Developing an Urban Land Use Simulator for European Cities

Roger White[1,2], Guy Engelen[2], Inge Uljee[2], Carlo Lavalle[3], Daniele Erlich[3]
[1]Memorial University of Newfoundland, Department of Geography
[2]Research Institute for Knowledge Systems
[3]European Commission—Joint Research Centre, Space Applications Institute

Introduction

Cities are among the most complex structures created by human societies, and unlike other complex man-made objects, like computers, we do not pretend to understand them fully. Furthermore, cities in Europe, as in most parts of the world, are currently undergoing rapid and profound changes which affect the quality of life for hundreds of millions of people—changes which should be managed to preserve or enhance the quality of life, and to ensure economic and environmental sustainability.

Effective planning and management requires both data on current conditions and an ability to foresee the likely consequences of proposed projects and policies. The Space Applications Institute (European Commission—Joint Research Centre) has as one of its objectives to enhance the value of Earth observation data by stimulating user oriented data services. In this context, the MURBANDY project is aimed specifically at enhancing the use of remotely sensed data for urban planning by implementing the following component projects: (1) building a data base of land use change over approximately the last 40 years for a representative sample of European cities, (2) developing a number of urban environmental indicators related to sustainability, and (3) developing a generic model of urban dynamics that will support the realistic exploration of urban futures under a variety of planning and policy scenarios—or a scenario of no planning. All three components make use of a recent extension to the CORINE land use/land cover database which significantly extends the number of urban land use classes and thus makes possible a much more detailed study of urban areas. In this paper we discuss work carried out under the third component, the development of a generic model of urban land use dynamics.

The goal of developing a generic model—one that will be applicable to essentially any European city—implies the rather strong assumption that at some level cities are fundamentally similar, evolving by the same processes. Especially in the European context, where cities grow out of and express a wide variety of cultural, economic, and historical contexts, this seems a bold hypothesis. However, the fractal analysis of land use patterns provides evidence in its favour (Frankhauser, 1994; White and Engelen, 1993), and suggests that the substantial differences among cities are largely due to city-specific boundary conditions. In particular, we hypothesise that it is such city-specific factors as local topography, the geometry of the transport network, and local planning regulations (which themselves represent a particular local intervention in what might otherwise be a generic process of urban development) that determine the observed differences among cities. The project in effect constitutes a test of this hypothesis.

The MURBANDY model of urban land use dynamics

Capturing urban land use dynamics in a realistic way requires an approach that satisfies several criteria. First, the model must function at a reasonably high spatial resolution. Second, it must contain an adequate representation of the spatial processes that determine the land use patterns. And finally, it must incorporate a suitable representation of relevant landscape features and legal constraints on land use. Constrained cellular automata running on inhomogeneous cell space have these attributes.

Cellular automata

Cellular automata (CA) are very simple dynamic spatial systems in which the state of each cell in an array depends on the previous state of the cells within a neighborhood of the cell, according to a set of transition rules. CA are very efficient computationally because they are discrete, iterative systems that involve interactions only within local regions rather than between all pairs of cells. Since they are efficient, it is possible to work with grids containing hundreds of thousands of cells. The good spatial resolution that can thus be attained is an important advantage when modeling land use dynamics, especially for planning and policy applications, since spatial detail represents the actual local features that people experience, and that planners must deal with.

A conventional cellular automaton consists of

- a *Euclidean space* divided into an array of identical cells;
- a cell *neighborhood* ;
- a set of discrete *cell states;*
- a set of *transition rules,* which determine the state of a cell as a function of the states of cells in the neighborhood;
- *discrete time steps*, with all cell states updated simultaneously.

Cellular automata (CA) were originally developed to provide a computationally efficient technique for investigating the general nature of dynamical systems. Now, however, they are being used as the basis of highly detailed models in specific domain applications. Tobler (1979) was the first to suggest applying CA to geographical problems. More recently Couclelis (1985, 1988) reintroduced the idea of using CA to model spatial dynamics, and Phipps (1989,1992) and Cecchini and Viola (1990) proposed CA models of various geographical processes. More recent applications have been directed at representing geographical systems realistically, both in terms of the processes modelled and the geographical detail (Couclelis, 1997; Engelen *et al*., 1993, 1996, 1997; Portugali and Benenson, 1995; Portugali *et al*., 1997; White and Engelen, 1993, 1994, 1997a, 1997b; White *et al*., 1997). These advances have been accompanied by an increase in the complexity of the models, and in the effort to build more realistic models, as Couclelis (1997, p.167) noted, “there is practically no defining characteristic of standard CA that researchers,..., have not been able to discard.”

