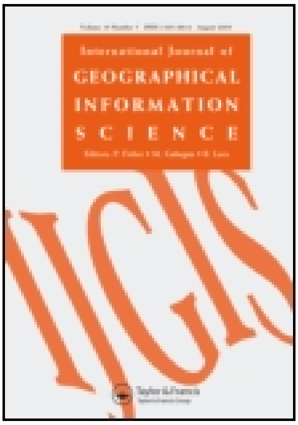

International Journal of Geographical Information Science

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tgis20

Defining agents' behaviour based on urban economic theory to simulate complex urban residential dynamics

Yimin Chen [a b] , Xia Li [a b] , Shujie Wang [a b] & Xiaoping Liu [a b]

[a] Guangdong Key Laboratory for Urbanization and Geo-simulation, School of Geography and Planning, Sun Yat-Sen University

[b] Guangdong Key Laboratory for Urbanization and Geo-simulation, Guangzhou, PR China

Version of record first published: 09 Jan 2012.

To cite this article: Yimin Chen, Xia Li, Shujie Wang & Xiaoping Liu (2012): Defining agents' behaviour based on urban economic theory to simulate complex urban residential dynamics, International Journal of Geographical Information Science, 26:7, 1155-1172

To link to this article: http://dx.doi.org/10.1080/13658816.2011.626780

International Journal of Geographical Information Science
Vol. 26, No. 7, July 2012, 1155–1172

Defining agents' behaviour based on urban economic theory to simulate complex urban residential dynamics

Yimin Chen, Xia Li*, Shujie Wang and Xiaoping Liu

Guangdong Key Laboratory for Urbanization and Geo-simulation, School of Geography and Planning, Sun Yat-Sen University, and Guangdong Key Laboratory for Urbanization and Geo-simulation, Guangzhou, PR China

(*Received 22 December 2010; final version received 20 September 2011*)

In recent years, agent-based models (ABMs) have become a prevalent approach for modelling complex urban systems. As a class of bottom-up method, ABMs are capable of simulating the decision-making as well as the multiple interactions of autonomous agents and between agents and the environment. The definition of agents' behaviour is a vital issue in implementing ABMs to simulate urban dynamics. Urban economic theory has provided effective ways to cope with this problem. This theory argues that the formation of urban spatial structure is an endogenous process resulting from the interactions among individual actors that are spatially distributed. However, this theory is used to explain urban phenomena regardless of spatial heterogeneity in most cases. This study combines GIS, ABM and urban economic models to simulate complex urban residential dynamics. The time-extended model is incorporated into an ABM so as to define agents' behaviour on a solid theoretical basis. A spatial variable is defined to address the neighbourhood effect by considering spatial heterogeneity. The proposed model is first verified by the simulation of three scenarios using hypothetical data: (1) single dominated preference; (2) varying preferences on the basis of income level; and (3) spatially heterogeneous environment. Then the model is implemented by simulating the residential dynamics in Guangzhou, China.

Keywords: agent-based model; urban economic theory; time-extended model; residential dynamics

1. Introduction

Cities are typical complex systems that are characterized as self-organized, non-linear and emergent. Studies have demonstrated that urban systems can be simulated by a group of 'bottom-up' simulation models, such as cellular automata (Tobler 1979, Clarke and Gaydos 1998, Wu and Webster 1998, Li and Yeh 2000, 2002, 2004, Li *et al.* 2011). However, it has been considered that these models have some limitations, such as the overstated emphasis on local interactions, the lack of economic foundation, the inability to explain human behaviour and the omission of external variables (Irwin and Geoghegan 2001). Agent-based models (ABMs), another class of bottom-up approach, are regarded as a promising tool for modelling complex urban systems, in which the landscape is organized through a cellular model and altered by the interacting individuals through a set of rules (Parker *et al.*

*Corresponding author. Email: lixia@mail.sysu.edu.cn; lixia@graduate.hku.hk

ISSN 1365-8816 print/ISSN 1362-3087 online

http://dx.doi.org/10.1080/13658816.2011.626780
http://www.tandfonline.com

2003, An *et al.* 2005). It is suggested that ABMs can better reveal the underlying interactions and dynamics of geographical phenomena and processes because of their strengths of explicitly simulating the process of individual decision-making (An *et al.* 2005, Jepsen *et al.* 2006, Xie *et al*. 2007).

As ABMs centralize around human action rather than landscape and transitions (Parker *et al.* 2003, An *et al.* 2005), a major problem in applying ABMs is how to solicit agents' behaviour using spatial information. Empirical knowledge and heuristics may be useful to define agents' behaviour. One can determine some of the parameters of an ABM using GIS and census data according to multicriteria evaluation (MCE) techniques (Li and Liu 2007). However, this cannot guarantee that agents can perform their functions in the same manner as human experts. Thus an essential requirement for building consistent ABMs is a well-defined theory, such as urban economic theory, for the definition of agents' behaviour. Urban economic theory demonstrates that the formation of urban spatial structure is an endogenous process which is the result of interactions among individual actors that are spatially distributed (Fujita 1989, Krugman 1991, Anas and Kim 1996). Therefore, the incorporation of urban economic theory can ensure that agents' behaviour is defined on a solid theoretical basis.

In this study, we integrate an ABM with urban economic theory to simulate urban residential dynamics. Literature on the simulation of residential dynamics can be traced back to as early as the 1970s, such as the basic segregation model proposed by Schelling (1971). Benenson (1998) is among those early authors to simulate residential dynamics using ABMs. In Benenson's model, the behaviour of inhabitants depend on their changing economic and cultural status as well as the properties of local and global residential environment. Kii and Doi (2005) proposed a multi-agent model for testing various development policies, but the resolution of the data they used was relatively coarse. Recently, Li and Liu (2008) incorporated sustainable development theories into an ABM to generate different scenarios of residential development. Caruso and colleagues (2007) combined cellular automata (CA) and microeconomic model to examine peri-urbanization. Although a CA model was utilized instead of an ABM, residential agents were implicitly involved and their behaviours were defined through the microeconomic model. Webster and Wu (1999a, 1999b) also embedded economic theory into CA to explore how urban development impacts environment under different types of regulation.

This study further elaborates on the definition of residential agents' behaviour based on urban economic theory for simulating complex urban residential dynamics. The proposed model incorporates more spatial details, such as the neighbourhood effect, accessibility and environmental amenity, to improve the simulation performance. First, agents' behaviour is defined through the equilibrium solution of the time-extended model (Fujita 1989). Second, the proposed model is verified by the simulation of three scenarios using hypothetical data: (1) single dominated preference; (2) varying preferences on the basis of income level; and (3) spatially heterogeneous environment. Finally, this model is tested by the simulation of residential dynamics in Guangzhou, China.

