Environment and Planning A 2002, volume 34, pages 1855 – 1876

DOI:10.1068/a3520

Urban expansion simulation of Southeast England using population surface modelling and cellular automata

Fulong Wu, David Martin
Department of Geography, University of Southampton, Southampton SO17 1BJ, England;
e-mail: F.Wu@soton.ac.uk, D.J.Martin@soton.ac.uk
Received 15 January 2002, in revised form 21st June 2002

Abstract. The question of where to accommodate future urban expansion has become a politically sensitive issue in many regions. Against the backdrop of 'urban compaction' policy, this study uses population surface modelling and cellular automata (CA) to conduct an empirical urban growth simulation for Southeast England. This implementation leads to a consideration of the proper balance between the theoretical abstraction of self-organised growth and empirical constraints to land development. Specifically, we use 1991 and 1997 postcode directories to construct population surfaces. From these, the distributions of developed and vacant (rural) land are derived. Development potential is assessed through accessibility surfaces, which are constructed from the travel/commuting time to major London rail termini, to motorway junctions, and to principal settlements. Through investigating the frequencies of land development in relation to the accessibility surfaces, we can begin to understand the distribution of land development in this region. Based on this empirical relationship, the transition rules of a CA simulation of future urban expansion are constructed. In addition, government population projections at the county level are used to constrain simulation to the year 2020. The study demonstrates the utility of empirical CA in urban growth modelling; in particular the importance of empirically informed CA simulation rules in characterising the distribution of land development.

1 Introduction

There is a growing literature on the applications of cellular automata (CA) to simulate the growth of urban settlement (for example Batty, 1998; Batty et al, 1997; 1999; Clarke and Gaydos, 1998; Clarke et al, 1997; Li and Yeh, 2000; White and Engelen, 1993; White et al, 1997; Wu and Webster, 1998). However, the tension between the theoretical abstraction of CA models and empirical policy constraints remains. CA allow researchers to view the city as a self-organising system in which the basic land parcels are developed into various land-use types. A model of the urban system is thus constructed by the aggregation of uncoordinated local decisionmaking processes. One of the most important potential uses for such simulations is their ability to model the impact of alternative policies on the development process. CA applications based on hypothetical urban forms can provide valuable insights, but the interpretation of such modelling is hampered by difficulties in relating the modelled form to empirical combinations of settlement and constraints. The use of CA methods to model the future development of real urban systems is made particularly complex by the tension between self-organisation and the application of empirical constraints. The advantages of using a GIS (geographic information system) environment have been widely documented because a variety of data formats and data-processing functions can be easily accessed (Wagner, 1997; Wu and Webster, 1998). Most realistic large-scale applications need to consider the use of various data sources such as historical land-use records (Batty and Xie, 1994), urban land-use maps (White and Engelen, 1993), historical maps (Clarke and Gaydos, 1998; Clarke et al, 1997), and remotely sensed images (Li and Yeh, 2000; Wu and Webster, 1998) to construct a geography of urban development. In this context, integrating data sources from different scales and dates becomes important. Surface modelling techniques can be applied to cross-reference data layers

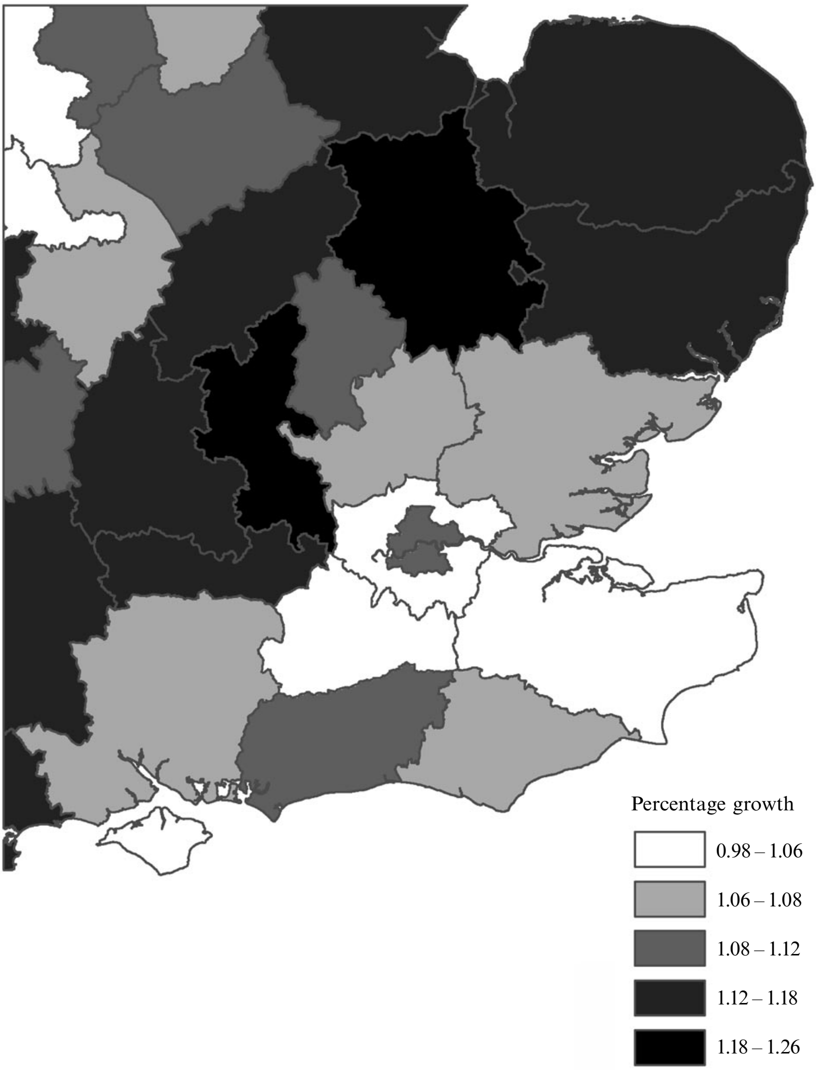

Figure 1. Southeast England study area, showing projected population growth rates to 2016, by 1991 county boundaries.

and construct an empirically informed CA simulation. More importantly, surface modelling can be used to identify the trend of development and thus inform the design of simulation models.

This paper describes the application of CA to the simulation of urban expansion in a 300×300 km area of Southeast England (figure 1), an area subject to considerable development pressure. The recent housing boom in the region has put further stress on greenfield sites. The population projection for the twenty-five years following 1991 is 4.4 million households. In 1996 the government raised its target to require that 60% of new developments should take place on 'brownfield' sites or 'previously developed land' (Breheny, 1997). This policy is often referred to as 'urban compaction' policy, which typically includes the objectives of promoting urban regeneration, the revitalisation of town centres, restraint on development in rural areas, higher densities, mixed-use development, promotion of public transport, and the concentration of urban development at public transport nodes (Breheny, 1997). Despite great efforts being made to develop a national land-use database (Harrison, 1999), the long-term trend of land development is unclear, simply because of the lack of empirical information on the trend of urban expansion. It has been estimated that the projected population growth would be likely to be translated into a total land transfer from 'rural' to 'urban' use of not less than 400 square miles between 1991 and 2016 (Bibby and Shepherd, 1997; Green, 1999, page 293). The Urban Task Force report—*Towards an Urban Renaissance* (1999)—highlights that the form of urban land development continues to be a politically sensitive issue. Against this backdrop, this work attempts to analyse the distribution of development and construct an empirical model of land conversion in relation to major accessibility measurements.

The main objective of this paper is to explore the use of CA in an applied context. The transition rules in CA are constructed according to empirical constraints. County-level population projections are used to constrain the growth rate in each county, and development sites (characterised as cells) are selected through combined consideration of the growth rate in the county and the intensity of development in the local neighbourhood. This is a process of self-organised growth because the neighbourhood effect tends to reinforce itself, but it is not a simple recursive model. With recursive local rules, the state of a complex system, involving nonlinear dependence, never reaches an equilibrium state. Although land development does not necessarily follow a strictly local rule, local interaction ensures that our settlements grow into a spatially connected system. In this model, land parcels are not scattered across all vacant land, but rather they are clustered together. The underlying mechanism for this is the feedback of previous development into the consideration of later development; even in the most sprawling urban development, residential plots rarely appear in totally isolated locations. The effect is not to smooth initial probabilities, but rather to reinforce the probability of development where development activity has previously occurred.

