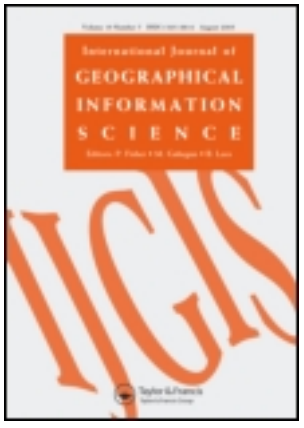


International Journal of Geographical Information Science

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

Modeling urban land-use dynamics in a fast developing city using the modified logistic cellular automaton with a patch-based simulation strategy

Yimin Chen[ab], Xia Li[ab], Xiaoping Liu[ab] & Bin Ai[bc]

[a] School of Geography and Planning, Sun Yat-sen University, Guangzhou, China

[b] Guangdong Key Laboratory for Urbanization and Geo-simulation, Sun Yat-sen University, Guangzhou, China

[c] School of Marine Sciences, Sun Yat-sen University, Guangzhou, China

Published online: 18 Sep 2013.

To cite this article: Yimin Chen, Xia Li, Xiaoping Liu & Bin Ai (2014) Modeling urban land-use dynamics in a fast developing city using the modified logistic cellular automaton with a patch-based simulation strategy, International Journal of Geographical Information Science, 28:2, 234-255, DOI: 10.1080/13658816.2013.831868

To link to this article: http://dx.doi.org/10.1080/13658816.2013.831868

International Journal of Geographical Information Science, 2014
Vol. 28, No. 2, 234–255, http://dx.doi.org/10.1080/13658816.2013.831868

Taylor & Francis
Taylor & Francis Group

Modeling urban land-use dynamics in a fast developing city using the modified logistic cellular automaton with a patch-based simulation strategy

Yimin Chen[a,b], Xia Li[a,b]*, Xiaoping Liu[a,b] and Bin Ai[b,c]

[a]*School of Geography and Planning, Sun Yat-sen University, Guangzhou, China;* [b]*Guangdong Key Laboratory for Urbanization and Geo-simulation, Sun Yat-sen University, Guangzhou, China;* [c]*School of Marine Sciences, Sun Yat-sen University, Guangzhou, China*

(*Received 26 March 2013; accepted 1 August 2013*)

Cellular automata (CA) have been used to understand the complexity and dynamics of cities. The logistic cellular automaton (Logistic-CA) is a popular urban CA model for simulating urban growth based on logistic regression. However, this model usually employs a cell-based simulation strategy without considering the spatial evolution of land-use patches. This drawback largely constrains the Logistic-CA for simulating realistic urban development. We proposed a Patch-Logistic-CA to deal with this problem by incorporating a patch-based simulation strategy into the conventional cell-based Logistic-CA. The Patch-Logistic-CA differentiates new developments into spontaneous growth and organic growth, and uses a moving-window approach to simulate the evolution of urban patches. The Patch-Logistic-CA is tested through the simulation of urban growth in Guangzhou, China, during 2005–2012. The cell-based Logistic-CA was also implemented using the same set of data to make a comparison. The simulation results reflect that the Patch-Logistic-CA has slightly lower cell-level agreement than the cell-based Logistic-CA. However, visual inspection of the results reveals that the cell-based Logistic-CA fails to reflect the actual patterns of urban growth, because this model can only simulate urbanized cells around the edges of initial urban patches. Actually, the pattern-level similarities of the Patch-Logistic-CA are over 18% higher than those of the cell-based Logistic-CA. This indicates that the Patch-Logistic-CA has much better performance of simulating actual development patterns than the cell-based Logistic-CA. In addition, the Patch-Logistic-CA can correctly simulate the fractal structure of actual urban development patterns. By varying the control parameters, the Patch-Logistic-CA can also be used to assist urban planning through the exploration of different development alternatives.

Keywords: cellular automata; Patch-Logistic-CA; urban simulation

1. Introduction

Cellular automata (CA), which originated from complexity sciences (Wolfram 1984), have been widely used for simulating urban phenomena. CA are attractive for urban modeling as they are simple and intrinsically spatial (White and Engelen 1993). Modifications and spatial constraints were also imposed by many researchers to make CA better for practical uses (White *et al*. 1997, Li and Yeh 2000). Various urban CA models were developed

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

over the last two decades for the simulation of realistic urban growth (White and Engelen 1997, Clarke and Gaydos 1998, Li and Yeh 2002, Wu 2002, Li and Yeh 2004, White *et al*. 2012). In these models, urban landscape is represented as raster space, in which a cell is the basic modeling unit for the simulation of land-use conversion. For a long time, accurate simulation of land-use conversion at cell level is considered fundamental for CA modeling. However, as actual land-use systems are full of stochastic processes and path-dependences, 'demanding that modelers get the locations right may be asking too much' (Brown *et al.* 2005). Moreover, models could be overfitting if they produce outcomes in which the locations of new developments are strictly consistent with the referenced land-use data (Brown *et al.* 2005).

From the perspective of investigating urban land-use systems, it is crucial for CA models to successfully replicate realistic spatial land-use patterns as well as predict the locations of new developments (Meentemeyer *et al.* 2013). Measures such as landscape metrics have been adopted to validate simulation models with respect to aggregate pattern similarity (Sui and Zeng 2001, Parker and Meretsky 2004, Liu *et al.* 2010). Some researchers even regard pattern similarity as the core criterion for calibrating CA models (e.g., Slope, Land use, Exclusion, Urban extent, Transportation, and Hillshade (SLEUTH)) (Clarke and Gaydos 1998, Silva and Clarke 2002, Dietzel and Clarke 2007). More recently, Li *et al* (2012) proposed a pattern-calibrated CA based on genetic algorithm to improve the goodness-of-fit between simulated and observed development patterns. However, most of the existing CA models, including SLEUTH (Clarke and Gaydos 1998) and the pattern-calibrated CA (Li *et al.* 2012), employ a cell-based simulation strategy without considering the spatial evolution of land-use patches (i.e., homogeneous cells that are spatially connected with each other). In this study, we take the logistic cellular automaton (Logistic-CA) (Wu 2002) as an example to illustrate the limitations of the cell-based simulation strategy for simulating actual urban development patterns. As we will demonstrate in Section 4.2, the cell-based Logistic-CA fails to produce realistic simulations of urban growth even though the model is well calibrated.

Logistic-CA is one of the popular CA models for simulating urban land-use dynamics (Fang *et al.* 2005, Han *et al.* 2009, Van Dessel *et al*. 2011, Jokar Arsanjani *et al.* 2013). The calibration of Logistic-CA is convenient based on logistic regression. Many improvements of the Logistic-CA focus on the derivation of more precise parameters through the use of advanced techniques, such as support vector machine (Yang *et al*. 2008), kernel-function (Liu *et al.* 2008), and genetic algorithm (Li *et al*. 2008). However, less attention is paid to the modification of the cell-based simulation strategy in the Logistic-CA. To address this issue, we proposed a new urban CA model, Patch-Logistic-CA, by incorporating a patch-based simulation strategy into the conventional Logistic-CA (referred to as cell-based Logistic-CA below). Recently, Meentemeyer *et al* (2013) also developed a patch-based urban CA model using the patch-growing algorithm (PGA). This method can allocate new developments around a fixed centroid to simulate the growth of urban patches. The patch-based simulation strategy we used is different from the algorithm proposed by Meentemeyer *et al* (2013). In this study, the patch-based simulation strategy differentiates new developments into spontaneous growth and organic growth, and adopts a moving-window approach to simulate the evolution of urban patches (see Section 2.2 for more details).

The proposed Patch-Logistic-CA was tested through the simulation of urban growth in Guangzhou, China, during 2005–2012. The cell-based Logistic-CA was also implemented with the same set of data to make a comparison. Additionally, the Patch-Logistic-CA was used to explore different development alternatives for Guangzhou from 2012 to 2020.

2. Patch-Logistic-CA

Figure 1 illustrates the flow of urban growth simulation using Patch-Logistic-CA. First, logistic regression is adopted to determine the weights of the input spatial variables for the subsequent calculation of development probability. Second, empirical land-use data are used to identify the actual new development patches and their size distribution. Third, new developments are iteratively simulated using the patch-based simulation strategy. This strategy simulates the growth of one new urban patch for each iteration. At the beginning of an iteration, the size of a new patch is estimated according to the observed size distribution

