
ELSEVIER
December 2007 Volume 85 Number 4
Journal of
Environmental
Management

Journal of Environmental Management 85 (2007) 1063–1075

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Defining agents' behaviors to simulate complex residential development using multicriteria evaluation

Xia Li*, Xiaoping Liu

School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, PR China

Received 8 March 2006; received in revised form 5 November 2006; accepted 10 November 2006
Available online 9 January 2007

Abstract

Cellular automata (CA) have been increasingly used to simulate complex geographical phenomena. These models may have limitations on reflecting individuals' behaviors which should be considered in urban simulation. Agent-based modeling can solve some of the problems in addressing individuals' influences in urban systems. However, there is a general lack of methodology on how to define agents' properties. This paper uses multicriteria evaluation techniques to determine some of the parameters for the agent-based model. Empirical data from GIS are used to define agent's properties. Sensitivity analysis is also carried out to assess the influences of parameters on simulation outcomes. This model has been applied to the simulation of the residential development in a fast growing city, Guangzhou, in south China.

Keywords: Multi-agent systems; Cellular automata; Urban systems; Residential development

1. Introduction

Fast residential development has been witnessed in many Chinese cities because of growing population. The residential development encroaching on valuable agricultural land has caused intensified land-use conflicts. Urban sprawl and fragmented use of land resources are a major problem for these fast growing cities (Yeh and Li, 2001). There are growing concerns on the side effects of growth such as sprawl, congestion, housing affordability, and loss of open space (Waddell, 2002). Urban models can be used to predict how a region will change according to the existing trajectory of development.

Cellular automata (CA) have attracted increasing attention as a powerful modeling tool in simulating geographical phenomena. Studies have demonstrated that very complex behaviors and global patterns can be generated by applying some simple local rules in CA models (Couclelis, 1997; Wu and Webster, 1998). Simulation of complex urban systems is one of the successful examples of using CA. Modeling urban systems can help to understand the mechanisms of urban evolution and examine existing urban theories. Urban systems involve complex processes which are difficult to predict by traditional 'top-down' models. These models are not suitable to reflect complex urban behaviors because of using strict mathematical equations. However, studies indicate that the 'bottom-up' approach of using CA is well adapted to the simulation of the evolution of urban systems (Batty and Xie, 1994; White and Engelen, 1993; Li and Yeh, 2000). CA have been applied to the simulation of artificial cities (Batty and Xie, 1994) as aids in thought experiments for verifying urban theories. They can be also used to simulate realistic cities by predicting future land-use changes for urban planning (White et al., 1997; Yeh and Li, 2001). Ward et al. (2000) use a constrained rule-based CA model to simulate the residential development of a single satellite city in the Gold Coast of eastern Australia.

CA models may have apparent limitations on reflecting the decisions and behaviors of individuals, such as governments, residents and investors, in shaping urban dynamics. The influences of human factors are difficult to implement in

*Corresponding author. Tel.: +86 20 33047500.
E-mail addresses: lixia@mail.sysu.edu.cn, lixia@graduate.hku.hk (X. Li).

0301-4797/$ - see front matter
doi:10.1016/j.jenvman.2006.11.006

traditional CA models. For example, CA are considered to be insufficient in dealing with mobile objects such as pedestrians, migrating households, or relocating firms (Benenson et al., 2002; Benenson and Torrens, 2004). Agents' spatial behaviors should be considered so that complex spatial interactions can be addressed in urban simulation. Actually, multi-agent systems (MAS) have been widely employed to represent individual decision-makers in social science research (Epstein, 1999; Kohler, 2000).

In the past, multi-agent models were used mostly in purely social contexts (Gilbert and Conte, 1995). They were used to validate or illustrate social theories (including biological, economic, and political theories), or predict the behavior of interacting social entities, such as actors in financial markets and consumer behaviors (Basu and Pryor, 1997). However, these types of agent-models have few spatial details. It is considered that these MAS methodologies and existing tools are over-general and underestimate, if not ignore, the importance of space and spatial behavior (Benenson and Torrens, 2004).

It is still an initial stage in the application of MAS in urban modeling. A major problem with MAS is how to define agents' properties (attributes) by using empirical data. Simulation results should be very sensitive to agents' decision behaviors. There is a general lack of methodology for defining agents' decision behaviors in a more consistent way. The trial and error approach is useful for this purpose. Some initial studies have been proposed to validate agent-based spatial models (Brown et al., 2005). More elaborated methods are still required for the fine turning of models' parameters.

This study proposes a multi-agent system integrated with CA and GIS to simulate residential development in urban systems. Agents make decisions under the constraints of the environment, which is heterogeneous in a two-dimensional landscape and subject to change in time. In this model, agents' decision behaviors are determined by utility functions which involve various spatial variables and weights. GIS provide the basic data for retrieving these spatial variables and defining agents' properties. An important part of this model is that the weights for various groups of agents are decided by multicriteria evaluation techniques. Saaty's pairwise comparison procedure is used to obtain these weights in a more consistent way.

2. Integration of multi-agent systems with CA and GIS

This integrated model consists of two major types of layers—immobile environment layers and mobile agent layers. The environment layers include the spatial data of land-use types, land price, surrounding environment, accessibility, general public facilities, and education. The environment is not uniform in spatio-temporal dimensions when using GIS. A GIS database can provide the basic spatial information for agents' location choices. The agent layers are used to accommodate the mobile entities that play a key role in shaping the evolution of urban structures. Agents do not act independently, but make decisions through negotiations and cooperation with each other. These agents not only adapt to the environment, but also change it.

2.1. The environmental layers

Land use: Land use is one of the most important factors that should be considered in urban simulation. This factor provides the basic environment for agents to make decisions, but it is also subject to changes with regard to agents' activities. Agents will have different decision behaviors related to land-use types. For example, development probability is very low for some land-use types, such as fishponds and forests. However, development probability may be much higher at the sites surrounded by a large number of built-up areas.

Land price: Land price is an important element in determining housing prices which are a major concern for a potential home buyer. Residents with high incomes can afford to buy good-quality homes in locations with high land prices. However, residents with low incomes can only afford to live in places with cheaper prices (land prices).

Surrounding environment: The attraction of a site for residential development is also related to its surrounding living environment. In this study, the amenity of the surrounding environment is mainly based on green space and water. These two factors are commonly considered by the Chinese in the choice of residence location. People would like to live in locations with more green land and water in the neighborhood. The qualities of green space and water are similar for the whole region. Therefore, the amenity is measured by using two indicators, the percentages of green land and water in the neighborhood. If the sites of green space or water are not of equal quality, some methods can be adopted to take this influence into account. For example, polluted waterways can be masked out in the calculation of the amenity.

A moving window is used to calculate the percentages of green land and water in the classified TM image which has a resolution of 30 m. A window size of 9×9 is adopted because this size seems to be appropriate to capture the spatial patterns of these two variables in the study area according to experiments. Finally, the utility (attraction) of a site related to this amenity is obtained by using the following equation:

$$B_{\text{env}} = \frac{1}{2} G_{\text{percent}} + \frac{1}{2} W_{\text{percent}}, \tag{1}$$

where B_{env} is the utility of the surrounding environment, and G_{percent} and W_{percent} are the percentages of green land and water, respectively. Equal weights are applied to the calculation since there is no prior knowledge.