2. Defining agents' behaviour based on urban economic theory

2.1. Time-extended model

The time-extended model (Fujita 1989) explains how residents optimally select their residential locations with respect to economic constraints. In this model, commuting costs involve not only money but also time. A resident who lives in a monocentric and geographically homogeneous city intends to maximize their residential utility, subject to the constraints of budget balance and time balance. The utility function is defined as $U(z, s, t_1)$, where the variables z, s, t_1 represent non-spatial commodity consumption, residential size and leisure time, respectively. The budget balance and time balance are represented by the following equations:

$$z + R(r)s + ar = Y_N + Wt_w \tag{1}$$

$$t_1 + t_w + br = \bar{t} \tag{2}$$

In Equation (1), $R(r)$ is the rent of a parcel with the distance r from the central business district (CBD), a represents the unit cost of commuting, t_w is the working time and Y_N and W represent the non-wage income and wage income, respectively. This equation indicates that a resident's overall cost of commodity consumption, rent payment and commuting equals the total income (including wage income and non-wage income). The variable $\bar{t}$ in Equation (2) is the total time spent by a resident for three purposes: leisure, working and commuting. The parameter b is the time cost per unit of distance. The objective of a resident is expressed as

$$\max_{r,z,s,t_1,t_w} U(z, s, t_1) \tag{3}$$

After transforming Equations (1) and (2), we can get

$$z + R(r)s + Wt_1 = I(r) \tag{4}$$

$$I(r) = Y_N + W(\bar{t} - br) - ar \tag{5}$$

where $I(r)$ is called the potential income.

The urban economic theory describes the urban residential pattern as the consequence of the process of interactions among individual residents. We depict the decision-making of the residents using the concept of bid-rent. Caruso *et al*. (2007) also used this concept to explain the formation of different residential structures in the urban fringe. Bid-rent refers to the maximum rent that a resident would like to pay for reaching the residential utility u at a specified location with the distance r from the CBD:

$$\Psi(r, u) = \max_{z,s,t_1} \left\{ \frac{I(r) - z - Wt_1}{s} | U(z, s, t_1) = u \right\} \tag{6}$$

The 'bid-rent' is relative rather than equivalent to the actual 'rent' in the real world. It is more appropriate to consider the 'bid-rent' as the residential cost (or more exactly, the highest cost one can afford), which can be the rent for an apartment or the housing mortgage payment, for example, Benenson's (1998) model. Benenson's model is concerned about the housing issue, in which the resident agents do not always 'fix' in the positions where they own their houses. It should be noted that the focus of this article is not on the rental market although the time-extended model is related to the concept of bid-rent. In the real rental market, the tenants are price (rent) takers, and the price is determined by the supply–demand relationship (or by policy), which needs other modules to address.

Generally, urban economic models neglect spatial heterogeneity to obtain an equilibrium solution; nonetheless, they offer promising potential for incorporating the effects of spatial heterogeneity at a disaggregate level (Irwin and Geoghegan 2001). Thus, in this study we introduce a spatial variable ρ to represent the neighbourhood effect in residential decisions. This variable suggests that the residents will consider the local condition of a parcel when evaluating its residential utility, for example, the compatibility to their neighbours, as well as budget their money and time during the decision-making process. The neighbourhood effect ρ is calculated using the following equation:

$$\rho = \exp\left[\sum_{j}^{n} \frac{\mathrm{Con}(H_j)}{n}\right] \tag{7}$$

where H_j represents the type of the neighbouring agents, which can be categorized as low, medium and high in terms of their income levels. The parameter n is the number of neighbours. $\mathrm{Con}(H_j)$ is the compatibility to a neighbouring resident of type j, which is calculated by the normalized difference of income level, that is, $|W_i - W_j|/(W_i + W_j)$.

The residential utility can be expressed as follows:

$$U(z, s, t_1, \rho) = \alpha \log z + \beta \log s + \gamma \log t_1 + \omega \log \rho \tag{8}$$

where α, β, γ and ω are the preference parameters to represent the weights an agent has for each component, that is, commodity, residential size, leisure time and neighbourhood effect, respectively; these parameters are subject to $\alpha > 0$, $\beta > 0$, $\gamma > 0$, $\omega > 0$ and $\alpha + \beta + \gamma + \omega = 1$.

Equation (8) is a revised version of the utility function in Fujita (1989). The original one is $U(z, s, t_1) = \alpha \log z + \beta \log s + \gamma \log t_1$. In Caruso *et al.* (2007), the Cobb–Douglas function was adopted to calculate residential utility. In this study we add the neighbourhood effect ρ in the utility function on the basis of the original utility function. As a result, the bid-rent can be rewritten as

$$\Psi(r, u) = \max_{z,s,t_1,\rho}\left\{\frac{I(r) - z - Wt_1}{s} | U(z, s, t_1, \rho) = u\right\} \tag{9}$$

In order to solve Equation (9), we can transform Equation (8) into

$$Z(s, t_1, \rho, u) = e^{u/\alpha} s^{-\beta/\alpha} t_1^{-\gamma/\alpha} \rho^{-\omega/\alpha} \tag{10}$$

From Fujita (1989), the maximization can be distilled as follows:

$$\frac{\partial\left(\frac{I(r) - Z(s, t_1, \rho, u) - Wt_1}{s}\right)}{\partial s} = 0 \tag{11}$$

$$\frac{\partial\left(\frac{I(r) - Z(s, t_1, \rho, u) - Wt_1}{t_1}\right)}{\partial t_1} = 0 \tag{12}$$

Then

$$-\frac{\partial Z}{\partial s} = \Psi(r, u) \tag{13}$$

$$-\frac{\partial Z}{\partial t_1} = W \tag{14}$$

Based on Equations (8), (10), (13) and (14), we can get

$$\Psi(r,u) = \left(\frac{\alpha}{1-\omega}\right)^{\alpha/\beta} \frac{\beta}{1-\omega}\left(\frac{\gamma}{1-\omega}\right)^{\gamma/\beta} W^{-\gamma/\beta} I(r)^{(1-\omega)/\beta} \rho^{\omega/\beta} e^{-u/\beta} \tag{15}$$

$$z = \frac{\alpha}{1-\omega} I(r) \tag{16}$$

$$t_1 = \frac{\frac{\gamma}{1-\omega} I(r)}{W} \tag{17}$$

$$s = \frac{\frac{\beta}{1-\omega} I(r)}{\Psi(r,u)} \tag{18}$$

$$u = V[I(r),\rho,\Psi] = \log\left[\left(\frac{\alpha}{1-\omega}\right)^{\alpha}\left(\frac{\beta}{1-\omega}\right)^{\beta}\left(\frac{\gamma}{1-\omega}\right)^{\gamma} W^{-\gamma} I(r)^{1-\omega} \rho^{\omega} \Psi^{-\beta}\right] \tag{19}$$

Equation (19) is called the indirect function of residential utility. The behaviours of residents are determined through the equations mentioned above, which illustrate how residents allocate their income and time to maximize their residential utility.

2.2. Simulating complex residential dynamics

Some assumptions (Kii and Doi 2005, Caruso *et al.* 2007) are adopted when implementing the proposed model to simulate residential dynamics: (1) residents (agents) are economically rational in that their objective is to maximize the residential utility at a certain parcel by optimally allocating their income and time, while the objective of the landlords is to maximize the land revenue, that is, the residential size multiplied by land rent; (2) for simplicity, the model omits the decision process of selecting a consumption site and the corresponding spatial cost of accessing the consumption site; (3) resident agents only take current conditions into account when deciding the residential location, regardless of the consequences; similarly, the landlords do not concern future change caused by the decision of re-allocating the parcel to a certain resident agent.