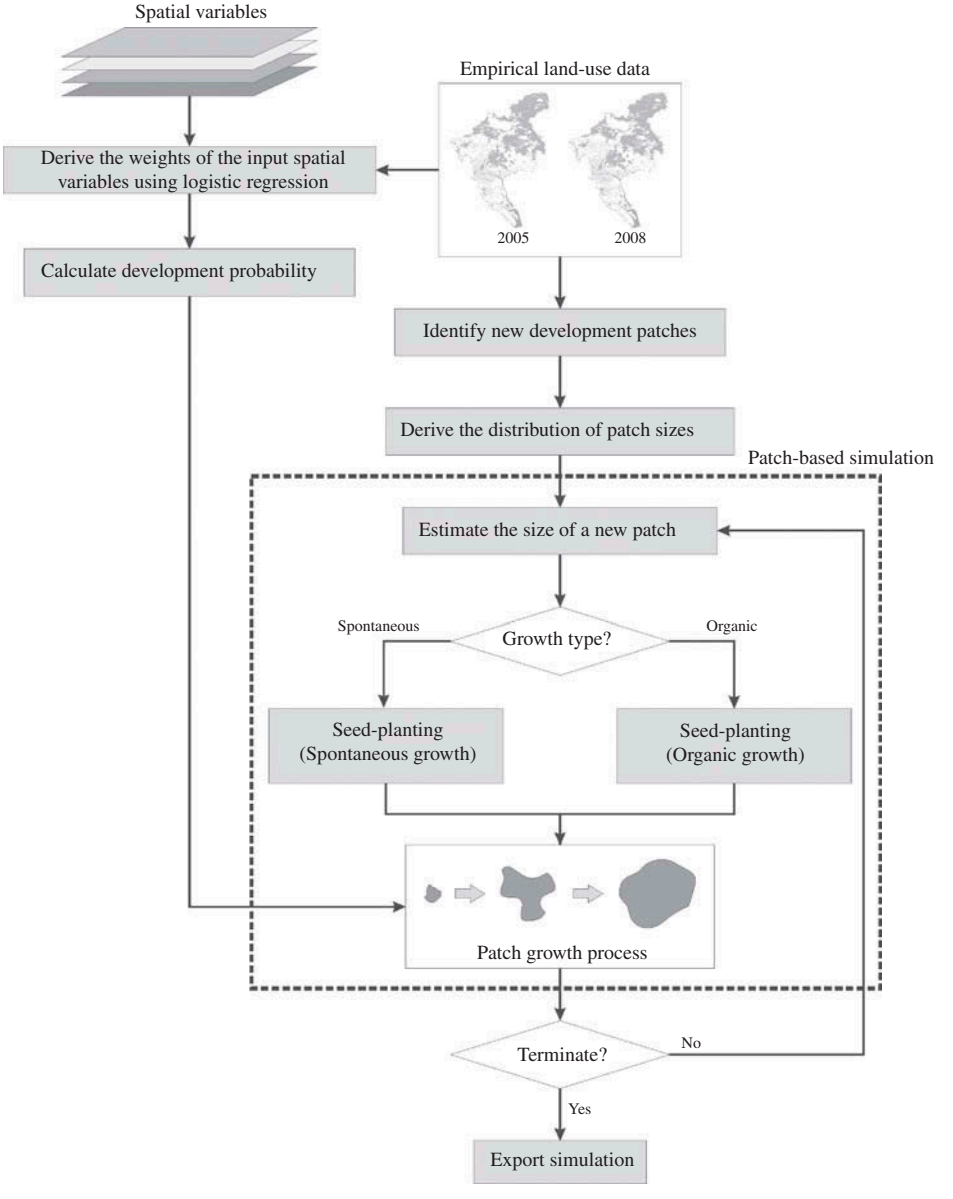


Figure 1. The methodology of the Patch-Logistic-CA.

derived in previous step. This patch is categorized either as spontaneous or organic growth depending on the comparison between a random pointer and a predefined threshold. Then, the patch growth is simulated based on a moving-window approach. Finally, the whole model terminates if the ending condition is satisfied (i.e., the quantity of simulated developments meets the given areal constraint); otherwise the model continues running. The following sections provide more detailed procedures.

2.1. Estimating development probability

The calculation of development probability considers three aspects: development potential, neighborhood development density, and suitability constraint. The development potential of a cell (i, j) is calculated based on a set of development factors, such as the proximity to urban centers or road networks. The development potential $p_{g,ij}$ of cell_{ij} is formulated in a logistic form as follows:

$$p_{g,ij} = \frac{\exp(z)}{1 + \exp(z)} = \frac{1}{1 + \exp(-z)} \tag{1}$$

where z is the combination score of development factors of cell_{ij}:

$$z = b_0 + \sum_k b_k x_k \tag{2}$$

where x_k is the development factors of cell_{ij}; b_0 is the intercept, and b_k are the weights of the development factors, which can be derived through logistic regression (Wu 2002, Li *et al*. 2008).

The development potential $p_{g,ij}$ only addresses the influences of static physical factors. Actual urban developments are also subject to the effects of dynamic factors. These effects are represented by the neighborhood development density:

$$\Omega_{ij}^t = \frac{\sum_{n^2} \text{con}(s_{ij} = \text{developed})}{n^2 - 1} \tag{3}$$

where Ω_{ij}^t is the development density within an $n \times n$ Moore neighborhood of cell_{ij} at time t; con() is a conditional function that returns 1 if the state of a cell within the neighborhood is developed.

Additionally, constraints are also included to represent the suitability for development. For instance, if a cell belongs to water, mountains or other types of restricted areas, then the suitability of this cell should be 0, which means that this cell will not be considered for development. Overall, the development probability of cell_{ij} at time t is calculated as follows:

$$p_{c,ij}^t = p_{g,ij} \Omega_{ij}^t \text{con}(s_{ij} = \text{suitable}) \tag{4}$$

where con() is a conditional function that returns 1 if cell_{ij} is suitable for development. As proposed by Wu (2002), a nonlinear transformation should be applied to $p^t{}_{c,ij}$ in order to promote the development probability in those cells with higher development potential:

$$p^t_{t,ij} = p^t_{c,ij} \exp[-\delta(1 - p^t_{c,ij}/p^t_{c,\max})] \quad (5)$$

where $p^t{}_{c,\max}$ is the maximum value of $p^t{}_{c,ij}$ at time t; and δ is called the dispersion parameter, ranging from 1 to 10.

Then, $p^t{}_{t,ij}$ is scaled into [0, 1] through the following equation:

$$p^t_{s,ij} = p^t_{t,ij} \Big/ \sum p_{t,i'j'} \quad (6)$$

where $p^t{}_{s,ij}$ is the development probability for simulating urban growth.

2.2. *Simulation of patch growth using a moving-window approach*

The size (the number of cells) of a new patch should be estimated before the simulation of its growth. A previous empirical study in the Pearl River Delta, which consists of Guangzhou and several other cities, reveals that the size distribution of urban patches in Guangzhou follows the power law (Fragkias and Seto 2009). Thus, we used a power function to estimate the size of a new patch

$$A_i = a_0(r_{\text{area}})^{a_1} \quad (7)$$

where A_i is the size of patch i; a_0 and a_1 are the parameters that can be determined according to the actual size distribution of new development patches. The meaning of the exponent a_1 is similar to that in the study by Fragkias and Seto (2009). Higher absolute values of a_1 represent a more uneven size distribution, while lower absolute values represent a more even size distribution (the value of zero suggests that all patches have equal size). The parameter r_{area} is a random value within (0, 1]. It represents the selection probability of a specific patch size from the actual size distribution.

After the patch size is estimated, the growth type of this patch is defined through a random pointer r_{type} ranging from 0 to 1. There are two possible growth types: spontaneous and organic. Figure 2 demonstrates the difference between these two types. Spontaneous growth refers to new developments that are disconnected from initial developments, while organic growth represents new developments that expand from initial developments. If r_{type} is less than a predefined threshold T_{spon}, then the growth type of this patch is categorized as spontaneous, otherwise the growth type is organic.

The seed planting is carried out after the determination of growth type. The implementation of seed planting is different between patches of spontaneous growth and those of organic growth. If a patch is considered as spontaneous growth, a seed is randomly planted

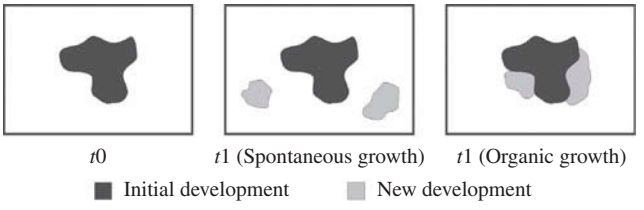


Figure 2. Spontaneous growth and organic growth.

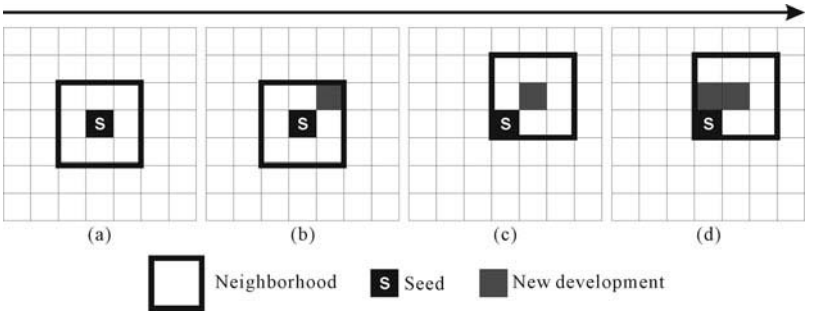


Figure 3. Simulation of patch growth based on a moving-window approach.

in the space but subject to the development potential of that location. Specifically, a non-urban cell is first randomly selected. If its development potential (Equation (1)) is greater than a random value within [0, 1], then this cell is developed as the seed of the patch. If the patch is of organic growth type, the development probability (Equation (6)) is used instead of the development potential during seed planting. This can force new development to only occur around the edge of initial urban patches as the neighborhood densities of these locations are always greater than 0.

The patch, either of spontaneous growth type or of organic growth type, starts its growth from the planted seed through a moving-window (Figure 3a). The size of the moving-window is set as 3×3 to ensure the contiguity of the simulated patch. Specifically, all non-urban cells in a 3×3 rectangle window (centered on the planted seed) are collected and sorted in ascending order according to their development probabilities (Equation (6)). The roulette wheel selection (Liu *et al.* 2012) is used to choose one of these non-urban cells for new development (Figure 3b). Then, the 3×3 window moves and centers on this developed cell (Figure 3c). The roulette wheel selection is again implemented to determine the next developed cell within the window (Figure 3d). This growth process continues until the number of new developed cells reaches the estimated patch size (Equation (7)), or the neighborhood is completely occupied by developed cells. The whole model will stop if the following condition is satisfied:

$$A = \sum_i A_i \tag{8}$$

where A is the observed or projected amount of new development (cell count).