Accessibility: Accessibility represents a locational characteristic that permits a place to be reached by the efforts of those at other places using various transport tools. It is related to its geographical location (e.g. distance to roads and town centers) and the conditions of road networks. Usually, a site can be accessed by expressways, highways, or railways. The site will be more likely to develop if it is easily accessed. The accessibility can be conveniently calculated by using GIS functions. The utility (benefits) related to accessibility can be represented as follows:

$$B_{\text{access}} = \frac{1}{3}\,\mathrm{e}^{-b_1 D_{\text{road}}} + \frac{1}{3}\,\mathrm{e}^{-b_2 D_{\text{express}}} + \frac{1}{3}\,\mathrm{e}^{-b_3 D_{\text{center}}}, \tag{2}$$

where B_{access} is the utility related to accessibility; D_{road}, D_{express}, and D_{center} are the variables in terms of Euclidian distances to roads, expressways and urban centers, respectively; b_1, b_2, and b_3 are the decay coefficients for these variables; The same weights are also applied to all these variables for simplicity.

General public facilities: The provision of public facilities is another important factor to affect the decision of a potential buyer. The sites will be more attractive if they are closer to general public facilities, such as hospitals, gardens, commercial centers and entertainment centers. Distance decay functions can be defined to reflect the utility of a site in terms of facility provision. The utility is calculated as follows:

$$B_{\text{facil}} = \frac{1}{4}\,\mathrm{e}^{-b_1 D_{\text{hospital}}} + \frac{1}{4}\,\mathrm{e}^{-b_1 D_{\text{garden}}} + \frac{1}{4}\,\mathrm{e}^{-b_1 D_{\text{commercial}}} + \frac{1}{4}\,\mathrm{e}^{-b_1 D_{\text{entertainment}}}, \tag{3}$$

where B_{facil} is the utility related to the provision of public facilities, such as hospitals, gardens, commercial centers and entertainment; D_{hospital}, D_{garden}, $D_{\text{commercial}}$ and $D_{\text{entertainment}}$ are the Euclidian distances to these facilities respectively. The same decay coefficient of b_1 in Eq. (2) is used since these facilities are mainly accessed by roads. All these variables are treated with equal weights in the calculation.

Education benefits: Education is a special type of public facility, and is the most important factor affecting home buying for the Chinese. Distance functions can also be used to represent the convenience of accessing education facilities, such as schools and libraries. More benefits can be achieved if there is a shorter distance to these facilities.

$$B_{\text{edu}} = \frac{1}{2}\,\mathrm{e}^{-b_1 D_{\text{school}}} + \frac{1}{2}\,\mathrm{e}^{-b_1 D_{\text{library}}}, \tag{4}$$

where B_{edu} is the utility related to the provision of educational facilities in terms of schools and public libraries; D_{school} and D_{library} are the Euclidian distances to these facilities, respectively. The same decay coefficient of b_1 in Eq. (2) is also used since these facilities are mainly connected by roads. Moreover, the same weights are also used for these two variables.

2.2. Defining agents' decision behaviors using multicriteria evaluation techniques

This model has three major types of agents—residents, property developers and governments. An agent can represent not only a single individual, but also a group of individuals. In this model, an agent represents a number of residents according to the total population. Each type of agent has unique features. For example, government agents can have global influences without spatial variations. It is also difficult to define the exact locations for developer agents. Resident agents are movable as they need to select locations for residency.

The definition of agents' decision behavior is essential to agent-based models. In most situations, agents' behaviors can be defined heuristically since there are not unique methods. However, it is more robust to define agents' decision behaviors based on GIS data. Each major type of agents can be divided into subtypes according to the agents' properties. For example, resident agents can be classified into a number of groups based on incomes and household size. Each group of residents has unique preferences of choosing proper sites for residency. This simplification procedure can help the model to be more operational by reducing computation time.

These agents will influence each other in making decisions. Resident agents choose the proper sites for home buying according to some utility functions. If there are too many buyers selecting the same location, housing prices will go up. If the prices go behind their affordable threshold, they will move to other places for residency. Property developers will adjust their investment strategies according to the purchasing behavior of residents. Their simple objective is to make as much profit as possible. However, they must get approval from the government before a site can be developed. Governments have a unique power to decide if a location can be developed by considering the broader public interests.

2.2.1. Resident agents

There are two kinds of residents—new residents moving in from outside and existing residents relocating to new places to live. The behaviors of these mobile residents will affect the investment strategies of property developers. The interactions between these agents are responsible for the formation and evolution of urban structures, such as social and ethnic segregation, self-organization and urban expansion.

A utility function is defined for a resident agent to assess the value of a potential site as a residence. His main objective is to maximize the utility function as much as possible in location decisions. The utility function of location (ij) for resident agent k can be represented as follows:

$$U(k, ij) = w_{\text{price}} B_{\text{price}} + w_{\text{env}} B_{\text{env}} + w_{\text{access}} B_{\text{access}} + w_{\text{facil}} B_{\text{facil}} + w_{\text{edu}} B_{\text{edu}} + \varepsilon_{tij} \tag{5}$$

where $w_{\text{price}} + w_{\text{env}} + w_{\text{access}} + w_{\text{facil}} + w_{\text{edu}} = 1$. B_{price}, B_{env}, B_{access}, B_{facil} and B_{edu} are the factors of land price, surrounding environment, accessibility, provision of general facilities and educational benefits for location (ij).

w_{price}, w_{env}, w_{access} and w_{edu} are the preferences (weights) for resident agent k for each factor. ε_{tij} is a stochastic term.

The utility function mainly affects the location behaviors of resident agents. In this equation, if these variables are treated equally, all the weights can be assigned with equal values. However, resident agents may have different preferences in choosing locations for residency, which can be reflected by the weights in the utility function. A higher value of the weight means that the variable will be treated more importantly. In this study, these weights are decided by using Saaty's pairwise comparison procedure (Eastman, 1999). The comparison is mainly based on experts' knowledge and preferences. Saaty (1990) proposes a consistency ratio to examine the consistency of the matrix. He suggests that the matrix should be re-evaluated if the ratio value is greater than 0.10.

A discrete selection model is used to facilitate the site selection for residency. The probability of selecting a site to reside in can be estimated by using the utility function in Eq. (5). For resident k, the probability of location (ij) to be selected is equal to the utility probability that the utility value at that location is greater than or equal to those at other locations (McFadden, 1978):

$$P^t_{\text{resident}}(k,ij) = P(U(k,ij) \geqslant U(k,i'j')) = \frac{\exp(U(k,ij))}{\sum_k \exp(U(k,ij))}, \tag{6}$$

where $P^t_{\text{resident}}(k,ij)$ is the development probability of location (ij) for resident agent k at time t.

The Monte Carlo method is used to decide the final selection of a location for residency (Wu and Webster, 1998). This can allow a stochastic variable to be added to the modeling process. After a satisfactory location has been identified by a resident agent, there are three situations:

(1) the location has been developed and occupied by another resident agent;
(2) the location has been developed and is available for residency;
(3) The location has not been developed.