At each time step, each resident agent calculates the potential income according to the commuting distance r_{xy}, wage W_i, non-wage income $Y_{i,N}$, monetary cost a and time cost b per unit of distance, that is, using Equation (5). Then by Equations (16)–(18), the corresponding amount of commodity consumption, residential size and leisure time are computed. The neighbourhood effect at parcel (x, y) is evaluated using Equation (7). Finally, the residential utility u^t_{xy} at time t of a specified position is computed by Equation (8).

Resident agents have to collect some basic information during the residential decision process. When land markets are competitive, residents are assumed to have perfect knowledge about the city they live in (Fujita 1989). However, in this study, we assume that agents have bounded knowledge instead of perfect knowledge. This assumption is based on the realistic behaviour of residents during the site selection process. In real life, people can usually obtain detailed information about their neighbourhood, but are incapable of acquiring complete information of a large geographical extent, for example, the whole city. The proposed model integrates two types of information for agents to make decisions: deterministic information at a local scale and random information at the global scale. The former guides each agent to search for a vacant parcel with the highest residential utility, denoted as *locally best parcel*, by traversing all parcels within a rectangle neighbourhood (e.g. 5 × 5). Besides, each agent selects a position randomly at a global scale, denoted as *randomly selected parcel*. The parcel that has higher residential utility between the *locally best parcel* and the *randomly selected parcel* is denoted as the *target parcel*, which is considered by agents for relocation. If the utility surplus δu, that is, the difference of utility between the *target parcel* and current location, is greater than the given threshold Δu_0 ($\Delta u_0 > 0$), the agent will apply to the landlord of the target parcel for moving there. By denoting the target parcel as (x', y') and the corresponding residential utility as $u^t_{x'y'}$, there are

$$\delta u = u^t_{x'y'} - u^t_{xy} = \log\left[C_1 \frac{\Psi^t_{xy}}{\Psi^t_{x'y'}}\right] \tag{20}$$

$$C_1 = \left(\frac{I_{x'y'}}{I_{xy}}\right)^{1-\omega} \left(\frac{\rho_{x'y'}}{\rho_{xy}}\right)^{\omega} \tag{21}$$

and

$$\Psi^t_{x'y'} = \Psi^t_{xy} C_1^{-\beta} e^{-\delta u/\beta} \tag{22}$$

$$\frac{\partial \Psi^t_{x'y'}}{\partial \delta u} = -\frac{1}{\beta} \Psi^t_{xy} C_1^{-\beta} e^{-\delta u/\beta} < 0 \tag{23}$$

Equation (23) indicates that $\Psi^t_{x'y'}$ increases when δu declines. A resident agent prefers to raise the bid-rent $\Psi^t_{x'y'}$ up to the highest value in order to outbid others. The maximum of $\Psi^t_{x'y'}$ can be obtained when $\delta u = 0$. This is called the short-run residential equilibrium (Caruso *et al.* 2007). After $\Psi^t_{x'y'}$ is determined, the residential size at parcel (x', y'), that is, $s_{x'y'}$, can be computed using Equation (18).

If the parcel is applied for by only one resident agent, then this agent can possess the parcel without bidding. If two or more resident agents are competing for the same parcel, the one who offers the highest land revenue for the landlord can eventually occupy it. The land revenue is defined as follows:

$$H_r = \max_{H_i} \left\{\Psi_{H_i} s_{H_i}\right\} \tag{24}$$

A resident agent relocates when it outbids other agents who claim the same parcel, otherwise remains at its current position. The immediate cost of agents making a bid is 0, but it does not mean that there is no risk of making a bid. This can be represented by a constraint that an agent has to leave the city if it fails to outbid for a couple of times.

The vacant parcels depreciate at a predefined rate until the land rents of these parcels decline to the lower bound (i.e. the initial land rent). The rent of a certain parcel occupied by a resident agent at time step $t + 1$, Ψ_{t+1}, is updated according to Equation (25):

$$\Psi_{t+1} = w_0\Psi_t + (1 - w_0)\bar{\Psi}_N \tag{25}$$

where Ψ_t and $\overline{\Psi}_N$ represent the current rent and the mean rent of the neighbourhood, respectively.

A resident agent will leave the city once any one of the following situations is encountered:

(1) The residential utility at the current location is less than or equal to 0, and no target parcel can be found by the agent.
(2) The agent fails to outbid and the residential utility at current location falls to 0 or less than 0.
(3) The number of times the agent fails to outbid exceeds a threshold.

3. Model implementation and results

3.1. Model configuration and data

First, the proposed model is tested by applying it to hypothetical data. We design three scenarios with a 50 × 50 artificial city to identify the performance of the model: (1) single dominated preference; (2) varying preferences on the basis of income level; and (3) spatially heterogeneous environment. Then this model is implemented for simulating the residential dynamics of Guangzhou in 2006 (scenario 4). This simulation outcome is evaluated by the actual housing price in 2006 (source: Guangzhou Land Resources and Housing Authority, http://www.laho.gov.cn/).

In the first experiment, all parcels in the artificial city are vacant with the initial land rent. We assume that all resident agents are working at the city centre, that is, the CBD. The resident agents are classified in terms of income level, graded as low, medium and high. At the beginning of the simulation process, the initial locations of the resident agents are randomly generated to avoid the bias caused by manually deciding their locations (Jackson *et al*. 2008, Yin 2009). At each iteration step, a number of new residents are added to the city until the population of residents reaches the predefined value, which depends on the land supply. The simulation process will be terminated when the residential pattern evolves into a stable stage in which the average residential utility converges to a certain value. Figure 1 illustrates the decision process of agents at each iteration step, which includes the following procedures: (1) the agent evaluates the location it currently possess; (2) the agent searches for better parcels; (3) the agent relocate when outbidding other competitors by offering the highest land revenue for the landlord; (4) the agent leaves the city under the situation specified in Section 2.2, otherwise still remains at the current position.

Table 1 shows the basic configuration for the first three scenarios using the hypothetical data. Since the simulation based on ABMs is sensitive to agents' preferences, the first two scenarios are used to evaluate the impact of agents' preferences. Therefore, we leave out other spatial constraints but only consider the distance to CBD so that the effects of agents' preferences can be better understood. The first scenario, that is, 'single dominated preference', assumes that all resident agents have the same consumption preference and is further divided into three sub-scenarios: dominated preference of commodity, residential size and leisure time, respectively. Table 2 lists agents' preferences for all sub-scenarios.