The initial pattern of development is derived through the use of surface modelling, using a technique developed for application to census centroid data (Martin, 1989) but here applied to unit-postcode locations. This offers greater spatial and temporal resolution than is available from the population census or conventional land-use mapping. Accessibility surfaces are constructed to represent the commuting time by rush-hour train to major London rail termini, travel distance to motorway junctions, and to the edge of the nearest settlement with a population of over 10 000, assembled using Arc/Info Grid data-processing functions. Microscopic location is assessed through development intensity. In this case we measure development intensity as the percentage of developed land within a neighbourhood, taking the presence of unit postcodes as the indicator of development. This location is measured *dynamically* during simulation to reflect the self-organised nature of urban growth. Although the development scenario is just one possible realisation of future urban growth, the empirically informed CA simulation provides valuable information on the distribution of urban development at the county and district levels. The main feature of this simulation is the consideration given to the empirical context of land development through the use of surface modelling techniques.

The remainder of this paper is arranged as follows. In section 2 we briefly discuss the position of this CA model in a typology of urban simulation. This is followed in section 3 by a review of the surface modelling technique used to reconstruct the geography of urban development. In section 4 we discuss in detail the context and methodology used in our simulation, and in section 5 the use of surface modelling in preparing data sources is introduced. In section 6 we analyse the empirical relationship between the accessibility surface and land development in Southeast England, and we discuss the implementation of the simulation in section 7. In section 8 the results of simulation are validated by the comparison of the actual development in 1997 with two simulated patterns. In section 9 the results of long-term simulation (to year 2020) are reported. In the final section we reflect on some lessons in the development of this empirical simulation.

2 Positioning CA in a typology of urban simulation

There has been a long tradition of urban modelling, but conventional urban models are built upon the neoclassical concept of equilibrium and hence are essentially static. Only recently have there emerged microscopic simulation approaches to understand urban dynamics (Batty et al, 1997). Dating back to the spatial diffusion phenomena modelled by Hägerstrand (1967), cellular models are drawing increasing research attention because the approach is essentially dynamic and thus appropriate to characterise urban change. As a modelling framework, CA have wide appeal because of their simplicity, intuitiveness, flexibility, and transparency (Webster and Wu, 1999). CA models adopt a computational approach, in the sense that they are only 'solvable' through computation, in particular through the medium of the computer. The recent emergence of computational power has contributed to the popularity of CA. In particular, graphically based systems like GIS provide a huge potential for implementing CA models to simulate changes in the urban built environment (Batty et al, 1999).

CA simulations have been widely applied. However, early attempts were typically more in the nature of metaphors of urban growth with little explicit relationship to underlying behaviour theory (Batty and Xie, 1994; Couclelis, 1985). It is now becoming clear that the CA approach is essentially heuristic and therefore attention should be drawn to the plausibility rather than the 'correctness' of models. With a better understanding of the technique, CA simulation is at the stage of exploring more complex behaviours. In the literature, a variety of ways of defining the transition rules of CA models have been reported (Batty and Xie, 1994; Clarke and Gaydos, 1998; Li and Yeh, 2000; White and Engelen, 1993; Wu and Webster, 1998). These exercises highlight the need for an integrated approach that combines the relatively simple abstraction of CA with the behaviourally richer models of urban processes found in the social sciences. The main obstacle to incorporating the richness of urban models is not one of technical difficulty but rather the theoretical justification for a complex model. The notion that CA reveals complex global patterns as they 'emerge' from a set of simple local transition rules is absolutely right, but urban development (for example, land-use conversions) is unlikely to be governed by such simple rules. This becomes increasingly apparent in studies of the political economy of urban land uses. With all these CA approaches a critical question remains about the way the transition rules should be interpreted in economic or other behavioural terms. The behavioural aspect of most CA simulation is still weak.

A two-dimensional matrix provides a simple modelling framework within which specific approaches may be located (figure 2). The two dimensions represent spectra in the trade-offs between global versus local and theoretical versus empirical characterisation. At each extreme, there are some well-known examples.

The strictly local rule associated with a pure theoretical configuration is often used in simple metaphoric models. These models follow some well-defined physical processes such as diffusion-limited aggregation (DLA) to simulate generic urban forms. Fractal properties are often presented as they can be seen in real-world cities (Batty and Longley, 1994). Classical CA models occupy a similar position in the matrix of figure 2. The insights generated from these models are not directly applied in the control of urban growth because of the abstraction of the model but they are useful

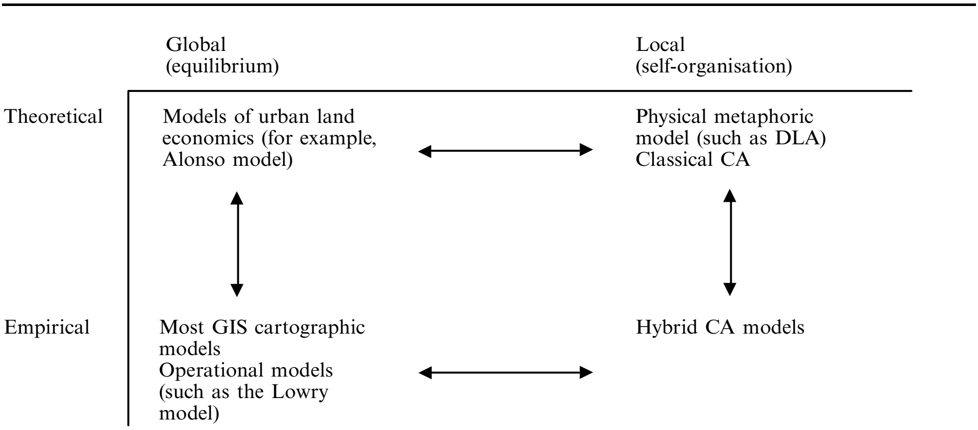

Figure 2. A matrix typology of urban simulation modelling approaches.

in the sense of *analogy*—the fundamental similarity between the morphology of theoretical and real cities suggests a similar process might in fact provide a plausible characterisation of urban growth. Lying at the other extreme are empirical and global models, which are often 'operational' models, developed in a GIS environment. These typically 'cartographic' models use such methods as map overlay and buffering. Factors affecting urban development such as access to roads, distance from the city centre, topography, and land uses stored as many layers are superimposed and manipulated to generate a final 'suitability' map. Often such a process is applied to the whole map area. The method is essentially static. More sophisticated cartographic models can be developed to include temporal dynamics, which may be formalised through geo-algebra (Takeyama and Couclelis, 1997). At this extreme are also empirical population-density models. These models are built up from disaggregated spatial units and calibrated through statistical methods. Although they may provide insights into how the population density is related, in a regression sense, to a bundle of locational factors, they do not describe the processes of population-density change because a simple extrapolation of existing relationships is problematic. As shown in recent studies on the dynamics of urban spatial structure, the population-density surface is an emergent phenomenon, in the sense that a monocentric structure can evolve into a polycentric one if the same relationship is applied repeatedly (Wu, 1998). The theoretical foundation of conventional urban models, however, is neoclassical urban economics that assumes that urban systems are always at equilibrium. The well-known theoretical model based on this assumption is the Alonso (1964) urban land-use model. Most models in this category are theoretical ones with limited connections to the practice of urban and regional planning. Just like the strict CA simulation, their intention is not to define the empirical pattern of growth but rather to reveal a metaphorical urban form. But the city is a self-organising system, as shown by the pioneering work of Peter Allen in the 1970s (Allen, 1997; Allen and Sanglier, 1981). Between these extremes are a vast number of hybrid models that mix the global and local rules and use different spatial resolutions and hence offer different degrees of realism.

We believe that the design of an appropriate simulation strategy should consider the purpose of modelling. Urban growth can be best articulated at a certain level of abstraction and balance of the empirical versus theoretical, global versus local considerations. In this research we propose a framework which allows such an appropriate simulation strategy to be implemented. We give explicit consideration to global and local factors affecting urban growth. The right positioning strategy for an empirical CA model is perhaps near to the hybrid one. In this work, we assess development potential according to a number of key factors, which will be discussed in detail below. Based on this assessment we use the Monte Carlo method to find suitable locations for development. This process reflects randomness, however, under the influence of global (and regional) constraints. To reflect the self-organised aspect or urban growth, the development situation is evaluated in a 3×3 local kernel, at each simulation time t. This is a standard CA rule definition. The strength of local growth is calculated as the ratio of developed land to undeveloped sites. It is worth noting that the development factors in this simulation are updated according to the changed land uses at time t. In particular, development factors such as local attractiveness are measured according to a gravity type of equation. In order to incorporate empirical data into the model, we will discuss the use of surface modelling in deriving fine-resolution data for the simulation in the following section.