The most important characteristic to be discarded is the homogeneous cell space, replaced by a space in which each cell has its own inherent set of attributes (as distinct from its single state) which represent its relevant physical, environmental, social, economic or institutional characteristics. This modification has allowed CA models to be linked both conceptually and practically with GIS (White and Engelen, 1994; Batty and Xie, 1994; Clarke *et al*., 1997): once the CA is running on an inhomogeneous cell space (essentially identical to what would be found in a raster GIS), the CA may even be thought of as a sort of dynamic GIS. Indeed, some current GIS provide, in principle, a sufficient set of operators to define CA: user defined filters, overlay, reclassification, and a scripting language to put the operations in a sequential order. Some researchers have recently proposed ways of building CA functionality into GIS, or conversely, GIS functionality into CA (Itami, 1994; Takayama, 1996; Wagner, 1997; Wu, 1998; de Savornin Lohman, 1998). At present, however, CA models developed in GIS remain simple, because GIS do not yet provide operators with sufficient flexibility to define complex CA transition rules, and in addition they lack the simulation engines needed to run complex models at practical speeds. At present the more practical approach is to couple GIS to special purpose CA software modules, and possibly other models as well. This is the approach taken in the generic urban simulation model described here.

The urban simulation CA

The model is adapted from a constrained CA-based simulation tool developed previously for modelling urban and regional systems (Engelen *et al*., 1996, 1997; White and Engelen, 1997, 1999; White *et al*., 1997). It is specified as follows.

The cell space

This consists of a rectangular grid of square cells each representing an area 100m square. Thus the minimum area represented is one hectare. This is the same size as the minimum area mapped in urban areas in the land use data set developed for the MURBANDY project, but in the latter case, the hectare-sized area may be of any shape. The grid size and shape varies according to the requirements of the city being modelled, but is typically less than 500 by 500 cells. The grid may be larger, but at the cost of long run times. At present the model executes at a rate faster than one second per iteration (representing one year) on a good PC.

The cell space is assumed not to be homogeneous. First, each cell is characterised by a vector of suitabilities, one suitability for each land use taking part in the dynamics. The suitabilities are defined as a weighted sum or product of a series of physical, environmental and institutional factors characterising each cell. They are normalised to values in the range of 0 – 1, and represent the inherent capacity of a cell to support a particular activity or land use. The suitability maps are calculated in a GIS and imported into the cellular simulation environment. They remain constant during the simulation unless the user interrupts the run and edits them manually. An editor is available in the simulation environment for that purpose.

Second, each cell is associated with a vector of accessibility factors, again one for each land use. These factors represent the importance of access to the transportation networks for the various land uses or activities; some activities, like commerce, require better accessibility than others, such as discontinuous low density residential use. The transportation networks are represented by cell-centred vectors and appear superimposed on the cell grid. Accessibilities are calculated as a function of distance from the cell to the nearest point on the network:

$$A_j = (1+D/a_j)^{-1} \quad (1)$$

where D = the Euclidean distance from the cell to the nearest cell through which the network passes, and

a_j = a coefficient representing the importance of accessibility to the network for land use j.

Finally, each cell has associated with it a set of numbers representing its zoning status for various land uses, and for various periods. The combined effect of suitablities, accessibilities, and zoning is that every cell is essentially unique in its qualities with respect to possible land uses. And it is on this highly differentiated cell space that the dynamics of the cellular automata itself unfold.

The cell neighbourhood

The cell neighbourhood is defined as the circular region around the cell out to a radius of eight cells. The neighbourhood thus contains 196 cells that are arranged in 30 discrete distance zones (1, √2, 2, √5,...). Since the resolution of the grid is 100m, the neighbourhood radius represents 0.8km; this distance delimits an area that is similar to what residents of a city commonly perceived to be their neighbourhood, and thus should be sufficient to allow local-scale spatial processes to be captured in the CA transition rules.

The cell states

The MURBANDY model uses 24 cell states, each representing a CORINE land use or land cover class. Since the purpose of the project is to model urban land uses, the most detailed level of classes is used for these, while agricultural and natural sand uses are represented by more highly aggregated classes (Level 1 and 2). Six of the classes (*road and rail networks, airports, mineral extraction sites, dump sites, artificial non-agricultural vegetated areas,* and *water bodies*) represent fixed features in the model—that is, land covers or uses which are assumed not to change (except by user intervention) and which therefore do not participate in the dynamics. They do, however, affect the dynamics of the active land uses, since in the cell neighbourhood they may represent an attractive or repulsive effect. Another eight (*arable land, permanent crops, pastures, heterogeneous agricultural areas, forests, shrub, sparsely vegetated areas,* and *wetlands*) are passive functions—that is, functions that participate in the land use dynamics, but the dynamics is not driven by an exogenous demand for land; they appear or disappear in response to land being taken or abandoned by the active functions. These latter, the urban land uses (*residential dense, medium dense, continuous, and discontinuous sparse*; *industrial areas; commercial areas; public and private services; port areas;* and *abandoned land*) are forced by demands for land generated exogenously to the cellular automaton, in response to the growth of the urban area. *Construction site* represents a transitional state between one function and another.

