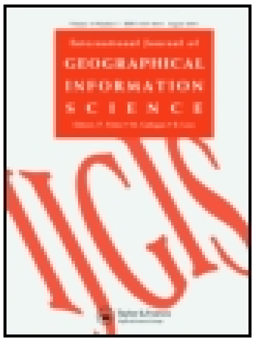

International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: http://www.tandfonline.com/loi/tgis20

Simulating urban land-use changes at a large scale by integrating dynamic land parcel subdivision and vector-based cellular automata

Yao Yao , Xiaoping Liu , Xia Li , Penghua Liu , Ye Hong , Yatao Zhang & Ke Mai

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE, 2017
https://doi.org/10.1080/13658816.2017.1360494

Taylor & Francis
Taylor & Francis Group

ARTICLE

Check for updates

Simulating urban land-use changes at a large scale by integrating dynamic land parcel subdivision and vector-based cellular automata

Yao Yao [a], Xiaoping Liu [a], Xia Li [a], Penghua Liu [a], Ye Hong [b], Yatao Zhang [c] and Ke Mai [c]

[a]School of Geography and Planning, Guangdong Key Laboratory for Urbanization and Geo-simulation, Sun Yat-sen University, Guangzhou, China; [b]School of Geography and Planning, Sun Yat-sen University, Guangzhou, China; [c]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China

ABSTRACT
Cellular automata (CA) have been widely used to simulate complex urban development processes. Previous studies indicated that vector-based cellular automata (VCA) could be applied to simulate urban land-use changes at a realistic land parcel level. Because of the complexity of VCA, these studies were conducted at small scales or did not adequately consider the highly fragmented processes of urban development. This study aims to build an effective framework called dynamic land parcel subdivision (DLPS)-VCA to accurately simulate urban land-use change processes at the land parcel level. We introduce this model in urban land-use change simulations to reasonably divide land parcels and introduce a random forest algorithm (RFA) model to explore the transition rules of urban land-use changes. Finally, we simulate the land-use changes in Shenzhen between 2009 and 2014 via the proposed DLPS-VCA model. Compared to the advanced Patch-CA and RFA-VCA models, the DLPS-VCA model achieves the highest simulation accuracy (Figure-of-Merit = 0.232), which is 32.57% and 18.97% higher respectively, and is most similar to the actual land-use scenario (similarity = 94.73%) at the pattern level. These results indicate that the DLPS-VCA model can both accurately split the land during urban land-use changes and significantly simulate urban expansion and urban land-use changes at a fine scale. Furthermore, the land-use change rules that are based on DPLS-VCA mining and the simulation results of several future urban development scenarios can act as guides for future urban planning policy formulation.

ARTICLE HISTORY
Received 22 February 2017
Accepted 24 July 2017

KEYWORDS
Vector cellular automata; large scale; urban land-use change; land parcel; urban planning

1. Introduction

Land use is an important issue to maintain the long-term stability of immediate human needs and ecosystems (Foley *et al.* 2005) and plays an essential role in environmental changes and sustainable development (Turner *et al.* 2007). Urban land-use changes are influenced by multiple factors, including government decision-making, residents'

CONTACT Xiaoping Liu liuxp3@mail.sysu.edu.cn; Xia Li lixia@mail.sysu.edu.cn

activities and urban development (Lambin and Geist 2008, Rahimi 2016). The demand for land resources by urban development and population growth also accelerates the alteration of land-use spatial patterns within cities (Sang *et al*. 2011). Hence, identifying the mechanisms of urban land-use changes and urban expansion is becoming increasingly important. Land-use simulation is an important means to study the inherent mechanisms of urban land use and predict urban land-use changes (Parker *et al*. 2003, Santé *et al*. 2010).