3 Surface modelling

In this work we have chosen to use a population-surface model as the basis for our simulations. Conventional choropleth (shaded area) representations of population distribution suffer from the significant disadvantage that they imply a population density is present at every location on the map, whereas the actual settlement pattern comprises relatively dense clusters of populated land separated by extensive unpopulated regions. Various approaches to population-surface construction have been proposed since Tobler (1979), but the variable kernel-redistribution algorithm used here was first presented in Martin (1989), and developed with the particular characteristics of UK census data in mind. This approach has been used to create a series of national surface models, described by Bracken and Martin (1995), and accessible to registered users at http://census.ac.uk/cdu/surpop/, which also provides more detailed background to the modelling technique than is appropriate here. The surface-construction technique is most recently reviewed in Martin (1996a) and Martin et al (2000), which also examine the conceptual and technical differences between surface and zonal models of population distribution.

This method requires centroid points for each small area for which population counts are available, and which may be considered to be ‘population weighted’. In the UK census context, population-weighted centroids are provided by the census offices for each enumeration district (ED), the smallest zone used for the publication of census data in recent censuses. The surface-construction algorithm visits each centroid in turn and examines its distance from other local centroids. This distance can be used as an indication of ED size in that region, and a distance-decay function is calibrated and used to redistribute the population total at the centroid into the surrounding cells of a raster output matrix. Thus in areas of high population density, with centroids located close together, population may be spread very short distances from the centroid, reflecting small ED sizes. In remote rural areas, population may be spread over larger distances up to a predetermined maximum, which is a parameter of the model. Thus cells in urban areas may receive population from several different centroids, while others which are remote from any centroid remain unpopulated. The resulting model embodies a representation of the settlement geography which is one of its most important advantages over zone-based representations, and it is for this reason that we have chosen to use this approach to construct the initial model from which to run our urban development simulations. Although total population has been used in this example, the model may be applied to any count data present at the zone-centroid locations. Other applications of census-based surface models produced using this approach, and again taking advantage of the reconstruction of the settlement pattern, may be found in Brainard et al (1997), Lovett et al (1997) and Mesev et al (1995).

4 Methodology of empirical CA

This simulation model uses the output of surface modelling as the input for the initial state and development evaluation. The measured accessibility surfaces are used to evaluate the potential for development. However, rather than directly using the evaluation score we actually use the frequency of land development observed through surface modelling to measure the probability of development (see section 6). This probability is therefore measured rather than arbitrarily defined. The frequency of development was incorporated into a grid of initial probability through a look-up table. Let this initial probability surface be defined as P_0. This initial surface, however, is constantly modified by new development, which reflects the changing development situation and self-organised nature of urban development. The vacant land that has just been developed is unlikely to be subject to immediate redevelopment; new development will increase the attractiveness in the buffer or neighbourhood by strengthening the agglomeration effect. This is measured at each iteration. The new probability surface is therefore a result of modified development change.

$$R_{ij}^t = P_0 V_{ij}^t L_{ij}^t ,$$

where R_{ij}^t is a modified probability surface, and V^t and L^t are two factors that redefine the initial probability surface P_0. V^t measures the land availability,

$$V_{ij}^t = \begin{cases} 1, & \text{if} \quad \text{if the site } ij \text{ is vacant,} \\ 0, & \text{otherwise}. \end{cases}$$

L^t measures the local growth strength, in a typical CA 3×3 neighbourhood,

$$L_{ij}^t = \frac{1}{9} \sum_{3 \times 3} S_{ij}^t ,$$

where S_{ij}^t is a flag, which equals 1 if the cell had been developed by time t.

Strictly speaking, through this modification, the probability surface no longer follows the original probability distribution. It is a potential surface or measured attractiveness surface. It is equivalent to the output from a multicriteria evaluation process, but it bears the empirical relationship between the factor score and potential.

The probability of development decreases with distance from the ideal site (which has the maximum potential $R_{\max}$). The ideal site changes with each iteration. Therefore the maximum potential value should be recalculated at each iteration, which is denominated as $R_{\max}^t$. The relationship between the potential score and probability should be nonlinear, and similar in form to a logistic or Poisson distribution. This is because better sites are limited in quantity and would receive disproportional chances of being developed. The equation given below means that the site generating the highest score at the time of development is treated as a benchmark. The probability of development decreases with decreasing scores. A nonlinear transformation (here negative exponential) is used to depress the probability away from the maximum score in order to achieve greater discrimination between cells in any one simulation:

$$P_{ij}^t = R_{ij}^t \exp\left[\alpha\left(\frac{R_{ij}^t}{R_{\max}^t} - 1\right)\right] ,$$

where P_{ij}^t is the probability of development at the site ij at time t; R_{ij}^t is the potentiality score at the time t at the site ij, and α is a dispersion parameter that governs the stringency of site selection, with a higher value reflecting a more stringent selection process, that is, only the sites with higher value will be selected. This value therefore controls randomness. The dispersion parameter is usually chosen as 5.0.

Most classical CA simulations consider only the local neighbourhood without taking the regional neighbourhood into account. Batty and Xie (1994) suggest that a nested neighbourhood could be used to reflect regional effects. In this study, however, we need to incorporate the regional population projection into one simulation scenario. This is achieved through regrouping the P_{ij}^t into regions. For each region, P_{ij}^t is 'standardised' into P_{ijr}^t according to the sum of the regional total:

$$P_{ijr}^t = \frac{P_{ij}^t}{\sum_{ij \in r} P_{ij}^t}.$$

The equation is applied to every region. This probability will ensure that in a Monte Carlo simulation the expected amount of development in a region is always equivalent to one site. In order to generate the expected number of development sites, this probability is then multiplied by the required amount of development, and the maximum probability value should be limited to 1.0. Finally, the stochastic simulation uses a random number grid to generate development sites and then the simulation moves to time $t+1$.

As mentioned earlier, three challenges confront the development of an empirical CA simulation: first, to incorporate the empirical pattern of observed development into the simulation, second to incorporate regional features into the simulation, and third to incorporate a truly stochastic process which permits unforeseeable changes. The model developed in this research is refined to meet these three requirements.

5 Data source and application of postcode geography

The surface modelling technique described in section 3 has been applied to UK census data in a number of contexts, but the representation of the detailed settlement pattern is limited by the geographical resolution of census data. Further, the decennial nature of the census makes it difficult to capture the continually evolving pattern of urban development. Although the census provides a rich range of socioeconomic variables, it is considered that spatial detail and timelines are of more importance to the current study, and this has led us to consider alternative data sources that might be used with the same modelling procedure. In the United Kingdom, postal geography is the most widely used georeferencing system for socioeconomic data outside the census, and we have therefore opted to apply the technique described above to the postcode system. In order to provide the initial model of population distribution, data have been taken from the directory of enumeration districts and postcodes (OPCS and GROS, 1992). Unit postcodes are the smallest component of the UK postcode system. EDs typically contain 200 households and 400 residents, and unit postcodes each refer to around 15 postal addresses. The directory of enumeration districts and postcodes was originally created in association with the 1991 Census. The file contains a record for each unique ED/postcode intersection, containing a 100 m grid reference for the postcode, a household count, and some additional information. Although these grid references are not population weighted, they represent the location of at least one of the properties known to fall within the smallest unit of the postal system, and the far greater number of data points permits a far more detailed representation of residential geography than is possible with the ED centroid data.