The neighbourhood effect

The fundamental idea of a CA is that the state of a cell at any time depends on the states of the cells within its neighbourhood. Thus a neighbourhood effect must be calculated for each of the 17 function states to which the cell could be converted. In the present model, the neighbourhood effect represents the attraction (positive) and repulsion (negative) effects of the various land uses and land covers within the neighbourhood. In general, cells that are more distant in the neighbourhood will have a smaller effect; a positive weight of a cell on itself (zero-distance weight) represents an inertia effect due to the implicit and monetary costs of changing from one land use to another. Thus each cell in a neighbourhood will receive a weight according to its state and its distance from the central cell. Specifically, the neighbourhood effect is calculated as

$$N_j = \Sigma_x \Sigma_d \, w_{kxd} I_{xd} \qquad (2)$$

where w_{kxd} = the weighting parameter applied to land use k at position x in distance zone d of the neighbourhood, and
I_{xd} = the Dirac delta function: $I_{xd} = 1$ if the cell is occupied by land use k; otherwise, $I_{xd} = 0$

The transition rules

A vector of transition potentials (one potential for each function) is calculated for each cell from the suitabilities, accessibilities, zoning, and neighbourhood effect, and the deterministic value is then given a stochastic perturbation (using a modified extreme value distribution), such that most values are changed very little but a few are changed significantly:

$$P_j = vA_jS_jZ_jN_j \qquad (3)$$

where P_j = the potential of the cell for land use j
v = a scalable random perturbation term,
A_j = accessibility of the cell to the road network,
S_j = the intrinsic suitability of the cell for land use j
Z_j = the zoning status of the cell for land use j
N_j = the neighbourhood effect on the cell for land use j

The transition rule is then to change each cell to the state for which it has the highest potential—subject, however to the constraint that the number of cells in each state must be equal to the number demanded at that iteration; cell demands are generated outside the CA. Thus all cells are ranked by their highest potential, and cell transitions begin with the highest ranked cell and proceed downward. When a sufficient number of cells of a particular land use have been achieved, the potentials for that land use are subsequently ignored in determining cell transitions; the result is that some cells are not in the state for which they have the highest potential. Each cell is subject to this transition algorithm at each iteration, although most of the resulting "transitions" are from a state to itself, that is, the cell remains in its current state. When a transition to another state does occur, the actual transition to the second state occurs after a one iteration (one year) delay; during this interval the cell is assigned the state *construction site*.

Land use demands

Land use demands are generated exogenously to the cellular model, since they largely reflect the growth of a city rather than the local configurational dynamics captured by the CA. In some models (e.g. White and Engelen, 1997, 1999) either a conventional system dynamics model or an integrated ensemble of models is used to generate estimates of activity levels which are then converted via productivity or density relations into demands for cell space. In the present model, however, this level of complexity is not considered suitable, since it would increase the difficulty of transferring the generic model to end users, and would not, according to current indications, substantially improve the land use forecasts generated by the cellular level. Thus in the present model, cell demands for each function land use are read from a file at each iteration. The schedule of demands for each land use may be easily generated using the spline tool built into the simulation software (Figure 1).

