Environment and Planning B: Planning and Design 2001, volume 28, pages 733 – 753

DOI:10.1068/b2740

A constrained CA model for the simulation and planning of sustainable urban forms by using GIS

Anthony Gar-On Yeh, Xia Li ¶
Centre of Urban Planning and Environmental Management, The University of Hong Kong, Pokfulam Road, Hong Kong SAR, PR China; e-mail: hdxugoy@hkucc.hku.hk, lixia@graduate.hku.hk
Received 2 June 2000; in revised form 26 October 2000

Abstract. A constrained cellular automata (CA) model based on 'grey cells' for the simulation of different types of urban forms and developments is developed within a raster GIS. Remote sensing and GIS are used to supply information on environmental constraints and locations of growth centers for the simulation. The model is able to consider different criteria, such as urban forms, environmental suitability, and land consumption for the planning of sustainable cities. Seven different types of urban forms and developments that range from compact – monocentric, compact – polycentric, compact – monocentric – environmental, compact – polycentric – environmental, dispersed, highly dispersed, to very highly dispersed developments were simulated by using the model. The model can generate urban forms and developments with fractal structures that are close to real cities. The simulated patterns were evaluated by using different cost indicators related to sustainable development principles.

1 Introduction

In recent years, cellular automata (CA) have been widely applied to the simulation of urban growth and forms (Batty and Xie, 1994; 1997; Deadman et al, 1993; White and Engelen, 1997; Wu and Webster, 1998). CA models can be set up to represent either a generic city or any particular city (White et al, 1997). Some CA models are intended just to investigate basic questions of urban forms and evolution of urban systems rather than to provide realistic simulations of the development of particular cities (White and Engelen, 1993). CA models are considered to be aids in thought experiments and provide important insights into the nature of geographical processes. There are other CA approaches that aim to simulate urban dynamics of real cities. Batty and Xie (1994) have illustrated how CA can deliberately articulate global patterns through some local processes. Their simulation is based on the suburban expansion of a peripheral municipality, the town of Amherst, in metropolitan Buffalo, NY. White et al (1997) also provide a realistic example of CA simulation for the land-use pattern of Cincinnati, OH. They have demonstrated that the CA approach is appealing because it enables us to think clearly about the dynamics of development in a geometric framework. Simulated outcomes are rigorously associated with changes in model structure, parameters, and constraints. The experiments suggest that CA may be useful in a planning context.

CA have provided powerful and convenient spatial modeling functions, especially when they are integrated with geographical information systems (GIS). CA have been popularly applied to urban simulation in order to discover the mechanisms of urban evolution. Modeling spatial complexity of real cities can now be easily realized by the assistance of GIS. Recent rapid development of GIS has provided the capabilities to satisfy the modeling environment and data requirements of CA simulations. Particularly, land-use information can be conveniently obtained from remote sensing, stored in GIS, and then passed to CA models as some kind of constraint for simulation

¶ Also affiliated to: Guangzhou Institute of Geography, Guangzhou, PR China.

(Li and Yeh, 2000). The integration with GIS can allow the CA models to access a large volume of spatial data and a variety of other social and economic data. As a result, the modeling results are closer to reality and are applicable for planning purposes.

The objective of this paper is to develop a CA model that can be used to simulate different urban forms and developments in the planning of sustainable cities. Alternative development patterns can be formulated by incorporating different 'sustainable' elements in the CA model. We also aim to develop methodology for comparing the performances of different urban forms and developments that are generated from the model.

2 Urban forms and sustainability

There are continuing debates on whether urban forms should be compact or dispersed (Ewing, 1997; Gordon and Richardson, 1997). Concentrated core-oriented metropolises or monocentric cities have emerged by agglomerating industry and employment in a single center and by packing the population around the center and the long, radiating transport networks (Alonso, 1964; Berry and Kim, 1993; Mills D, 1967; Mills E, 1981; Muth, 1969). This type of urban form was at its peak by World War 2, but since then it has been replaced by suburbanization and dispersion. Contemporary urban growth is characterized by dispersal and decentralized patterns (Gordon and Richardson, 1997). Restructuring has been witnessed by the decentralization of jobs, services, and residences from traditional urban centers to suburban settings and 'edge cities' within expanded metropolitan areas (Garreau, 1991). The new urban regions are multicentered with more than one core (Fishman, 1990). This trend is the result of a variety of factors, but perhaps the most important ones are the advancements in transportation and communication technology (Archer and Smith, 1993). These types of land development can be regarded as a form of urban sprawl where land is scattered in isolated land parcels, separated from other areas by vacant land (Ottensmann, 1977). It is also often referred to as leapfrog development (Gordon and Richardson, 1997). Urban sprawl has often been criticized for inefficient use of land resources and energy and for large-scale encroachment on agricultural land. One criticism of urban sprawl is that it wastes and misallocates land. It also contributes to public-sector inefficiency as infrastructure costs are greater when leapfrog, scattered, and mixed developments occur (Mills, 1981).

Urban areas are growing rapidly in some fast developing countries. This has triggered off a lot of environmental and resource problems because urban growth is usually accompanied by agricultural land loss and wasteful use of land resources. Urban sprawl has become a major problem in many rapidly growing cities. Although the trend of urban sprawl continues in many Western countries, the debate on urban sprawl is not over and is still the subject of controversy. Most urban planners still remain convinced that urban sprawl is undesirable (Ewing, 1997). This is further supported by the recent concept of sustainable development and sustainable cities. Compact development has increasingly been promoted by urban planners for containing urban development and conserving land resources. The relationship between urban form and sustainability is currently one of the most hotly debated issues. Although it is not simple and straightforward, research has shown that urban forms can be significantly linked to sustainable development. It is found that compact development patterns can help to reduce energy consumption and resource depletion, although the relationships may be complicated (Banister et al, 1997; Ewing, 1977; Owens, 1992).

The planning of urban form will be central to the promotion of sustainable development (Breheny, 1996). There have been many debates on how to confine urban sprawl and conserve agricultural land resources (Bryant et al, 1982; Daniels, 1997; Ewing, 1997; Howard, 1972). An essential part of planmaking is to develop simulation models in generating regional and subregional growth alternatives. The California Urban Future Model (CUF model) is such a model, which is valuable for presenting and comparing the details of different development scenarios (Landis, 1995). The model is the first large-scale metropolitan simulation model to use a GIS for data integration and spatial analysis. Clarke et al (1997) also provide a CA model that can simulate urban dispersion using self-modifying rules. This model has been applied to the simulation of urban growth in the San Francisco Bay area. The behavior of the system is mainly controlled by five factors, which can be calibrated by historical data. The model is used to generate three predictions of urban growth in the region—uncontrolled rapid growth, sustained slow growth, and a growth which stabilizes at a desirable or sustainable level. Deadman et al (1993) applied a CA-based model to predict patterns in the spread of rural residential development. They demonstrated the potential of the model to be run 'into the future' to predict the outcome of policy decision. There is a lot of potential in using CA with GIS data to simulate possible urban forms for the planning of sustainable urban development. These models can enable planners to explore various options and predict possible environmental impacts.

3 Transition rules for sustainable cellular cities

Transition rules are central to CA simulation. The transition rules for standard CA are neighborhood-based as the transition potential of a central cell is determined by the states in the neighborhood. There are numerous ways of defining transition rules for CA simulations. A common way is to use the probability of conversion which is estimated from the hierarchy of local (neighborhood), regional, and global factors. However, any type of information obtained from the hierarchy should return exact values to the central cell as the simulation is cell-based. The information can be regarded as some kind of attractiveness and constraints which are essential to the CA simulation. Wu (1998) defines a utility score to indicate the attractiveness of the site before estimating the probability of development. The utility of selecting a cell for development is calculated by the linear combination of various types of attractiveness. A logistic function is further used to transform the utility score into probability.

The appropriate configuration of CA models can enable various types of urban forms and structures to be explored. The bottom-up representation from basic spatial units can generate realistic and desired global development patterns (Batty and Xie, 1994; Li and Yeh, 2000). White and Engelen's (1993) study also indicates that cellular approaches can achieve a high level of spatial detail and realism. Similar fractal dimensions are found between cellular cities and actual cities.