The third situation is the essential part of this study. If the development of an undeveloped site is considered desirable by resident agents, developer agents will assess the potential profits from the development. They will like to develop the site if the profit is greater than a threshold value. The project can be carried out when the application has been approved by government agents. This interaction process is the basis for simulating complex residential development in urban systems.

2.2.2. Developer agents

Property developers play an important role in influencing residential development in fast growing cities. They need to consider the preferences of residents in home buying and the policies of governments in managing land resources. The main criterion is to achieve a certain amount of profit above expectations. This criterion can be used to determine the decision behaviors of developer agents. The following equation is used for the assessment of development potential:

$$D^t_{\text{profit}} = H^t_{\text{price}} - L^t_{\text{price}} - D^t_{\text{cost}}, \tag{7}$$

where D^t_{profit} represents the investment profit, H^t_{price} is housing price, L^t_{price} is land price, D^t_{cost} is development cost. They are measured in a currency unit (e.g. RMB).

The development probability related to developer agents can thus be represented as follows:

$$P^t_{\text{developer}}(k,ij) = \frac{D^t_{\text{profit}} - D_{\text{tprofit}}}{D_{\text{mprofit}} - D_{\text{tprofit}}}, \tag{8}$$

where $P^t_{\text{developer}}(k,ij)$ is the development probability related to developer agents, D_{tprofit} is a threshold value, and D_{mprofit} is the maximum value of the investment profit.

A developer agent will apply for investing in the site if the assessment is in favor of the development according to Eq. (8). In this model, all the developer agents have the same attributes (status) because the detailed information is unavailable. Therefore, the model adopts a simplified procedure in which the first applicant will be considered for the evaluation of site selection. The authorities will evaluate the application before the approval can be given according to the criteria described in the following section.

2.2.3. Government agents

Government agents will decide if an application for land development is approved or not according to a number of factors. They make decisions not only by considering environmental factors, but also by communicating with resident agents and developer agents. Government agents will examine the development suitability of a site according to land use, surrounding environment, transportation, general facilities, and educational benefits. The initial development probability of each site can be estimated based on these environmental factors.

Firstly, existing land use is a major factor in determining land use conversion. Different land uses will have different approval probabilities for land development. For example, land development is not allowed in ecologically sensitive areas. The probability for land development in wetland areas or mountainous areas is extremely low. Secondly, the approval probability is also related to development plans. It is more likely that an application will be approved if there are no conflicts with existing planned land use.

The behaviors of government agents are also affected by the behaviors of resident agents and developer agents. For example, an incremental development probability will be added to a site when it has been requested for development by resident agents or developer agents in simulation. This means that government agents should consider the willingness of residents in choosing the place to live. When a site has been requested for residency more often by resident agents in the simulation, it has a higher probability of development. This interaction reflects the communication and negotiation processes in urban simulation.

The probability of development approval from government agents is based on land use types. The initial approval probability is decided by the government, but it is subject to changes due to the influence of residents and property developers. The following equation can be used to represent this type of negotiation between government agents, resident agents and developer agents in affecting the development probability of a cell (*ij*):

$$P^t_{\text{gov}}(ij) = P^{t-1}_{\text{gov}}(ij) + g\Delta P_1 + +h\Delta P_2 \quad (If\ P^t_{\text{gov}}(ij) > 1,\ Then\ P^t_{\text{gov}}(ij) = 1), \tag{9}$$

where $P^0_{\text{gov}}(ij)$ is the initial approval probability, g and h are the total numbers of applications for development at cell (ij) by resident agents and developer agents, respectively, and ΔP_1 and ΔP_2 are the incremental probabilities for each application by resident agents and developer agents, respectively.

2.3. Integrating cellular automata with multi-agent systems

CA is an important component in this integrated simulation system. CA mainly focuses on local interactions of physical factors whereas agent-based systems pay much attention to individual's behaviors. The integration of these two models is essential for the simulation of residential development. In this study, the development probability related to physical factors is estimated by using a standard-CA. It is based on a logistic-CA model by taking s number of proximity variables into account (Wu, 2002):

$$P^t_{\text{ca}}(ij) = \frac{1}{1 + \exp[-(d + \sum_h D_h x_h)]} \text{con}^t(ij)\Omega^t(ij), \tag{10}$$

where $P^t_{\text{ca}}(ij)$ is the development probability determined by the neighborhood function of cellular automata for location (ij), d is a constant from the logistic regression model, x_h is the hth spatial variable, D_h is the weight of the variable, con(ij) is the combined physical constraint, and $\Omega(ij)$ is the percentage of developed cells in the neighborhood.

The parameters d and D_h are obtained by the calibration procedure of logistic regression (Wu, 2002). con(ij) is used to address the combined physical constraint for residential development. A site will be associated with a small value of con(ij) if it is unsuitable for development. For example, the value of zero will be assigned to all the cells of water and gardens.

The final decision is based on the interactions between government agents, resident agents, developer agents and the environment. The probability for residential development at a cell is obtained by the product of these probabilities:

$$P^t_{ij} = AP^t_{\text{resident}}(k, ij)P^t_{\text{developer}}(k, ij)P^t_{\text{gov}}(ij)P^t_{\text{ca}}(ij), \tag{11}$$

where A is an adjusted coefficient.

The procedure of modeling between t and $t+1$ follows this order: Estimating the potential annual sprawl with CA → Application for residence by resident agents → Estimating the demand and benefit for developer agents before investment → Evaluation of the applications by government agents before issuing the permit.

3. Application

3.1. Study area and data

The Haizhu district of Guangzhou city has been selected for testing the proposed model. The model is used to simulate the land-use dynamics from 1995 to 2004 for this fast growing region. The land use types in this study area include green land, farmland, water, road, old-developed residential land and newly developed residential land. The test area had a large proportion of farmland, but most of it has been converted into urban land-use. The simulation and prediction of the fast urban expansion can provide useful information for land-use planning and management.

The spatial information for the simulation includes remote sensing data and GIS data. Landsat TM images dated on 30 December 1995 and 13 June 2004 were used to obtain training data about actual land-use conversion, which can be used to calibrate the logistic-CA model (Wu, 2002). GIS data were also used to represent the independent factors that determine land-use changes. These spatial data include the maps of urban planning, land price, land use, and the distribution of public facilities (e.g. hospitals, schools, and parks). Social and economic data were also obtained from statistical yearbooks and the *Fifth National Census*. These social and economic data are only available for street-blocks, which are the smallest administrative unit in Chinese cities.

A raster GIS is used for handling these spatial data in a convenient way. All the original raster data are re-sampled into the resolution of 100 × 100 m to reduce the computation time for the agent-based model. Fig. 1 shows the spatial patterns of some utility variables (e.g. land price, surrounding environment, accessibility, general facilities and education) which are used in this model. Fig. 2 displays the constraints of initial land use and the plan scheme for simulation. However, detailed housing prices and development costs are unavailable for Eq. (7) in the study area. Some assumptions are adopted to avoid this data problem. Usually, housing prices are closely related to land prices. Therefore, H^t_{price} can be proportional to L^t_{price}. It is also fair to assume that development costs are uniform in space. The costs can then be ignored in the equation.