A cellular automaton (CA) is a discrete model applied in simulating complex science, such as geographical process modeling and urban expansion simulation (Batty *et al*. 1999, Li and Yeh 2002, Liu *et al*. 2014). Raster-based geographical CA models have been widely used in urban land-use simulations (Li and Yeh 2000, 2002, Santé *et al*. 2010, Arsanjani *et al*. 2013, Liu *et al*. in press). Various factors that affect urban development, such as land suitability (Barredo *et al*. 2003, Mitsova *et al*. 2011), accessibility (White and Engelen 2000, Lau and Kam 2005) and socio-economic factors (Caruso *et al*. 2005, Wu *et al*. 2010) are incorporated into urban simulation models so urban CA models more closely match reality. Artificial intelligence and fuzzy logic models (Al-Ahmadi *et al*. 2009) such as neural networks (Li and Yeh 2002, Basse *et al*. 2014), support vector machines (Yang *et al*. 2008, Rienow and Goetzke 2015, Feng *et al*. 2016), random forests (Kamusoko and Gamba 2015), simulated annealing (Feng and Liu 2013) and other methods are also used to mine urban development rules. Meanwhile, some studies have also discussed the definition of neighborhoods (Kocabas and Dragicevic 2006, Wu *et al*. 2012) and growth constraints (Li and Yeh 2000, Guan *et al*. 2011) for CA models. In these studies, urban expansion simulations that were based on the CA model have yielded good results, proving that this approach is feasible in complex urban system simulations (Santé *et al*. 2010, Li *et al*. 2017).

Ground objects, such as the fine-scale land parcels, buildings, trees, etc., that are represented by rasters cannot accurately reflect reality (Stevens and Dragićević 2007), especially the expression of urban morphology (Lez *et al*. 2015). Raster-based CA models are very effective at simulating large-scale urban expansion, but defects in the expressive ability of geometric entities by using raster cell has become increasingly obvious with the refinement of the simulation scale and the complexity of the simulation system (Barreira-González *et al*. 2015, Lez *et al*. 2015, Pinto and Antunes 2010). Meanwhile, effectively determining the size (resolution) of a single cell and its actual geographical meaning in an urban expansion simulation remains an unsolved problem (Chen and Mynett 2003, Moreno *et al*. 2008, Dahal and Chow 2015, Abolhasani *et al*. 2016).

Along with the development of raster-based CA, patch-based CA integrates the idea of plaques into raster-based CA, which has exhibited favorable simulation performance for urban land-use patterns. The land patch refers to a collection of adjacent cells that, when combined together, represent an entity differing from its surroundings in nature or appearance (Wang and Marceau 2013, Chen *et al*. 2016, Tepe and Guldmann 2017). Along with the refinement of the resolution, raster cells can not express the features completely, thus transformation of urban land use is usually manifested as a collection of neighboring cells that are simultaneously transformed (Wang and Marceau 2013). By using patches to represent urban land uses, the newly developed urban cells can be deployed around the pre-stage developed urban cells, which are combined into patches. On the one hand, this operation contribute to eliminate the phenomenon of 'salt and pepper', while on the other hand, in comparison to the raster-based CA models, more realistic and reliable development pattern and classification structure can be simulated (Chen *et al*. 2014, 2016, Tepe and Guldmann 2017). Nevertheless, patch-based CA is simply a raster-based CA that incorporates morphological processing. Additionally, patch-based CA models contain too many stochastic factors from a morphological point of view, which lower the simulation accuracy (Chen *et al*. 2014, 2016) and thus are not conducive for fine land-use change simulations with land parcels as basic units.

In this case, vector-based CA (VCA) becomes an important method to solve the problems about the relevance of each cell and its corresponding ground objects (Abolhasani *et al*. 2016, Chen *et al*. 2016, Lez *et al*. 2015). Although early CA models that were based on Voronoi polygons (Shi and Pang 2000, Pang and Shi 2002), Delaunay triangulation (Semboloni 2000) and graph-based CA (O'Sullivan 2001a, 2001b) were beyond the limitations of regular cells, these models could hardly reveal the actual state of the ground objects (Moreno *et al*. 2009, Lez *et al*. 2015). To overcome these drawbacks, VCA models that are based on ground objects (Benenson *et al*. 2002, Hu and Li 2004), census blocks (Pinto and Antunes 2010) or land or cadastral parcels (Abolhasani *et al*. 2016, Lez *et al*. 2015, Moreno *et al*. 2008, Stevens *et al*. 2007, Stevens and Dragićević 2007) are applied to model land-use changes. Among all VCA models, those that are based on land or cadastral parcels play a strong supporting role in urban planning (Lez *et al*. 2015) and can more realistically reveal ground objects (Abolhasani *et al*. 2016, Barreira-González *et al*. 2015, Dahal and Chow 2015, Moreno *et al*. 2008, Stevens *et al*. 2007); these models are thus widely used and developed. For instance, Stevens *et al*.'s (2007) iCity model and Moreno *et al*.'s (2008) VecCity model introduced the geometric transformation of parcels to improve the simulation performance of VCA and thereafter introduced graph theory to reduce the calculation cost (Barreira-González *et al*. 2015). In short, VCA models have a strong advantage in land-use change modeling at a very fine scale.