Different types of development, including sprawl and compact patterns, can be simulated by the use of different model parameters. However, many existing CA approaches in urban simulation are rudimentary monocentric-based. Usually, these models do not pay enough attention to the structure of urban systems. Wu (1998) develops a CA model to simulate the formation of subcenters from a monocenter through a series of stochastic 'errors'. The clusters of development are generated by the cumulative and aggregated self-organizing process.

Real cities usually grow in a polycentric pattern although their urban forms can be either sprawl or compact. There are a lot of alternative urban forms. In this study we will demonstrate that different types of urban forms and developments can be generated by using the proposed CA model. The simulated urban forms and developments can then be evaluated to find which type of forms and developments can better fulfil the criteria of sustainability. CA models do not only enable the prediction of future urban development, they can also be used to find feasible alternatives for the planning of sustainable cities.

In the general CA models, there is only a binary value for the selection of a cell—selected or not for conversion (development). The value is usually decided by comparing the probability of conversion with a threshold value or a random number. The cell is selected when the probability is greater than the threshold value or the random number. However, the binary value has limitations when it is used to define transition rules. It does not allow selecting cells for development based on a cumulative process. Constraints are also not easily incorporated in the transitional rules in this type of model. A cell will not 'suddenly' mature for development. It will be more appropriate to select a cell for conversion gradually through a couple of iterations in simulation. A 'grey cell' can be defined to address the state of this continuous selection process. It can be used to deal with local, regional, and global constraints in constrained CA, discussed in the paper by Li and Yeh (2000). In this paper, 'grey cells' are further developed and used to simulate different types of urban forms and developments.

The state of a cell in the 'grey-cell' method is expressed by a continuous value for development or conversion. The value indicates the cumulative degree of development for a candidate cell before it is completely selected for development or conversion. An iteration formula can be defined for the cumulative process:

$$G_{xy}^{t+1} = G_{xy}^{t} + \Delta G_{xy}^{t} \;, \tag{1}$$

where G is the 'grey value' for development which falls within the range of $0-1$; t is the simulation time; and xy is the location of the cell. A cell will not be regarded as a developed cell until the value reaches 1. The value should be assigned to 1 when it is greater than 1 during the calculation. ΔG^t is the gain of the 'grey value' at each loop.

The essential part of the model is to calculate the value of ΔG^t, which can be defined by using the neighborhood function which is the basis of CA simulations. According to the neighborhood function, the probability of conversion at a cell is dependent on the states of its neighboring cells. There is a higher chance for the conversion at a cell if it is surrounded by more cells which have already been converted. The increase of 'grey value' should be determined by the amount of developed cells in the neighborhood. ΔG^t can be simply defined by the following neighborhood function:

$$\begin{aligned} \Delta G_{xy}^{t} &= \mathrm{f}_{xy}(q, N) \\ &= \frac{q}{\pi l^2} \;, \end{aligned} \tag{2}$$

where q is the total number of developed cells in the neighborhood; l is the radius of the circular neighborhood. A circular neighborhood is used because it has no bias in all directions (Li and Yeh, 2000). The neighborhood has a radius of 2 pixels.

This transition function depends only on the amount of developed cells in the neighborhood. The evolution of real cities in influenced by a series of complicated factors which can be obtained on various local (neighborhood), regional, and global levels. The neighborhood function cannot address the issue of urban structures and environmental problems. Some kinds of constraints should be used to regulate the simulation to improve modeling accuracy. Without the constraints, the CA simulation will generate patterns as usual based on historical trends. The constraints should be added into the model to address environmental and sustainable development considerations. They are the important factors for the formation of idealized patterns.

By taking environmental and other factors into consideration, equation (2) can be revised as:

$$\Delta G_{xy}^{t} = \mathrm{f}_{xy}(q, N) \times \prod_{i}^{m} \delta_{ixy}(I, N, R, G) \tag{3}$$

where δ_{xy} is the function to address the constraints from local, regional, and global influences which should be normalized to the range of $0-1$. It can be regarded as a scaling factor to readjust the increase of 'grey value'. I is the information for obtaining the constraints from various sizes of neighboring areas. N, R, and G are local (neighborhood), regional, and global areas, respectively.

Local (neighborhood) factors calculate the number of development cells (state) or suitability in the small set of cells around a location. Regional factors measure the influences away from a distance. They can be measured by the use of some kind of distance decay functions. The values of global factors are invariant spatially, but changeable temporally. In this model, three types of constraints—urban forms, environmental suitability, and land consumption—are used. However, more constraints can easily be added to the model if needed. Table 1 lists the different factors used in defining the transition rules of local, regional, and global constraints of the constrained CA simulation.

Table 1. Factors in the transition rules for the constrained CA simulation.

Factors	Local (neighborhood)	Regional	Global
Neighborhood function	Amount of developed cells (q)		
Urban form		Distances from centers (d_{R}, d_{r}); weights (w_{R}, w_{r})	
Environmental suitability	Agricultural suitability	Distances from protected sources, such as drinking water (rivers), forest, wetland.	
Land consumption			Density control

R, the main center; r, the nearest subcenter.

A stochastic disturbance term can also be added to the model to represent the unknown errors during the simulation. This can allow the generated patterns to be closer to reality. The random error term (R_{e}) can be given by (White and Engelen, 1993):

$$R_{\mathrm{e}} = 1 + (-\ln \gamma)^{\alpha}, \tag{4}$$

where γ is a uniform random variable within the range {0, 1}, and α is a parameter to control the size of the stochastic perturbation (α can be used as a dispersion factor in this simulation).

Equation (3) becomes:

$$\Delta G_{xy}^{t} = [1 + (-\ln \gamma)^{\alpha}] \times \mathrm{f}_{xy}(q, N) \times \delta_{xy}(\mathsf{FORM}) \times \delta_{xy}(E_{\mathrm{T}}) \times \delta_{xy}(\mathsf{LAND}) , \tag{5}$$

where $\delta_{xy}(\mathsf{FORM})$ is the constraint for urban forms, $\delta_{xy}(E_{\mathrm{T}})$ is the constraint for environment suitability (the total environmental suitability score), and $\delta_{xy}(\mathsf{LAND})$ is the global constraint that uses land consumption to control development density.

3.1 Urban forms

The function of $\delta_{xy}(\mathsf{FORM})$ is determined by the relationships between urban growth and urban centers. Land development can be concentrated around the main center to promote the growth of large cities, or it can be shifted to around subcenters to promote polycentric growth. There are many possible urban growth patterns which are the main concerns of urban planning. Land development and urban forms are closely related to the efficient use of energy, capital, and land resources (Banister et al, 1997; Burchell et al, 1998). The constraint is decided by location factors in terms of the distance to urban centers. Urban centers play an important role in urban growth as they provide the support for the requirements of energy, materials, capital, and techniques for development. The influence of urban centers can be measured by a distance decay function. The classical measure of urban structure is the density gradient from the CBD (Muth, 1969). Although the density gradient is related to the monocentric concept of urban form, nevertheless it gives us an index of the degree of decentralization. Two distances can be defined to capture the hierarchy of urban structures that consist of a major center and many subcenters. The constraint score which indicates the attractiveness related to urban centers can be expressed by the following function:

$$\delta_{xy}(\mathsf{FORM}) = \exp\left[-\frac{(w_{\mathrm{R}}^2 d_{\mathrm{R}}^2 + w_{\mathrm{r}}^2 d_{\mathrm{r}}^2)^{1/2}}{(w_{\mathrm{R}}^2 + w_{\mathrm{r}}^2)^{1/2}}\right], \tag{6}$$

where d_{R} is the distance from a cell to the main center, d_{r} is the distance from the cell to its closest subcenter, and w_{R} and w_{r} are the weights for the two distances.

A higher value of $w_{\mathrm{R}}/w_{\mathrm{r}}$ indicates that more weight is given to the main center. In contrast, a lower value of $w_{\mathrm{R}}/w_{\mathrm{r}}$ will put more weight to the subcenters. The growth rate of the 'grey value' at a location is affected by the constraint. Different types of urban forms can emerge by simply changing the value of w_{R} and w_{r}. A higher value of $w_{\mathrm{R}}/w_{\mathrm{r}}$ will give a monocentric development and a lower value will give a dispersed polycentric development.