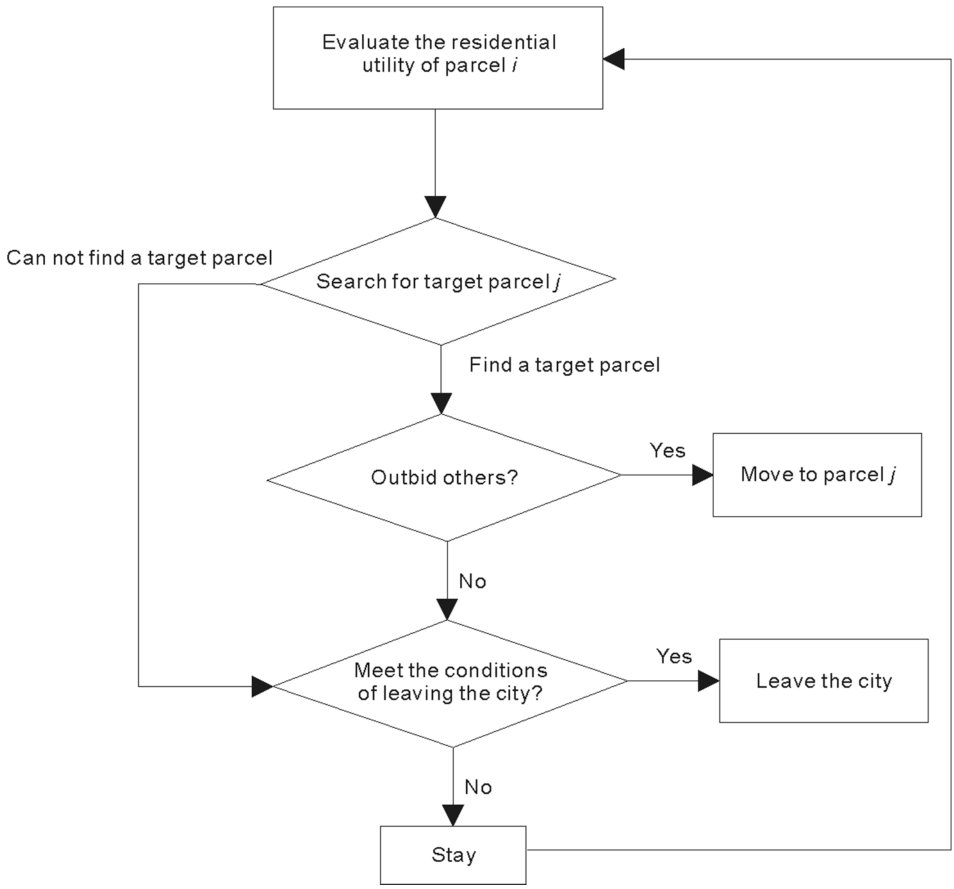

Figure 1. Flow chart of agents' decision process.

Table 1. Model configuration for all three scenarios.

	Income level				Income level		
	Low	Medium	High		Low	Medium	High
Proportion	0.4	0.35	0.25	Non-wage income	1000	500	0
Wage	10	30	50	$\bar{t}$	1000	500	0
b	4	2.5	1	Agent count	1500	1500	1500
a	1	45	110				

Table 2. Agents' preferences in scenario 1.

Dominated preference of	α	β	γ	ω
Commodity consumption	0.85	0.05	0.05	0.05
Residential size	0.05	0.85	0.05	0.05
Leisure time	0.05	0.05	0.85	0.05

The second scenario represents 'varying preferences on the basis of income level', which assumes that low-income agents are more likely to spend their money on commodity while high-income agents are prone to enlarge their residential size and have more leisure time. The medium-income agents' preferences are moderate between those of the low-income and high-income agents. Table 3 shows agents' preferences of this scenario.

The artificial city is in a rather homogenous pattern which cannot represent the geographical complexity of the real world. In scenario 3, three spatial variables, distance to roads, distance to stations and distance to environmental entities, are incorporated into the proposed model so as to involve more spatial heterogeneity (Figure 2).

Generally, the accessibility to the CBD is better for parcels close to major roads. This effect can be transformed into the sense of 'shrinking' the distance between a parcel and CBD:

$$s_d = k_r d_r + b_0 \tag{26}$$

$$r_s = s_d r \tag{27}$$

where s_d is the shrinking parameter. As shown by Equation (26), s_d is the linear function of d_r (the normalized distance to roads), in which k_r is the slope and b_0 is the intercept. k_r, b_0 are subject to $k_r \geq 0$, $b_0 \geq 0$ and $k_r + b_0 = 1$. These conditions ensure that s_d ranges from 0 to 1. Equation (26) indicates that locations closer to roads have stronger shrinking effects (lower value of s_d). In Equation (27), r is the original distance to the CBD and r_s is the shrunken distance to the CBD.

Table 3. Agents' preferences in scenarios 2–4.

	Scenario 2				Scenario 3				Scenario 4 (Guangzhou)			
Income level	α	β	γ	ω	α	β	γ	ω	α	β	γ	ω
Low	0.65	0.15	0.1	0.1	0.65	0.15	0.1	0.1	0.65	0.1	0.1	0.15
Medium	0.3	0.3	0.25	0.15	0.3	0.3	0.2	0.2	0.3	0.2	0.25	0.25
High	0.1	0.3	0.45	0.15	0.1	0.3	0.35	0.25	0.15	0.25	0.3	0.3

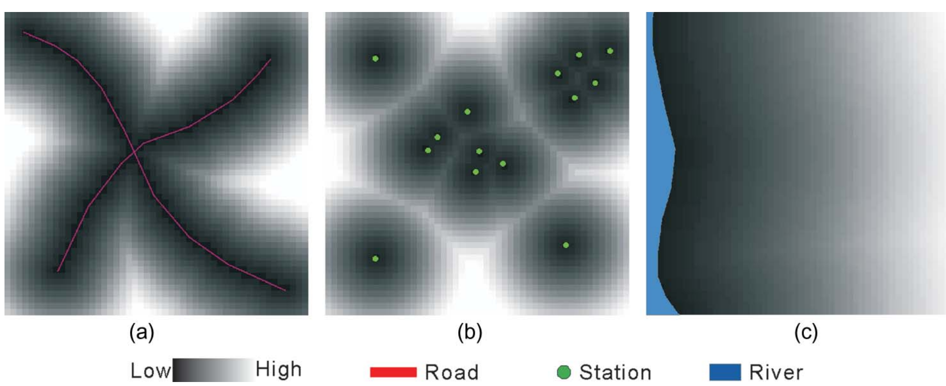

Figure 2. Hypothetical data for scenario 3: (a) distance to roads; (b) distance to stations; and (c) distance to river.

The public transit system is important in urban transport systems, particularly for the low-income residents. Therefore this scenario assumes that time cost per unit of distance decreases at a rate that correlates to the normalized distance to stations d_s:

$$s_t = k_s d_s + b_0 \tag{28}$$

$$b_s = s_t b \tag{29}$$

where s_t is the reduction parameter; it is calculated using a linear function in which k_s and d_0 are slope and intercept, respectively. They are also subject to $k_s \geq 0$, $b_0 \geq 0$ and $k_s + b_0 = 1$. In Equation (29), b and b_s are the original and reduced time cost per unit of distance, respectively.

Environmental amenity is also a crucial factor for agents to make residential decision. Locations are more attractive when they are near mountains covered with a large area of well-grown natural vegetation or those close to rivers/lakes with clean water. In scenario 3, the environmental amenity E is represented through the exponential function of normalized distance to environmental entities d_e:

$$E = \exp(-\lambda d_e) \tag{30}$$

Equation (30) indicates that environmental amenity decreases as the distance to environmental entities increases. λ is the scaling factor, which will magnify or weaken the effect. Originally, the neighbourhood effect variable ρ only takes into account the homogeneity of neighbouring agents, which can be denoted as S. The environmental amenity E can be further embedded into ρ, and consequently ρ can be reformulated as follows:

$$\rho = w_s S + w_e E \tag{31}$$

w_s and w_e are subject to $w_s + w_e = 1$.