As VCA models are emerging simulation models in the studies of urban developments, many issues remained to be discussed (Barreira-González *et al*. 2015, Dahal and Chow 2015, Chen *et al*. 2016). First, the cells of VCA are irregular polygons, so the method to define neighborhoods for raster-based CA is no longer applicable and must be redefined in accordance with the actual ground objects (Moreno *et al*. 2008). Moreover, VCA models do not eliminate the sensitivity of the neighborhood's type and size, so the accuracy of the simulation results significantly depends on the configuration (Stevens *et al*. 2007, Dahal and Chow 2015). Previous studies proposed various methods to define neighborhoods. Stevens and Dragićević (2007) defined neighborhoods based on the topology of adjacent cells. Crooks (2010) defined neighborhoods in terms of buffer distances from the center of the cell by considering the geographic element-blocking effect. Moreno *et al*. (2009) eliminated the parametric sensitivity of neighborhood configurations by defining dynamic neighborhoods. Ballestores and Qiu (2012) used a certain buffer distance from the cell boundary to delimit the neighborhood space. The above neighborhood definition methods have their own strong points and weaknesses and apply to different environments. Dahal and Chow (2015) defined 30 neighborhood configurations to compare the parameter sensitivity of the simulation results based on previous studies. These results demonstrated that center-buffer neighborhoods with feature blocks in VCA models reached the highest simulation accuracy.

Additionally, urban expansion is a highly fragmented process (Su *et al*. 2012), so the appearance of other types of land use can destroy the homogeneity of a VCA cell, where each cell is an irregular unit with homogeneity (Wang and Marceau 2013). Additionally, land parcels are too coarse to be used as the basic simulation units, which directly affects the simulation accuracy, so a reasonable land division method must be developed. To our knowledge, the existing methods of land division are mostly used for visualization, assisting urban planners in assessing land-use regulations and environmental protection policies, and so on (Vanegas *et al*. 2009). For example, Wickramasuriya *et al*. (2011) developed a tool that was based on ArcGIS to automatically split land parcels, where larger street plots were split into neater small plots and street layouts. The Parcel-Divider tool, which was developed by Dahal and Chow (2014), can split land parcels into a variety of finer layouts depending on the shape, size and orientation of the original plots. Vanegas *et al*. (2009) used binary recursive partitioning to split large urban structures into small parcels and new streets to achieve a satisfactory visualization effect based on high-resolution remote sensing images and vector data sets.

Nevertheless, only a few studies have introduced land division mechanisms into VCA models. For example, Moreno *et al*.'s (2008,2009) VCA model split land parcels by the geometric transformation of cells, which was achieved by determining their buffer distance. Abolhasani *et al*. (2016) introduced a VCA model where land division was conducted during data preprocessing. However, the division process of the former model could barely be conducted on the edges of the land parcels compared to within these spaces. Although the latter model avoided this issue, the lack of an iterative dynamic land division process was not conducive to simulating dynamic changes in urban land use. Therefore, we must introduce an accurate dynamic land division method in VCA models.

Previous studies on VCA and urban land-use conversion were conducted on a limited scale because of the complexity of the data structures and thus could not meet the needs of urban planning under real scenarios. For example, Stevens and Dragićević (2007) only simulated the land-use changes in a slice of a downtown area in a western Canadian city. Barreira-González *et al*.'s (2015) VCA model only simulated the land-use changes in a small community in the Iberian Peninsula. Abolhasani *et al*. (2016) used their ParCA model to study two districts in Tehran, Iran, and simulate land-use changes in approximately a thousand plots.

Therefore, this study builds a VCA-based framework to simulate urban land-use changes at a large scale by integrating dynamic land parcel subdivision based on previous studies of urban land-use change simulations via CA. Moreover, we adopt a novel random forest algorithm (RFA) regression model to explore the rules of urban land-use changes at the level of land parcels. Finally, our proposed VCA model is applied to simulate urban land-use changes from 2009 to 2014 in Shenzhen, the fourth largest metropolitan city in China. The validity and rationality of this model are demonstrated through a comparison with several state-of-the-art CA models.