3.2 Environmental suitability

Environmental suitability is an important constraint in the model. There is growing concern on environmental issues related to urban growth in the world. Urban development should be determined not only by pure economic factors, but also by environmental constraints. Environmental consciousness should be reflected in the model so that idealized urban development patterns can be formulated. It is convenient to obtain environmental constraints and embed them in CA simulation based on the integration of CA and GIS technologies.

Environmental suitability is used to indicate whether a piece of land is suitable for development or not with regard to environmental considerations. Urban development is associated with some amount of environmental loss, such as agricultural land loss. There is a need to protect agricultural land and other important ecological areas in urban planning. The model uses environmental suitability to facilitate the implementation of these environmental factors. The total environmental suitability score can be calculated by combining the agricultural suitability score and other environmental scores for protecting resources and the environment. Agricultural suitability reflects the potential of agricultural production which can be used to address the need to reserve important agricultural land. Higher costs could be associated with encroachment onto sites of good-quality agricultural land. Other environmental scores can be used to address the disturbance of development when it comes close to some environmentally sensitive or protected areas, such as drinking-water supply rivers, forest, wetland, and other protected ecological areas. Development in the neighborhood of these types of land use can bring about environmental degradation and ecological disturbance. The environmental scores can be defined based on the buffer distances to these environmentally sensitive areas. The influences should be in the form of distance-decay functions.

Multicriteria evaluation (MCE) techniques can be employed to obtain the total score of environmental suitability based on these environmental factors. It is necessary to standardize the scores because they may be measured on different scales. A typical method of standardization is to use the minimum and maximum values as scaling points for a simple linear transformation (Voogd, 1983). However, other types of nonlinear transformation can provide more plausible results by achieving greater discrimination between cells (Li and Yeh, 2000; Wu and Webster, 1998). The CA simulation based on a linear transformation cannot generate typical development patterns. The nonlinear transformation can be defined in an ad hoc way because a unique transformation does not exist. The transformation can be in exponential (Wu and Webster, 1998), logistic (Wu, 1998), or power forms (Li and Yeh, 2000). Usually, the values of the adjusted scores should fall within the range of 0 to 1 for comparison. It is rather arbitrary to specify the parameters in these forms of nonlinear transformation. However, it is this flexibility that allows CA models to be able to simulate various global patterns. The parameters then have meanings, by linking the values to the forms. It is simple and convenient to adjust the values of parameters rather than change the configuration of transition rules.

A location with a high environmental suitability score (E_{T}) means that there is more of a need to protect it from development. The constraint value of $\delta_{xy}(E_{\mathrm{T}})$ will then be smaller to ensure that the growth rate of the 'grey value' is smaller. This can be accomplished by using the following expression:

$$\delta_{xy}(E_{\mathrm{T}}) = \sum_{i=1}^{n} w_i (1 - E_{ixy})^k \ , \tag{7}$$

where E_{T} is the total environmental suitability score, E_{ixy} is the environmental suitability score and w_i is the weight for the ith environmental criterion. Each factor should be normalized between the range 0 to 1. A higher value of k (the parameter for the nonlinear transformation) will ensure that the environmentally sensitive land can be strictly protected, but the patterns may be fragmented.

3.3 Land consumption

The global function $\delta_{xy}(\mathsf{LAND})$ controls the amount of land consumption in the model. A higher density of land use can be achieved when less amount of land is developed for the same population. The model will stop simulation based on the global control function.

4 Simulation of urban forms and developments

The constrained CA model was implemented within a GIS package, ARC/INFO GRID by using the Arc Macro Language (AML). The data set for the simulation was the same as the previous study (Li and Yeh, 2000). The study area is located in Dongguan in the Pearl River Delta region of southern China which is a fast growing region with a tremendous amount of land-use changes and urban sprawl in recent years. The focus of the previous study is mainly on protecting agricultural land by using CA simulation. In this study, the constrained CA is further developed based on the concept of 'grey cells' in simulating different urban forms and developments.

The model has been substantially improved because the model structure is changed and the factors in this model are more generic. The simulation of various urban forms is one of the new focuses in this study. A stochastic disturbance term is incorporated in the 'grey cells' to generate the patterns of urban sprawl. Different urban forms and developments can be obtained by using different parameters in the 'grey-cell' CA model.

The initial land-use information for the simulation was obtained by using Landsat TM images (Li and Yeh, 1998). Land-use information, including the growth centers, was detected and further imported into ARC/INFO GRID as the basis of simulation. Other physical and economic data are also stored in the database, such as roads, agricultural suitability, and administrative boundary maps. The spatial data are converted into grids in ARC/INFO format. The grids used in the simulation have a dimension of 709×891 pixels with a resolution of 50 m on the ground. The original TM images have a resolution of 30 m on the ground. The resolution was reduced to 50 m for faster simulation. The initial grid for the simulation is based on the urban areas classified from the 1988 TM image (figure 1). There was a major center (the city proper) in the northwest part and many subcenters (towns) in rural areas across the region in the initial stage. In this study, we attempt to simulate various types of urban forms and development for the period of 1988–93 based on the major center and subcenters. All the simulated patterns consume the same amount of land that was actually consumed in 1988–93. The simulated results will be compared with the actual development pattern which was obtained from the classification of the 1993 TM image (figure 2).

The simulation of alternative urban forms and developments is important to urban planning. The CA model can be used to simulate alternative patterns by changing the model parameters. Seven different types of urban forms and developments ranging from compact–monocentric, compact–polycentric, compact–monocentric–environmental,

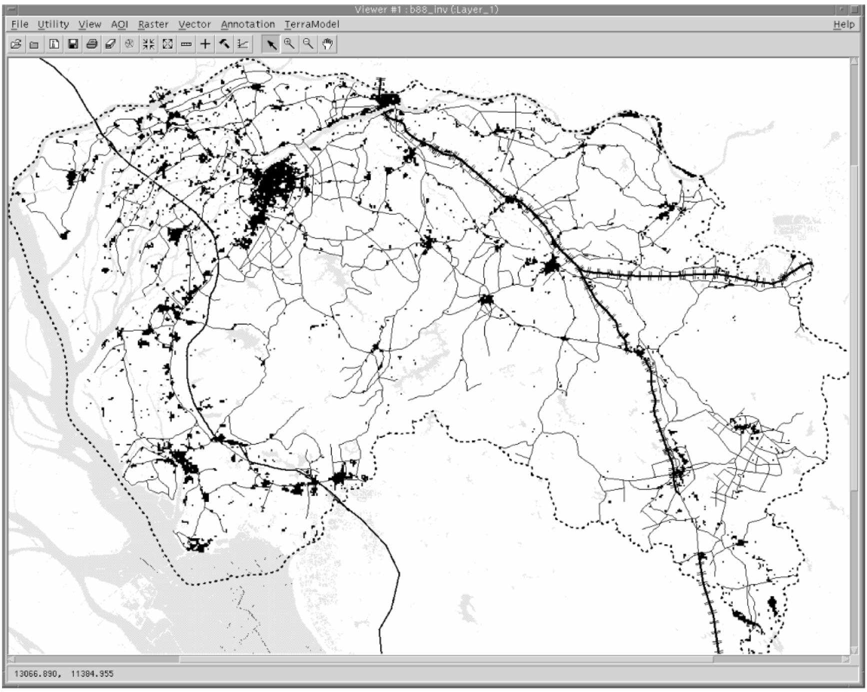

Figure 1. Urban areas of Dongguan classified from the 1988 TM image (initial).

compact – polycentric – environmental, dispersed, highly dispersed, to very highly dispersed developments were simulated and evaluated by using the model. Other mixed urban forms can easily be generated by using the same method and changing the parameters. The parameters used for the simulation of the seven types of urban forms and developments are listed in table 2. The same form of transition function [equation (5)] is applied for all the simulations. The global density control acts to ensure that the total amount of land conversion in the simulations is the same as the actual land conversion. The dispersion parameter (α) is used to explore various dispersion and sprawl of land development.