Additionally, the economic status of residents will be updated at each step. The wage of an agent at time t is renewed according to Equation (32) (Benenson 1998):

$$W_{t+1} = W_t + \varepsilon R\left(\frac{W_t}{K}\right)\left(1 - \frac{W_t}{K}\right) \tag{32}$$

where W_t and W_{t+1} are the respective wages at time t and t+1, R is the growth rate, ε is the random effect and K is the capacity of wage, that is, the maximum value of wage. Agents' preferences and model configuration for scenario 3 are shown in Tables 3 and 4, respectively.

In the second experiment, the proposed model is applied to the simulation of residential dynamics in Guangzhou, China (scenario 4). The spatial variables which are used for implementing this model are shown in Figure 3. The study area is approximately 1355 km^2. There are 449 × 435 pixels with the resolution of 100 m. These spatial variables include distance to city centres, distance to stations, urban extent, distance to major roads and distance to environmental entities. According to the analysis presented by Lu *et al*. (2006), there are dual centres in Guangzhou: one is in the old city proper and another is in the so-called new urbanized areas that are mainly within Tianhe District and Huangpu District. Therefore, the variable of distance to city centres (Figure 3a) is calculated based on the

Table 4. Model configuration in scenario 3.

	Income level				Income level		
	Low	Medium	High		Low	Medium	High
k_r	0.5	0.2	0	λ	5	5	5
k_s	0.5	0.5	0.5	w_s	0.7	0.5	0.3
R	0.2	0.2	0.2	w_e	0.3	0.5	0.7
K	100	100	100				

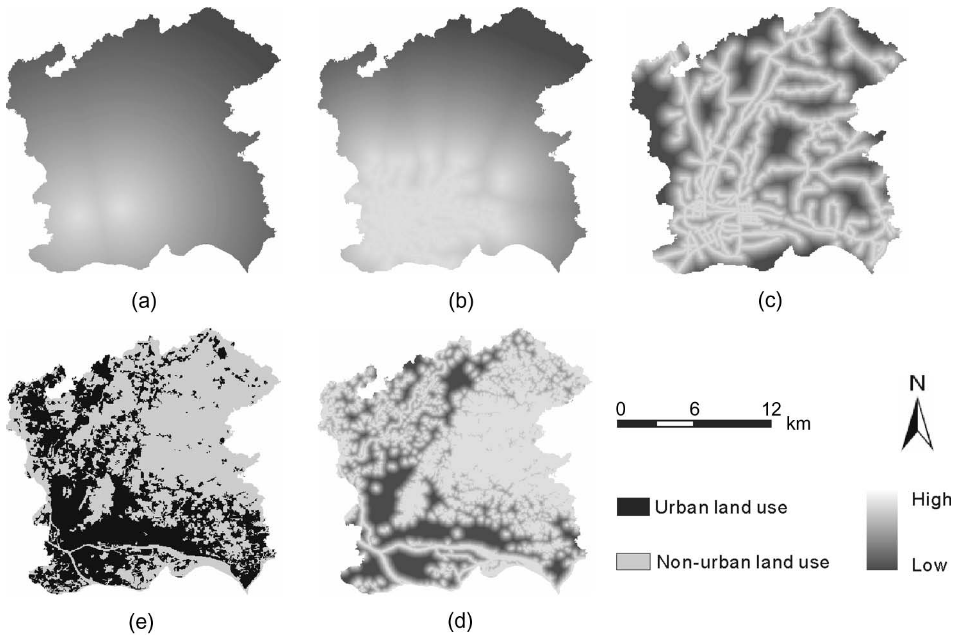

Figure 3. The required data for simulating the residential dynamics in Guangzhou, China: (a) distance to city centre; (b) distance to stations; (c) distance to major roads; (d) distance to environmental entities; and (e) urban extent.

locations of the dual centres. The study conducted by Li and Liu (2007) demonstrates that the distance to major roads is a significant factor when residents in Guangzhou consider relocation. As shown in Figure 3c, the distance to roads represents the accessibility of a location, which is generated based on the spatial distribution of the major roads rather than every road or street in Guangzhou. The distance to environmental entities is confirmed to be important for evaluating a location's residential utility in accordance with the fact that the rent/housing price is the highest along the Pearl River in Guangzhou.

The MCE method (Li and Liu 2007) is adopted to estimate the agents' preferences in each group (Table 3). The parameters k_r, k_s and λ are specified by a trial-and-error approach through extensive simulations. This is accomplished by running the model using different combinations of parameters and comparing the simulated land rent to the actual housing price. The values of k_r, k_s and λ are set according to the combination which generates the most similar result with the actual housing price. The number of agents is set according

Table 5. Model configuration in residential modelling of Guangzhou, China.

	Income level				Income level		
	Low	Medium	High		Low	Medium	High
Proportion	0.35	0.35	0.3	k_r	0.3	0.1	0
Agent count	4550	4550	3900	k_s	0.35	0.35	0.35
Wage	10	30	50	w_s	0.7	0.5	0.3
b	10	7	5	w_e	0.3	0.5	0.7
a	10	20	30	R	0.2	0.2	0.2
Non-wage income	0	0	0	K	100	100	100
$\bar{t}$	600	600	600	λ	15	15	15

to the quantity of urban land use in 2006. The values of other parameters are determined through statistical data for Guangzhou (http://www.gzstats.gov.cn/pchb/) (Table 5).

3.2. Results and discussion

Figure 4 displays the simulation of the three sub-scenarios in scenario 1. The residential patterns are similar in the sub-scenarios 'commodity consumption' and 'leisure time' that both residential patterns exhibit concentric structures. The majority of high-income resident agents congregate around the CBD, and conversely, almost all low-income agents locate at the periphery of the city. The space between the parcels occupied by the two

