one, as $S_{ij}(t+1) = S_{ij}(t) + s_{ij}(t+1)$ and in this way, the track gains in capacity. Walkers who are still in search of resources then react to the gradient formed by this track $\nabla S_{ij}(t)$ following the route from cell i to cell j. Ultimately, as tracks or network channels emerge, this reflects the relative nearness of the origins to the destinations.

The formal mechanism is little different from what has been stated already, but for any walker in search of a resource

$$w_i^k(t) \rightarrow w_j^k(t+1), \text{ where } j \leftarrow \max_j\{\nabla S_{ij}(t) + \varepsilon_j^k(t)\}\,. \tag{21}$$

This procedure works in a trackless landscape in which movement is entirely dictated by random noise. Eventually all the walkers discover all the resources and the network landscape begins to stabilise in its morphology. When a track is formed as walkers who have discovered resources head back to base, it is usual simply to set $s_{ij}(t)$ to a constant which reflects a simple addition to the capacity from the actions of one agent.

An application of this network formation is illustrated in figure 6 which simulates the tracks formed between ten fixed-walker and thirteen resource locations in which 1000 walkers have been randomly assigned to the ten origins. Figure 6(a) shows the origins and destinations, and 6(b) to (d), the distribution of the agents in the 201×201 cellular space, paths taken in the landscape, and the tracks formed as a subset of all paths taken, at times $t = 50$, $t = 500$, and $t = 5000$. The convergence from random walks to nodal structure in the landscape is impressive. There is more information, however, contained in this simulation. It has been assumed that the numbers of agents visiting each resource location are unknown even though the amount of resources there may be known. It is, however, possible to compute the numbers of walkers visiting these locations in a cumulative manner which would give some indication of their size as

$$\bar{R}_J(T) \sim \sum_{\tau=1}^{T} \sum_k w_J^k(\tau)\,. \tag{22}$$

As the network stabilises, so will the numbers attracted to each resource destination and to express these in terms of the total number of walkers, it is a simple matter to scale these totals as

$$\hat{R}_J(T) = \varphi \bar{R}_J(T) = \bar{R}_J(T) \frac{\sum_k w_I^k(0)}{\sum_J \bar{R}_J(T)}\,. \tag{23}$$

$\hat{R}_J(T)$ can in fact be regarded as a measure of potential—network potential—of the resource node which can then be used to condition an extension of the model to incorporate growth in different locations.