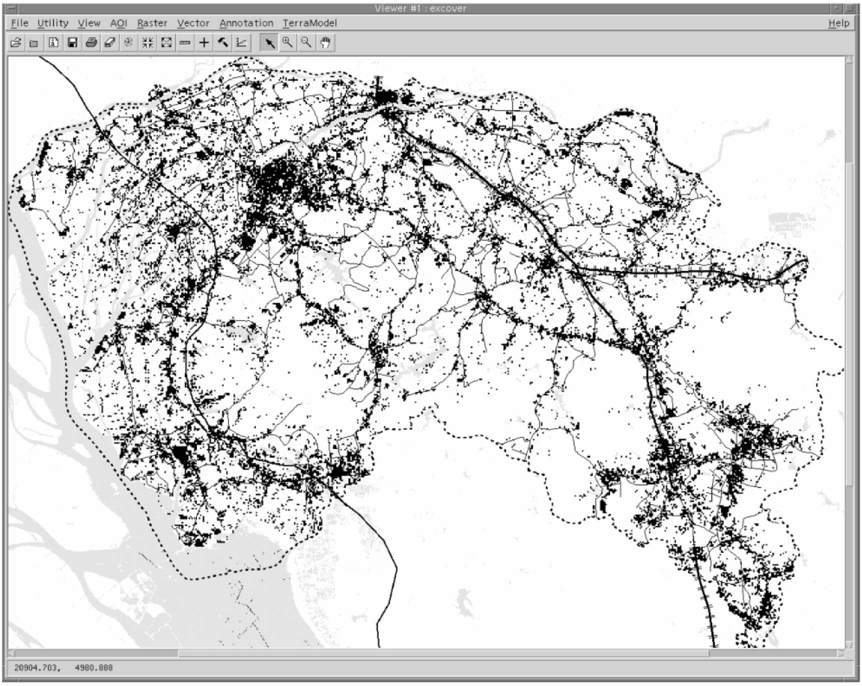

Figure 2. Sprawl pattern of Dongguan classified from the 1993 image (actual).

Table 2. Model parameters for simulating different types of urban forms and developments.

Urban forms and developments	Dispersion factor	Urban form	Environmental consideration
(1) Compact – monocentric development	$\alpha = 0$	$w_R = 1$; $w_r = 0$	nil
(2) Compact – polycentric development	$\alpha = 0$	$w_R = 0$; $w_r = 1$	nil
(3) Compact – monocentric – environmental development	$\alpha = 0$	$w_R = 1$; $w_r = 0$	$\delta_{xy}(E_T)$
(4) Compact – polycentric – environmental development	$\alpha = 0$	$w_R = 0$; $w_r = 1$	$\delta_{xy}(E_T)$
(5) Dispersed development	$\alpha = 1$	nil	nil
(6) Highly dispersed development	$\alpha = 5$	nil	nil
(7) Very highly dispersed development	$\alpha = 10$	nil	nil

α, dispersion parameter; w_R, w_r, weighting parameters for the main center and the closest subcenter, respectively.

4.1 Compact – monocentric development

This emphasizes the role of the main center in supporting urban growth. In this scenario, the growth rate is affected by the distance to the city proper, but it is not affected by the distances to subcenters. Higher growth rates are allowed only around the city proper. There is very limited growth around those subcenters that are farther away from the city proper. The dispersion factor α is set to 0. As a result, these settings can promote compact development which contains a large part of development within a small area around the city proper (the northwest part, as shown in figure 3). This pattern is generated without taking environmental factors and constraints into consideration.

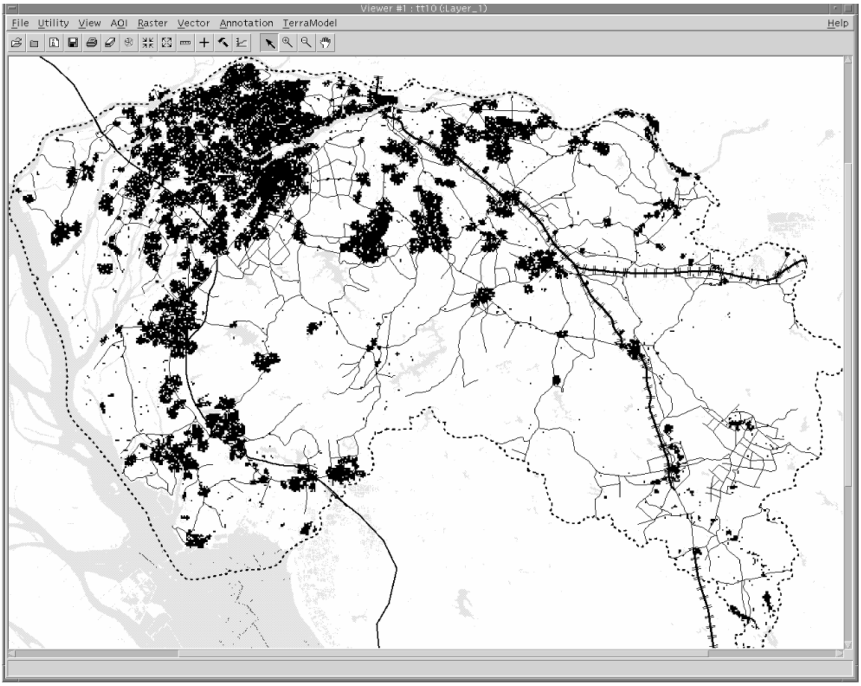

Figure 3. Simulation of compact – monocentric development.

4.2 Compact – polycentric development

This assumes that development is dominated by the process of decentralization. Towns will be allowed to grow faster than the major urban area—the city proper. The patterns of polycentric growth are encouraged by using the constraint function of $\delta_{xy}(\mathsf{FORM})$ in which w_R is set to 0 and w_r is set to 1. Higher growth rates are applied to the areas which are closer to subcenters. The simulation generates a polycentric growth pattern (figure 4). However, it is still rather compact as compared with the actual pattern (figure 2). Similar to the compact – monocentric development, it does not take environmental factors and constraints into account.

4.3 Compact – environmental development

Environmental suitability should be an important factor for sustainable urban development. A series of factors can be identified for obtaining the suitability. Multicriteria evaluation (MCE) techniques are needed to obtain the total score. In this simulation, agricultural suitability is used as an important constraint factor which is obtained from land evaluation. A nonlinear transformation of agricultural suitability into the constraint score is applied to protect good agricultural land. The value of k in equation (7) is set to 3 to produce greater effects on protecting agricultural land.

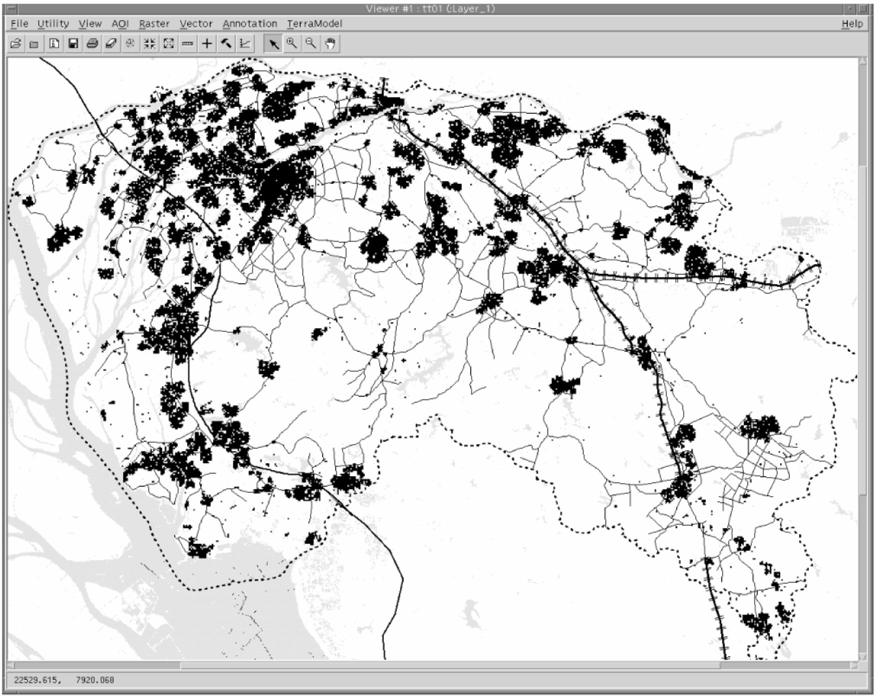

Figure 4. Simulation of compact – polycentric development.