As the postcode geography has evolved over time, the directory has been kept up to date and periodic revisions published which relate 1991 Census EDs to contemporary postcode geography. No postcodes are removed from the file, but terminated and reused codes are flagged. New codes are assigned grid references and may hence be associated with an ED. The 1995 and 1997 versions of the file also contain a large-user or small-user indicator for each postcode. Large-user postcodes typically receive over twenty-five items of mail per day and are usually commercial addresses. Only the current postcodes have been used from each directory, allowing the resulting models to approximate to postal geography in 1995 and 1997. A household count of 15 has been assigned to any new small-user postcodes for which household counts are unavailable, representing the typical number of addresses per postcode. There are thus around 850 000 data points available for surface modelling. These locations have been used to provide centroids for the construction of surface models for the study area in figure 1 with a cell size of 200 m and kernel width of 250 m, resulting in models with 1500×1500 cells. The use of unit-postcode locations as an input to surface modelling and selection of the relevant parameters are discussed in Martin et al (2000). More precise geographical locations for surface modelling could now be obtained from Ordnance Survey's Code-Point product (which is based on the spatial averaging of the addresses within each postcode) rather than the directory of EDs and postcodes used here, but Code-Point is not available for the time period required by this study.

Household counts have been used as the variable for redistribution, so that each output model is effectively a household-density surface, in which we treat all cells with population estimates of 0 as undeveloped, and all those with estimates greater than 0 or containing large-user postcodes as already developed. We thus infer a simple two-category land-use classification from the postcode-based surface modelling. Ideally, we would seek to use a multicategory classification of land use that was able separately to identify land uses such as residential, commercial, and industrial, and to detect development at different densities within these categories. In the absence of a geographically detailed land-use database however, as discussed above, it is necessary to use proxy data that combine some measure of development with a reasonable level of geographical and temporal resolution. A potential extension to the current study would be to treat large-user and small-user postcodes as two different land-use classes in order to produce a classification into undeveloped/residential/commercial types, but this would require careful validation of the large-user postcodes against some other geographically referenced indicators of economic activity.

In order to constrain the amount of population growth occurring in the simulation models, a series of official population estimates based on 1993 mid-year population estimates at county level have been used (OPCS, 1995). This set of official estimates falls closest in time to the 1991-based census and postcode products on which the rest of this study is based. The study area covers all or part of thirty counties, for which population projections to 2016 are available. A county map has therefore been created with associated growth rates to constrain the simulation within the officially projected population-growth levels. A map of projected population change by county is shown in figure 1.

Major factors affecting the attractiveness of residential areas throughout the region include commuter travel times to London, accessibility to the national motorway network, and accessibility to a medium or large centre of population, which may be expected to provide a good range of services and facilities. Commuters travel to London from everywhere within the study region, and the dominant mode of transport for this long-distance commuting is train. A London travel-time surface has therefore been devised by taking fastest travel times to the appropriate London terminus arriving between 0800 and 0900 on a weekday, as this represents the timing most likely to be used by commuters. Times have been obtained from 130 major stations. Commuter-route information was extracted from McGhie (1992) and times were obtained from the Railtrack travel enquiry site at http://www.railtrack.co.uk/. Travel times for each individual cell have been calculated by estimating travel time to the nearest station by crow-fly distance assuming a mean travel speed of 60 km h^{-1}, and these times have been added to the fastest available train time. Special treatment has been given to the Isle of Wight, in the centre of the south coast, which is the only genuine island in the study area containing a significant population, and for which travel times have been increased by an amount equivalent to the necessary passenger-ferry crossings. In deriving the full travel-time surface, the assumption has been adopted that, if a major station is located within 10 minutes' drive time of a given cell, the commuter journey will be routed via that station. At greater distances, the shortest travel time to London is applied, regardless of the distance to the station that must be used. Although simple, these general assumptions better reflect commuter behaviour than a simple Euclidean allocation of cells to their nearest stations.

A further aspect affecting development potential in this region is accessibility to the national motorway network, and a complete set of motorway-access points has been digitised, and distances calculated to each cell. Only sections of motorway which are connected to the principal motorway network are included (thus no account is taken of short isolated lengths of urban motorway). Further analysis of the postcode-based surface model allows the identification of settlements with populations in excess of 10 000. These are represented in the surface model by contiguous clusters of populated cells whose combined population is greater than 10 000, each cluster being surrounded by undeveloped land. A sequence of conventional GIS grid-modelling functions may be used to identify these settlements, as described in Martin (1996b). In Arc/Info, region grouping of contiguous populated cells allows the extraction of distinct clusters, whose populations are then obtained by summing the individual cell population estimates. Clusters with populations below the required size are then filtered out.

6 Profiling development distribution

In order to understand the 1991 – 97 distribution of development, we use our grid models of development between 1991 and 1997 to examine the relationship between development and accessibility. We measure the rail commuting times in 5-minute bands, and distances from motorway intersections and large settlements in kilometre bands. The number of cells that have seen development within the period 1991 – 97 is summed within each band. The scatter plots (figures 3 – 5) characterise the distribution of development in terms of accessibility.

The amount of development declines with increasing distance from existing settlements. Figure 3 represents a classical distance-decay relationship. It can be seen that a majority of growth takes place fairly close to existing settlements (within 10 km). However, the geographical area of successive bands increases with distance from the

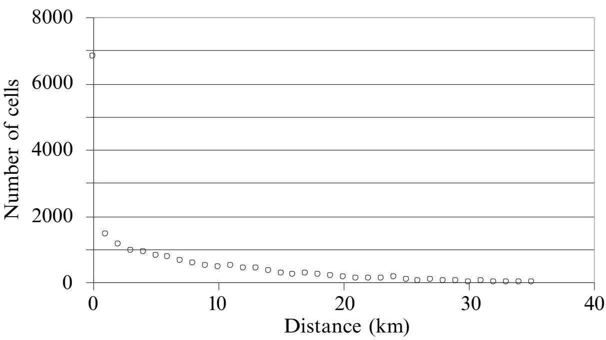

Figure 3. Development profile measured in 1 km distance bands to edge of nearest settlements with population in excess of 10 000.

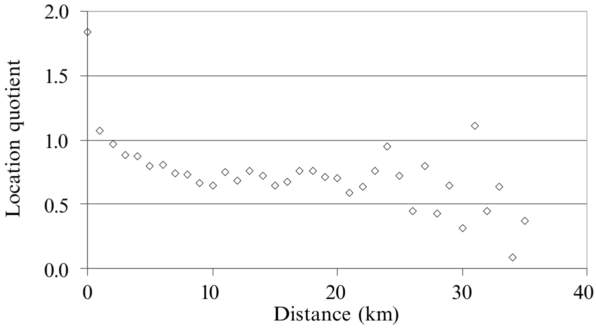

Figure 4. Location quotient measured in 1 km distance to edge of nearest settlements with population in excess of 10 000.

centre. In order to assess the intensity of development, we use a 'location quotient' to recalculate the development distribution. The location quotient is a ratio of the band's share of total development to the same band's share of total area. Figure 4 shows the distribution of the location quotient; values over 1.0 indicate that the band accommodates development above the overall quantity expected in terms of its area. From figure 4, it can be seen that the first 3 km receive above or near the expected level of development, and the value decreases until around 10 km; then the location quotient varies and slightly increases. In summary, the distribution profile shows a concentration of development close to existing settlements.

Similarly, the distribution of development in terms of the accessibility to London is represented in figure 5. The figure clearly shows that within 1 hour commuting distance (12×5-minute bands), the amount of development increased; then beyond 1 hour the quantity of development began to decrease. Over 1.5 hours, the amount of development significantly decreased. The pattern suggests that within the central area (because most land is already developed) the amount of new development in 1991 – 97 is lower than in the fringe area (around 1 hour by train). Another important factor influencing the amount of development is the small size of the central area. The intensity of land development is better seen in the profile measured by location quotient as shown in figure 6 (see over). It can be seen that there are several distinct regions of location quotient values. Up to a commute time of 1 hour the value is over 1.0, suggesting a concentration of development in central London; from 1 to 2 hours the location quotient is near 1.0, indicating that development intensity is as expected; beyond 2 hours development significantly declined. The figure clearly reveals the London factor in the regional development of Southeast England. Empirical evidence strongly

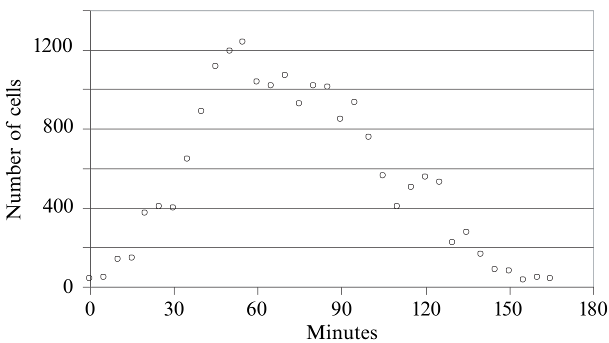

Figure 5. Development profile measured in 5-minute bands to London rail termini.