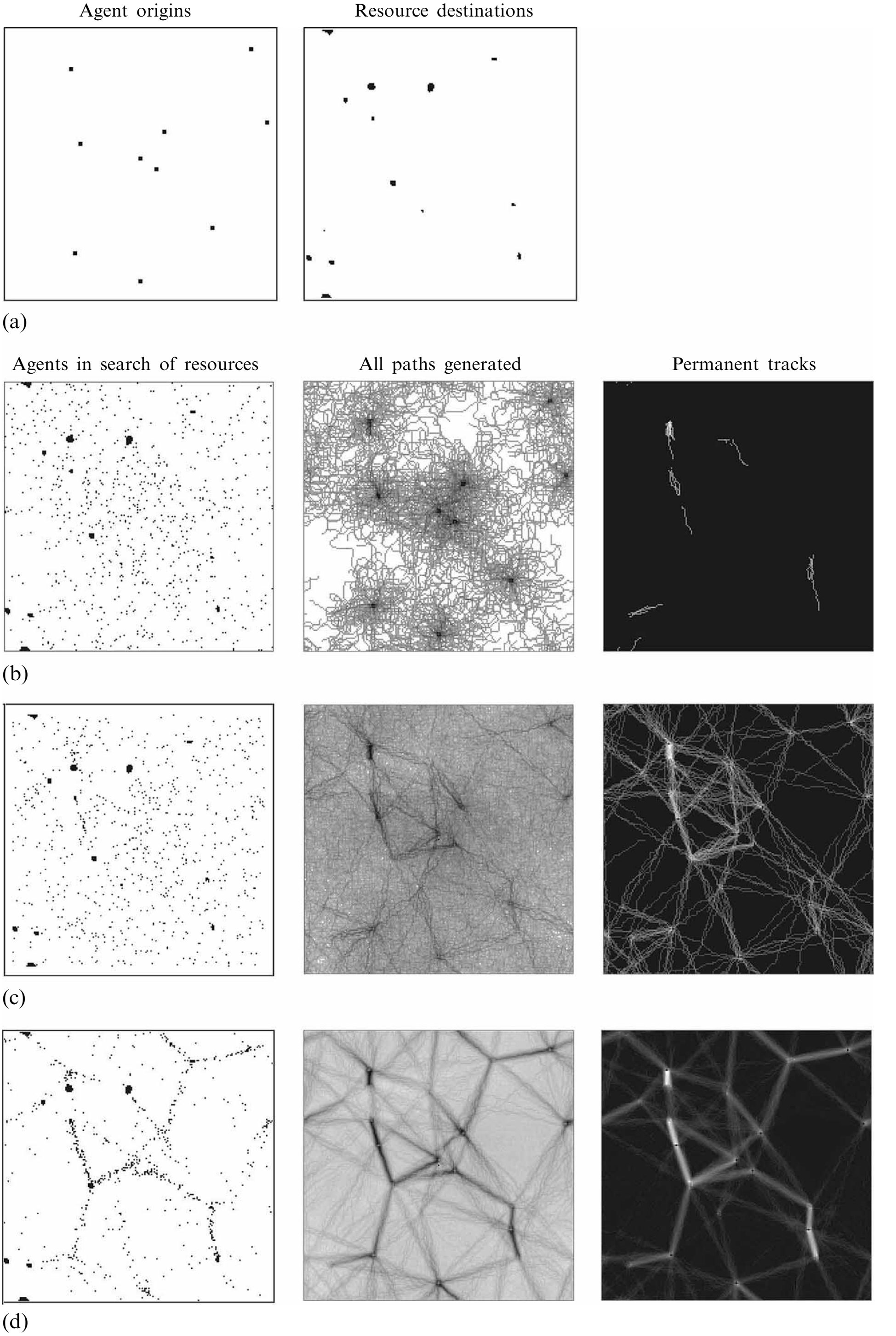


Figure 6. Network formation between walker origins and resource destinations. (a) The initial conditions for simulation are based on 10 walker origins and 13 resource destinations. Walkers search randomly, moving out from origins in search of resources which, when found, provide the rationale for laying direct tracks back to destinations. Distribution of the agents, paths taken, and permanent tracks at times (b) $t = 50$, (c) $t = 500$, (d) $t = 5000$.

Imagine we now wish to grow the number of agents from the initial base. One way of locating them would be to form a measure of potential and then to seek locations for new walkers in which this potential is maximised. We form a generic potential by using a reaction–diffusion equation

$$P_i(t+1) \sim P_i(t) + \omega\nabla^2 P_i(t) + \hat{w}_i(t+1) + \varepsilon_i(t)\,, \tag{24}$$

where ω is a weight on the diffusion term and $\hat{w}_i(t+1)$ is the location of one new agent in each time period, reflecting uniform growth through time whose location is chosen so that $\hat{w}_i(t+1)$ is determined from the cell given by $\max_j P_j(t)$. If we start with one walker, then what happens is that the first walker is located randomly as the potential surface is uniform. Reaction and diffusion ensure that this initial location survives and a path dependence then sets in which can be broken only if the noise in the system $\varepsilon_i(t)$ is large enough. In such applications, it is likely that the initial cluster will be reinforced. However, if we make the connection between generic potential and the potential interaction at the node J as $\hat{R}_J(T) = P_J(t)$, we set in motion a process in which population becomes a function of resource potential which in turn is a function of the way populations discover resources through the emergence of their networks. To make the structure more elaborate, we might introduce a second potential equation in which new resources are located as a function of population, developing a structure in which agents and landscapes interact in all possible ways.

The simpler structure in equation (24) will be illustrated in which we begin with twenty fixed resource locations and 100 walkers randomly located. As walkers begin their random walk using equation (21), new walkers are introduced, one in each time period using equation (24). The tracks in space, the potential, and the location of the