2.3. A multilevel validation approach

The performance of the Patch-Logistic-CA is assessed from three aspects: cell-level agreement, pattern-level similarity, and fractal dimension. The cell-level agreement is evaluated through the indicator of 'Figure of merit' (Pontius *et al.* 2007). 'Figure of merit' is a ratio, where the numerator is the number of instances that are observed developed and correctly simulated as developed, while the denominator is the total number of instances excluding persistently non-changed instances. A higher value of 'Figure of merit' indicates a higher cell-level agreement. The indicator of 'Figure of merit' is calculated using the following equation (Pontius *et al.* 2008):

$$F = B/(A + B + C + D) \times 100\% \tag{9}$$

where F is the 'Figure of merit', A is the error due to observed developed and simulated as persistence, B is the agreement due to observed developed and simulated as developed, C is the error due to observed developed and simulated as incorrect gaining category, and D is the error due to observed persistence and simulated as developed. As the Patch-Logistic-CA only simulates the change of states from non-urban to urban, the value of C should be equal to 0.

The pattern-level similarity is estimated through the comparison between the simulated and observed landscape metrics. A total of four landscape metrics are selected to delineate the development patterns from different aspects (Dietzel *et al.* 2005, Seto and Fragkias 2005). These metrics are: (1) number of urban patches (NP) and largest-patch index (LPI), which are size metrics; (2) mean Euclidean nearest-neighbor distance (ENN_MN), which measures the distribution of patches; (3) mean perimeter-area ratio (PARA_MN), which measures the shape complexity of the patches. The landscape metrics were calculated using FRAGSTATS 4.1 (University of Massachusetts, Amherst) (McGarigal *et al.* 2012). The pattern-level similarity (a_l) is estimated as follows:

$$a_l = 1 - \frac{1}{8}\sum_i \Delta l_i \tag{10}$$

$$\Delta l_i = \begin{cases} \left|l_{i,s} - l_{i,o}\right|/l_{i,o} \times 100\%, l = NP, ENN\_MN, PARA\_MN \\ \left|l_{i,s} - l_{i,o}\right|, l = LPI \end{cases} \tag{11}$$

where $l_{i,s}$ and $l_{i,o}$ are the values of ith landscape metrics derived from the simulated pattern and the observed pattern, respectively; Δl_i is the normalized difference of the ith pair of simulated and observed landscape metrics. The Δl_i for LPI is calculated as the absolute differences because the original units of LPI are already percent.

The urban growth process can 'generate structures that grow outward from a nucleating center' (White 2006). Therefore, a city may consist of a large primary urban patch in the city center and small patches far from the center (White and Engelen 1993). Such a spatial structure can be measured using the indicator of fractal dimension (White and Engelen 1993, White *et al*. 2012). The fractal dimension is estimated through the following area–radius relationship:

$$\mathrm{Log}(A_i) = D_{\mathrm{frac}}\mathrm{Log}(r_i) + c \tag{12}$$

where A_i is the total number of urbanized cells in the ith zone with the distance of r_i from the city center, D_{frac} is the fractal dimension, and c is a constant.

3. Study area and data

The proposed Patch-Logistic-CA was tested through the simulation of urban growth in Guangzhou, China. Guangzhou is a fast developing city located in the Pearl River Delta. It is expected that the proposed model can provide a better understanding of the historical urban dynamics and assist urban planners to explore future development alternatives for Guangzhou.

Three scenes of Landsat Thematic Mapper (TM)/Enhanced Thematic Mapper Plus (ETM+) images (122–44) in 2005, 2008, and 2012 with a spatial resolution of 30 m were

used to obtain the land-use data for Guangzhou. The failure of the Landsat ETM+ scan-line corrector (SLC) has caused the serious problem of data gaps on the Landsat ETM+ images. In this study, the data gaps of the Landsat ETM+ images were repaired using a gap-filling product developed by Scaramuzza *et al.* (2004) and United States Geological Survey (USGS 2004). The land-use types include natural water, fishpond, forest, farmland, built-up area, and bare land. The land-use classification was implemented using the technique of object-based classification (Definiens Developer 7.0 2003). The land-use types of fishpond, farmland, and bare land were aggregated as 'non-urban' type; while natural water was reclassified as 'restricted area' since no development was allowed during simulation. In addition, the land-use type of forest was also considered as 'restricted area' because most of the forests locate on mountains and are protected by the planning bureau of Guangzhou.

Manual editing was proceeded for the land-use classifications based on concurrent fine-resolution remote-sensing images, such as SPOT (e.g., 284–304, 3 December 2007; 285–304, 4 January 2008; 286–303, 23 December 2006) and Google Earth images, to ensure the quality of land-use data for subsequent simulation. Google Earth provides time series of fine-resolution images (0.5 m) for the study area. These images are particularly useful for the correction of classification errors. For instance, the bright area in Figure 4a is usually classified as bare land by the object-based classification method because of the high reflectance. But based on the interpretation of the concurrent Google Earth image (Figure 4b), this area actually is a built-up area with several metal-roofed buildings that

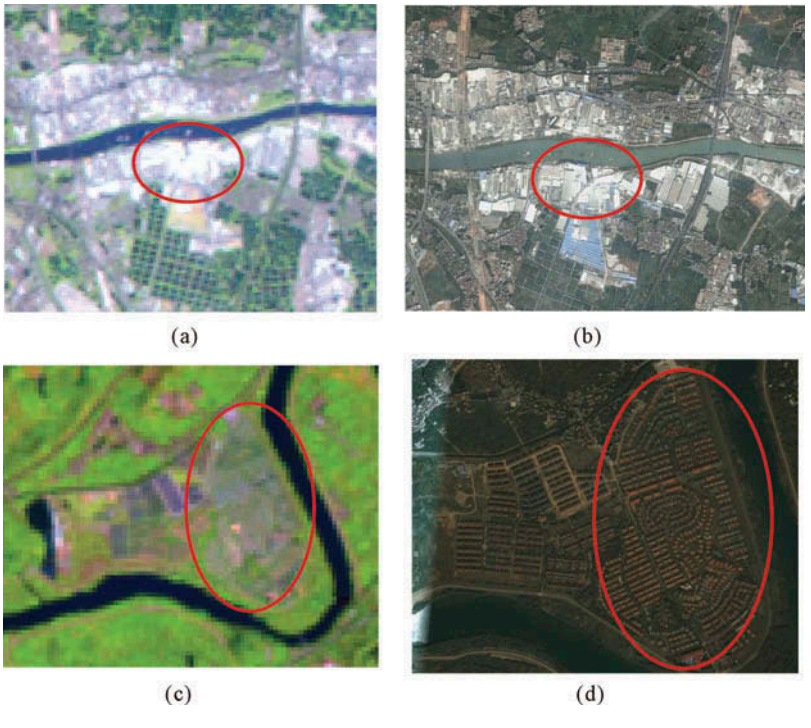


Figure 4. Manual interpretation using Google Earth images. (a) Landsat image (122/44, 23 November 2005). (b) Google Earth (21 November 2005). (c) Landsat image (122/44, 2 November 2012). (d) Google Earth (22 October 2012).

Table 1. Confusion matrices of the classification of urban areas.

		2005		2008		2012	
	Ref.	Urban	Non-urban	Urban	Non-urban	Urban	Non-urban
Cls.	Urban	195	13	280	9	323	9
	Non-urban	22	1770	14	1697	11	1657

Notes: Ref. = Reference; Cls. = Classification.

cause the high reflectance. Figure 4c and d provide another example of how Google Earth images support the correction of misclassifications. The highlighted area in the Landsat image (Figure 4c) exhibits similar spectral features to those of vegetation, and thus this area is frequently identified as farmland in the land-use classification. In fact, it is a misclassification because this area is a residential area with low development density and moderate vegetation coverage (Figure 4d). Overall, the data quality can be effectively improved through such manual interpretation of fine-resolution images. The classification accuracies of built-up area were assessed based on field survey. Table 1 shows the confusion matrices of the classification of built-up areas. The accuracies are 93.73%, 96.89%, and 97.29% for the classifications of 2005, 2008, and 2012, respectively.

The land-use data reveal a rapid urbanization process in Guangzhou during 2005–2012. In this period, the amount of new developments is 512.35 km^2, accounting for 68.36% of the total built-up area in 2005. The land-use data in 2005 and 2008 were used to calibrate the Patch-Logistic-CA, while the land-use data in 2012 were treated as the reference for validating the prediction accuracy of the proposed model.

The spatial variables for the calculation of urban development potential (Equation (1)) were prepared using a raster database with a 30-m resolution (Figure 5). They are: (1) variables representing the proximity to centers at different levels, including distance to the city center, distance to major business centers, distance to town and district centers, and distance to local business centers; (2) variables representing the proximity to transportation networks, including distance to expressways and avenues, distance to railways, and distance to major roads; and (3) slope, representing the physical condition of a location for development.

4. Results and discussion

4.1. Calibration

The calibration of the Patch-Logistic-CA follows four steps. First, Equation (7) was calibrated to estimate the patch size based on the actual size distribution of new development patches between 2005 and 2008. These patches were sorted in descending order according to their sizes (the number of cells). The frequencies and proportions were also obtained for each patch size. The cumulative proportion was then calculated, as shown by the blue area in Figure 6a. Equation (7) was used to fit the cumulative proportion and estimate the values of a_0 and a_1 for the period of 2005–2008 (Figure 6a). This procedure was also executed using the land-use data of years 2008 and 2012. The estimated values of a_0 and a_1 for this period are shown in Figure 6b.