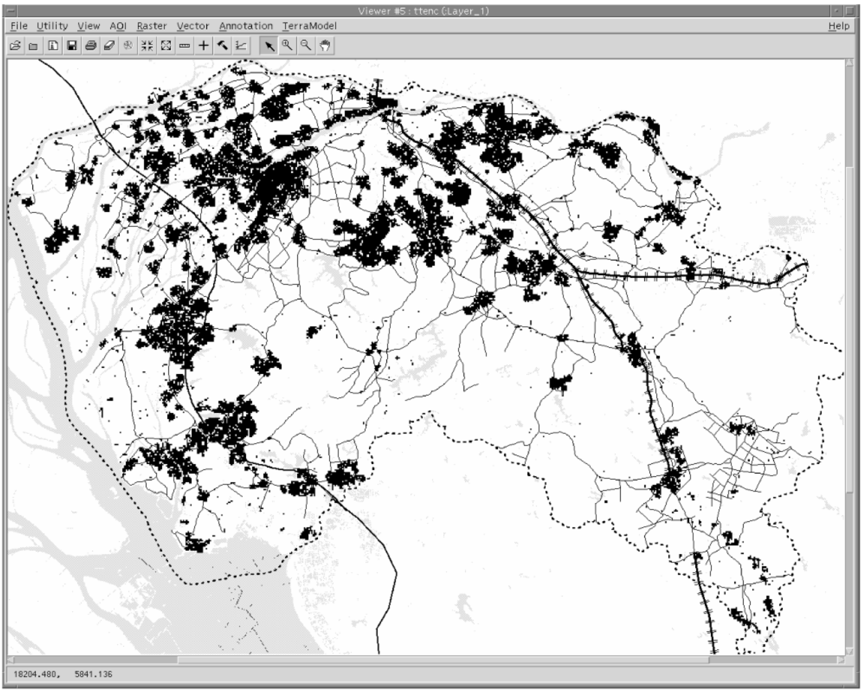

Figure 5. Simulation of compact – monocentric – environmental development.

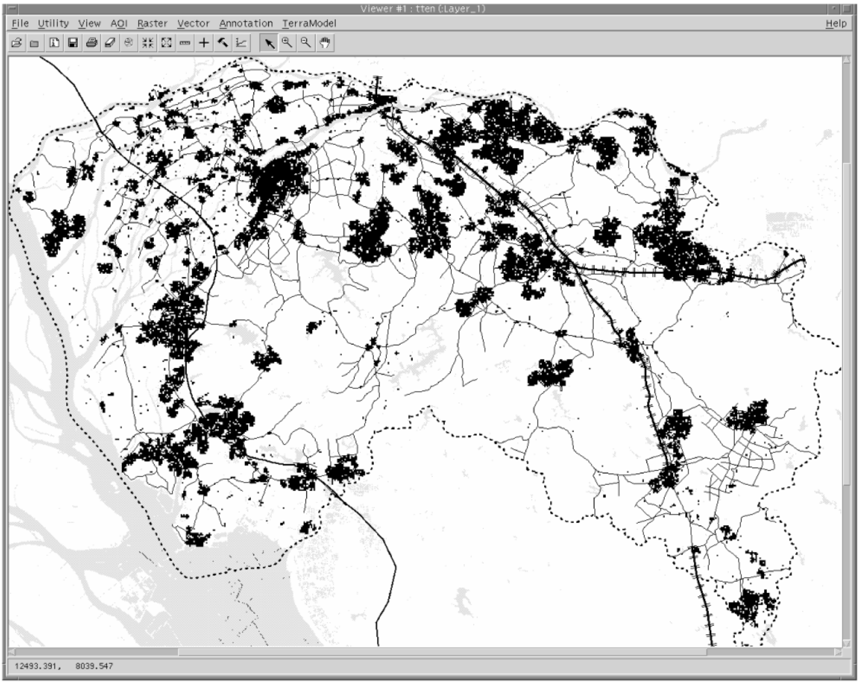

Figure 6. Simulation of compact – polycentric – environmental development.

The constraint factor for protecting forest is also used to confine land development away from forest areas. An inverse exponential function is needed to transform the distance into a suitability score. The distance decay function can ensure that the growth rate will be substantially reduced when development is close to forest areas. Other environmental factors can also easily be embedded in the model, if needed. Environmental constraints are used to generate two development alternatives based on compact – monocentric and compact – polycentric forms (figures 5 and 6). The effects on protecting good agricultural land are apparent as the land development is shifted from the important agricultural areas (the northwest part) to other, less important areas. The environmental objective may be related to different forms of development in terms of either monocentric or polycentric emphasis.

4.4 Dispersed development

The satellite image (figure 2) has revealed that actual development is rather dispersed. The 'grey cell' model can be used to simulate dispersed patterns by using various values of the dispersion factor (α). In this experiment, the values of α are set at 1, 5, and 10, respectively, to explore various sprawl patterns (figures 7, 8, and 9). The constraints of $\delta_{xy}(\mathsf{FORM})$ and $\delta_{xy}(E)$ are not used for this option. The experiment indicates that a higher value of α can result in a more dispersed pattern. It is quite interesting to see that a very high value ($\alpha = 10$) is needed to produce a pattern that is close to the actual development (see figure 8 and figure 2). This means that actual development may be equal to the highly dispersed pattern. These settings can be used to predict the impacts of future urban growth provided that the present trend of urban development continues and that there is no planning intervention. The highly dispersed pattern is the result of a series of economic factors, such as land speculation, rural industrialization, rise of localism, investments from Hong Kong, and transport improvement (Yeh and Li, 1999).

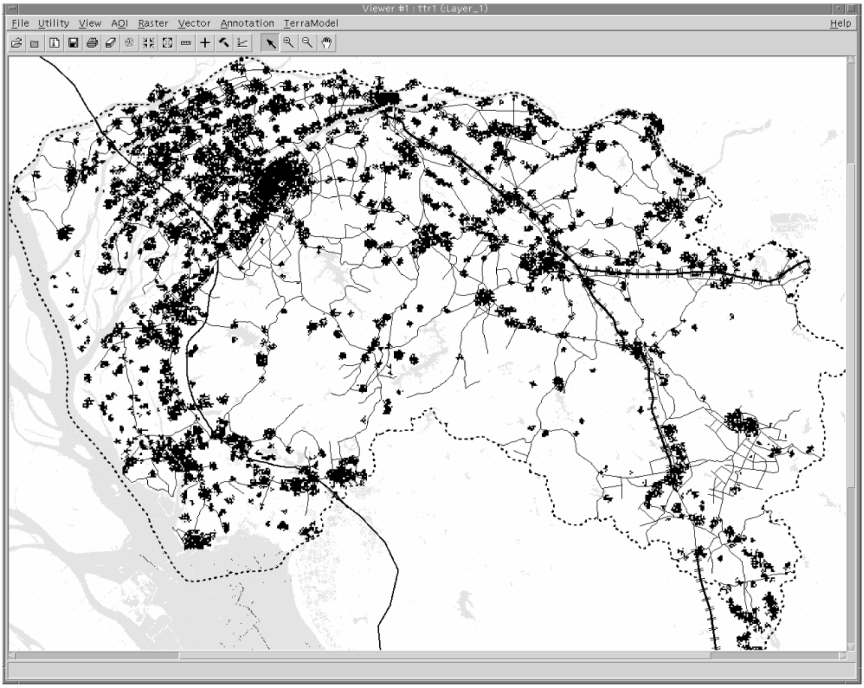

Figure 7. Simulation of dispersed development ($\alpha = 1$).