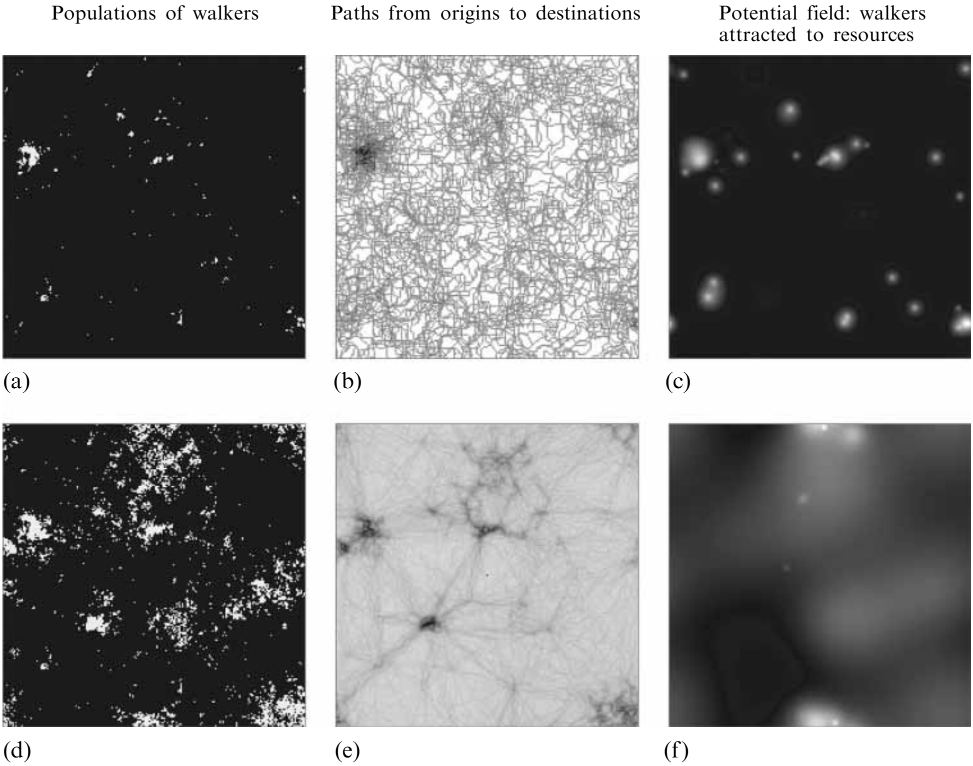


Figure 7. The generation of a coupled active urban landscape: (a)–(c) $t = 100$, (d)–(f) $t = 2000$.

populations at an early stage ($t = 100$) are shown in figures 7(a) to (c), and then at a much later stage ($t = 2000$) in figures 7(d) to (f). Although there is considerable persistence in the spatial structures generated over quite long time periods, over thousands of time periods, clusters of activity can change quite radically. The patterns produced do mirror real systems of cities in terms of their social physics (Batty, 2001; Manrubia and Zanette, 1998).

7 Cities at the mesoscale: metropolitan dynamics and urban sprawl

Between the very small and the very large scale lie cities. Many urban models at this mesoscale have recently been introduced, fashioned around aggregating agents into cells, building on the traditional idea that land use and related urban activities take place in zones. These developments have been driven by geographic information systems, treating space as a raster or pixel grid often inspired by remotely sensed digital data. CA methods provide the workhorse for these new models. Structurally they all reflect the generic reaction–diffusion process in equation (4) adapted to CA form in equations (9) to (11) but in these models, any single land use can interact with any other. Thus, for any two land uses, there can be a reaction–diffusion equation. Typically, a land use k in cell i at time t, $P_i^k(t)$, has the potential to spawn a new land use l in cell j at time $t+1$, $P_j^l(t+1)$, where the reaction and diffusion depend not only on the location and neighbourhood around i but also on the particular composition of land uses.

The model introduced here was initiated by Xie (1994) under the acronym DUEM (Dynamic Urban Evolutionary Model) and has continued to evolve through several versions (Batty et al, 1999). There are currently five land uses in the model—residential (population), manufacturing and primary uses, services and commerce, streets, and vacant land—and, in principle, twenty-five separate reaction–diffusion equations govern the way land uses spawn and constrain each other. Only residential, industrial, and service uses in fact spawn each other, whereas streets are generated by these uses but do not generate land use themselves. The rules are too specific to formalise in the simple manner illustrated previously but I can sketch the way they operate. In each time period, the probability of converting a land use k in cell i to land use l in cell j is computed as $p_{ij}^{kl}(t+1) = \mathrm{F}\{P_i^k(t),\ P_j^l(t+1)\}$, where the functional form is based on a series of 'if $\rightarrow$ then rules' of the conventional CA type as reflected in equations (9) to (11). For any land use k in cell i, the probability is first determined by the distance from cell i to cell j in the field around cell i. Any distortions in direction are added, and then this probability is checked for legitimacy against a series of density constraints on the occurrence of different land uses in the neighbourhood around cell i. If this growth probability survives, this is tested against the presence of some street pattern in the neighbourhood, for it is essential that any new use be 'near' some transport. Streets are land uses too and these are grown in a similar but slightly more restrictive manner. In this way, land development and transportation are coordinated in physical terms. This process of selecting what land uses are generated through the various tests eventually leads to a final probability matrix $[p_{ij}^{kl}(t+1)]$ from which a new land use is chosen according to some random mechanism.