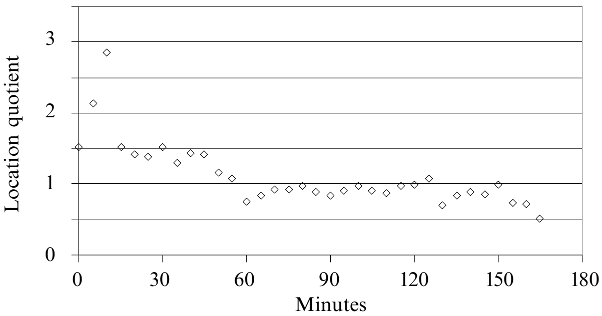

Figure 6. Profile of location quotient measured in 5-minute bands to London rail termini.

suggests that the creation of new postcodes is related to accessibility to London; this finding is consistent not only with experience but also with the distribution of property values in the region.

The profile of development in terms of access to major motorway junctions is represented in figure 7 by 1 km distance bands. In the closest bands (1 – 3 km) development does not reach the highest intensity; however, measured in terms of location quotient accessibility to a motorway is a distinct advantage (figure 8). A significant decrease in development is seen within the first 25 km, after which the relationship becomes irregular, perhaps reflecting diminishing influence of motorway access and an increase in the importance of the nonmotorway road network. Up to 13 km from motorway junctions, the development quotient is over or close to 1.0, suggesting this to be an attractive area for development. This is in keeping with the observation that retail and business parks are very often located close to the motorway network.

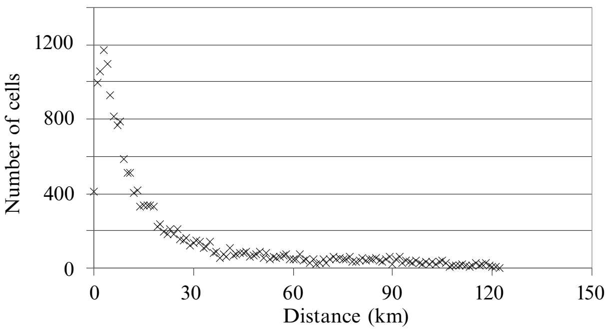

Figure 7. Profile of development measured in 1 km distance bands to motorway junctions.

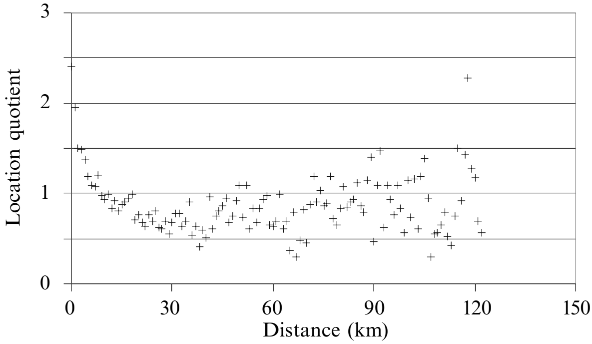

Figure 8. Profile of location quotient measured in 1 km distance bands to motorway junctions.

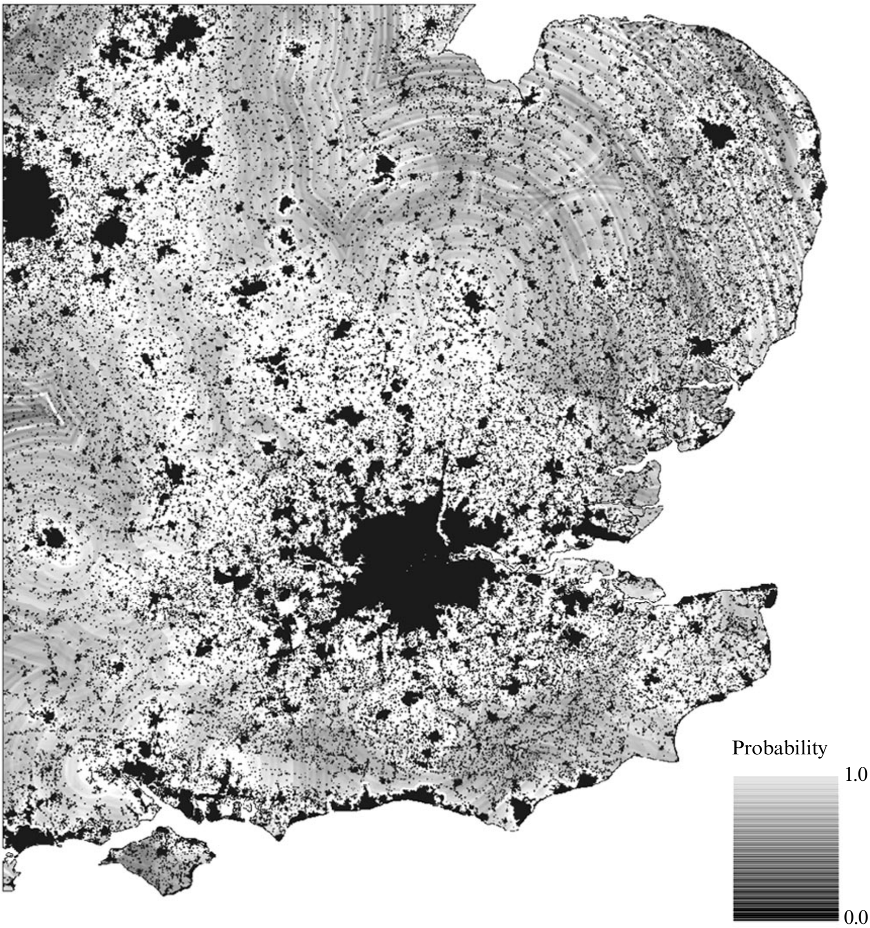

Figure 9. Development probability surface derived from initial accessibility surfaces.

A comparison of figures 6 and 8 suggests that the location quotient for motorway access decreases more rapidly than that for rail travel time. This is consistent with the understanding that rail travel remains important for long-distance commuting whereas private cars are primarily used for short distances in extended metropolitan areas.

Profile development in this way is useful in the sense that it provides empirical relationships for configuring the simulation rules. The frequencies in each time/distance band can be used to calculate the probability of development. The composite probability surface is shown in figure 9. The composition process uses the multicriteria evaluation approach, that is, summation of all standardised scores. The method used here is to link frequency of land-use change with distance. We then standardise the frequency scores. Equal weights are used for the three resulting frequency surfaces to calculate a single probability surface. This probability surface is compared with a random number surface and sites with probability greater than the random number are selected for development. The development as simulated, however, immediately modifies the value of the probability. The area that has just been developed is not considered for redevelopment. We use a conditional function to identify recently developed areas and a neighbourhood function to measure the intensity of development within a kernel. The intensity of local development reflects its self-organised nature, and is thus used to compose a *dynamic* probability surface, which changes along with the simulation. This approach differs importantly from conventional overlay of potential scores that are neither locally refined nor updated at each iteration.

7 Simulation implementation

The initial state of the land use is derived from the processed 1991 postcode coverage as described in section 5. We assume that the expansion of the urban built-up area is at the same rate as projected population growth. This does not take into account 'densification' or decreasing urban densities within each already developed 200 m × 200 m cell. The population in the Southeast region was 30.617 million in 1993; it is estimated that in 2016 it will reach 33.241 million, that is, an overall increase of 8.57% from 1993 to 2016. Moreover, we use the county-based projected growth rates for simulation. The target growth rates are then translated into the increase of the number of developed cells per iteration. The number of cells for new development is then used as the constraint. The simulations are run over 29 years from 1991, which gives a final scenario for the year 2020. Areas that are not suitable for urban development such as sea and coastal water are excluded from the site-selection process.

We use the Arc/Info GIS as a data-management tool, together with Arc Macro Language (AML) to permit automatic iteration of grid calculations. The grid-modelling functions available in this environment facilitate the manipulation both of input surface models and of the iterative computation. For example, we consider the open space within the existing built-up area (cemeteries, parks, and other protected urban spaces) to be a special type of use. These areas surrounded by existing development within the urban area are frequently subject to planning controls and are therefore unlikely to be developed. We use a conventional grid-modelling function to identify and mask undeveloped cells that are fully surrounded by developed land. (This is the same function used to fill sinks in an elevation model during hydrological modelling.) Thus, although these areas are very attractive according to the multicriteria evaluation function discussed above, they are excluded from the consideration of further development. The local growth strength has been evaluated in a 3×3 cells neighbourhood. The value is detected and updated in each iteration, and then used to modify the start probability surface.