Some coefficients need to be determined for calculating the utilities in Eqs. (2), (3) and 4. The values of b_1, b_2, and b_3 in Eq. (2) can be estimated according to traffic flows since these coefficients are related to transport conditions. It is assumed that a transport tool will have a smaller decay of influence (a smaller value of the coefficient) if it carries a larger traffic flow. For example, expressways which have

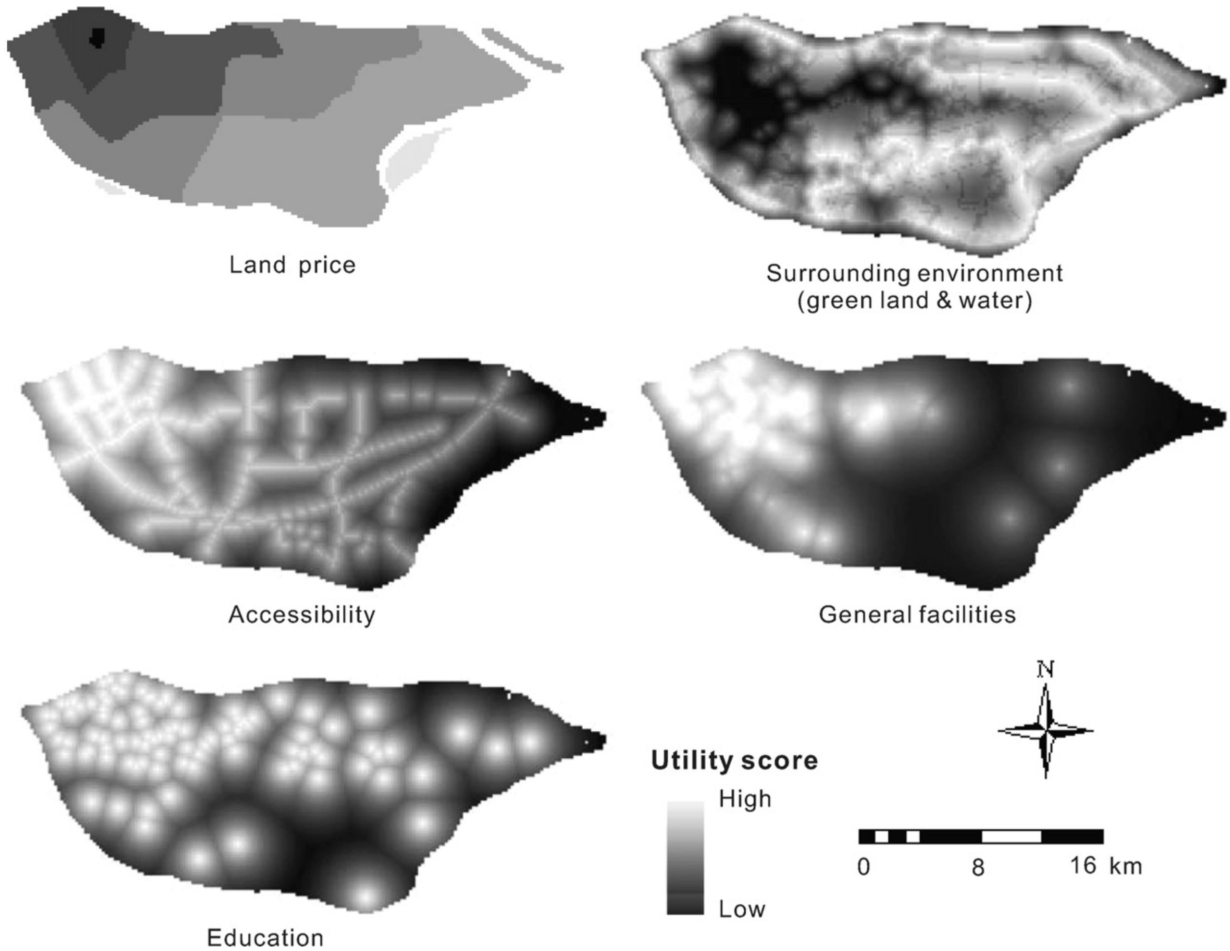

Fig. 1. The utilities of various spatial variables prepared by a raster GIS.

larger traffic flows should be assigned smaller values of the coefficient than roads. The following equation can be used to represent this relationship:

$$b_1/b_2 = f_{\text{express}}/f_{\text{road}}, \tag{12}$$

where f_{road} and f_{express} are the average traffic densities for roads and expressways, respectively.

The same method can be applied to the estimation of the value of b_3. It is assumed that there are z_1 roads and z_2 expressways connected to urban centers. The equation becomes:

$$b_1/b_3 = (z_1 f_{\text{road}} + z_2 f_{\text{express}})/f_{\text{road}}. \tag{13}$$

Table 1 shows the traffic flows of roads and expressways according to statistical data. When b_1 is set to 0.00100, b_2 and b_3 become 0.00023 and 0.000125, respectively, according to above equations. The same value of b_1 is also used in Eqs. (3) and (4).

The initial values of the environmental factors in GIS should be normalized within the range of [0, 1] so that these factors are comparable in the decision making. The normalization is different for the following two types of factors—positive (higher scores are better) and negative (lower scores are better). Positive factors (e.g. surrounding environment, accessibility, general public facilities, and education benefits) and negative ones (e.g. land price) were normalized by using the following functions:

Positive factors

$$x' = \frac{x - \text{Min}}{\text{Max} - \text{Min}}, \tag{14}$$

Negative factors

$$x' = \frac{\text{Max} - x}{\text{Max} - \text{Min}}, \tag{15}$$

where x is the original data, Max and Min are its maximum and minimum values, x' is the normalized value.

The value of D_{mprofit} (the maximum value of the investment profit) in Eq. (8) is estimated according to local knowledge. It is 12,000 RMB/m^2 in the study area.[1] The values of the incremental probabilities (ΔP_1 and ΔP_2) in Eq. (9) can be also determined by experiments. In this study, ΔP_1 is set to 0.005 and ΔP_2 is set to 0.1. Sensitivity analysis can be carried out to assess the impacts of these parameter values on the simulation outcomes. Table 2 lists

[1]US$ is about equivalent to 8.02 RMB in March 2006.