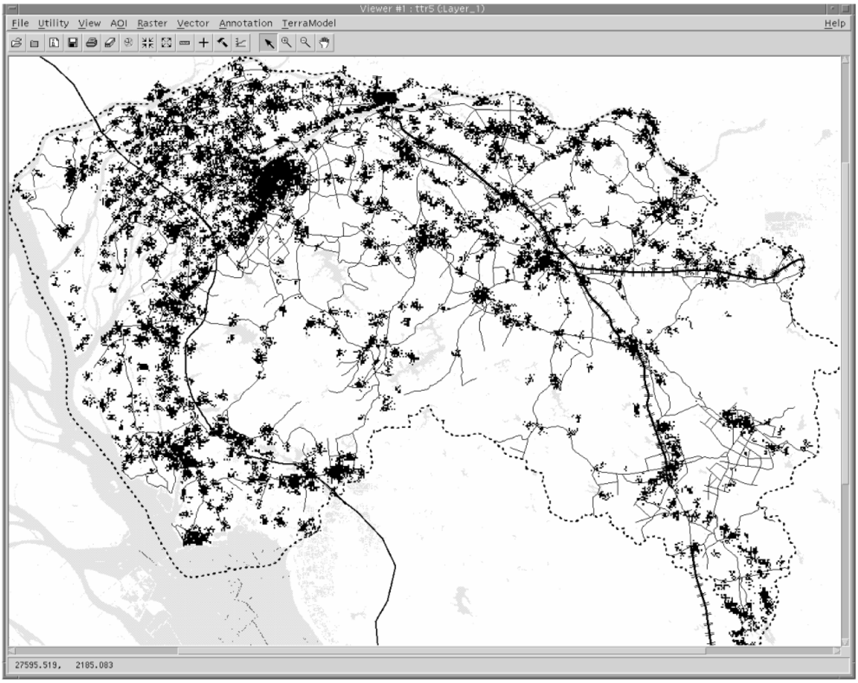

Figure 8. Simulation of highly dispersed development ($\alpha = 5$).

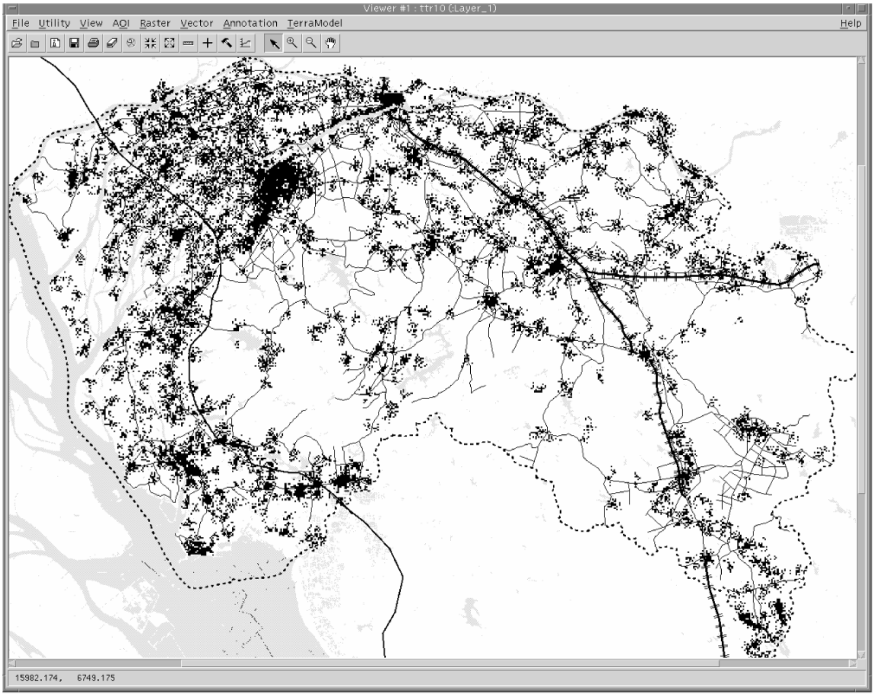

Figure 9. Simulation of very highly dispersed development ($\alpha = 10$).

5 Evaluation of the model and urban forms

It would be desirable to have a quantitative measure of the degree of similarity between the simulations and the actual urban growth (White et al, 1997). The similarity can be measured from a coincidence matrix which is generated by cell – cell comparison of two maps. However, urban systems are rather complicated and their exact evolution systems is unpredictable. It is argued that CA simulations should not be assessed just on the goodness of fit but also on feasibility and plausibility, that is, on possible urban forms (Batty and Xie, 1997; Wu, 1998).

5.1 Validity of the CA simulation

Fractal analysis has been manifested as a useful tool for testing the validity of CA simulation (Batty, 1997; White and Engelen, 1993; Wu, 1998). It is expected that CA simulations should be able to generate spatial patterns which are similar to those of actual cities. Urban forms are usually irregular and complicated. Recent research shows that complexity may be an inherent and important feature of form and is not noise (White and Engelen, 1993). Although land-use patterns can be revealed by some standard measures, such as density gradient, the assumed form of these measures is a rather simple one. It is considered that fractal dimension is one of the few concepts that are directly relevant to the problem or urban complexity (White and Engelen, 1993). Fractal dimension can be used to reflect how much space is filled. A larger value of the dimension means that the city is more compact. If a city were completely compact, it would have the maximum value of 2 for the dimension. Studies have shown that the fractal dimensions of real cities usually lie between 1.4 and 1.8 (Batty and Kim, 1992; White and Engelen, 1993; Wu, 1998). It means that real cities tend to be dispersed rather than completely compact. However, a relatively compact urban form is still possible as there are substantial variations of the fractal dimension.

A useful way of estimating the fractal dimension of a city can be based on the density – radius relationship. The density is calculated by the ratio of the urban area to the total area. The density at any radius, R, can be written as (Batty and Longley, 1994):

$$\rho(R) = \frac{A(R)}{\hat{A}(R)} \propto \frac{\pi R^D}{\pi R^2} = \xi R^{D-2} , \qquad (8)$$

where R is the radius from the urban center, $\rho(R)$ is the density related to the radius R, $A(R)$ is the size of the city, $\hat{A}(R)$ is the total area containing the city, ξ is a coefficient, and D is the fractal dimension.

Based on equation (8), a regression model can be used to estimate the fractal dimension. The regression model is written as:

$$\ln \rho(R) = a + (D - 2) \ln R , \qquad (9)$$

where a is the intercept.

Regression analysis can yield the slope and hence the fractal dimension. The fractal dimension can provide useful information about the spread patterns of urban areas. It reflects the pattern that cities consist of a more-or-less dense scattering of urban activities in the space which contains them. A smaller value of the fractal dimension means that urban development is more dispersed. However, the comparison is valid only for the same scale of development; for example, the same size of urban areas.

The fractal dimension of the actual development in the study area is 1.204 in 1988 and 1.544 in 1993. This does not mean that the latter is more compact as the size of the urban areas is not the same. The fractal dimension of a city may increase as it develops. It is because urbanization in the inner zone will become essentially complete, although some vacant cells still remain. The increase in the fractal dimension is quite rapid in 1988 – 93. The increase rate is as high as 5.6% per annum. This is comparatively much higher than that of other cities. White and Engelen (1993) found that the increase in fractal dimension in Berlin is only 1.0% per annum over its fast growing period of 1910 – 45. This means that the city of the study area is expanding very rapidly.

Figure 10 (over) shows a plot of the density against the radius for the comparison between various simulations and the actual pattern. The simulations will have curves similar to the actual (1993) curve if higher values of disturbance are used. When the density and radius are transformed by a logarithmic function, the curves become linear (figure 11, over). The fractal dimension can then be estimated by a linear regression. Significant relationships are found from the regression analyses. It shows that CA simulation based on the 'grey-cell' model can produce fractal structures similar to that of the real development. The fractal dimensions are about 1.5 for the actual development and the simulated patterns.

5.2 Evaluation of different simulated urban forms

Two indicators can be used to evaluate the performance of different simulated urban forms and developments in relation to sustainable development. Environmental suitability should be an important concept for measuring sustainable cities. The constraints in the CA model can allow development sites to be shifted to suitable locations to reduce environmental costs. The conversion of natural or agricultural land into urban use will bring about environmental costs because there is a loss of food production, natural habitat, and ecological systems. Environmental suitability can be obtained to estimate these types of costs by using GIS. The potential environmental costs can be estimated by the overlay of environmental suitability and land conversion.