Second, the weights of spatial variables were determined through logistic regression. A total of 20% of samples (including developed and non-developed cells) were randomly selected for logistic regression. A correlation analysis was first implemented for the spatial variables. The results indicate that there are several correlated variables (Table 2), which

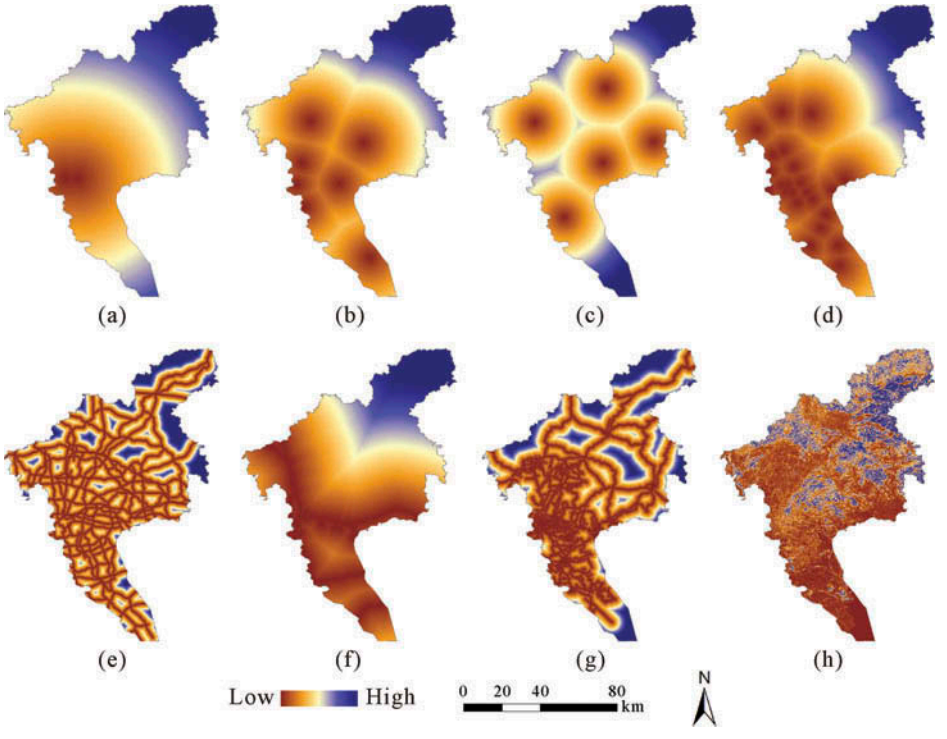
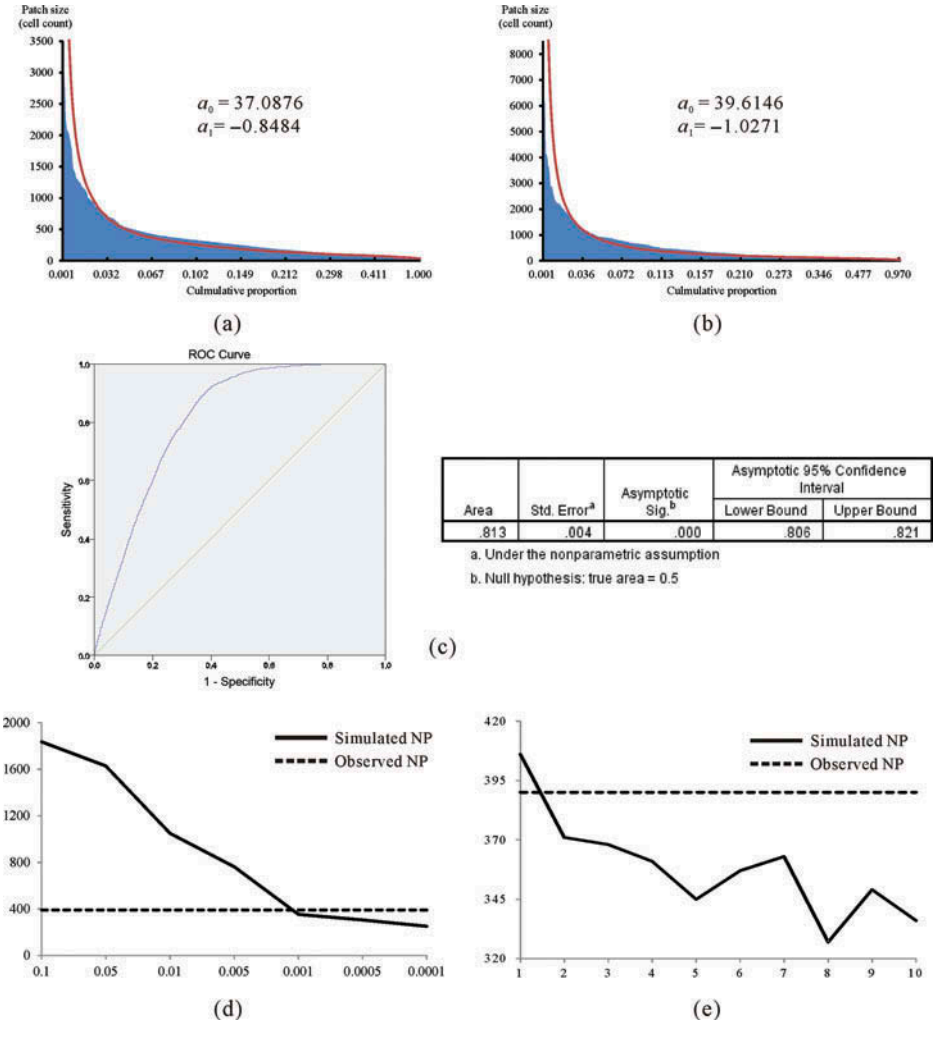


Figure 5. Spatial variables for the simulation of urban growth in Guangzhou: (a) distance to the city center; (b) distance to the major business center; (c) distance to town and district center; (d) distance to local business center; (e) distance to expressway and avenue; (f) distance to railway; (g) distance to road; (h) slope.

are redundant and may cause errors in subsequent analysis. Thus, we chose the stepwise (forward) method when performing the logistic regression so as to exclude those redundant variables (including distance to major business center, distance to local business center, and distance to expressway). The estimated weights for the rest of the spatial variables are listed in Table 3. The *P*-values for these weights are all 0.000. The Receiver operating characteristic (ROC) curve of the logistic regression and associated statistics are shown in Figure 6c.

Third, the configuration of neighborhood for Equation (3) was determined through a sensitivity analysis. We ran the model with five neighborhood configurations: 3×3, 5×5, 7×7, 9×9, and 11×11. Table 4 shows that the fragmentation of the simulated patterns increases as the neighborhood becomes larger, as indicated by the increased NP. The simulation most fits the observed pattern when the neighborhood is set as 3×3. Thus, this neighborhood configuration is selected for subsequent applications.

Finally, the values of T_{spon} and the dispersion parameter δ (Equation (5)) were defined through a trial-and-error approach. This was accomplished by running the model with different combinations of parameters and comparing the simulated development patterns with the observed ones. The threshold T_{spon} represents the tendency of dispersed development. Figure 6d shows the curves of the observed NP and the simulated NP with declining values of T_{spon} from 0.1 to 0.0001. It can be seen that the value of NP decreases if the value of T_{spon} reduces. These two curves intersect when T_{spon} almost reaches 0.001. This indicates that T_{spon} should be set with a value close to 0.001 so as to fit the observed development pattern.



Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.813	.004	.000	.806	.821

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Figure 6. (a) and (b) are observed size distributions (blue areas) of the new development patches and their fitting curves (red lines) for the periods of 2005–2008 and 2008–2012, respectively; (c) ROC curve and associated statistics of the logistic regression; (d) the simulated NP with declining values of T_{spon} from 0.1 to 0.0001; (e) the simulated NP with increasing values of δ from 1 to 10.

The parameter δ controls the shape of the development probability surface. A higher value of δ represents a steeper gradient of the development probability. As depicted by Figure 6e, the simulated NP decreases if the value of δ increases. The simulated NP most fits the observed NP when δ is in the interval of [1, 2] as these two curves intersect in this range. We further ran the model with refined values of T_{spon} and δ, and finally decided to use the values of 0.0013 and 1.5 for T_{spon} and δ, respectively (Table 3).

4.2. Results and discussion

The calibrated Patch-Logistic-CA was first tested through the simulation of urban growth in Guangzhou during 2005–2008, and further validated by predicting the urban development

Table 2. Correlations of the selected spatial variables.

	d_{city}	$d_{m\_business}$	d_{town}	$d_{l\_business}$	$d_{express}$	$d_{railway}$	d_{road}	Slope
d_{city}	1.00							
$d_{m\_business}$	0.53	1.00						
d_{town}	0.09	−0.02	1.00					
$d_{l\_business}$	0.51	0.42	0.01	1.00				
$d_{express}$	0.23	0.15	0.10	0.15	1.00			
$d_{railway}$	0.39	0.39	0.06	0.36	0.15	1.00		
d_{road}	0.29	0.20	0.17	0.25	0.21	0.16	1.00	
Slope	0.07	0.12	0.02	0.17	0.07	0.22	0.08	1.00

Notes: d_{city} = distance to the city center; $d_{m\_business}$ = distance to the major business center; d_{town} = distance to town and district center; $d_{l\_business}$ = distance to local business center; $d_{express}$ = distance to expressway and avenue; $d_{railway}$ = distance to railway; d_{road} = distance to road.

Table 3. The calibrated parameters of the Patch-Logistic-CA.

b_0	b_{city}	b_{town}	$b_{railway}$	b_{road}	b_{slope}	T_{spon}	δ
3.101	−4.799	−3.499	−0.892	−8.765	−4.722	0.0013	1.5

Note: P-values are all 0.000.

Table 4. Sensitivity analysis of different neighborhood configurations for the Patch-Logistic-CA.

	NP	LPI	PARA_MN	ENN_MN
Observed	390	8.80	424.30	392.04
3 × 3	371	8.08	418.96	622.59
5 × 5	421	8.05	395.22	628.37
7 × 7	495	7.88	369.56	779.67
9 × 9	483	7.93	368.53	716.87
11 × 11	470	7.96	374.61	747.30

Notes: $T_{spon} = 0.001$ and $\delta = 1.0$.

in 2012. The quantity of new development during 2005–2008 is a total of 239,178 cells (equivalent to 215.25 km^2), which is derived using Guangzhou's land-use data in 2005 and 2008. The actual and simulated development patterns are shown in Figure 7a and b.