2. Study area and data

Shenzhen (Figure 1) is located in Guangdong Province in South China and has a total area of 1996.850 km^2 and a residential population of approximately 10.779 million. This city, which had an overall GDP of 1750.299 billion RMB yuan in 2015 (http://www.sztj.gov.cn/xxgk/tjsj/tjnj/), has been considered one of the largest international metropolitan

Figure 1. Case study area: Shenzhen, Guangdong Province. The background data are the high spatial-resolution (HSR) remote sensing image that was provided by Tianditu.cn and has a spatial resolution of 5 m.

cities and economic centers in China. As illustrated in Figure 1, Shenzhen has 10 administrative county-level districts (Futian, Luohu, Nanshan, Yantian, Baoan, Guangming, Longhua, Longgang, Pingshan and Dapeng). Futian, Luohu, Nanshan and Yantian are the oldest special economic zones in Shenzhen and still the most populated and developed districts. Together, these districts comprised 52.48% of the city's total GDP. The land-use pattern of Shenzhen, which is the most developed immigrant city in South China, is extremely complex, and the land-use types are still rapidly changing (Yao *et al*. 2016).

According to the governmental land-use data from the *Bureau of Land and Resources of Shenzhen*, the urban structures in the study area have high complexity and contain a large amount of land-use types, such as public management-service land (P), residential land (R), commercial land (C), industrial land (I) and non-construction land (N). Among all, public management-service land, residential land, commercial land and industrial land are urban lands, and the non-construction land represents non-urban land that mainly contains agricultural land (farmland and woodland) and land that has not yet been developed (bare land and mountains). The spatial distribution and quantitative proportion of each land-use type in the study area in the years 2009 and 2014 are illustrated in Figures 2 and 3, respectively. Based on official statistical data, the total number of land-use blocks in 2009 and 2014 were 104,608 and 123,325, respectively. In particular, Figure 3 and Table 1 show that the proportions of public management-service land, residential land, commercial land and industrial land, which all belonged to the urban area, gradually increased with time from 2009 to 2014. Few alterations in land-use types occurred within this urban area, with more than 85.51% converted from non-construction land from 2009 to 2014 (Table 1). In particular, approximately 8.803 km^2 of non-construction land was occupied by different urban land types annually at the basic

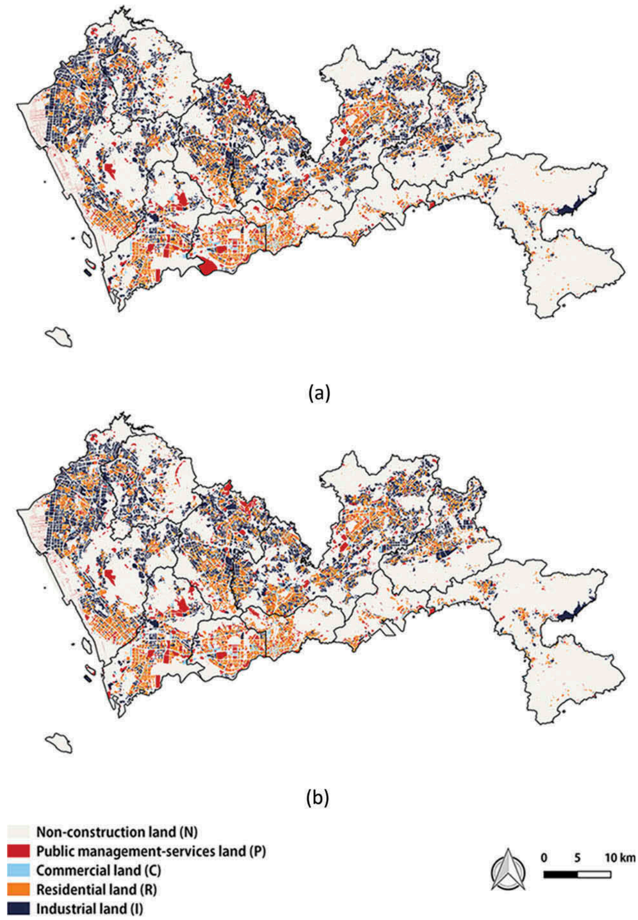

Figure 2. Urban land-use data in the study area in (a) 2009 and (b) 2014. The black lines at the front are the borders of administrative regions.

unit of land parcels. Therefore, this study's simulation and prediction were mainly based on the translation of non-construction land into urban land-use types.