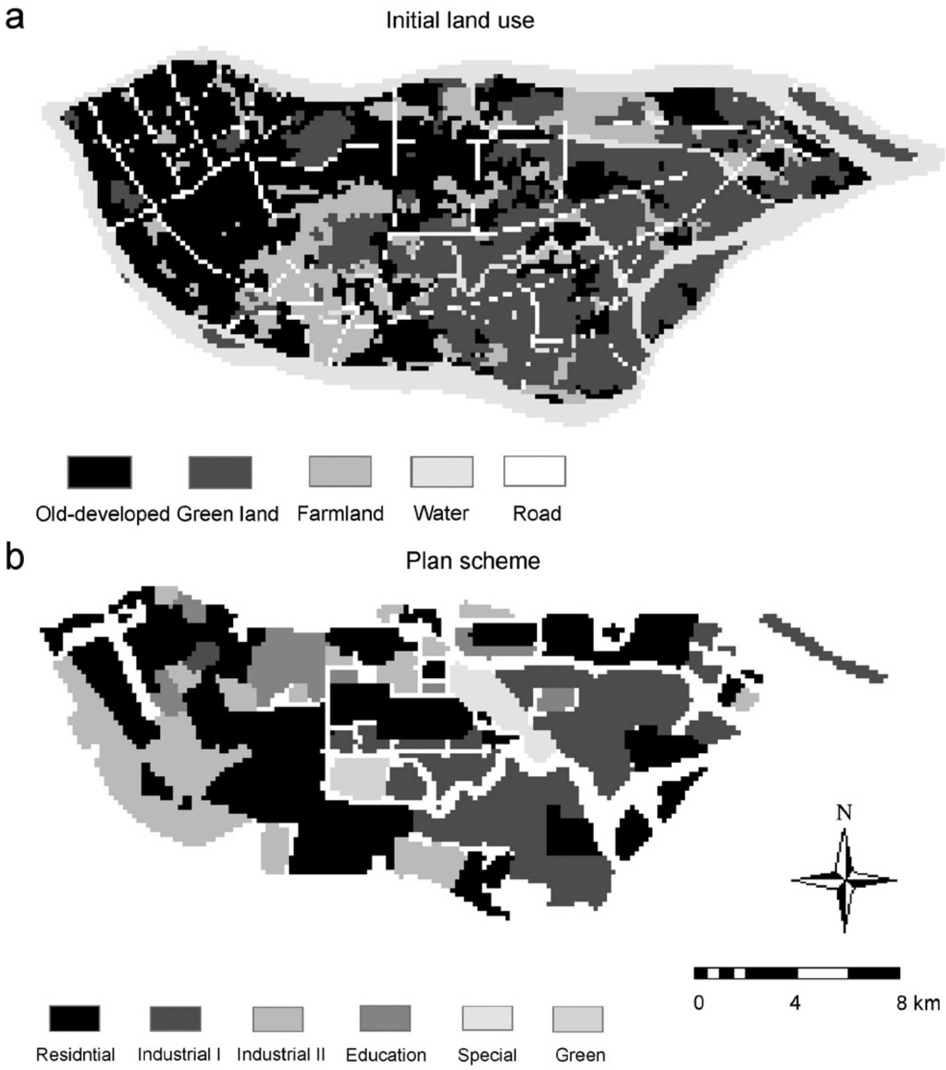

Fig. 2. Constraints of initial land use and plan scheme for simulation.

Table 1
Traffic flows of roads and expressways according to statistical data

Transport types	Length (km)	Traffic (1000 persons)	Average traffic density (1000 persons/km)
Roads	4637.2	147,330	31.77
Expressways	382.8	52,310	136.7

Source: Guangdong Statistical Yearbook (2000).

Table 2
The coefficients of the logistic-CA model

d	D_1 Distance to main centers	D_2 Distance to sub-centers	D_3 Distance to main roads	D_4 Distance to roads	D_5 Distance to expressways
0.625	−0.002	0.005	−0.009	−0.006	0.002

the coefficients of the logistic-CA model in Eq. (10) based on the regression analysis.

3.2. Defining resident agents' properties

Detailed information for each agent is unavailable in most situations. Agents should be classified into a few categories so that their properties can be defined. The attributes for the aggregated agents are obtained by using social and economic data. This study considers two major attributes, income and household size, which are obtained from the statistical yearbook of Guangzhou for 2004, and the *Fifth National Census*, respectively. These attributes can be used to define the decision behaviors for the resident agents.

Resident agents were first classified into two groups, without children and with children, according to the household size. Each group was further classified into three subgroups, the low-income class (income < 9,600 RMB/year), the middle-income class (9600 RMB/year < income < 60,000 RMB/year), and the high-income class (income > 60,000 RMB/year). Six classes of residents were then obtained by using these two indicators. The actual percentages for these six groups were obtained from the statistical yearbook of Guangzhou for 2004, and the *Fifth National Census* (Table 3). These percentages were used to determine the agent numbers for various groups of resident agents in the simulation.

Each group of resident agents has distinct behaviors or preferences in the choice of residence location. In this model, their preferences are reflected by the weights in the utility function as described in Eq. (5). These weights are given according to experts' experiences. It is expected that different groups of resident agents have different concerns about the factors of land price, surrounding environment, accessibility, provision of public facilities, and educational benefits for determining the residence location.

Their concerns are represented by the weights in the utility function. Saaty's pairwise comparison procedure was applied for estimating these weights. The pairwise comparison was completed by consulting the opinions of urban planners and professionals. Table 4 gives the results of the weights derived from this multicriteria evaluation technique.

A set of weights can represent the unique preferences of location choice for a group of resident agents. For example, residents of high income are very sensitive to living quality (surrounding environment), whereas those of low income are very sensitive to housing price (land price). Residents with larger family sizes (with children) are more concerned about education than those without children.

3.3. Simulation results

Satellite TM images were used to provide the empirical data for simulating residential development in the study area. Land-use classification was carried out for the 1995 and 2004 TM images. The classification accuracy will definitely affect the simulation accuracy. The accuracy assessment for land-use classification was carried out with reference to available land-use maps, air photographs and field investigation (Li and Yeh, 2004a). The kappa coefficient is 0.83, and the total accuracy is 0.87 according to the accuracy assessment.

The land use in 1995 was used as the initial stage and the land use in 2004 was used to verify the simulation. The locations for resident agents were randomly selected at the initial stage. They will choose the residence site based on

Table 3
Proportion of each group of resident agents

Groups of resident agents						
Household size	Without children			With children		
Income	Low income	Middle income	High income	Low income	Middle income	High income
Proportion	9%	39%	9%	6%	31%	6%

Table 4
Weights for different groups of resident agents obtained by using the Saaty's method

Types of residents	Weights					Total	CR
	Land price	Surrounding environment	Accessibility	Public facilities	Education		
Low income without children	0.443	0.093	0.206	0.155	0.103	1	0.042
Low income with children	0.401	0.081	0.154	0.081	0.283	1	0.087
Middle income without children	0.175	0.379	0.165	0.194	0.087	1	0.057
Middle income with children	0.220	0.276	0.142	0.140	0.222	1	0.094
High income without children	0.048	0.526	0.194	0.141	0.091	1	0.072
High income with children	0.084	0.434	0.171	0.076	0.235	1	0.064