We also ran the cell-based Logistic-CA, originally developed by Wu (2002), to make a comparison. The same set of data (Figure 5) and the calibrated parameters shown in Table 3 (except for T_{spon}) were used for the cell-based Logistic-CA. The calculation of development probability is the same as that of the Patch-Logistic-CA. Urban growth is then simulated using the cell-based simulation strategy (Wu 2002). Specifically, a non-urban cell is first randomly selected. The development probability $p^t{}_{s,ij}$ is then compared with a random value within the range of [0, 1]. If $p^t{}_{s,ij}$ is greater than this random value, the selected cell is converted to urban land-use; otherwise this cell remains unchanged. Figure 7c shows the simulated development pattern produced by the cell-based Logistic-CA.

As illustrated by Figure 7d, the actual urban development is a process of patch birth and its growth, either of spontaneous or organic types. The Patch-Logistic-CA successfully yields a development pattern that is similar to the actual one (Figure 7e). The cell-based Logistic-CA model, however, can only simulate new developed cells that are connected

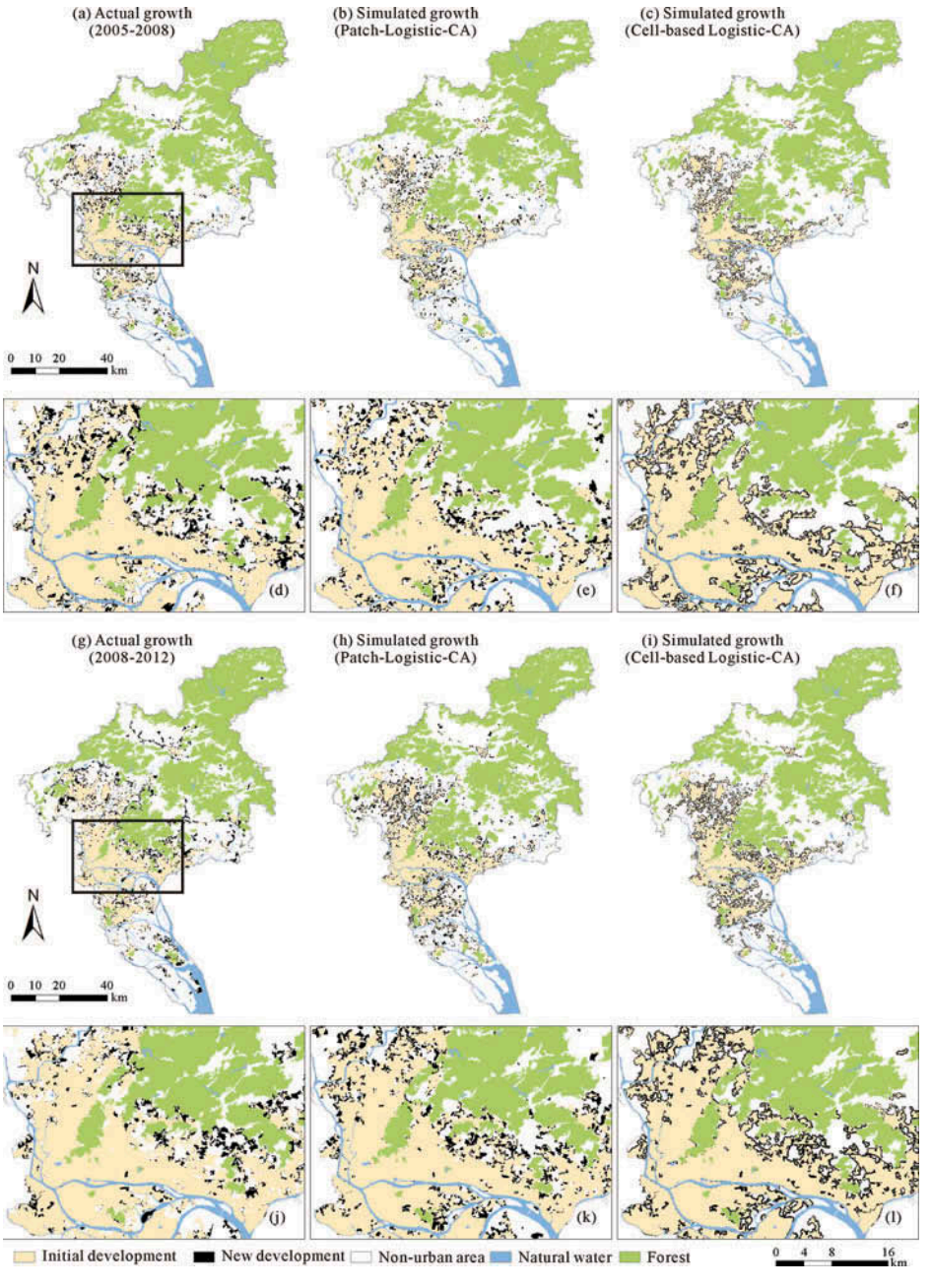


Figure 7. Comparison between the actual growth and the simulated growth during 2005–2012. (d)–(f) are the respective zoom-in views of (a)–(c). (j)–(l) are the respective zoom-in views of (g)–(i).

with initial urban cells. As a result, all simulated new developed cells distribute along the edges of initial urban patches (Figure 7f). Obviously, such simulated growth does not match the actual growth shown in Figure 7d.

The Patch-Logistic-CA produces better simulations than the cell-based Logistic-CA due to two reasons. First, the patch-based simulation strategy can better represent the

actual land development process. In the real world, urban development usually occurs in the units of parcels and land lots. These new development patches are usually composed of multiple cells if spatial data with a 30-m resolution are used. Thus, the model should reflect not only the state conversion of single cells but also the formation process of patches. This can be better fulfilled by using the patch-based simulation strategy. Second, the Patch-Logistic-CA can simulate spontaneous growth (Figure 7f) while the cell-based Logistic-CA cannot. García *et al.* (2012) also reported this drawback of the cell-based Logistic-CA in their comparative analysis involving several CA models. In fact, such an ability is especially important for urban modeling in China because many fast developing Chinese cities expand in the way of spontaneous growth.

Because CA are stochastic models with uncertainties, we ran the Patch-Logistic-CA for 100 times for validating its simulation accuracies. Such procedure was also implemented for the cell-based Logistic-CA. The spatial variations of these simulations are shown in Figure 8. The cell-level agreements of these two models were validated using the 'Figure of merit'. We overlaid the observed development pattern with the respective results of the Patch-Logistic-CA and the cell-based Logistic-CA to identify four groups of cells (i.e., persistent non-change, observed non-change simulated change, observed change simulated non-change, and observed change simulated change) for the calculation of 'Figure of merit' (Figure 9).

For the 100 outcomes of the Patch-Logistic-CA, the mean 'Figure of merit' ranges from 16.07% to 17.70%, with a mean value of 16.79%. For the results of the cell-based Logistic-CA, the 'Figure of merit' is 25.74–25.88%, with a mean value of 25.81% (9.02% higher than that of the Patch-Logistic-CA). The higher 'Figure of merit' of the cell-based Logistic-CA is owing to its cell-based simulation strategy. Such strategy can in effect evenly distribute the new developed cells around the edges of initial urban patches and

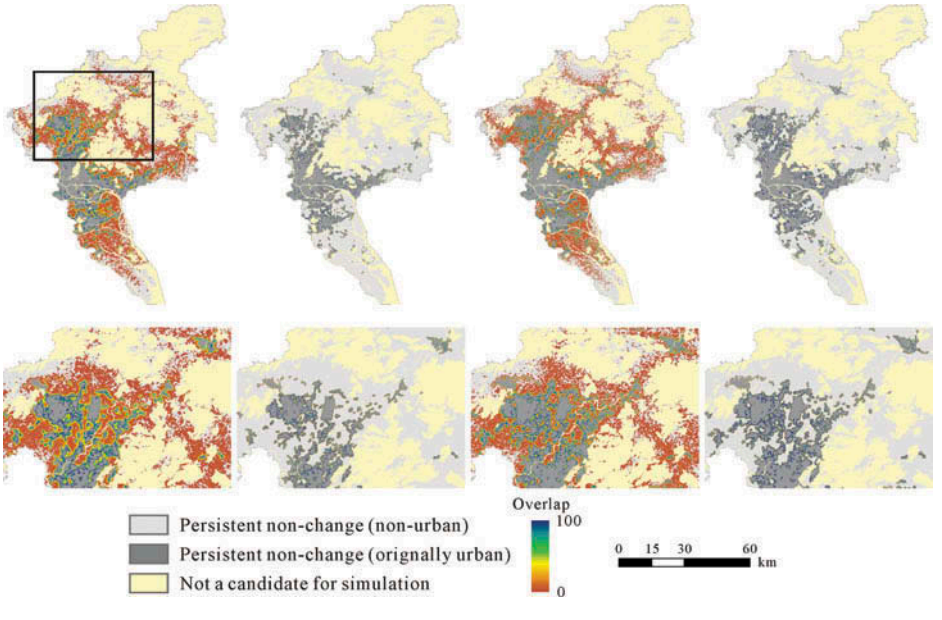


Figure 8. Spatial variations of the 100 simulations produced by Patch-Logistic-CA and cell-based Logistic-CA. (a) Patch-Logistic-CA (Simulation 2008). (b) Cell-based Logistic-CA (Simulation 2008). (c) Patch-Logistic-CA (Simulation 2012). (d) Cell-based Logistic-CA (Simulation 20012).