There is no symmetric process to simulate decline, for this is determined by the way life cycles and aging enter the model. The model defines a life cycle for each land use, with the possibility that any land use in its early life will spawn new land uses in the manner just suggested. Uses are classified into initiating, mature, or declining (and then vacant) with a regular transition between each controlled by parameters specific to the average time each land use is in each cycle. Initiating land uses have the potential to grow new land uses of any type within a field which is wider than the neighbourhood within which they are located. Constraints posed by distance and direction first determine the probability of new land uses, but these are subject to a series of land-use-type and density constraints within the narrower neighbourhood. Overall, regional constraints dictate whether or not a cell can be occupied for a particular land use.

The version of the model introduced here enables both real and hypothetical examples to be developed, but I will simply present the hypothetical. A series of land uses of these five types have been randomly planted in a 350×250 pixel space and the default rules have been used to grow the system through time to the point at which the entire space is occupied. As the space fills up, the total quantities of land use grow logistically to upper limits but, as aging takes place, land uses become vacant and the trajectories of growth begin to oscillate. This is indicative of a simple capacitated system. As the rules are specified locally, we have no idea in advance how these will combine together to produce realistic structures, and thus the experimentation shown in figure 8 is essential for tuning the transition probabilities governing the evolution of the system.

Growth has also been simulated in real systems, such as metro Detroit, and Ann Arbor (Xie and Batty, 2005) where the focus is on population decline and jobless growth in an economy which is dominated by the continuing decline of the central city and wider peripheralisation through sprawl. There are now upwards of fifty applications of similar CA/CS models of urban growth with sustained effort taking place in some half dozen places, namely at the Research Institute for Knowledge Systems (White and Engelen, 2000), Santa Barbara (Clarke et al, 1997), Southampton and Cardiff (Wu and Martin, 2002), Hong Kong (Li and Yeh, 2000), Tel Aviv (Portugali, 2000), and Brisbane (Ward et al, 2000), as well as work at the Centre for Advanced Spatial Analysis. Most but not all these models are inspired by applications involving urban sprawl but some such as the Tel Aviv models are geared to segregation and polarisation. These applications are all subject to

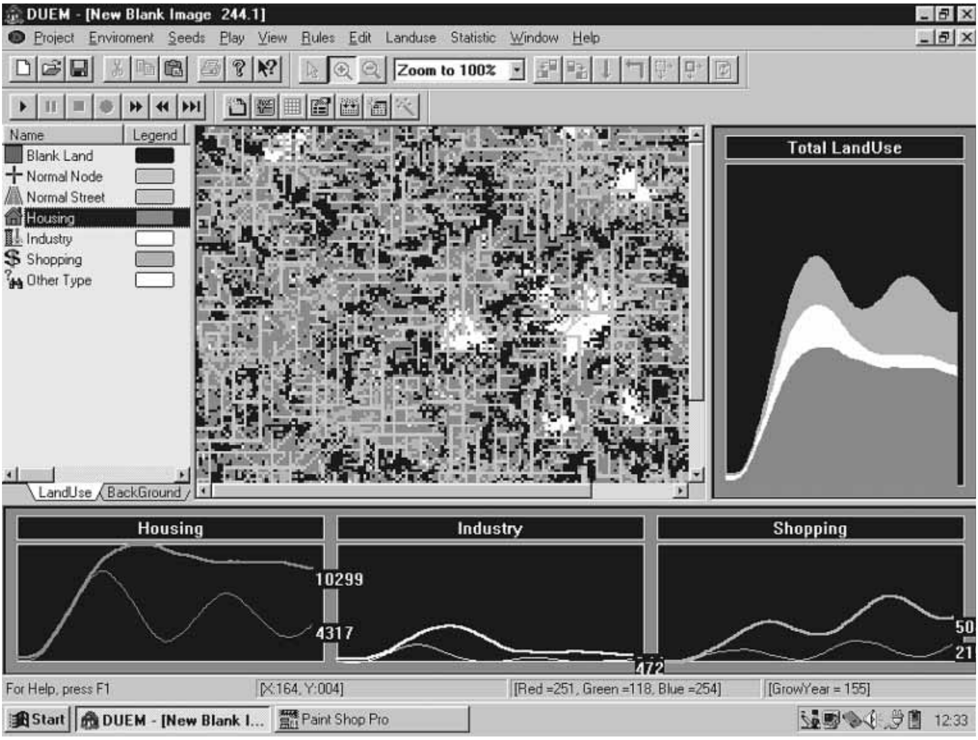


Figure 8. Logistic growth of a capacitated urban system.