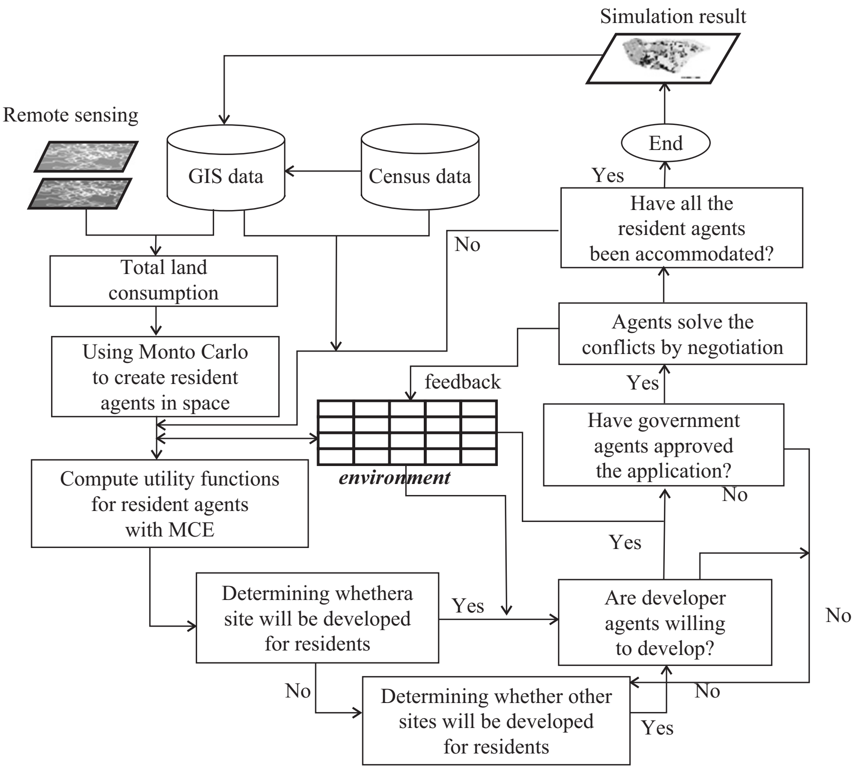

Fig. 3. Procedures of the integrated CA-agent model for simulating complex residential development.

the interaction between three kinds of agents. The procedures for simulating residential development are as follows (Fig. 3):

(1) Determining the total number of resident agents according to the actual amount of residential development in 1995–2004 from the classification of remote sensing data;
(2) Using the Monte Carlo method to create resident agents in space using the actual proportion of various types of residents based on census data (Table 3);
(3) Using Eq. (5) and Table 4 to compute the location utility for each type of resident agents, selecting the locations with the highest utility values and estimating the development probabilities for these places based on the interactions between residents, property developers and governments by using Eq. (11);
(4) Determining whether the location of the highest utility value will be developed by using the Monte Carlo method. If yes, the location will be marked and the next round of site selection will be started by going back to step 1. If no, the next site will be evaluated by choosing the second highest utility value.
(5) This procedure continues until all the residents have been accommodated.

Fig. 4 shows the simulated patterns of residential development of the study area in 2004. The simulated residential development (Fig. 5a) can be compared with the actual residential development (Fig. 5b) to reveal the agreement and disagreement visually. Very plausible results have been observed by comparing these two patterns.

Fig. 6 shows the changes in the total amounts of various land-use types in the study area from 1995–2004. The increase in residential areas is at the cost of losing green land and farmland (vegetable fields). This will gradually result in a decrease of available land stock for future land development. The landscape of the region will also change as a result of residential development. This model allows urban planners to forecast future land-use changes and predict possible land demand for planning purposes.

3.4. Model validation

Model validation is usually required when urban models are applied to the simulation of real cities. It is preferable that a model can produce a high goodness of fit in terms of pattern accuracy and process accuracy (Brown et al., 2005). The assessment of pattern accuracy has been carried out for many simulation studies (Wu, 2002; Li and Yeh, 2004b). It is impossible to reproduce the exact spatial patterns of cities because of unexplained variables. In this study, the proposed model is assessed in two ways: (1) comparing the simulated patterns with the actual ones; (2) comparing the simulation patterns between the integrated CA-agent model and the pure CA model.

First, the simulated patterns are compared with the actual ones obtained from the classification of remote sensing imagery. A simple method to assess the goodness-of-fit is to compare the simulated patterns with the actual

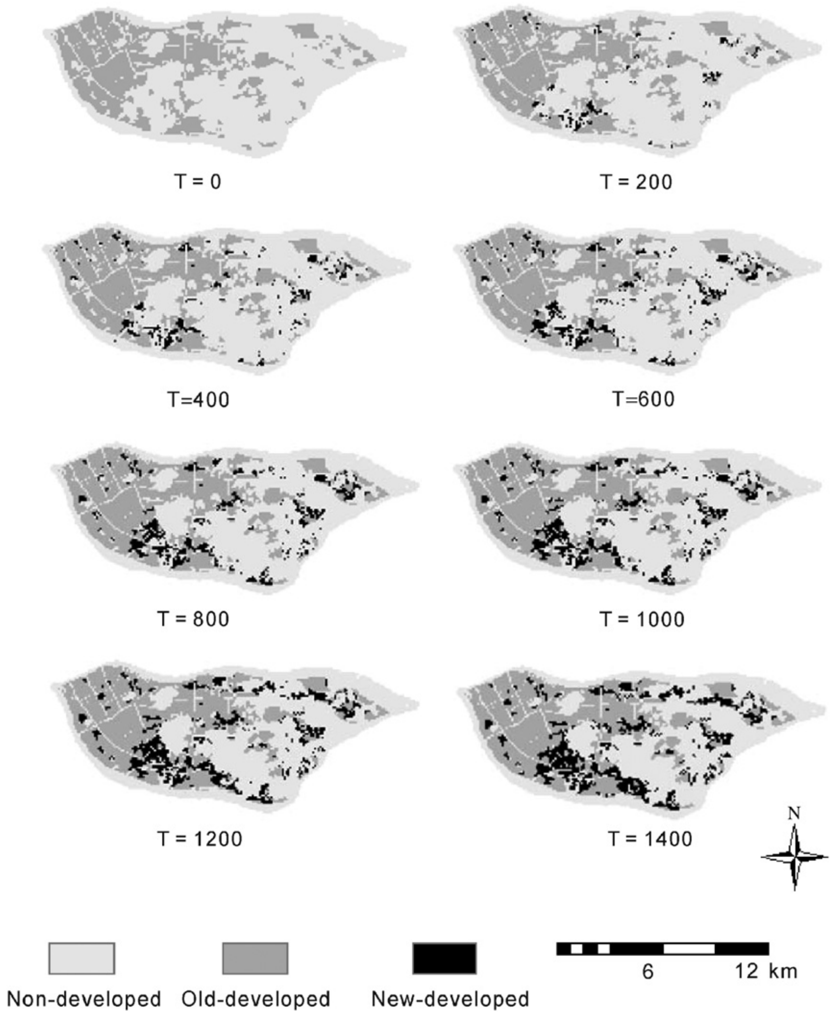

Fig. 4. Simulating complex residential development in Guangzhou in 1995–2004.

patterns visually (Fig. 5). The visual comparison can provide a rough estimation about the accuracy of this proposed model for simulating urban development.

A further quantitative analysis is to produce a confusion matrix indicating the concordance between the simulated and the actual development patterns. It is based on the spatial overlay of these two patterns cell-by-cell. Table 5 is the comparison of these two patterns in 2004. The overall accuracy is 0.789 and the kappa coefficient is 0.530 for the simulation.

The cell-by-cell comparison has limitations in reflecting pattern accuracy because the comparison does not consider characteristics of morphology. Many geographical applications are concerned with the characteristics of morphology, such as connectivity, fractals, and compactness. These features can be measured by using some aggregated indicators, such as compactness (Yeh and Li, 2001), fractal dimension (White and Engelen, 1993), and Moran's I (Wu, 2002).