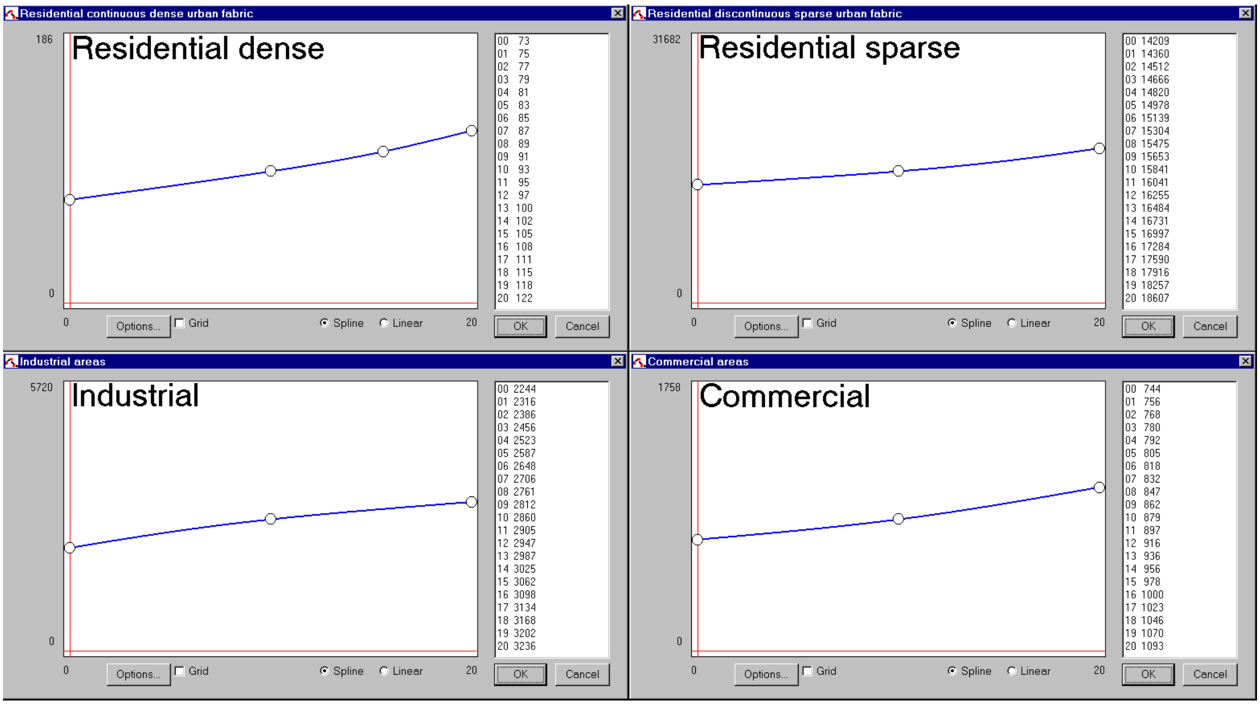

Figure 1: *Dublin: total cell demands for four urban land uses, yearly from 1988 to 2008*

An application to Dublin

In the initial phase of the MURBANDY project the generic model is to be calibrated and tested on five cities: Dublin, Bilbao, Milano, Wien, and one other to be selected. None of the calibrations is complete, because in each case some data is as yet unavailable. However, we present here some preliminary results of an application to Dublin. The calibration is carried out over the period 1988-1998 using the CORINE urban land use data sets developed by the Space Applications Institute, and then run forward to 2008. The calibration makes no use of suitabilities or zoning, since these data are not yet available, and the transport network does not yet contain information on motorway access points or railway stations. Nevertheless, the preliminary results are a useful demonstration of the extent to which the urban form is determined by neighbourhood effects and local accessibility.

The urbanised area of Dublin has expanded considerably over the past 40 years. Figure 2 shows the land use in 1988, the initial year for the simulation, with the road network used in the accessibility calculations superimposed. While the network shows essentially all roads and streets in the urbanised area, such detail is not necessary for the model. A map showing only the more important routes, similar to what appears in the rural areas on the Figure 2, would be sufficient. However, in a future version of the simulation, the road network will be disaggregated into three networks, representing secondary roads, primary routes, and motorways (with access points) in order to achieve more nuanced accessibility indices.