the limitations that plague this area generally: the lack of explicit transportation, poor command over control totals, and rather stylised representations of land uses in cells. Nevertheless, progress is being made and there are many new developments in the pipeline.

8 Conclusions: next steps

This long review warrants a short conclusion. Currently the extent to which CA/CS and agent-based models of urban systems can be fully implemented for policy applications is quite limited. As the scale gets finer and the agents and their cells become more like real objects, their operationality increases to the point at which substantive policy applications are possible. Such is the case with the pedestrian models in which the definitions of objects are quite unambiguous in comparison with larger scales. At the largest scale, these approaches stretch the field in terms of theory and such is their role. But the biggest problem with all these models is their lack of parsimony. The richness of the data required makes calibration and estimation difficult and predictive accuracy hard to assess in terms of past simulations. Thus many of these models are demonstrated in pedagogic fashion to show emergence and path dependence, and their use in forecasting is largely restricted to very long-term futures.

It is interesting that, apart from a few notable exceptions (such as Sanders et al, 1997), hardly any agent-based models have yet been developed at the mesoscale, although there is active development for more environmentally based land-cover systems (Parker et al, 2003). Applications will doubtless increase but at the end of this review, we are still left with the perennial question which dominates all discussion of science in public affairs: to what extent can formal models be built which will provide robust enough forecasts for real policy analysis? These new approaches provide only a part of this answer. Although promising in that new forms of representation clearly get to grips somewhat more effectively with the way contemporary problems are articulated, this perspective raises a new set of questions which limit their applicability in rather different ways from traditional urban models. This is a recurrent feature of this field which suggests that not one but many different approaches will always be required.

Acknowledgements. The ESRC NEXSUS Project (L326-25-3048) has provided partial support for this research. An early version of this paper was presented to the Conference on "Framing Land Use Dynamics", held at the University of Utrecht, The Netherlands, 16–18 April 2003.

References

Albin P, 1975 *The Analysis of Complex Socio-economic Systems* (Lexington Books, Lexington, MA)
Alonso W, 1964 *Location and Land Use: Toward a General Theory of Land Rent* (Harvard University Press, Cambridge, MA)
Batty M, 1971, "Modelling cities as dynamic systems" *Nature* **231** 425–428
Batty M, 1976 *Urban Modelling: Algorithms, Calibration, Predictions* (Cambridge University Press, Cambridge)
Batty M, 2001, "Polynucleated urban landscapes" *Urban Studies* **38** 635–655
Batty M, Xie Y, Sun Z, 1999, "Modeling urban dynamics through GIS-based cellular automata" *Computers, Environments and Urban Systems* **23** 205–233
Batty M, Desyllas J, Duxbury E, 2003, "Safety in numbers? Modelling crowds and designing control for the Notting Hill Carnival" *Urban Studies* **40** 1573–1590
Berry B J L, 1964, "Cities as systems within systems of cities" *Papers and Proceedings of the Regional Science Association* **13** 147–163
Borgers A, Timmermans H, 1986, "A model of pedestrian route choice and demand for retail facilities within inner-city shopping areas" *Geographical Analysis* **18** 115–128
Camazine S, Deneubourg J-L, Franks N R, Sneyd J, Theraulaz G, Bonabeau E, 2001 *Self-organization in Biological Systems* (Princeton University Press, Princeton, NJ)
Clarke K C, Hoppen S, Gaydos L, 1997, "A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area" *Environment and Planning B: Planning and Design* **24** 247–261