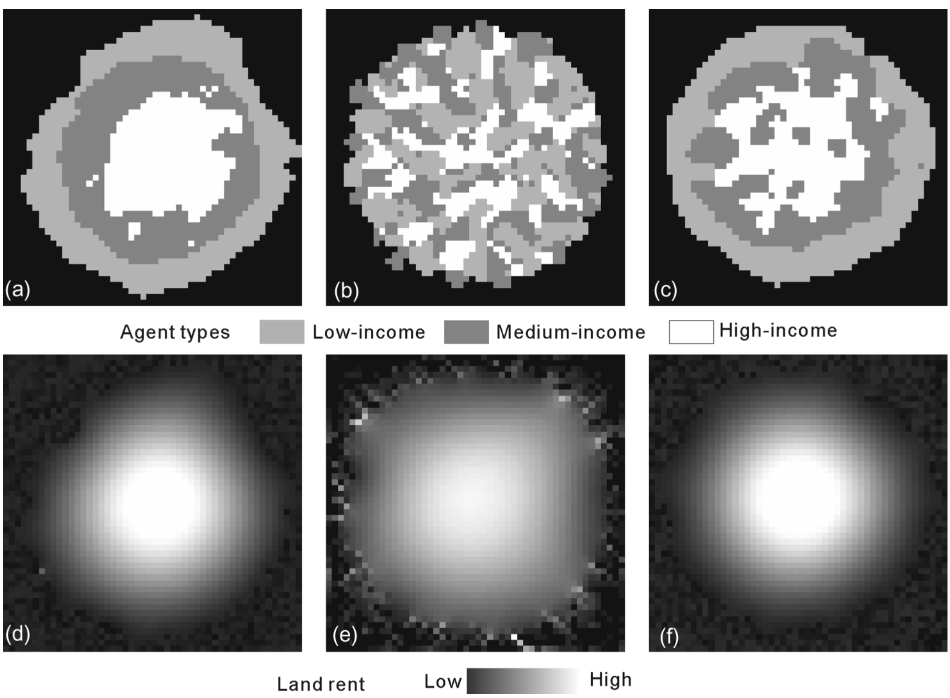

Figure 4. Simulation results of scenario 1 (divided into three sub-scenarios).

previous agent groups is possessed by the medium-income agents. The values of average distance to CBD of high-income agents are lowest compared with other resident groups, while of low-income agents are the highest in both sub-scenarios mentioned above. Unlike these two sub-scenarios, the sub-scenario 'residential size' exhibits a striped residential pattern. The stripes have irregular spatial distribution and each stripe is occupied by the same type of agents. Moreover, the spatial distribution of different resident agents is distinct from those of the previous two sub-scenarios: the values of average distance to CBD of three resident groups are extremely identical, and the values of standard deviation are the highest among three sub-scenarios (Table 6). Despite the difference in the residential patterns, the patterns of land rent shares similar spatial distribution in all sub-scenarios. The simulated land rent in all sub-scenarios consistently tends to decrease gradually from the CBD to the urban fringe. These patterns are very classical in urban economic theory (Fujita 1989). However, these classical patterns will be 'broken' when spatial heterogeneity is involved in the following scenarios.

As illustrated by Figure 5, the value of mean residential utility in the sub-scenario 'leisure time' increases rapidly in the early stage of the simulation process then gradually declines and eventually converges to a value close to 5.5. The convergence processes of mean residential utility are quite similar among the three sub-scenarios, so we just use the sub-scenario 'leisure time' as a visual example.

Figure 6 shows the results of scenarios 2 and 3. In scenario 2, most of the medium-income and high-income residents are living at the urban periphery while low-income residents are gathering around the central area (Table 7), which is contrary to the residential

Table 6. Average and standard deviation of distance to the CBD in scenario 1.

	Income level		
Dominated preference of	Low	Medium	High
Commodity consumption	19.69 (1.85)	13.88 (2.31)	7.77 (3.21)
Residential size	14.80 (5.03)	15.66 (5.01)	12.77 (5.22)
Leisure time	19.32 (1.82)	13.54 (3.28)	8.48 (3.51)

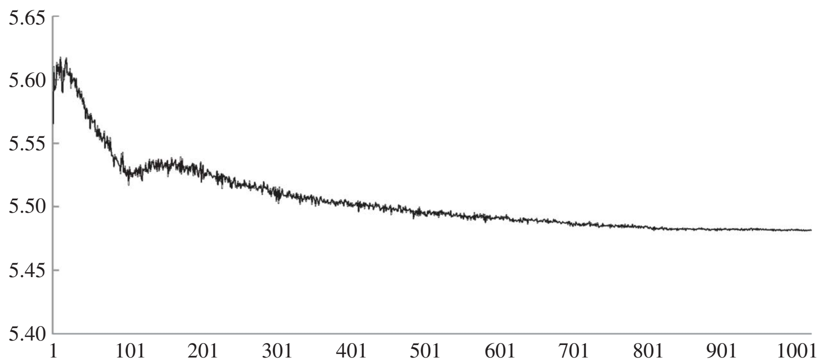

Figure 5. The convergence of mean residential utility (sub-scenario 3 in scenario 1). (a)-(c) are simulated residential patterns of sub-scenario commodity consumption, residence size and leisure time, respectively. (d)-(f) are simulated land rent of sub-scenario commodity consumption, residence size and leisure time, respectively.

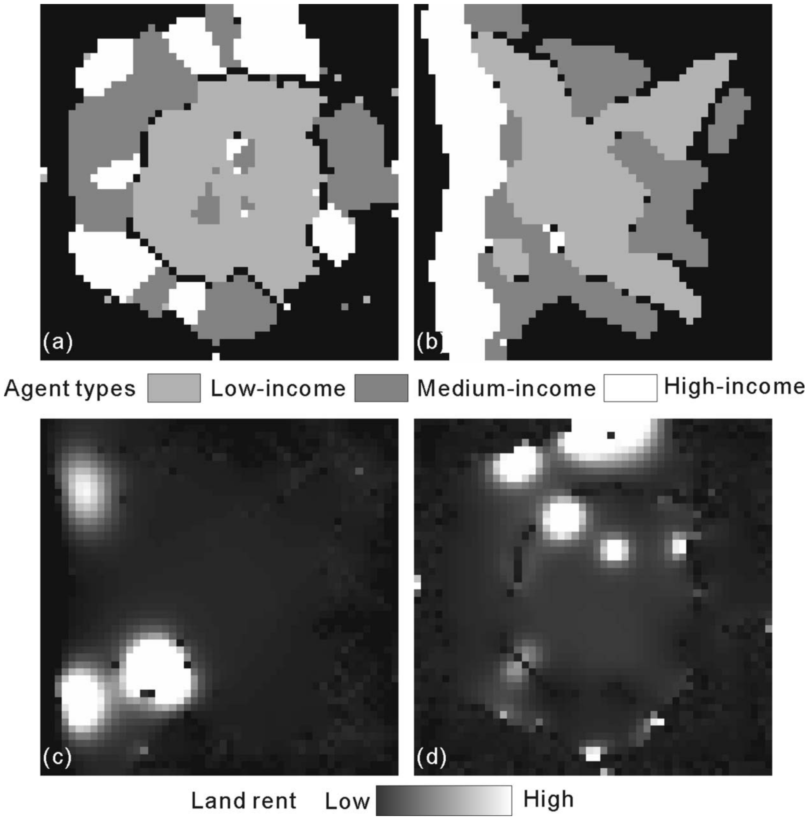

Figure 6. Simulation results of scenarios 2 and 3. (a) and (b) are simulated residential patterns of scenario 2 and 3, respectively. (c) and (d) are simulated land rent of scenario 2 and 3, respectively.

Table 7. Average and standard deviation of distance to the CBD in scenarios 2 and 3.

	Scenario 2			Scenario 3		
Income level	Low	Medium	High	Low	Medium	High
Average	10.24	17.68	19.36	11.87	15.98	22.75
Standard deviation	3.95	4.14	3.68	5.76	5.26	4.98

pattern in scenario 1. Furthermore, the land rent is highest in the urban fringe where there are different types of residents. This is different from the result that the land rent peaked in the city centre in scenario 1.