In this study, the simple indicator Moran's I is applied to the measurement of land-use patterns. It is quite easy to calculate the Moran's I values in the GIS package, ARC/INFO GRID. Moran I is a useful spatial indicator that can reveal the degree of spatial autocorrelation (Goodchild, 1986). The indicator is able to estimate how close the simulated land-use pattern is to the actual pattern (Wu, 2002). The maximum value is one which indicates the most compact form of a land-use type. A smaller value, which can be below zero, indicates a more even distribution of the land-use type. The analysis indicates that the simulated patterns are very close to the actual patterns based on the Moran I indicator. The Moran I values are 0.688 and 0.671 for the actual 2004 patterns and the simulated 2004 patterns, respectively.

Multi-agent systems usually involve numerous parameters. How to define these parameters and calibrate these models is still a problem. Automatic calibration procedures have been used to assist the search for suitable parameter values of cellular automata (Li and Yeh, 2004b). However, these procedures have difficulties in solving the calibration problems of multi-agent systems, of which the parameters may be in different hierarchies. These parameters are usually heuristically defined according to the approach of trial and error. In this study, ΔP_1 is set to 0.005 and ΔP_2 is

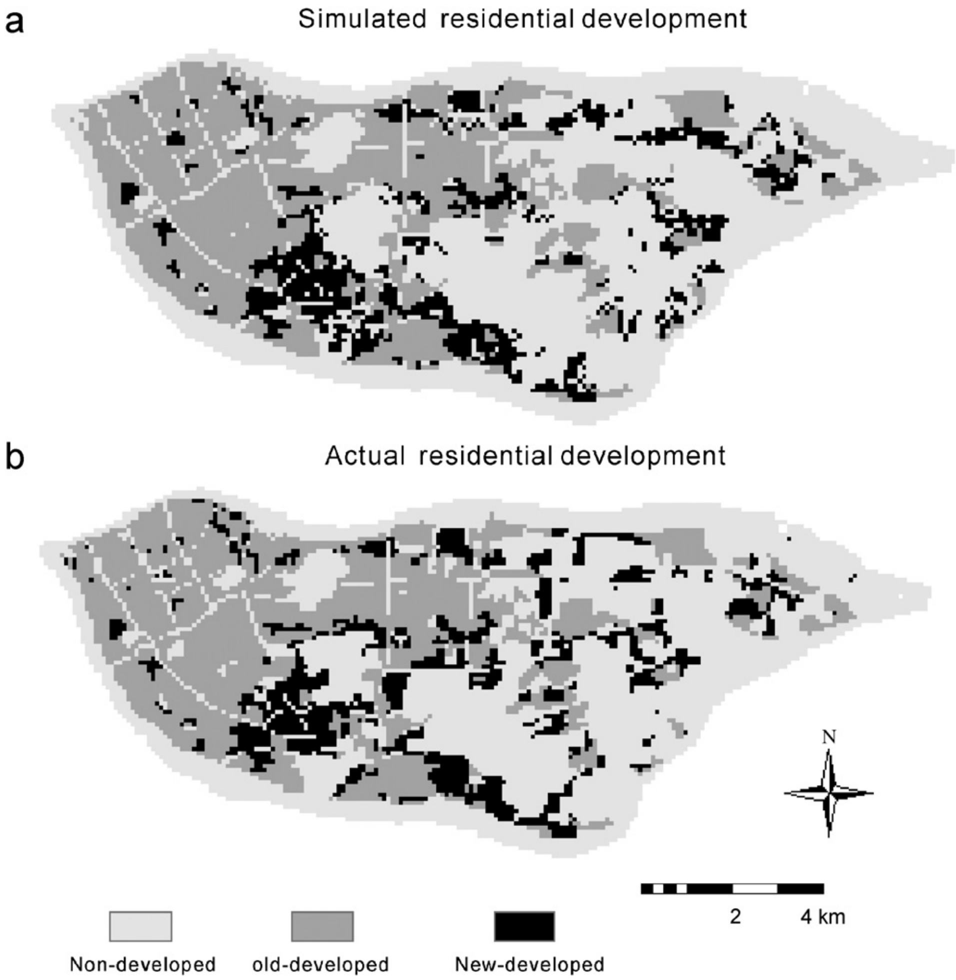

Fig. 5. Comparison of simulated patterns (a) with actual patterns (b) for residential development in Guangzhou in 2004.

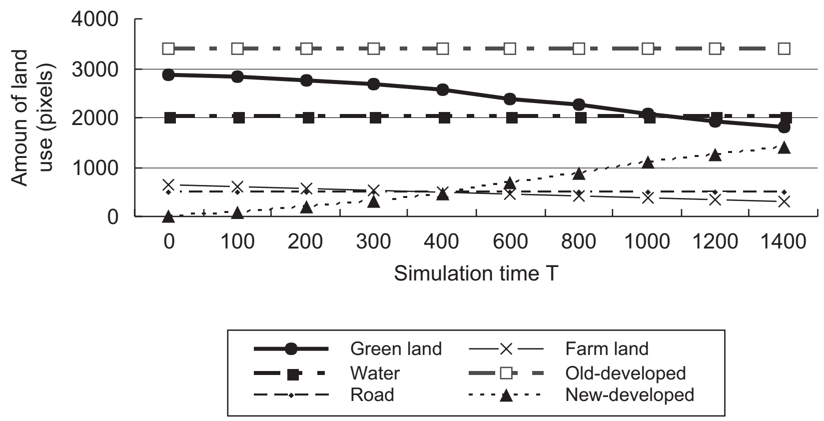

Fig. 6. Changes in the total amounts of various land-use types in Guangzhou in 1995–2004.

set to 0.1. However, different sets of parameters can be defined and experiments are required to assess the model's sensitivity to parameters. Table 6 lists five possible combinations of the parameters. Fig. 7 shows the overlap of the simulation outcomes from these five parameter settings. It is clear that the simulation patterns can be largely repeated for these different parameter settings. The uncertainties only exist in the fringe areas of development clusters with a small percentage of disagreement.

The performances of this integrated model were compared with those from the pure CA models. A logistic-CA model (Wu, 2002) was used to simulate the residential development in the study area. The overall accuracy was 0.510, and the Moran's I was 0.597 for this CA model. This indicates that the agent-based model can produce better outcomes for modeling residential development in this study area. This is because the behavior of various players in actual urban systems can be well addressed by using this agent-based approach.

4. Conclusion

Recent years have witnessed fast urbanization and intensive land-use conflicts in some developing countries. The simulation and prediction of urban growth is important for urban planners to formulate sustainable development strategies. The simulation of residential development can provide useful information about future land demands and landscape changes. However, cities are complex systems which are difficult to represent by using mathematical equations. There is a growing amount of researche on using "bottom-up" approaches, such as cellular automata (CA), to simulate complex urban systems.

One problem with CA models is that human and social factors are difficult to incorporate in the simulation. This kind of model cannot reflect the complex interactions between individuals and the environment. Agent-based modeling techniques can be used to simulate complex residential development which involves various players in shaping urban morphology. It is critical to define agents' behavior in a more consistent way. In this study, the characteristics of agents are defined by using GIS data to reflect various decision behaviors. Agents can be classified into a few groups according to data availability. However, the model cannot be too generalized if the spatial details are maintained. This study distinguishes six groups of resident agents by using census data. Weights are used to calculate the utility functions so that their decision behavior can be defined. The Saaty method can be used to facilitate the determination of weights in a more robust way.