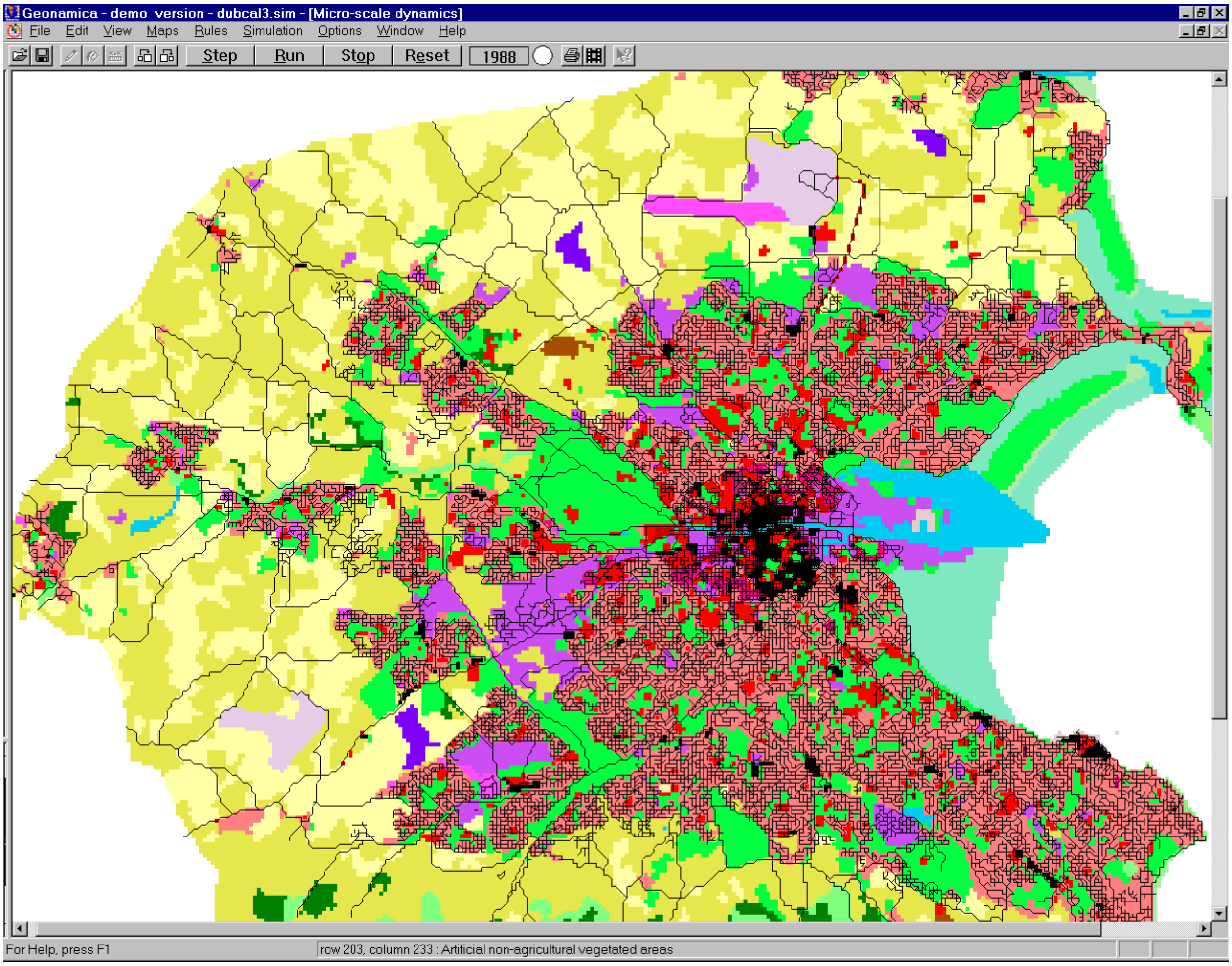

Figure 2: *Dublin: actual 1988 land use, showing transportation network*

The weighting parameters of eq. 2 which define the neighbourhood attraction and repulsion effects were calibrated to minimise differences between the simulated map for 1998 and the actual land use map for that year. The pattern of weighting parameters for any one pair of land uses is unsurprising—for example, *discontinuous sparse residential* land use is attracted to itself, much more so at close distances, and also, but less strongly, to *commerce* (except in immediately adjacent cells exert no attraction) and to *artificial non-agricultural vegetated areas* (primarily parks), while it is repelled by *industry*, especially at close distances. The pattern of weights for the various land use pairs is remarkably similar to that observed in a similar model calibrated to the city of Cincinnati, USA, thus supporting the idea that cities are relatively similar in their underlying processes.

The scaling of these parameter sets with respect to each other is also important, since it determines the magnitudes of the transition potentials for the various land uses relative to each other, and thus the relative power of one land use to displace another. The magnitude of the random perturbations is also a calibrated parameter. In general it is set to reproduce the radial (fractal) dimension of the urbanised area measured for the actual land use map. In this case it is fine tuned also to generate a sufficient number of new "seed" cells of various land uses in new locations, e.g. rural areas, which will subsequently grow into, for example, new industrial, commercial, or residential areas. In a manifestation of stochastic resonance, when the model is finely calibrated, these seed cells tend to appear near the transportation network.