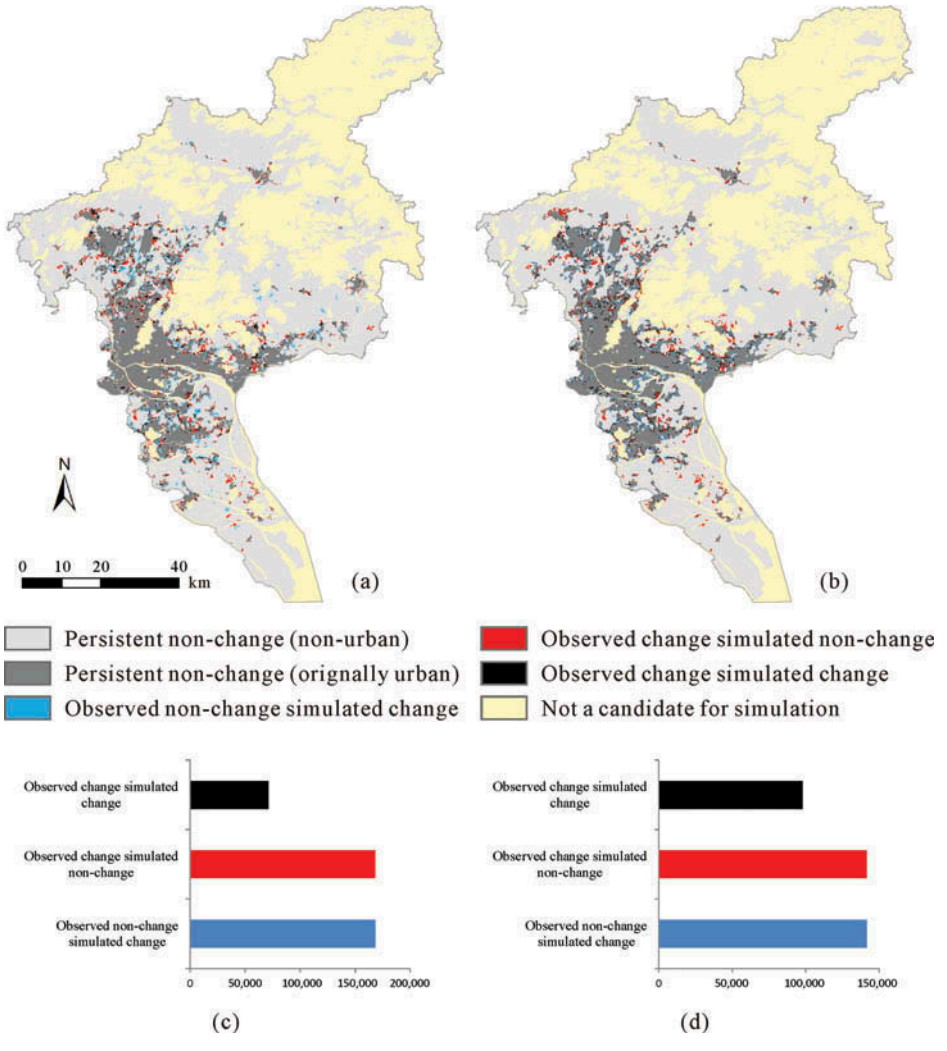


Figure 9. Spatial distribution of successes and errors of the simulations in 2008 produced by (a) Patch-Logistic-CA and (b) cell-based Logistic-CA. (c)–(d) Mean quantities of successes and errors for Patch-Logistic-CA and cell-based Logistic-CA, respectively.

hence increase the possibility of hits (i.e., cells observed change and simulated change). The Patch-Logistic-CA, however, adopts a patch-based simulation strategy which allocates new developed cells less evenly in the space. Consequently, the possibility of hits reduces while the risk of cell-level disagreement increases.

Table 5 shows the validation of pattern-level similarity for the simulations in 2008. The results of Patch-Logistic-CA have a mean pattern-level similarity of 79.66% to actual development patterns, while those of the cell-based Logistic-CA have only 58.32%. These results indicate that the Patch-Logistic-CA has better performance of replicating actual development patterns than the cell-based Logistic-CA. For these two models, the largest disagreements are in the simulated ENN_MN. However, the cell-based Logistic-CA also has very large errors in the simulated NP and PARA_MN. As aforementioned, the

Table 5. Comparison between observed and simulated values of landscape metrics for year 2008.

	NP	LPI	PARA_MN	ENN_MN
Observed	390	8.80	424.30	392.04
Patch-Logistic-CA (pattern similarity: 75.65–84.26%)				
Minimum	401	7.61	380.11	560.19
Median	423	7.92	391.15	654.11
Maximum	446	8.29	412.88	746.77
Cell-based Logistic-CA (pattern similarity: 56.30–60.70%)				
Minimum	166	8.13	306.51	704.40
Median	169	8.18	320.39	714.83
Maximum	173	8.48	333.65	753.02

cell-based Logistic-CA can only change the state from non-urban into urban for those cells that are connected with the edges of the initial urban patches. Then these new developed cells might bridge those initially separating urban patches during the simulation process. As a result, the fragmentation and irregularity of the pattern decrease, as indicated by the lower NP and PARA_MN.

The prediction performance of the Patch-Logistic-CA was validated through the simulation of urban growth from 2008 to 2012 (Figures 7h and k, 10a and c) using the a_0 and a_1 shown in Figure 4b and the calibrated parameters listed in Table 3. The cell-based Logistic-CA was also run in a similar way to make a comparison (Figures 7i and l, 10b and d). The values of 'Figure of merit' is 12.23–13.81% (mean = 14.24%) for the Patch-Logistic-CA, and 16.92–17.06% for the cell-based Logistic-CA (mean = 17.00%). The difference between the mean 'Figure of merit' of these two models is much lower compared with that in the simulations of year 2008. The pattern-level similarity of the Patch-Logistic-CA, as shown in Table 6, is still over 18% higher than that of the cell-based Logistic-CA. Such results reflect the better performance of the Patch-Logistic-CA for predicting development patterns.

The results of the Patch-Logistic-CA were also validated by comparing the simulated and observed fractal dimensions. A set of radii ranging from 5 km to 60 km (in 5 km increments) from the city center were used to estimate the area–radius relationships and obtain the fractal dimensions. As shown by Figure 11, the area–radius relationships are kinked, with steep slopes for the inner zone and flat ones for the outer zone. These kinked relationships are similar to those in the cases of White and Engelen's (1993) study. It can be found that the fractal dimensions of the simulated patterns are close to those of the actual patterns. This reflects that the proposed Patch-Logistic-CA can capture the fractal features of actual urban systems.

The experimental results shown above reveal a advantage of the Patch-Logistic-CA: the ability to produce multiple simulations given the same set of parameters. This is indicated by comparing the spatial variations of the simulation results of these two models (Figure 8). A similar result can also be found by comparing the variations of 'Figure of merit' and pattern-level similarity for the Patch-Logistic-CA and the cell-based Logistic-CA (Tables 5 and 6). The ability to produce multiple simulations is also referred to as the ability of 'always getting different things right' by Brown *et al* (2005). In fact, such ability is important as it allows policy makers to explore possible development paths and evaluate the influences of human intervention to initial development process.

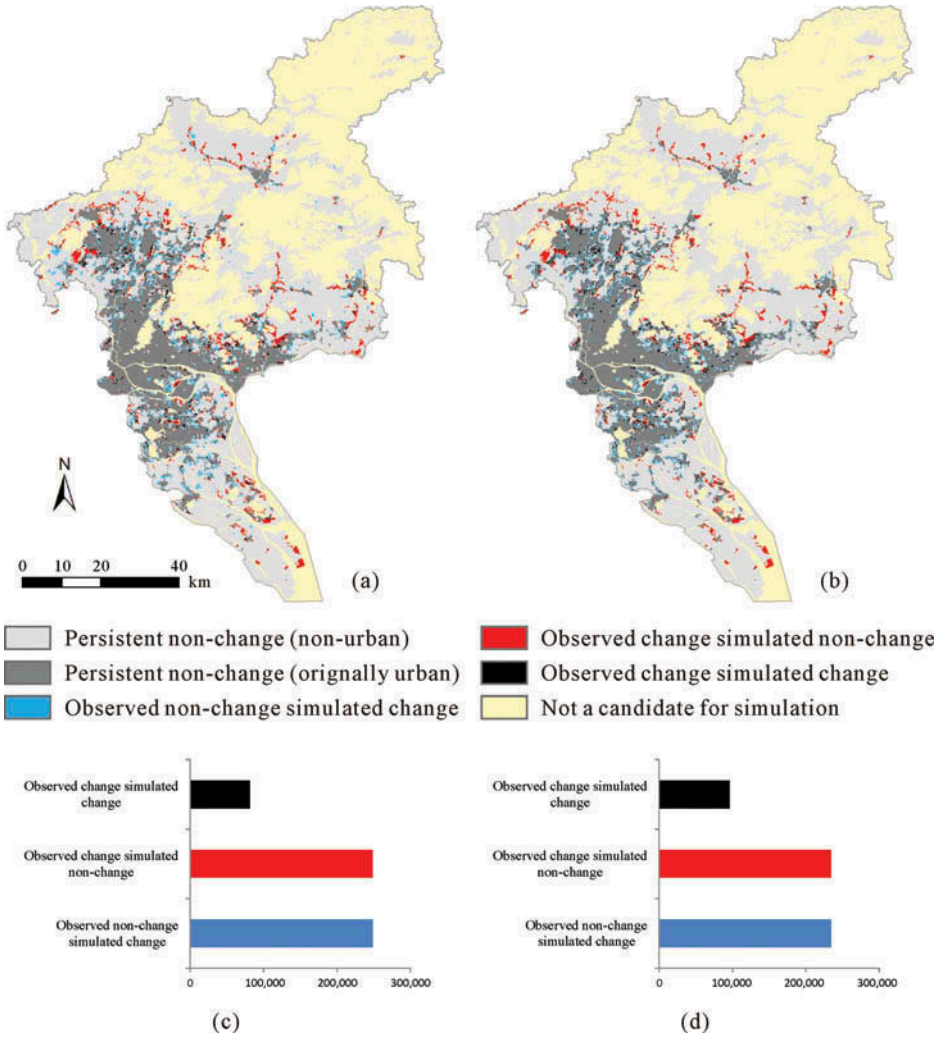


Figure 10. Spatial distribution of successes and errors of the predictions in 2012 produced by (a) Patch-Logistic-CA and (b) cell-based Logistic-CA. (c)–(d) Mean quantities of successes and errors for Patch-Logistic-CA and cell-based Logistic-CA, respectively.