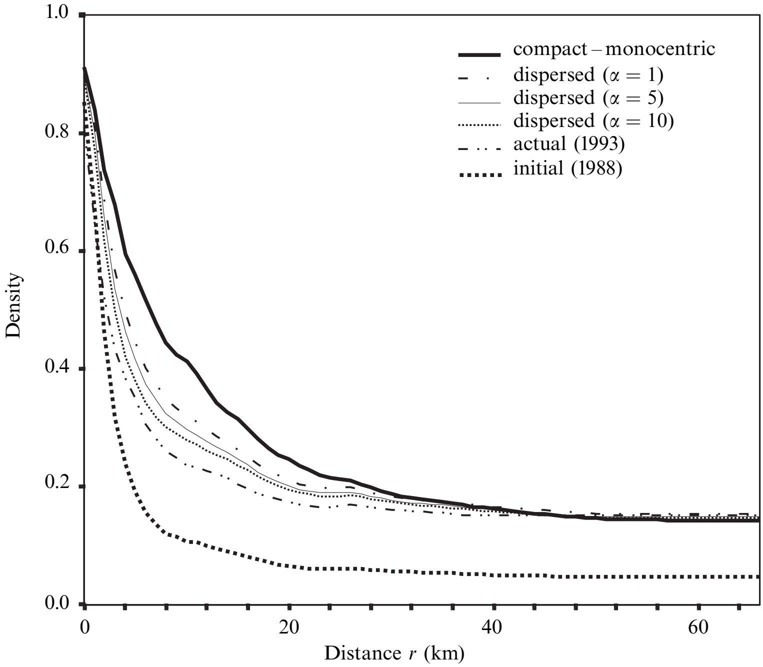

Figure 10. The relationship between density and distance for various development patterns.

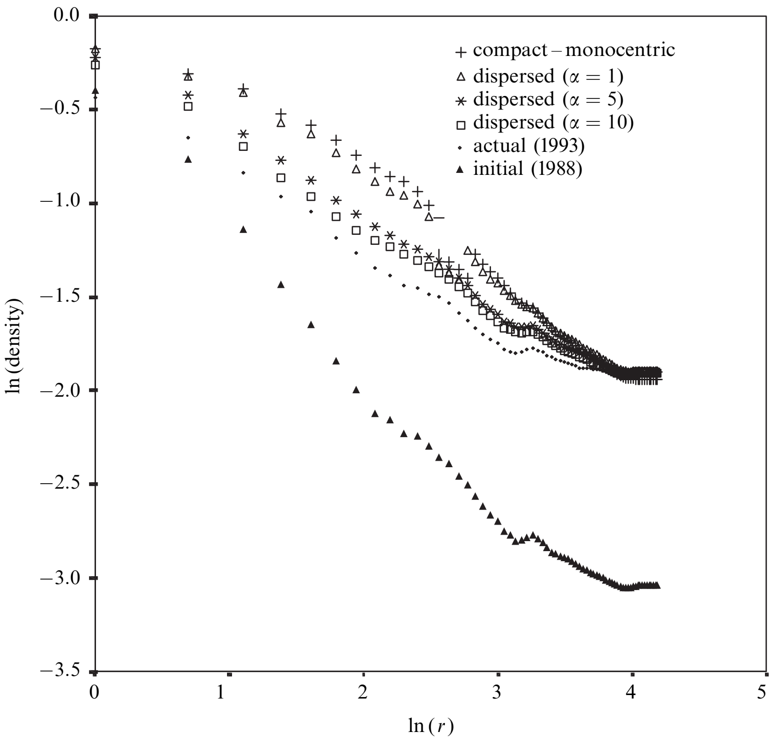

Figure 11. Linear relationships between the logarithm of density and the logarithm of distance.

The index can be written by:

$$C_{\mathrm{E}} = \frac{\sum_{i} \sum_{\{x,y\} \in \Omega} E_{ixy}}{\sum_{i} A}, \tag{10}$$

where C_{E} is the average amount of the environmental costs; E_i is the environmental suitability for the ith criterion; Ω is the set of development sites; and A is the size of Ω. The index has been normalized by the total area, falling in the range of 0 – 1. A higher value indicates that urban development is associated with higher environmental costs, such as the consumption of large proportions of good agricultural land.

Another useful indicator is the agglomeration index which measures whether the development is compact. The index is based on the area:perimeter ratio and is defined as:

$$I_{\mathrm{A}} = \frac{\left(\sum_{j} S_j\right)^{1/2}}{\sum_{j} P_j}, \tag{11}$$

where I_{A} is the agglomeration index, and S_j and P_j are the area and perimeter of developed parcel, j. The area and perimeter of each developed cluster can easily be calculated by using ARC/INFO GRID. The larger the value of I_{A}, the more compact is the development.

Table 3 shows that the environmental costs and agglomeration indexes can provide useful information about different forms of urban development. It can easily be seen that the actual development in 1993 is similar to the 'very highly dispersed' pattern which is generated by using a very large disturbance value. This reinforces that actual development in the study area is in a chaotic, highly dispersed pattern which has been pointed out in other studies (Yeh and Li, 1997). The table shows that the simulated compact forms (models 1, 2, 3, and 4) have much higher compact values. It can also

Table 3. Evaluation of environmental costs and compactness.

	Urban forms and developments	Average environmental costs	Agglomeration index
CA models	(1) Compact – monocentric development	0.85	0.00735
	(2) Compact – polycentric development	0.83	0.00717
	(3) Compact – monocentric – environmental development	0.80	0.00717
	(4) Compact – polycentric – environmental development	0.74	0.00715
	(5) Dispersed development	0.82	0.00530
	(6) Highly dispersed development	0.81	0.00390
	(7) Very highly dispersed development	0.81	0.00313
Actual development	(8) Actual (1993)	0.79	0.00308

be seen from models 1 and 2 that merely achieving compact development will not minimize environmental costs unless environmental considerations are also taken into consideration. Models 3 and 4 show the effects of combining both the compactness and the environmental criteria. Environmental costs have been reduced but compactness has also been slightly reduced. It is only under very special circumstances—in other words, special resources and environment settings of a city—that highly compact development can also produce low environmental costs. In most cases, like the study area, the minimization of environmental costs may reduce compactness, and vice versa.

Table 3 shows that the performances in compactness and environmental costs are very similar whether urban forms are monocentric (models 1 and 3) or polycentric (models 2 and 4). This is because they measure only local characteristics and do not take transport costs into account. Infrastructure costs, especially transport, can affect the energy consumption of the urban system and thus should be considered in the planning of sustainable development (Banister et al, 1997). They should be considered in evaluating different types of urban forms and developments.

There are two types of infrastructure costs. The first type of cost is for the connection of various types of infrastructure between new development sites and their closest networks. The second type of cost is the improvements of networks to accommodate the flows of population and goods among different zones. Different urban forms will have different interactions among the zones and different interzonal infrastructure requirements.

The first type of cost is estimated by GIS overlay analysis. It is assumed that the basic infrastructure has already been built along the main transport networks. The infrastructure development costs of each developed land parcel are the costs for connecting the land parcel to the main provider along the main transport network. The development costs (C_{D}) of each land parcel can be calculated by using the following equation:

$$C_{\mathrm{D}} = \sum_{k} \sum_{\{j\} \in \Omega} C_k D_{j(N)} A_j \ , \tag{12}$$

where C_k is the unit price for each type of infrastructure k, $D_{j(N)}$ is the distance between the developed land parcel j and its closest network N, A_j is the area of the developed land parcel, and Ω is the set of all developed land parcels.

The second type of cost is related to the improvement of network facilities connecting the zones. The distribution of land development within the city around different subcenters will affect the network development and operation costs. A more dispersed development, such as polycentric development, will lead to more interactions and thus higher costs in providing the network connecting the subcenters. A gravity model is used to estimate the interactions among the zones of different types of simulated urban forms and developments for the calculation of the network costs. The interactions are calculated by using the subcenters (towns) as the basic zonal unit. The total amount of developed land in a town is used to measure the attractiveness of each town in the model. An unconstrained gravity model is used to estimate the interactions (T_{ij}) between town i and town j:

$$T_{ij} = \frac{kO_i D_j}{d_{ij}^2} \ , \tag{13}$$

where O_i and D_j represent the total developed land area in the origin town, i, and the destination town, j, respectively; d_{ij} is the network distance between the towns i and j; and k is the universal constant to make the total amount of interactions to the total developed land area in the city.

The sum of the interzonal interactions multiplied by the interzonal distances ($\sum_i \sum_j T_{ij} d_{ij}$, when $i \neq j$) is used to estimate the amount of relative network costs for each model. The monetary network costs are then obtained by using the real costs of the network improvements in 1993 as the reference. The ratio between the relative network costs of a model and those of the actual development was multiplied by the real network improvement costs of 1993 to obtain the monetary network costs of different models. The real costs for the improvement of the network infrastructure in 1993 were obtained from the Statistical Yearbook of Dongguan (Statistical Bureau of Dongguan, 1994).