Scenario 3 displayed in Figure 6 produces a segregation pattern. Low-income residents prefer the parcels along the roads and around the city centre. The corridors close to rivers are more attractive to high-income residents. Medium-income residents scatter all over the region, including the urban fringe, locations with higher accessibility and parcels with high-quality environmental amenities. Due to the improved accessibility, the values of average distance to CBD in this scenario are the highest among all tested scenarios (Table 7). The highest land rent is observed in the northwest and southwest parts of the city which have greater environmental amenity and better transportation conditions.

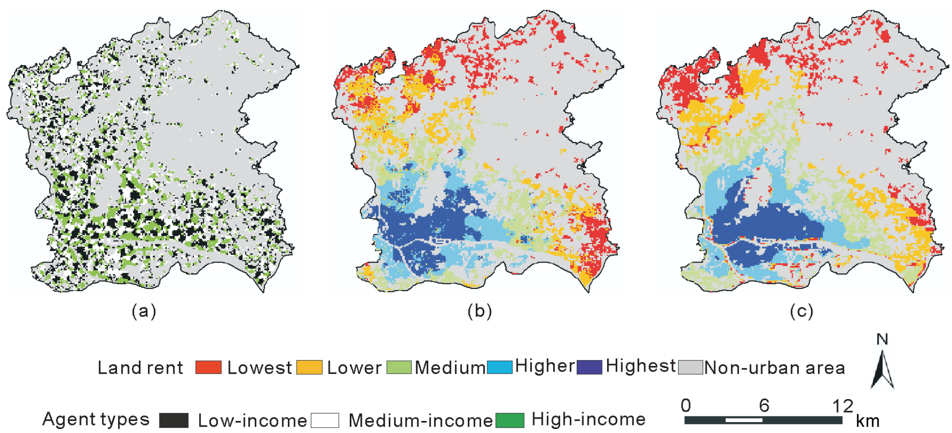

Figure 7. Simulating both the residents' distribution and land rent distribution of Guangzhou in 2006: (a) residents' distribution; (b) simulated land rent; and (c) actual housing price.

Figure 7 illustrates the simulation results using a real data set of Guangzhou in 2006. Table 5 lists the parameters of this simulation. Agents' preferences are determined by the MCE method. The use of the MCE method is quite popular when various spatial variables are involved to model geographical phenomena (Jankowski and Richard 1994, Jiang and Eastman 2000). This method will not produce unrealistic cluster patterns because of using regionalized preferences. One example is the simulation of sub-scenario 2 in scenario 1. In this sub-scenario, we only consider the distance to the CBD and the weight of neighbourhood effect is set to a very low value. As a result, the model produces a distinct residential pattern (a striped pattern) compared with those in sub-scenarios 1 and 3.

As depicted in Figure 7a, the locations with better environmental quality, for example, river banks, are mainly occupied by high-income residents. Low-income residents choose to live along major roads and stations to reduce commuting cost. The residential pattern of the medium-income agents is a trade-off between environmental quality and accessibility. The spatial pattern of agents in Figure 7a is more irregular, compared with that in Figure 6b. This is because much more spatial complexity exists in real spatial data, so the clusters are not as large as those shown in Figure 6b.

Figure 7b and c shows the simulated land rent and actual housing price of Guangzhou. A visual test reveals that the pattern of simulated land rent can well accord with the actual pattern of housing price. It is more convincing to use a quantitative method to assess the agreement between these two patterns. The kappa coefficient is generally used to quantify the similarity between two binary maps for model validation (Li and Liu 2007). We employ the Pearson's r to validate the model because it is more suitable for the comparison involving continuous values:

$$r = \frac{\sum_{i=1}^{n}\left(l_{s,i} - \bar{l}\right)_s \left(l_{a,i} - \bar{l}_a\right)}{\sqrt{\sum_{i=1}^{n}\left(l_{s,i} - \bar{l}\right)_s^2}\sqrt{\sum_{i=1}^{n}\left(l_{a,i} - \bar{l}_a\right)^2}} \tag{33}$$

where $l_{s,\,i}$ and $l_{a,\,i}$ are the simulated land rent and actual housing price at cell i.

The calculation of r is accomplished using a GIS package, ArcGIS 9.3 (ESRI, Redlands, CA). The result of r is 0.81 (the 95% confidence interval is [0.805, 0.815]), and thus the coefficient of determination (r^2) is 0.6561. This indicates a strong correlation between the simulated pattern of land rent and the actual housing price. Besides, we use the t-test to assess the significance of the correlation coefficient:

$$t = \frac{r}{\sqrt{\frac{1-r^2}{N-2}}} \text{ with } df = N - 2 \tag{34}$$

where N is the number of observation pairs and df is the degree of freedom.

The value of t is 157.473 (with $df = 12{,}998$) $>$ t (0.000, 12,998). This indicates that the strong correlation is significant. Therefore, the proposed model can yield the simulation result which well fits to the actual one.

4. Conclusions

ABMs have become a prevalent method for modelling complex systems which are characterized as emergent. Emergent phenomena, such as residential patterns, are aggregate outcomes that cannot be predicted by examining the elements in isolation (Parker *et al.* 2003). Studies have indicated that these emergent phenomena can be practically modelled using ABMs. However, the definition of agents' behaviour is not an easy job when ABMs are implemented to model complex natural systems. This problem can be alleviated by well-defined theories, such as urban economic theory.

This research developed an ABM which incorporates urban economic theory for simulating complex residential dynamics. The incorporation of urban economic theory can ensure that agents' behaviour is defined on a solid theoretical basis and also allow the production of unexpected results because of the uncertainties of the ABM. One previous example can be seen in Webster and Wu (1999a, 1999b)'s studies. These authors used economic theory to define the transition rules of a CA to simulate urban growth. Although the theory is deterministic, there are additional unknowns due to the uncertainties introduced by the CA model. For instance, CA models are trend dependent, and thus it is hard to predict that the outcome of particular economic equilibrium will be in cellular space (Webster *et al*. 1999b). In this study, our simulation model can also yield some unexpected results compared with those classical patterns predicted by urban economic theory. One example is the simulation of sub-scenario 2 in scenario 1 (see Section 3.2).

First, we used an artificial city to examine the performance of the proposed model. When agents' preferences are set the same (scenario 1), the produced residential patterns have concentric structure and striped structure. The concentric residential structure is observed when all residents concentrate on the preference of commodity consumption or leisure time to achieve high residential utility. When the preference to residential size is set to be dominant, the residential pattern presents the striped structure. Yet in all the three subscenarios, the values of land rent consistently decline from the CBD to the urban fringe. In scenario 2, both the residential pattern and the land rent pattern differ from the results in scenario 1 significantly, on the condition that agents' preferences are varying in accordance with income level. We have taken roads, stations and rivers into account in scenario 3 with the purpose of representing real spatial heterogeneity. Compared with scenarios 1 and 2, scenario 3 has achieved broader residential extent because of the improved accessibility.