The proposed model includes three main types of agents—resident agents, developer agents, government agents. Different types of agents can interact and negotiate with each other in influencing residential development. The actions of these agents are also subject to the constraints of the surrounding environment, which can be defined and updated by using remote sensing and GIS data.

The Haizhu district of Guangzhou city was selected for testing this proposed model. The experiment indicates that fast residential development can be simulated by the integration of cellular automata, multi-agent systems and GIS. Plausible simulation results have been confirmed by comparing the simulated patterns with the actual ones, which are classified from remote sensing data. The validation is based on cell-by-cell comparison and Moran's I indicator. The analysis indicates that the model can produce satisfactory results. This model also performed better than pure-CA models in simulating the complex residential development in the study area. This is because the behavior of various players in actual urban systems can be well addressed by this agent-based approach.

Chinese cities are experiencing fast transitions and quick changes in physical sizes and functional structures. There are difficulties in the simulation of complex land-use dynamics because of the uncertainties brought by these fast changes. Moreover, there is a general lack of detailed social and economic data at a fine spatial scale. This has created problems in defining agents' properties in a more accurate way. For example, governments' decisions are often made in a back-box manner which is hard to predict. There is also a general lack of detailed spatial information for property developers. Meanwhile, inherent data errors in

Table 5
Accuracies of the simulation according to the cell-by-cell comparison

	Simulated developed	Simulated non-developed	Accuracy
Actual developed	966	440	0.687
Actual Non-developed	433	2309	0.842
Overall Accuracy			0.789
Kappa coefficient			0.530

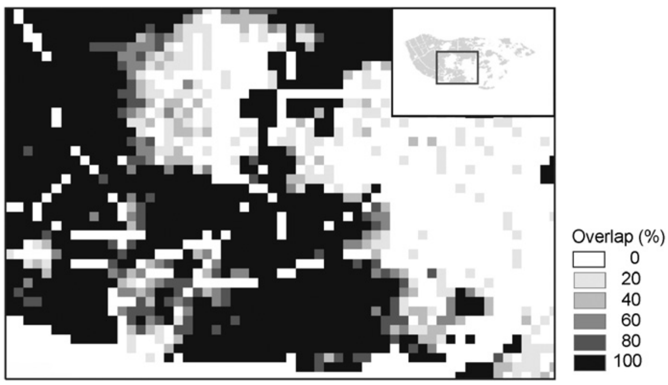

Fig. 7. Investigation of the model's sensitivity to parameters.

Table 6
Some possible combinations of the parameters

Experiments	1		2		3		4		5	
	ΔP_1	ΔP_2	ΔP_1	ΔP_2	ΔP_1	ΔP_2	ΔP_1	ΔP_2	ΔP_1	ΔP_2
Values	0.005	0.1	0.004	0.09	0.006	0.11	0.003	0.12	0.007	0.08

GIS and uncertainties from models will also affect the application of these agent-based models. The issues of data errors and model uncertainties have been well addressed in the GIS literature. The above problems should be addressed in future studies.

Acknowledgments

This study was supported by the National Outstanding Youth Foundation of China (Grant no. 40525002), the National Natural Science Foundation of China (Grant no. 40471105), the PhD development program from the Ministry of Education of China (Project no. 20040558023), and the "985 Project" of GIS and Remote Sensing for Geosciences from the Ministry of Education of China (Grant no. 105203200400006).

References

Basu, N., Pryor, R.J., 1997. Growing a Market Economy. Technical Report SAND-97-2093, Sandia National Laboratories, Albuquerque, NM, USA.

Batty, M., Xie, Y., 1994. From cells to cities. Environment and Planning B: Planning and Design 21, 531–548.

Benenson, I., Torrens, P.M., 2004. Geosimulation: Automata-Based Modeling of Urban Phenomena. Wiley, Chichester, 287 pp.

Benenson, I., Itzhak, Omer, Erez, Hatna, 2002. Entity-based modeling of urban residential dynamics: the case of Yaffo, Tel Aviv. Environment and Planning B 29, 491–512.

Brown, D.G., Page, S., Riolo, R., Zellner, M., Rand, W., 2005. Path dependence and the validation of agent-based spatial models of land use. International Journal of Geographical Information Science 19 (2), 153–174.

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

Eastman, J.R., 1999. Multi-criteria evaluation and GIS. In: Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W. (Eds.), Geographical Information Systems. Wiley, New York, pp. 493–502.

Epstein, J.M., 1999. Agent-based computational models and generative social science. Complexity 4 (5), 41–60.

Gilbert, N., Conte, R., 1995. Artificial Societies: The Computer Simulation of Social Life. UCL Press.

Goodchild, M.F., 1986. Spatial Autocorrelation: Concepts and Techniques in Modern Geography. Geo Books, Norwich, 47pp.

Kohler, T.A., 2000. Putting social sciences together again: an introduction to the volume. In: Kohler, T.A., Gumerman, T.A. (Eds.), Dynamics in Human and Primate Societies. Oxford University Press, New York, pp. 1–18.

Li, X., Yeh, A.G.O., 2000. Modeling sustainable urban development by the integration of constrained cellular automata and GIS. International Journal of Geographical Information Science 14 (2), 131–152.

Li, X., Yeh, A.G.O., 2004a. Analyzing spatial restructuring of land use patterns in a fast growing region using remote sensing and GIS. Landscape and Urban Planning 69 (4), 335–354.

Li, X., Yeh, A.G.O., 2004b. Data mining of cellular automata's transition rules. International Journal of Geographical Information Science 18 (8), 723–744.

McFadden, D., 1978. Modeling the choice of residential location. In: Karlqvist, A., Lundqvist, L., Snickars, F., Weibull, J. (Eds.), Spatial Interaction Theory and Planning Models. North-Holland Publishing Co., Amsterdam, pp. 75–96.

Saaty, T.L., 1990. The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation. University of Pittsburgh, Pittsburgh, 287 pp.

Waddell, P., 2002. UrbanSim—modeling urban development for land use, transportation, and environmental planning. Journal of the American Planning Association 68 (3), 297–314.

Ward, D.P., Murray, A.T., Phinn, S.R., 2000. A stochastically constrained cellular model of urban growth. Computers, Environment and Urban Systems 24, 539–558.

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., Uijee, I., 1997. The use of constrained cellular automata for high-resolution modeling of urban land use dynamics. Environment and Planning B: Planning and Design 24, 323–343.

Wu, F., 2002. Calibration of stochastic cellular automata: the application to rural-urban land conversions. International Journal of Geographical Information Science 16 (8), 795–818.

Wu, F., Webster, C.J., 1998. Simulation of land development through the integration of cellular automata and multicriteria evaluation. Environment and Planning B 25, 103–126.

Yeh, A.G.O., Li, X., 2001. A constrained CA model for the simulation and planning of sustainable urban forms by using GIS. Environment and Planning B: Planning and Design 28, 733–753.