After loading the initial probability surface, including the protected land (figure 9), the iteration procedure begins, which involves the following steps:

(1) Constrain the initial probability grid by excluding any newly developed land in previous iterations.
(2) Construct new probability distribution.
(3) Standardise modified probability surface.
(4) Compute local development intensity.
(5) Update dynamic probability grid.
(6) Transform dynamic probability surface using a negative exponential function.
(7) Constrain probabilities to zonal growth rates to ensure that overall growth in each zone matches the target growth rates.
(8) Simulate development by Monte Carlo testing of probability grid against a random number grid.
(9) Update the state grid with newly developed sites.
(10) Return to step 1.

The above procedure has been refined several times during code development. The aim of refinement has been to reduce 'add-ons' to a minimum. For example, we have used two different approaches to computing local growth strength. The first model simply counts the number of developed sites in a neighbourhood. The second model incorporates a spatial interaction equation to calculate the local attractiveness. This considers the size of each continuously extended settlement and the shortest distance to the edge of that settlement. To reflect a nonlinear relationship we use the logarithm of the size of the settlements. However, the distribution of the score, in the second method, is not controllable at each iteration because the attribute is calculated on the basis of a local kernel and it further depends on the distribution of developed sites or the shape of settlements, which differ from iteration to iteration. Through experimentation we found that the second method, despite its apparent sophistication, does not enhance the performance of the model. Therefore in the final simulation we adopt the simpler method. We believe that, unless there is sufficient justification and evidence to support the use of a more complex method, the simpler version is probably more appropriate. This is in fact in keeping with the principle of CA, that is, the use of simple rules to drive complex patterns of development. A further advantage of using a simpler version is that the structure can be fully tested, with a better understanding of the impact of model configuration on the results.

8 Initial validation

In order to validate CA performance we simulated urban expansion from 1991 to 1997 and then compared the simulated development with the actual development. Because the overall development accounts for only a very small proportion of the total land area, direct comparison based on each cell is not appropriate. Further, stochastic simulation is bound to generate different site locations from those actually chosen. The measurement therefore should be based on some 'structural' comparison. Because the simulation also takes account of the regional population projections, we use the development profile measured in terms of the accessibility surface to compare the simulated and observed development.

Figures 10 – 12 are comparisons with the three development profiles presented in section 7. The simulation captures the basic features of the development distribution, for example, the decrease of development with increasing distance from the edge of major existing settlements and motorway junctions and the inverse-U curve of travel-time bands to London rail termini. Note that this comparison considers only new development and excludes the major settlement pattern represented by the area that was already developed in 1991. The overall simulated land-development profile seems to be below the actual development profile. This is because the simulation is constrained

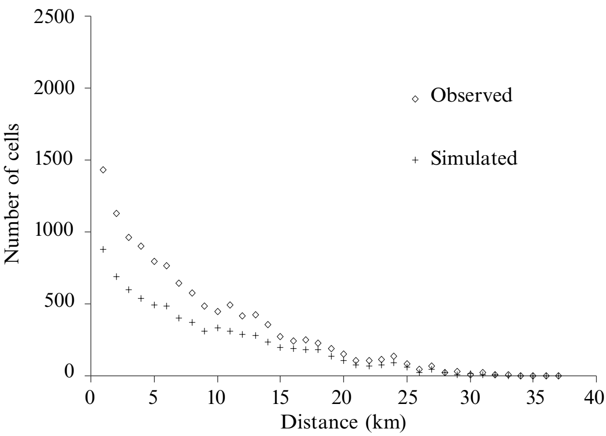

Figure 10. Profile of 1991 – 97 development by distance from nearest settlement with population in excess of 10 000.

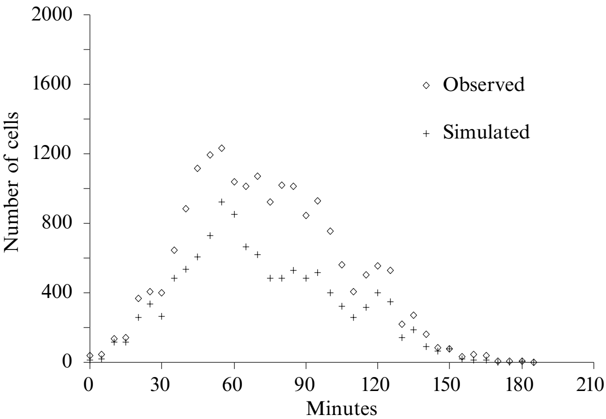

Figure 11. Profile of 1991 – 97 development by rail travel time from London termini.

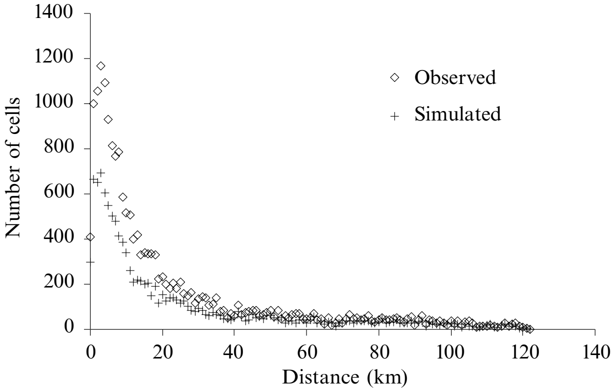

Figure 12. Profile of 1991 – 97 development by distance from motorway junctions.

by the regional population projection (which is intended for the simulation from the second part from 1997 to 2020). The number of cells simulated is 11 708, whereas the observed number of developed cells is 18 688. This is not a problem, however, as what we intend to examine is the shape of the development profile. The overall rate is controlled in this model and can be easily adjusted to produce different scenarios (as shown in the simulation from 1997 to 2020).

We began by identifying empirical relationships between development and three accessibility measures in the Southeast region. These were combined into a single standardised probability surface that provided the starting point for our simulation. The purpose of this initial validation has been to investigate the extent to which the relationships are preserved through the subsequent iterative simulation process, involving both local functions and zonal constraints. Although the absolute numbers differ, the individual relationships between development and each of the original accessibility measures are maintained to a high degree. This suggests that we should be able to use this technique based on the combined probability surface to project future development patterns while respecting known relationships. The probability surface thus provides a mechanism for building our existing understanding into the simulation process without the need for multiple complex constraints, and allowing the process to remain fundamentally one of local self-organised growth.

Validation is an important issue that has not really been satisfactorily addressed in CA research. For microscopic models such as CA, a cell-by-cell comparison is not appropriate, but neither is the use of a spatial confusion matrix. This is because these models involve a very large number of spatial units. A perfect simulation result, but with the entire distribution shifted by one cell, would result in low values of conventional goodness-of-fit statistics (Wu and Webster, 1998). Instead, some measure of the 'signature' or profile is more appropriate, but selection of the best signature depends on the problem domain (Wu, 2000; Wu and Webster, 2000). For example, Wu and Webster (2000) propose some economic indicators in a simulation related to economic efficiency and social cost. In this analysis we use profiles of various accessibility measures because our concern is with the question of whether the model can produce a profile of land development similar to that observed in reality rather than with the exact location of particular developments.

9 Simulation scenarios

The procedure introduced above has been applied to the simulation of two scenarios from 1991 to 2020. The first of these adopts the county-level population-growth rates (from 1993 to 2016 projections). In this simulation new greenfield development in each county is proportional to the total projected population growth. The second scenario respects the overall population projection for Southeast England, without constraining development at the county level.