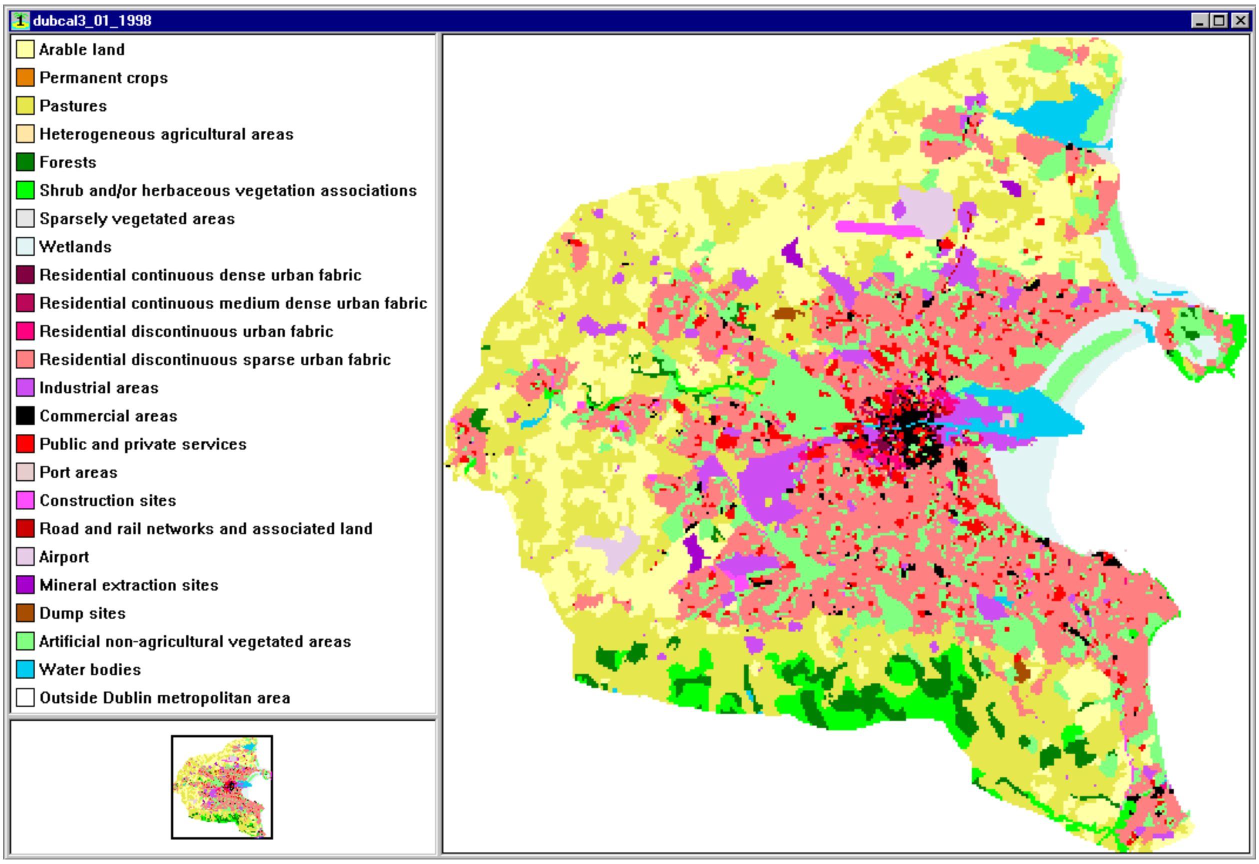

Figure 3: *Dublin: simulated land use, 1998*

In a comparison of the 1998 data with the 1998 simulation shown in Figure 3, the Kappa values for the dynamic land uses range from 0.75 to 1.0, with most values over 0.8. However, the Kappa statistic is of limited value in comparing the maps. On the one hand it overstates the performance of the model, since it is based on total coincidences of cells, rather than on the degree to which additional cells placed by the model coincide with additional cells appearing on the actual land use map. In another sense, however, it understates the degree of agreement, since the model (like reality) is stochastic, and so whether a cell placed by the simulation actually coincides with a cell placed in reality is of little interest. What is important is that the locations be *similar*, and that the overall *patterns* of location, in various relevant aspects, resemble each other. Thus in judging the quality of a simulation, we pay much more attention to similarities in locational patterns between the two maps than to measures like the Kappa statistic. The importance of pattern similarity has prompted work to find ways of representing and quantifying the similarities between configurations using fuzzy set theory and polygons rather than cells to generate comparison measures (Power et al., 1999).