4.3. Scenario simulation of future development alternatives

In addition to the simulation of realistic urban growth, the proposed Patch-Logistic-CA was also used to explore different development alternatives for Guangzhou from 2012 to 2020. As demonstrated by Figure 6a and b, the parameters a_0 and a_1 can vary over time. Because we focus more on the model's performance in scenario simulation, for simplicity, we used a linear extrapolation approach to predict the values of a_0 and a_1 from 2012 to 2020 (Figure 12). As a result, $a_0 = 42.1416$ and $a_1 = -1.2058$ for the period of 2012–2016; while $a_0 = 44.6686$ and $a_1 = -1.3845$ for the period of 2016–2020. We also assumed that the land demand from 2012 to 2020 follows the same growth rate in the period of 2005–2012 (i.e., 73.19 km^2 per year).

Table 6. Comparison between observed and predicted values of landscape metrics for year 2012.

	NP	LPI	PARA_MN	ENN_MN
Observed	436	12.65	385.03	436.50
Patch-Logistic-CA (pattern similarity: 72.89–80.35%)				
Minimum	318	13.26	327.26	647.61
Median	336	13.74	351.68	728.91
Maximum	372	14.07	365.11	773.65
Cell-based Logistic-CA (pattern similarity: 54.41–57.96%)				
Minimum	177	12.10	213.85	752.40
Median	182	12.11	232.23	757.45
Maximum	186	12.17	243.61	779.83

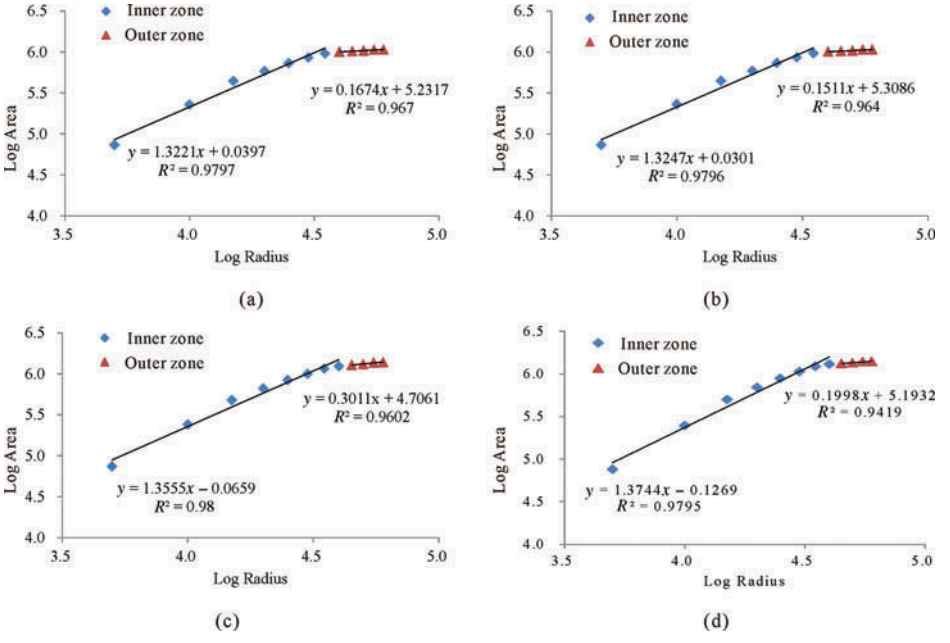


Figure 11. The actual and simulated area–radius relationships. (a) and (c) are actual development patterns in 2008 and 2012, respectively; (b) and (d) are simulations in 2008 and 2012, respectively.

Three scenarios were designed and simulated according to different assumptions: (1) *Diffusion scenario*. In this scenario, the parameter T_{spon} was set as 0.013 (ten times of the calibrated T_{spon} shown in Table 3) to represent a loose control on the spatial distribution of new development. (2) *Business-as-usual* (BAU). This scenario assumes that the city follows the historical development path during the period of 2008–2020. Therefore, the calibrated T_{spon} was used for this simulation. (3) *Coalescence scenario*. In this scenario, the parameter T_{spon} was set as 0.00013 to represent the planning objective of enhancing the spatial connections between initial developments and new developments.

The results of these three scenarios are shown in Figure 13a–f. Landscape metrics (Table 7) were calculated to illustrate the differences among the simulated patterns. Sprawl

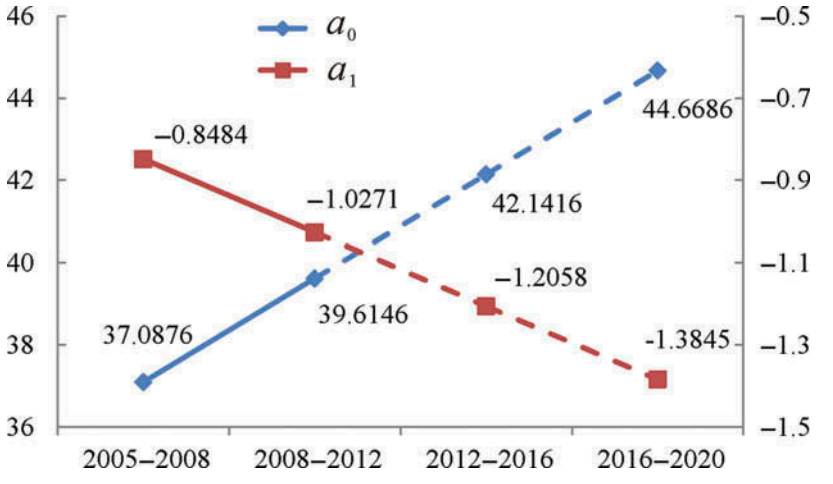


Figure 12. The predicted a_0 and a_1 for the periods of 2012–2016 and 2016–2020.

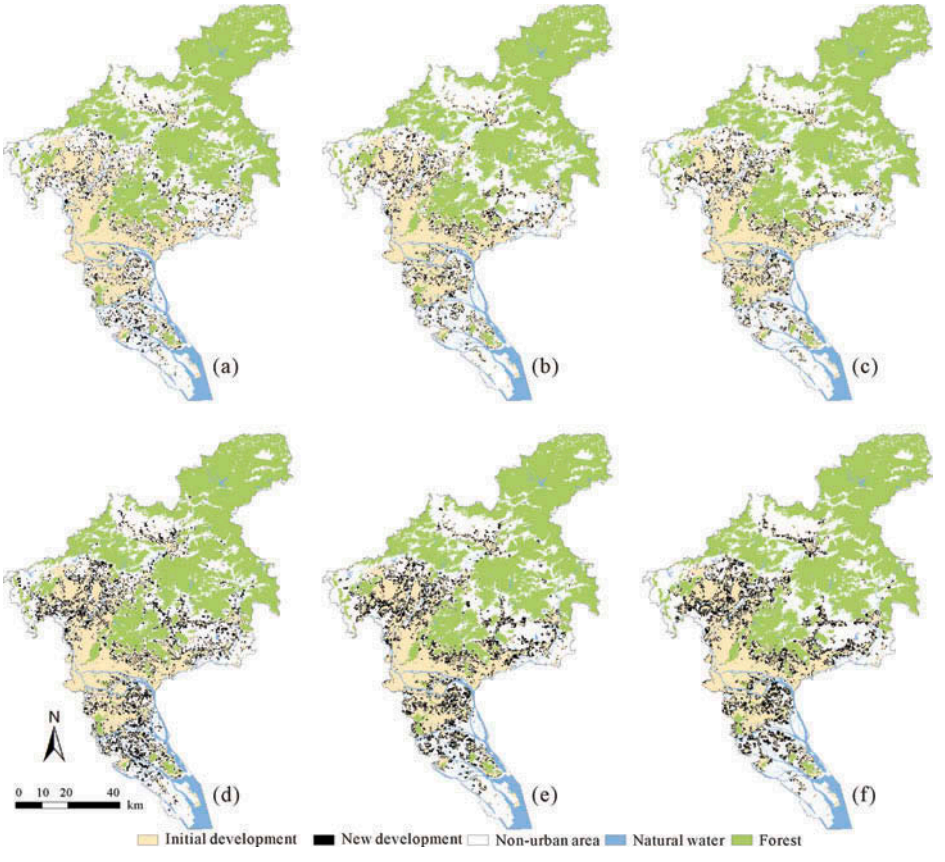


Figure 13. Scenario simulation of urban growth in Guangzhou from 2012 to 2020. (a) and (c) are diffusion scenarios for years 2016 and 2020, respectively; (b) and (d) are business-as-usual scenarios for years 2016 and 2020, respectively; (e) and (f) are coalescence scenarios for years 2016 and 2020, respectively.

Table 7. Comparison of scenario simulations in terms of landscape metrics.