Second, we implemented this integrated model to simulate the residential dynamics in Guangzhou (scenario 4). The simulation result shows that high-income residents prefer the places with high environmental quality, while low-income residents live in the inner city to save commuting cost. Medium-income residents choose the trade-off places for dwelling by considering environmental quality and accessibility simultaneously. The Pearson's r was used to validate the similarity of patterns between the simulated land rent and the real housing price. The result of r is 0.81 ($r^2 = 0.6561$) that suggests strong correlation. Meanwhile the result of t-test also reveals that such correlation is significant. Therefore, the proposed model is capable of simulating residential patterns which are very close to those of the real world.

This article has demonstrated the potential of combining ABM with urban economic models for simulating residential dynamics. This integrated model can be used to verify the assumptions of urban theory because of its ability to reproduce the classic pattern of land rent, as the simulation shown in scenario 1. Moreover, our experiments have also indicated that the proposed model can serve as a useful exploratory tool for evaluating development policies in fast growing regions. For instance, the proposed model can be used to assess the impacts of new infrastructure on residential dynamics. In this context, more spatial constraints are needed, for example, the distance to the infrastructure or other corresponding attributes. Then the model is run under two scenarios: before and after the construction of the new infrastructure. Finally, the impacts of such projects can be assessed by comparing the simulation outcomes.

The objective of this article is to incorporate urban economic theory to define agents' behaviour. Further study may need to test the performance of this model for assisting policy making since this model has the capability to help spatial decision-making. Meanwhile the top-down planning process usually plays an important role in the resident spatial pattern except for the individual decision processes. To address this issue, it is necessary to introduce the agents representing planners/decision-makers and model the interactions between different stakeholders. For example, Ligtenberg *et al*. (2001) have proposed an ABM to model the process of land use planning involving multi-actors. We will consider that game theory (Parsons *et al*. 2002), which may do an interesting job explaining the interactions between different groups of agents (Epstein 1999), can be used for modelling urban residential development using ABMs in our future studies.

Acknowledgements

This study was supported by the National Basic Research Program of China (973 Program) (Grant No. 2011CB707103) and the Key National Natural Science Foundation of China (Grant No. 40830532). We thank the anonymous reviewers for their useful comments, especially the reviewer who mailed us the marked hard copy with the detailed suggestions.

References

An, L., *et al.,* 2005. Exploring complexity in a human–environment system: an agent-based spatial model for multidisciplinary and multiscale integration. *Annals of the Association of American Geographers*, 95 (1), 54–79.

Anas, A. and Kim, I., 1996. General equilibrium models of polycentric urban land use with endogenous congestion and job agglomeration. *Journal of Urban Economics*, 40 (2), 232–256.

Benenson, I., 1998. Multi-agent simulations of residential dynamics in the city. *Computers, Environment and Urban Systems*, 22 (1), 25–42.

Caruso, G., *et al.*, 2007. Spatial configurations in a periurban city. A cellular automata-based microeconomic model. *Regional Science and Urban Economics*, 37 (5), 542–567.

Clarke, K. and Gaydos, L., 1998. Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore. *International Journal of Geographical Information Science*, 12 (7), 699–714.

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

Fujita, M., 1989. *Urban economic theory: land use and city size*. Cambridge: Cambridge University Press.

Irwin, E. and Geoghegan, J., 2001. Theory, data, methods: developing spatially explicit economic models of land use change. *Agriculture, Ecosystems & Environment*, 85 (1–3), 7–24.

Jackson, J., Forest, B., and Sengupta, R., 2008. Agent based simulation of urban residential dynamics and land rent change in a gentrifying area of Boston. *Transactions in GIS*, 12 (4), 475–491.

Jankowski, P. and Richard, L., 1994. Integration of GIS-based suitability analysis and multicriteria evaluation in a spatial decision support system for route selection. *Environment and Planning B: Planning and Design*, 21 (3), 323–340.

Jepsen, M., *et al.*, 2006. Agent-based modelling of shifting cultivation field patterns, Vietnam. *International Journal of Geographical Information Science*, 20 (9), 1067–1085.

Jiang, H. and Eastman, J.R., 2000. Application of fuzzy measures in multi-criteria evaluation in GIS. *International Journal of Geographical Information Science*, 14 (2), 173–184.

Kii, M. and Doi, K., 2005. Multiagent land-use and transport model for the policy evaluation of a compact city. *Environment and Planning B: Planning and Design*, 32 (4), 485–504.

Krugman, P., 1991. Increasing returns and economic geography. *Journal of political economy*, 99 (3), 483–499.

Li, X. and Liu, X.P., 2007. Defining agents' behaviors to simulate complex residential development using multicriteria evaluation. *Journal of Environmental Management*, 85 (4), 1063–1075.

Li, X. and Liu, X.P., 2008. Embedding sustainable development strategies in agent-based models for use as a planning tool. *International Journal of Geographical Information Science*, 22 (1), 21–45.

Li, X. and 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 (2), 131–152.

Li, X. and Yeh, A.G.O., 2002. Urban simulation using principal components analysis and cellular automata for land-use planning. *Photogrammetric Engineering and Remote Sensing*, 68 (4), 341–351.

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

Li, X., *et al.*, 2011. Concepts, methodologies, and tools of an integrated geographical simulation and optimization system. *International Journal of Geographical Information Science*, 25 (4), 633–655.

Ligtenberg, A., Bregt, A., and Van Lammeren, R., 2001. Multi-actor-based land use modelling: spatial planning using agents. *Landscape and Urban Planning*, 56 (1–2), 21–33.

Lu, L., Wang, J., and Wei, Y., 2006. Spatial restructuring in Guangzhou City in the context of globalization and the new economy. *Acta Geographica Sinica*, 61 (8), 798–808.

Parker, D., *et al.*, 2003. Multi-agent systems for the simulation of land-use and land-cover change: a review. *Annals of the Association of American Geographers*, 93 (2), 314–337.

Parsons, S., Gmytrasiewicz, P.J., and Wooldridge, M.J., 2002. *Game theory and decision theory in agent-based systems*. Norwell, MA: Kluwer Academic.

Schelling, T., 1971. Dynamic models of segregation. *Journal of Mathematical Sociology*, 1 (2), 143–186.

Tobler, W., 1979. Cellular geography. *In*: S. Gale and G. Olssen, eds. *Philosophy in geography*. Dordrecht: Reidel Publishing, 379–386.

Webster, C. and Wu, F., 1999a. Regulation, land-use mix, and urban performance. Part 1: theory. *Environment and Planning A*, 31 (8), 1433–1442.

Webster, C. and Wu, F., 1999b. Regulation, land-use mix, and urban performance. Part 2: simulation. *Environment and Planning A*, 31 (9), 1529–1545.

Wu, F. and Webster, C., 1998. Simulation of land development through the integration of cellular automata and multicriteria evaluation. *Environment and Planning B: Planning and Design*, 25 (1), 103–126.

Xie, Y.C., Batty, M., and Zhao, K., 2007. Simulating emergent urban form using agent-based modeling: Desakota in the Suzhou-Wuxian region in China. *Annals of the Association of American Geographers*, 97 (3), 477–495.

Yin, L., 2009. The dynamics of residential segregation in Buffalo: an agent-based simulation. *Urban Studies*, 46 (13), 2749.