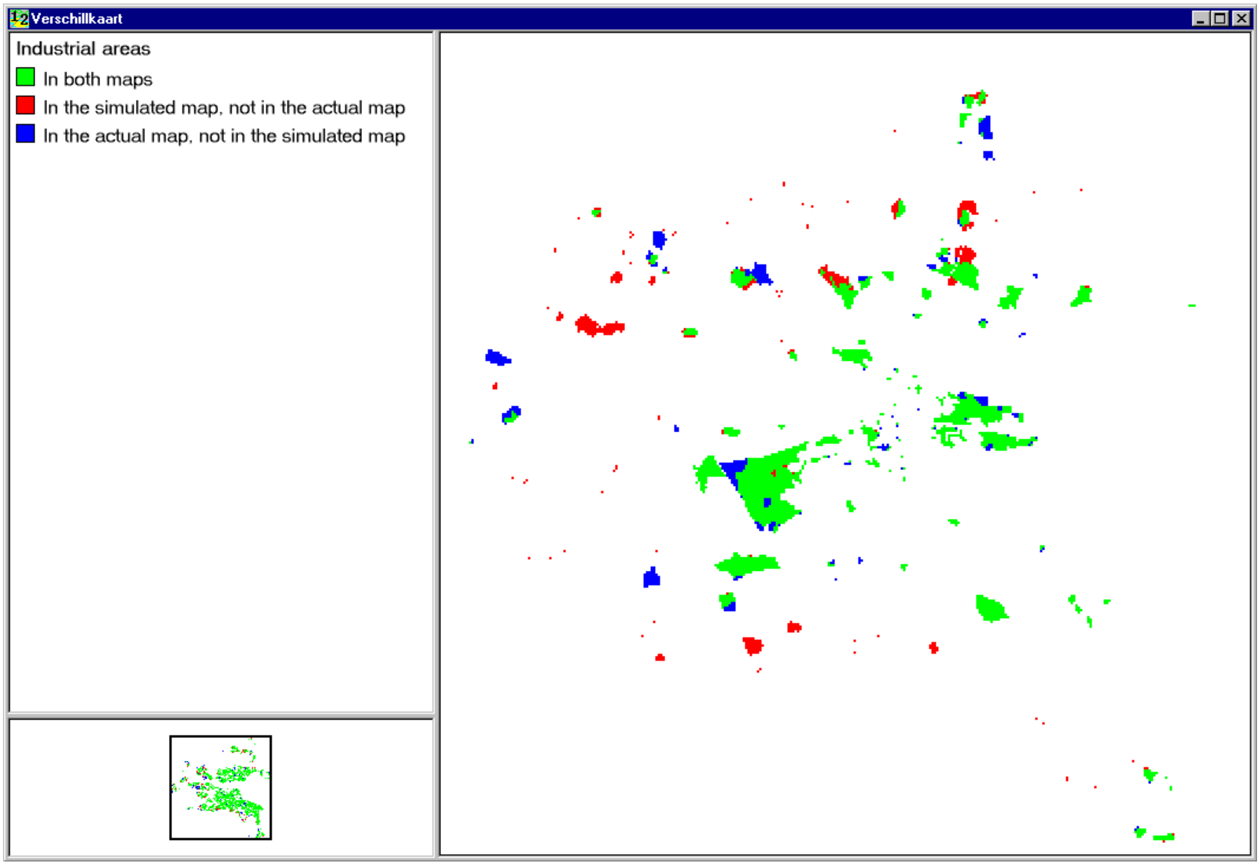

Figure 4: *Comparison of simulated and actual location of* industry*, 1998*

Returning to the approach of a cell-by-cell evaluation, Figure 4 shows a comparison of the actual and simulated locations of *industry*. While most new industry cells placed by the simulation model do not coincide with those that actually appeared in the period 1988-1998, the sizes of the clusters are quite similar in the two cases, and so are the types of locations observed. The two patterns would be much more similar if suitability and zoning were included, since zoning, especially, is important in determining which of several sites of roughly equal potential will actually be occupied by industry. Without the zoning constraint, the choice among the sites becomes essentially random, as will be seen when the calibrated model is re-run, since the stochastic perturbation ensures a different result at each run.

Conclusions

The current calibrations give surprisingly good results in view of the small amount of data used to constrain the simulation. This is an important advantage, since many situations in which such a model might be used may be relatively data-poor. Nevertheless, the addition of suitabilities, zoning, and a more complete network representation should significantly improve the quality of the simulation results. Since the modelling framework is open, it is possible to include almost any additional data that is relevant to the evolution of the urban land use structure, including the output of other models. Thus as the simulation tool is moved closer to end users such as urban planners who have access to a wealth of data, it will be possible to extend the richness of the modelling framework to provide better estimates of future conditions under various scenarios.