	Year	NP	LPI	PARA_MN	ENN_MN
Diffusion scenario	2016	747	15.89	292.59	490.60
	2020	815	18.90	269.48	424.16
Business-as-usual	2016	344	16.64	361.85	584.67
	2020	336	20.05	322.04	618.30
Coalescence scenario	2016	272	16.97	431.67	599.75
	2020	215	20.11	404.63	714.97

patterns are observed in the diffusion scenario because of the weak spatial restriction on development. Thus, this scenario has the largest values of NP compared with other scenarios. The values of ENN_MN are the lowest in this scenario. This is because new urban patches tend to appear near existing patches or in the gaps between them. As a result, the closeness between urban patches increases. In the coalescence scenario, new developments tend to connect with those initial urban patches. This simulates a typical coalescence process observed in actual urban developments (Dietzel *et al.* 2005). The values of LPI are the highest among all scenarios. Moreover, the average shape complexity increases as the urban patches grow and merge together. This is reflected by the highest values of PARA_MN in this scenario. Overall, the proposed Patch-Logistic-CA can generate the development patterns as expected.

5. Conclusions

This study has demonstrated that the conventional cell-based Logistic-CA have limitations for simulating realistic urban growth. This model can only simulate new developed cells that are connected with initial developments (Figure 7f and l). Such simulations are unrealistic compared with actual urban developments (Figure 7d and g). To address this issue, we developed a Patch-Logistic-CA by incorporating a patch-based simulation strategy into the conventional cell-based Logistic-CA.

The proposed Patch-Logistic-CA was applied to the simulation of urban growth in Guangzhou during 2005–2012. The results show that although the Patch-Logistic-CA has lower cell-level agreements, it can generate more reliable development patterns than the cell-based Logistic-CA (Figure 7, Tables 5 and 6). The Patch-Logistic-CA also can correctly simulate the fractal structure of actual urban development patterns (Figure 11). By varying the control parameters, this model can support the 'what-if' experiments. This is useful for urban planners to evaluate the potential impacts of different urban development strategies (Figure 13 and Table 7).

In this study, we discussed the performance of the Patch-Logistic-CA with the 30-m spatial data. This is actually a scale in which many existing urban CA models are applied due to the convenience in deriving land-use data from the popular Landsat TM/ETM+ images (Liu *et al.* 2008, Santé *et al.* 2010, Li *et al.* 2012). The change of scale can have substantial impacts on CA-based urban simulations (Samat 2006, Ménard and Marceau 2012). We will further examine the scale sensitivity of the proposed model in future work. Another limitation of this model is that it cannot simulate the development of multiple patches at the same time, which may cause simulation biases. We will modify this in the

future by adding a model component to represent the connections and interactions among patches. Additionally, the cell-level agreement of the proposed model needs to be improved from the perspective of practical uses. Perhaps extra spatial constraints should be included into the model for reducing the randomness of the location of simulated developments. But the amount of spatial constraints should be decided with caution so that the problem of overfitting can be avoided (Brown *et al.* 2005).

Acknowledgements

We thank the anonymous reviewers for their useful comments and suggestions.

Funding

This research was supported by the National Basic Research Program of China (973 Program) [grant number 2011CB707103], the National Natural Science Foundation of China [grant numbers 41171308 and 41171308], and Foundation for the Author of National Excellent Doctoral Dissertation of P. R. China (3149001).

References

Brown, D.G., *et al*., 2005. Path dependence and the validation of agent-based spatial models of land use. *International Journal of Geographical Information Science*, 19 (2), 153–174.

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.

Definiens Developer 7.0, 2003. *Definiens Developer Version 7* [online]. http://www.ecognition.com/document/definiens-developer-version-7 [Accessed 8 June 2012].

Dietzel, C. and Clarke, K., 2007. Toward optimal calibration of the SLEUTH land use change model. *Transactions in GIS*, 11 (1), 29.

Dietzel, C., *et al*., 2005. Diffusion and coalescence of the Houston Metropolitan Area: evidence supporting a new urban theory. *Environment and Planning B: Planning and Design*, 32 (2), 231–246.

Fang, S., *et al*., 2005. The impact of interactions in spatial simulation of the dynamics of urban sprawl. *Landscape and Urban Planning*, 73 (4), 294–306.

Fragkias, M. and Seto, K., 2009. Evolving rank-size distributions of intra-metropolitan urban clusters in South China. *Computers, Environment and Urban Systems*, 33 (3), 189–199.

García, A.M., *et al*., 2012. A comparative analysis of cellular automata models for simulation of small urban areas in Galicia, NW Spain. *Computers, Environment and Urban Systems*, 36 (4), 291–301.

Han, J., *et al*., 2009. Application of an integrated system dynamics and cellular automata model for urban growth assessment: a case study of Shanghai, China. *Landscape and Urban Planning*, 91 (3), 133–141.

Jokar Arsanjani, J., *et al*., 2013. Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion. *International Journal of Applied Earth Observation and Geoinformation*, 21, 265–275.

Li, X., *et al*., 2012. Calibrating cellular automata based on landscape metrics by using genetic algorithms. *International Journal of Geographical Information Science*, 27 (3), 594–613.

Li, X., Yang, Q.S., and Liu, X.P., 2008. Discovering and evaluating urban signatures for simulating compact development using cellular automata. *Landscape and Urban Planning*, 86 (2), 177–186.

Li, X. and Yeh, A.G., 2002. Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographical Information Science*, 16 (4), 323–343.

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., 2004. Data mining of cellular automata's transition rules. *International Journal of Geographical Information Science*, 18 (8), 723–744.

Liu, X., *et al*., 2012. An integrated approach of remote sensing, GIS and swarm intelligence for zoning protected ecological areas. *Landscape Ecology*, 27 (3), 447–463.
Liu, X., *et al*., 2008. Simulating complex urban development using kernel-based non-linear cellular automata. *Ecological Modelling*, 211 (1–2), 169–181.
Liu, X., *et al*., 2010. Simulating land-use dynamics under planning policies by integrating artificial immune systems with cellular automata. *International Journal of Geographical Information Science*, 24 (5), 783–802.
Ménard, A. and Marceau, D.J., 2012. Exploration of spatial scale sensitivity in geographic cellular automata. *Environment and Planning B: Planning and Design*, 32 (5), 693–714.
McGarigal, K., *et al*., 2012. *FRAGSTATS v4: spatial pattern analysis program for categorical and continuous maps. Computer software program produced by the authors at the University of Massachusetts, Amherst*. [online]. Available from: http://www.umass.edu/landeco/research/fragstats/fragstats.html [Accessed 2 February 2013].
Meentemeyer, R.K., *et al*., 2013. FUTURES: multilevel simulations of emerging urban–rural landscape structure using a stochastic patch-growing algorithm. *Annals of the Association of American Geographers*, 103 (4), 785–807.
Parker, D. and Meretsky, V., 2004. Measuring pattern outcomes in an agent-based model of edge-effect externalities using spatial metrics. *Agriculture, Ecosystems & Environment*, 101 (2–3), 233–250.
Pontius, R., *et al*., 2007. Accuracy assessment for a simulation model of Amazonian deforestation. *Annals of the Association of American Geographers*, 97 (4), 677–695.
Pontius, R., *et al*., 2008. Comparing the input, output, and validation maps for several models of land change. *The Annals of Regional Science*, 42 (1), 11–37.
Samat, N., 2006. Characterizing the scale sensitivity of the cellular automata simulated urban growth: a case study of the Seberang Perai Region, Penang State, Malaysia. *Computers, Environment and Urban Systems*, 30 (6), 905–920.
Santé, I., *et al*., 2010. Cellular automata models for the simulation of real-world urban processes: a review and analysis. *Landscape and Urban Planning*, 96 (2), 108–122.
Scaramuzza, P., Micijevic, E., and Chander, G., 2004. SLC gap-filled products: phase one methodology [online]. Available from: https://landsat.usgs.gov/documents/SLC_Gap_Fill_Methodology.pdf [Accessed 20 March 2013].
Seto, K. and Fragkias, M., 2005. Quantifying spatiotemporal patterns of urban land-use change in four cities of China with time series landscape metrics. *Landscape Ecology*, 20 (7), 871–888.
Silva, E. and Clarke, K., 2002. Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. *Computers, Environment and Urban Systems*, 26 (6), 525–552.
Sui, D. and Zeng, H., 2001. Modeling the dynamics of landscape structure in Asia's emerging desakota regions: a case study in Shenzhen. *Landscape and Urban Planning*, 53 (1–4), 37–52.
USGS, 2004. *Phase 2 gap-fill algorithm: SLC-off gap-filled products gap-fill algorithm methodology* [online]. Available form: http://www.ga.gov.au/image_cache/GA4861.pdf [Accessed 20 March 2013].
Van Dessel, W., Van Rompaey, A., and Szilassi, P., 2011. Sensitivity analysis of logistic regression parameterization for land use and land cover probability estimation. *International Journal of Geographical Information Science*, 25 (3), 489–508.
White, R., 2006. Pattern based map comparisons. *Journal of Geographical Systems*, 8 (2), 145–164.
White, R. and 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, 1175–1175.
White, R. and Engelen, G., 1997. Cellular automata as the basis of integrated dynamic regional modelling. *Environment and Planning B*, 24, 235–246.
White, R., Engelen, G., and Uljee, I., 1997. The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. *Environment and Planning B*, 24, 323–344.
White, R., Uljee, I., and Engelen, G., 2012. Integrated modelling of population, employment and land-use change with a multiple activity-based variable grid cellular automaton. *International Journal of Geographical Information Science*, 26 (7), 1251–1280.
Wolfram, S., 1984. Cellular automata as models of complexity. *Nature*, 311 (4), 419–424.
Wu, F.L., 2002. Calibration of stochastic cellular automata: the application to rural-urban land conversions. *International Journal of Geographical Information Science*, 16 (8), 795–818.
Yang, Q.S., Li, X., and Shi, X., 2008. Cellular automata for simulating land use changes based on support vector machines. *Computers & Geosciences*, 34 (6), 592–602.