Table 4 shows the results of the estimated costs for the connection of development sites to the nearest network and the improvement costs of the networks. It compares the development costs for various types of simulated urban forms and developments with the actual development in 1993. The costs of the actual development in 1993 for connecting development sites to the nearest network are estimated from equation (12), but the costs for improving networks among zones are obtained from the statistical yearbook. It can be found from table 4 that compact – monocentric developments (models 1 and 3) are relatively cheaper than compact – polycentric developments (models 2 and 4). This is mainly because the former have lower interzonal interactions and thus lower network costs. The most compact development (model 1—compact – monocentric development) can save about 24% of the total infrastructure costs as compared with the actual dispersed development in 1993. This has not taken into account the annual savings in transport maintenance and energy consumption.

Table 4. Comparison of infrastructure costs between actual development and the CA simulations (in million US$).

Urban forms and developments	Costs for connection to the nearest network					Cost for network improvements	Total
	electricity	water	gas	telecom-munication	roads		
(1) Compact – monocentric development	336.6	56.8	227.3	227.3	1128.0	1204.8	2332.8
(2) Compact – polycentric development	334.6	56.5	225.9	225.9	1121.0	1287.6	2408.6
(3) Compact – monocentric – environmental development	355.6	60.0	240.2	240.2	1191.5	1183.7	2375.2
(4) Compact – polycentric – environmental development	359.7	60.7	242.9	242.9	1205.3	1330.7	2536.0
(5) Dispersed development	415.9	70.2	280.9	280.9	1393.5	1379.5	2772.9
(6) Highly dispersed development	420.9	71.1	284.2	284.2	1410.2	1393.1	2803.3
(7) Very highly dispersed development	445.9	75.3	301.1	301.1	1494.0	1466.1	2960.1
(8) Actual (1993)	528.3	89.2	356.8	356.8	1770.2	1313.3	3083.5

6 Conclusion

A 'grey-cell' constrained CA model has been developed to simulate different urban forms by using various constraints in a raster GIS environment. The constrained model is based on 'grey cells' which can reflect local, regional, and global constraints in the simulation. These constraints address the factors of urban form and environmental suitability in generating alternative patterns.

In addition to simulating a future pattern of urban growth when the prevailing urban process continues, the model can be used to generate different types of urban forms and developments according to different planning objectives (such as protecting good agricultural land and promoting compact development) and monocentric versus polycentric urban form. The simulation of urban forms is possible by the use of distance-decay functions as the constraints. A stochastic disturbance term can be incorporated in the 'grey cells' to simulate the dispersed patterns. A larger size of the disturbance can result in more dispersed patterns. It is interesting to note that a very large size of disturbance is needed to generate the pattern that is similar to the actual development in the study area. This indicates that actual land development in the study area is highly dispersed and chaotic, as identified by other studies (Yeh and Li, 1997; 1999).

The assessment of environmental and development costs can be carried out by a GIS overlay analysis and gravity model. The comparison between various types of urban forms and developments has indicated that compact development is most preferable for the planning of sustainable development because it can significantly reduce environmental and development costs. It is found that the existing development in the study area is very dispersed, which is associated with high development and environmental costs.

In this paper we have shown that 'grey cells' can incorporate various kinds of constraints from remote sensing and GIS that supply a series of spatial information, such as environmental factors and locations of subcenters for the simulation. The integration of the CA model with GIS is useful to address the factors of urban forms, environmental suitability, and land consumption for the planning of sustainable cities. Planners and government officials can use the 'grey-cell' constrained CA model as a planning support system to formulate strategic urban development plans to meet the objectives of sustainable development.

Acknowledgements. This study was supported by funding from the Croucher Foundation, Hong Kong.

References

Alonso W, 1964 *Location and Land Use* (Harvard University Press, Cambridge, MA)
Archer W R, Smith M T, 1993, "Why do suburban offices cluster?" *Geographical Analysis* **25** 52 – 64
Banister D, Watson S, Wood C, 1997, "Sustainable cities: transport, energy, and urban form" *Environment and Planning B: Planning and Design* **24** 125 – 143
Batty M, 1997, "Cellular automata and urban form: a primer" *Journal of the American Planning Association* **63** 266 – 274
Batty M, Kim K S, 1992, "Form follows function: reformulating urban population density functions" *Urban Studies* **29** 1043 – 1070
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, Xie Y, 1997, "Possible urban automata" *Environment and Planning B: Planning and Design* **24** 175 – 192
Berry B J K, Kim H M, 1993, "Challenges to the monocentric model" *Geographical Analysis* **25** 1 – 4

Breheny M, 1996, "Centrists, decentrists and compromisers: views on the future of urban form", in *The Compact City: A Sustainable Urban Form?* Eds M Jenks, E Burton, K Williams (Spon, London) pp 13 – 35

Bryant C R, Russwurm L H, McLellan A G, 1982 *The City's Countryside: Land and Its Management in the Rural – Urban Fringe* (Longman, New York)

Burchell R W, Shad N A, Listokin D, Phillips H, Downs A, Seskin S, Davis J S, Moore T, Helton D, Gall M, 1998 *The Costs of Sprawl—Revisited* (National Academy Press, Washington, DC)

Clarke K C, Gaydos L, Hoppen S, 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

Daniels T L, 1997, "Where does cluster zoning fit in farmland protection?" *Journal of the American Planning Association* **63** 129 – 137

Deadman P D, Brown R D, Gimblett H R, 1993, "Modelling rural residential settlement patterns with cellular automata" *Journal of Environmental Management* **37** 147 – 160

Ewing R, 1997, "Is Los Angeles-style sprawl desirable?" *Journal of the American Planning Association* **63** 107 – 126

Fishman R, 1990, "America's new city" *The Wilson Quarterly* **14** 24 – 48

Garreau J, 1991 *Edge City: Life on the Frontier* (Doubleday, New York)

Gordon P, Richardson H W, 1997, "Are compact cities a desirable planning goal?" *Journal of the American Planning Association* **63** 95 – 106

Howard J F, 1972, "The impact of urbanization on the prime agricultural lands of Southern Ontario", MA thesis, Department of Geography, University of Waterloo, Waterloo, Ontario

Landis J D, 1995, "Imagining land use futures: applying the California urban futures model" *Journal of American Planning Association* **6** 438 – 457

Li X, Yeh A G O, 1998, "Principal component analysis of stacked multi-temporal images for monitoring of rapid urban expansion in the Pearl River Delta" *International Journal of Remote Sensing* **19** 1501 – 1518

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

Mills D E, 1981, "Growth, speculation and sprawl in a monocentric city" *Journal of Urban Economics* **10** 101 – 226

Mills E S, 1967, "An aggregative model of resource allocation in a metropolitan area" *American Economic Review* **57** 197 – 210

Muth R F, 1969 *Cities and Housing* (University of Chicago Press, Chicago, IL)

Ottensmann J R, 1977, "Urban sprawl, land values and the density of development" *Land Economics* **53** 389 – 400

Owens S, 1992, "Energy, environmental suitability, and land-use planning", in *Sustainable Development and Urban Form* Ed. M J Breheny (Pion, London) pp 79 – 105

Statistical Bureau of Dongguan, 1994 *Statistical Yearbook of Dongguan (1993)* Statistical Bureau of Dongguan, Dongguan, Guangdong Province, China (in Chinese)

Voogd H, 1983 *Multicriteria Evaluation for Urban and Regional Planning* (Pion, London)

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 – 1199

White R, Engelen G, 1997, "Cellular automata as the basis of integrated dynamic regional modelling" *Environment and Planning B: Planning and Design* **24** 235 – 246

White R, Engelen G, Uijee 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, 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

Yeh A G O, Li X, 1997, "An integrated remote sensing and GIS approach in the monitoring and evaluation of rapid urban growth for sustainable development in the Pearl River Delta, China" *International Planning Studies* **2** 193 – 210

Yeh A G O, Li X, 1999, "Economic development and agricultural land loss in the Pearl River Delta, China" *Habitat International* **23** 373 – 390