The results of these two simulations are given as figures 13 and 14 (see over), which show the simulated growth to 2020 expressed as a percentage of developed land in 1991. The extent to which these maps diverge from actual development will in large part be attributable to the success of the government's current objective that 60% of development should take place on brownfield sites (Urban Task Force, 1999). In order to illustrate more clearly the pattern of development that results in each case, we have chosen to present the development totals aggregated at the level of the local authority district or London borough. These are administrative divisions at a level below the counties, but serve to show more clearly the urban–rural differences in development, as each major outlying town tends to be represented by a single district. In both cases, it is apparent that the existing urban areas generally display low growth rates, being already fully developed, whereas the rural districts display moderate to high growth rates, with those adjacent to established urban centres (and thus highly accessible) experiencing the most development. The differences between the two scenarios primarily relate to the ways in which the officially projected population growth interacts with attractiveness for development as measured by our empirical accessibility factors. The official projections take into account both natural change and migration averaged across the years immediately preceding the base year, and were reached through consultation with local authorities (Armitage, 1986). They may therefore be considered to include metropolitan-to-nonmetropolitan migration patterns already established in the early 1990s, but not subsequent policy changes which may have resulted in new locational trends in development.

The constrained model (figure 13) strongly reflects the projected pattern at the county level with most growth occurring in an arc from the southwest to the northeast of London. Even within this arc it is apparent that development is concentrated in the more rural districts. The unconstrained model (figure 14) is more strongly influenced by the accessibility relationships that were observed in our initial analysis. This scenario is indicative of the pattern of development that would result directly from following the access-based development-attractiveness measures. The model experiences much greater development close to London and an even greater contrast between urban and rural areas. However, this analysis takes no account of policy devices such as the greenbelt, which effectively shifts development outwards from the most accessible sites.

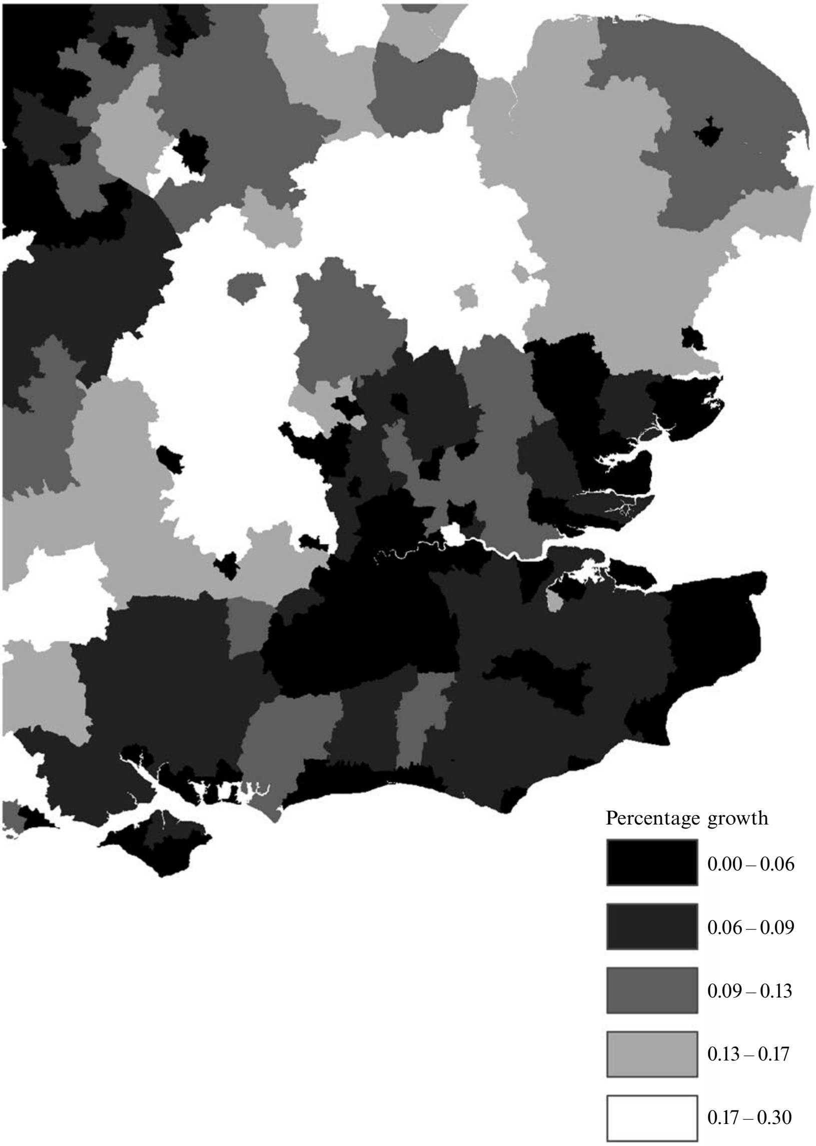

Figure 13. 1991 – 2020 district-level development rates under scenario 1 (constrained to county-level projections).

10 Conclusions

Urban compaction policy as advocated in the Urban Task Force report will certainly have an impact on the form of urban development if the target is implemented in day-to-day development control. Similarly, in the USA, 'Smart Growth' aims to achieve higher densities, more mixed land uses and redevelopment, and constrained suburban sprawl (Katz, 2000). However, development scenarios over ten to twenty years are not well understood. The value of urban expansion simulation is to provide reference scenarios. The attractiveness of empirical CA combined with surface modelling lies in its relatively transparent structure—the modelling is dependent upon relatively simple information (land use and population derived from surface modelling) and heuristically plausible rules (for example, that development is related to accessibility).

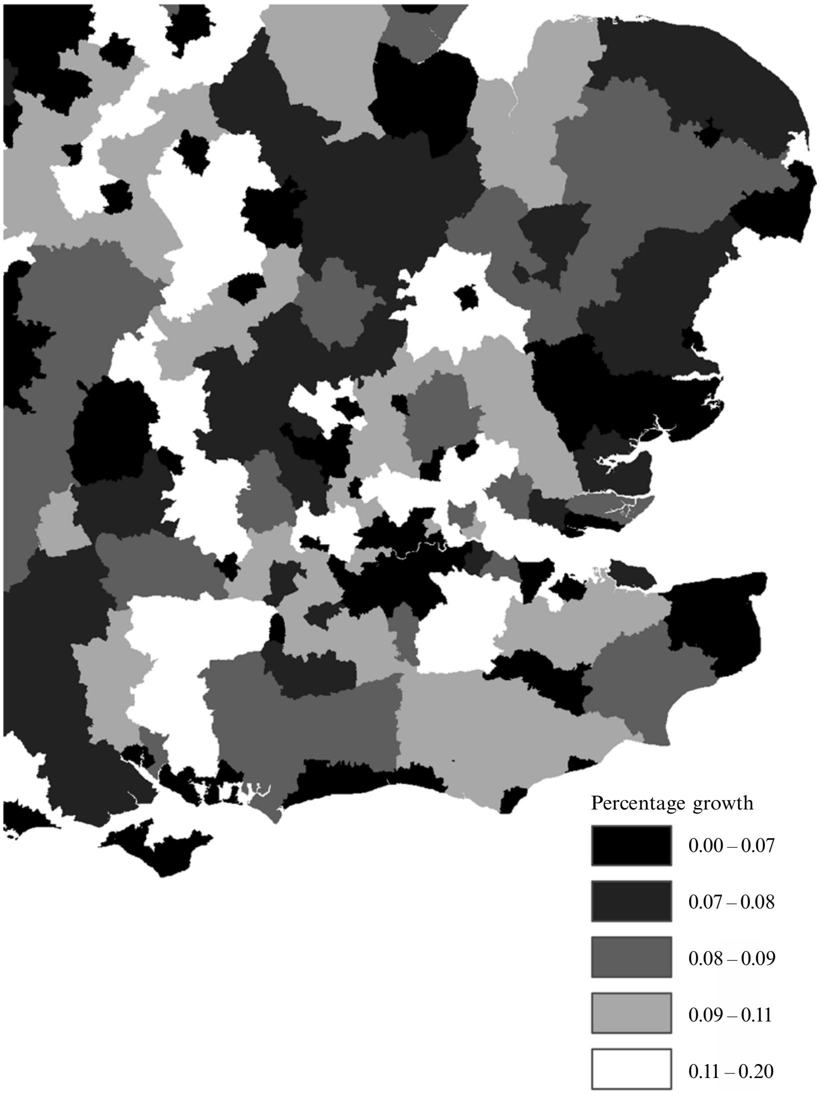

Figure 14. 1991 – 2020 district-level development rates under scenario 2 (unconstrained at county level).