References

Batty, M. and Xie Y. “From cells to cities”, *Environment and Planning B*, Vol. 21, pp. s31-s48, 1994

Cecchini, A. and Viola, F., “Eine Stadtbausimulation”, *Wissenschaftliche Zeitschrift der Hochschule fur Architektur und Bauwesen*, Vol.36, pp. 159-162, 1990

Clarke, K., Hoppen S. and Gaydos I., “A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area”, *Environment and Planning B*, Vol. 24, pp. 247-261, 1997

Couclelis, H., “Cellular worlds: a framework for modelling micro-macro dynamics”, *Environment and Planning A*, vol. 17, pp. 585-596, 1985

Couclelis, H., “Of mice and men: what rodent populations can teach us about complex spatial dynamics”, *Environment and Planning A*, Vol.20, pp. 99-109, 1988

Couclelis, H., "From cellular automata to urban models: new principles for model development and implementation", *Environment and Planning B*, Vol. 24, pp. 165-174, 1997

de Savornin Lohman, A., "Landgebruik in Beweging: Het Gebruik van Cellular Automata voor Ruimtelijke Planning", Doctoraalverslag, Wageningen: Landbouwuniversiteit, 1998

Engelen, G., White, R., and Uljee, I., "Exploratory Modelling of Socio-Economic Impacts of Climate Change", in Maul, G. (ed), *Climate Change in the Intra-Americas Sea*, London: Edward Arnold, 1993

Engelen, G., White, R., Uljee, I., and Wargnies, S, "Numerical Modeling of Small Island Socio-Economics to Achieve Sustainable Development", in Maul G.A. (ed.) *Small Islands. Marine Science and Sustainable Development*, Coastal and Estuarine Studies, *51,* Washington DC: American Geophysical Union, pp.437-463, 1996

Engelen, G., White, R. and Uljee, I., "Integrating constrained cellular automata models, GIS and decision support tools for urban planning and policy-making", in Timmermans H. (ed.), *Decision Support Systems in Urban Planning,* London: E&FN Spon, pp. 125-155, 1997

Frankhauser, P., *La fractilite des structures urbaines*. Paris: Economica, 1994

Itami, R., "Simulating spatial dynamics: cellular automata theory", *Landscape and urban Planning*, Vol. 30, pp. 27-47, 1994

Phipps, M., "Dynamical behaviour of cellular automata under constraints of neighbourhood coherence," *Geographical analysis*, Vol. 21, pp. 197-215, 1989

Phipps, M., "From local to global: the lesson of cellular automata", in DeAngelis, d. and Gross, L. (eds.), *Individual based models and approaches in ecology: populations, communities and ecosystems* New York: Routledge, Chapman and Hall, pp. 165-187, 1992

Portugali, J. and Benenson, I., "Artificial planning experience by means of a heoristic cell-space model: simulating international migration in the urban process". *Environment and Planning A,* Vol. 27, pp. 1647-1665, 1995

Portugali, J., Benenson, I. and Omer, I., "Spatial cognitive dissonance and sociospatial emergence in a self-organizing city", *Environment and Planning B,* Vol.24, pp. 263-285, 1997

Power, C., Simms, A. and White, R., *Hierarchical fuzzy pattern matching for the regional comparison of land use maps.* Maastricht: Research Institute for Knowledge Systems, 1999

Takeyama, M., *Geo-Algebra: A Mathematical Approach to Integrating Spatial Modeling and GIS*, PhD. Thesis, Santa Barbara: University of California, 1996

Tobler, W., "Cellular geography", in Gale S. and Olsson G. (eds.), *Philosophy in Geography*, Dordrecht: Reidel, pp. 379-386, 1979

Wagner, D., "Cellular automata and geographic information systems", *Environment and Planning B*, Vol 24, pp. 219-234, 1997

White, R. and Engelen, G., "Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land use patterns", *Environment and Planning A*, Vol.25, pp. 1175-1199, 1993

White, R. and Engelen, G., "Cellular dynamics and GIS: modelling spatial complexity", *Geographical Systems*, Vol. 1, pp. 237-253, 1994

White, R. and Engelen, G., "Cellular automata as the basis of integrated dynamic regional modelling", *Environment and Planning B*, Vol.24, pp. 235-246, 1997a

White, R. and Engelen, G., "Multi-scale spatial modelling of self-organizing urban systems", in Schweitzer, F. and Haken, H. (eds), *Self-Organization of complex structures: from individual to collective dynamics*, London: Gordon and Breach, pp. 519-535, 1997b

White, R. and Engelen, G., "High Resolution Integrated Modelling of the Spatial Dynamics of Urban and Regional Systems", *Computers, Environment, and Urban Systems*, in press, 1999

White R., Engelen G. and Uljee I., "The use of constrained cellular automata for high-resolution modelling of urban land use dynamics", *Environment and Planning B*, Vol. 24, pp. 323-343, 1997

Wu, F., "SimLand: a prototype to simulate land conversion through the integrated GIS and CA with AHP-derived transition rule", *International Journal of Geographical Information Science*, Vol 12, pp. 63-82, 1998