Couclelis H, 1985, "Cellular worlds: a framework for modeling micro – macro dynamics" *Environment and Planning A* **17** 585 – 596
Epstein J M, Axtell R, 1996 *Growing Artificial Societies: Social Science from the Bottom Up* (MIT Press, Cambridge, MA)
Forrester J W, 1969 *Urban Dynamics* (MIT Press, Cambridge, MA)
Gardner M, 1970, "The fantastic combinations of John Conway's new solitary game of 'Life' " *Scientific American* **223** 120 – 123
Haklay M, Thurstain-Goodwin M, O'Sullivan D, Schelhorn T, 2001, " 'So go downtown': simulating pedestrian movement in town centres" *Environment and Planning B: Planning and Design* **28** 343 – 359
Hansen W G, 1959, "How accessibility shapes land use" *Journal of the American Institute of Planners* **25** 73 – 76
Helbing D, Farkas I, Vicsek T, 2000, "Simulating dynamical features of escape panic" *Nature* **407** 487 – 490
Helbing D, Molnar P, Farkas I J, Bolay K, 2001, "Self-organizing pedestrian movement" *Environment and Planning B: Planning and Design* **28** 361 – 383
Kayser D R, Aberle L K, Pochy R D, Lam L, 1992, "Active walker models: tracks and landscapes" *Physica A* **191** 17 – 24
Li X, Yeh A G-O, 2000, "Modelling sustainable urban development by the integration of constrained cellular automata and GIS" *International Journal of Geographic Information Science* **14** 131 – 152
Lowry I S, 1964 *Model of Metropolis* RM-4035-RC, Rand Corporation, Santa Monica, CA
Manrubia S C, Zanette D H, 1998, "Intermittency model for urban development" *Physical Review E* **58** 295 – 302
Niemeyer L, Pietronero L, Wiesmann H J, 1984, "Fractal dimension of dielectric breakdown" *Physical Review Letters* **52** 1033 – 1036
Parker D C, Manson S M, Janssen M A, Hoffman M J, Deadman P, 2003, "Multi-agent systems for the simulation of land-use and land-cover change: a review" *Annals of the Association of American Geographers* **93** 314 – 337
Portugali J, 2000 *Self-organization and the City* (Springer, Berlin)
Poundstone W, 1985 *The Recursive Universe* (William Morrow, New York)
Sanders L, Pumain D, Mathian H, Guerin-Pace F, Bura S, 1997, "SIMPOP: a multiagent system for the study of urbanism" *Environment and Planning B: Planning and Design* **24** 287 – 305
Schelling T C, 1978 *Micromotives and Macrobehavior* (W W Norton, New York)
Schreckenberg M, Sharma S D (Eds), 2002 *Pedestrian and Evacuation Dynamics* (Springer, Berlin)
Schweitzer F, 2003 *Brownian Agents and Active Particles: Collective Dynamics in the Natural and Social Sciences* (Springer, Berlin)
Stanley H E, Ostrowksy N (Eds), 1986 *On Growth and Form* (Martinus Nijhoff, Dordrecht)
Stewart J Q, Warntz W, 1958, "Physics of population distribution" *Journal of Regional Science* **1** 99 – 123
UAS, 1996, "Tate Gallery, Millbank: a study of the existing layout and new masterplan proposal", Unit for Architectural Studies, Bartlett School, University College, London, London
Ward D P, Murray A T, Phinn S R, 2000, "A stochastically constrained cellular model of urban growth" *Computers, Environment and Urban Systems* **24** 539 – 558
Wegener M, 1994, "Operational urban models: state of the art" *Journal of the American Planning Association* **60** 17 – 30
White R W, Engelen G, 2000, "High-resolution integrated modelling of the spatial dynamics of urban and regional systems" *Computers, Environments and Urban Systems* **24** 383 – 400
Wilson A G, 1970 *Entropy in Urban and Regional Modelling* (Pion, London)
Witten T A, Sander L M, 1981, "Diffusion-limited aggregation: a kinetic critical phenomena" *Physical Review Letters* **47** 1400 – 1403
Wolfram S, 1994 *Cellular Automata and Complexity: Collected Papers* (Addison Wesley, Reading, MA)
Wu F, Martin D, 2002, "Urban expansion simulation of Southeast England using population surface modelling and cellular automata" *Environment and Planning A* **34** 1855 – 1876
Xie Y, 1994 *Analytical Models and Algorithms for Cellular Urban Dynamics* unpublished PhD dissertation, State University of New York at Buffalo, Buffalo, NY
Xie Y, Batty M, 2005, "Integrated urban evolutionary modelling", in *Geodynamics* Eds P M Atkinson, G M Foody, S E Darby, F Wu (CRC Press, Boca Raton, FL) pp 273 – 293