Although the computation is intensive, involving iterative processing of large grids, the model itself is straightforward. More importantly, this approach may be driven by empirical relationships such as that between development and accessibility. This study demonstrates the enhancement of CA by the explicit inclusion of an empirical relationship between urban form and development factors. The model itself is flexible, potentially allowing the role of alternative relationships to be explored in the dynamics of regional growth. In a sense, the value of any empirical CA model lies in how well it captures these dynamics. The simulation technique, that is, the self-organisation approach embedded in the iterative computation, does *not* guarantee realistic prediction.

Various improvements could be made to the model presented here. The study would benefit from more detailed information on greenbelt land and other planning restrictions. However, we suggest that, because the development of a national land-use database is complex and currently incomplete, surface modelling, in particular when available from successive censuses, can provide a useful source of data on urban form. Only the simplest characteristics of the available datasets have been exploited here. In particular and as noted above, the distinction between large-user and small-user postcodes has not been used in any attempt to separate out residential and commercial land uses. The postcode dataset is not ideal for this purpose, but it may be possible to generate an acceptable geographical model for commercial properties if a differentiated land-use map were required. Significant improvements in the postcode datasets are anticipated as part of 2001 Census outputs. Separate treatment of commercial land in this way would require the acquisition of plausible county-level growth scenarios for commercial activity, as this is not directly governed by population change, and further work would be required in order to accommodate mixed occupancy within cells. Assumptions have also been made in the present models about population density, with all new development taking place at uniform density and on greenfield sites. A more sophisticated model could involve varying the development density according to the location and size of development taking place, in order to capture a more realistic range of values, and would allow for redevelopment within the existing urban areas. Clearly, a more sophisticated use of these types of simulation tool would require more careful formulation of the empirical constraints on development. Empirical research on the determinants of land prices for development and the attractiveness of competing sites would allow a more direct quantification of the effects of commuting distance and local property markets, although the currently available data are incomplete.

The objective of the simulations described in this paper has not primarily been to provide a single predictive model of urban development to the year 2020. Rather, our interest has been in the incorporation of a range of both theoretical and empirical development constraints in the simulation process. We suggest that CA simulation is useful in exploring future urban growth by understanding the impact of different development conditions. This extends classical CA built upon abstract rule definition to hybrid CA based on detailed real-world data and constraints. In the terms of figure 2, such hybrid CA models occupy the reign of the matrix that falls at the intersection of empirical data and local rules. Considering computational efficiency, a system that combines GIS functionality and dynamic process modelling seems to be an appropriate environment for the development of such work. Future research should incorporate more realistic constraints as discussed above, and focus on the impacts of changing development behaviour such as increasing brownfield development, mixed land use, and higher development density. Through such research, hybrid CA models have the potential to make a valuable contribution to the evaluation of urban policy responses, such as those suggested by the Urban Task Force report.

References

Allen P, 1997, "Cities and regions as evolutionary, complex systems" *Geographical Systems* **4** 103 – 130

Allen P M, Sanglier M, 1981, "Urban evolution, self-organization, and decisionmaking" *Environment and Planning A* **13** 167 – 183

Alonso W, 1964 *Location and Land Use* (Harvard University Press, Cambridge, MA)

Armitage R I, 1986, "Population projections for English local authority areas" *Population Trends* **43** 31 – 40

Batty M, 1998, "Urban evolution on the desktop: simulation with the use of extended cellular automata" *Environment and Planning A* **30** 1943 – 1967

Batty M, Longley P A, 1994 *Fractal Cities: A Geometry of Form and Function* (Academic Press, London)

Batty M, Xie Y, 1994, "From cells to cities" *Environment and Planning B: Planning and Design* **21** s31 – s48

Batty M, Couclelis H, Eichen M, 1997, "Urban systems as cellular automata" *Environment and Planning B: Planning and Design* **24** 159 – 164
Batty M, Xie Y, Sun Z, 1999, "Modeling urban dynamics through GIS-based cellular automata" *Computers, Environment and Urban Systems* **23** 205 – 233
Bibby P, Shepherd J, 1997, "Projecting rates of urbanization for England 1991 – 2016: method, policy applications and results" *Town Planning Review* **68** 93 – 124
Bracken I, Martin D, 1995, "Linkage of the 1981 and 1991 UK Censuses using surface modelling concepts" *Environment and Planning A* **27** 379 – 390
Brainard J S, Lovett A A, Bateman I J, 1997, "Using isochrone surfaces in travel-cost models" *Journal of Transport Geography* **5** 117 – 126
Breheny M, 1997, "Urban compaction: feasible and acceptable?" *Cities* **14** 209 – 217
Clarke K C, Gaydos L J, 1998, "Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore" *International Journal of Geographical Information Science* **12** 699 – 714
Clarke K C, Hoppen S, Gaydos L, 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
Green R, 1999, "Compaction and greener cities" *Town and Country Planning* October, page 293
Hägerstrand T, 1967 *Innovation Diffusion as a Spatial Process* (University of Chicago Press, Chicago, IL)
Harrison A, 1999, "The national land use database: from concept to implementation", in *Proceedings of GIS'99* (Association for Geographic Information, London) pp 371 – 374
Katz B, 2000, "The federal role in curbing sprawl" *Annals of the American Academy of Political and Social Science* **572** 66 – 77
Li X, Yeh A G O, 2000, "Modelling sustainable urban development by the integration of constrained cellular automata and GIS" *International Journal of Geographical Information Science* **14** 131 – 152
Lovett A A, Parfitt J P, Brainard J S, 1997, "Using GIS in risk analysis: a case study of hazardous waste transport" *Risk Analysis* **17** 625 – 633
McGhie C, 1992 *The Royal Insurance London Commuter Guide* (Good Books, Whitley, N Yorks)
Martin D, 1989, "Mapping population data from zone centroid locations" *Transactions of the Institute of British Geographers, New Series* **14** 90 – 97
Martin D, 1996a, "An assessment of surface and zonal models of population" *International Journal of Geographical Information Systems* **10** 973 – 989
Martin D, 1996b, "Depicting changing distributions through surface estimation", in *Spatial Analysis: Modelling in a GIS Environment* Eds P Longley, M Batty (GeoInformation International, Cambridge) pp 105 – 122
Martin D, Tate N J, Langford M, 2000, "Refining population surface models: experiments with Northern Ireland Census Data" *Transactions in GIS* **4** 343 – 360
Mesev V, Longley P, Batty M, Xie Y, 1995, "Morphology from imagery: detecting and measuring the density of urban land use" *Environment and Planning A* **27** 759 – 780
OPCS, 1995, "1993-based subnational population projections", PPP3 number 9, Office for Population Censuses and Surveys, now Office for National Statistics, 1 Drummond Gate, London SW1V 2QQ
OPCS, GROS, 1992 *ED/Postcode Directory: Prospectus* 1991 Census User Guide 26, Office for Population Censuses and Surveys, Titchfield, Fareham, Hants
Takeyama M, Couclelis H, 1997, "Map dynamics: integrating cellular automata and GIS through geo-algebra" *International Journal of Geographical Information Science* **11** 73 – 91
Tobler W R, 1979, "Smooth pycnophylactic interpolation for geographical regions" *Journal of the American Statistical Association* **74** 519 – 530
Urban Task Force, 1999 *Towards an Urban Renaissance* (E&FN Spon, London)
Wagner D F, 1997, "Cellular automata and geographic information systems" *Environment and Planning B: Planning and Design* **24** 219 – 234
Webster C J, Wu F, 1999, "Regulation, land-use mix, and urban performance. Part 2: simulation" *Environment and Planning A* **31** 1529 – 1545
White R, 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 – 1189

White R, Engelen G, Uljee I, 1997, "The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics" *Environment and Planning B: Planning and Design* **24** 323 – 343
Wu F, 1998, "An experiment on the generic polycentricity of urban growth in a cellular automatic city" *Environment and Planning B: Planning and Design* **25** 731 – 752
Wu F, 2000, "A parameterised urban cellular model combining spontaneous and self-organising growth", in *GIS and Geocomputation: Innovation in GIS 7* Eds P Atkinson, D Martin (Taylor and Francis, London) pp 73 – 85
Wu F, Webster C J, 1998, "Simulation of land development through the integration of cellular automata and multicriteria evaluation" *Environment and Planning B: Planning and Design* **25** 103 – 126
Wu F, Webster C J, 2000, "Simulating artificial cities in a GIS environment: urban growth under alternative regimes" *International Journal of Geographical Information Science* **14** 625 – 648