To our knowledge, urban land-use change is a complicated phenomenon that is caused by interactions between urban planning and human activities (Long *et al*. 2012, Yao *et al*. 2016, Chen *et al*. 2017). Several basic geographic information and social media data sets, including points-of-interest (POIs) and OpenStreetMap road nets, were applied in our study to reasonably simulate urban land-use changes at a fine scale. We fetched the POIs of approximately 211,076 records with eight categories in the study area via Gaode Maps APIs, including business establishments, commercial sites, medical facilities, entertainment facilities, shopping malls, parks, factories, etc. Gaode Maps possesses fine geocoding accuracy and has been used in several previous urban studies at the level of traffic analysis zones (Yao *et al*. 2016). Additionally, spatial auxiliary variables (Figure 4) were classified into four major categories, including natural factors (elevation and slope), traffic factors (roads, highways and railways), location factors (distance to district centers) and urban environmental factors. Previous studies indicated that these spatial variables

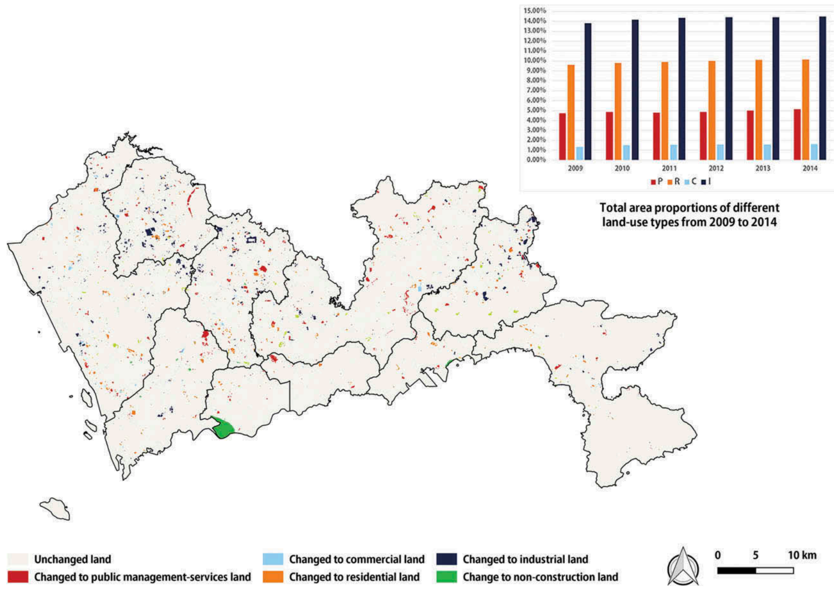

Figure 3. Spatial distribution of land-use change and total area proportions of different land-use types in the study area from 2009 to 2014.

Table 1. Total transition area (unit: km^2) between different land-use types in the study area from 2009 to 2014.

2009 \ 2014	N	P	C	R	I
N	1333.423	13.322	5.154	11.134	14.407
P	4.772	87.189	0.015	0.000	0.269
C	0.000	0.000	26.379	0.000	0.010
R	0.670	0.016	0.001	187.292	0.003
I	1.959	0.015	0.000	0.000	268.308

could reasonably reflect urban planning and human activity characteristics (Yuan *et al.* 2012, Jiang *et al.* 2015, Long and Liu 2016). In this study, the bandwidth of the Gaussian function-based kernel density analysis of POIs was automatically determined according to the MISE criterion (Wand and Jones 1994, Yuan *et al.* 2012).

3. Methodology

A flowchart of the proposed RFA-based VCA model with dynamic land parcel subdivision (DLPS-VCA) is illustrated in Figure 5. By using DLPS-VCA, the purpose of our research is to integrate RFA and vector unit automation to consider dynamic land subdivision during urban land-use change processes and to simultaneously simulate urban expansion and land-use changes at the scale of land parcels. This study used four steps to simulate urban land use. (1) Each vector land parcel was divided based on the minimum area boundary rectangle (MABR) to obtain a reasonable block distribution along city roads. (2) We created auxiliary spatial variables based on multi-source geospatial data

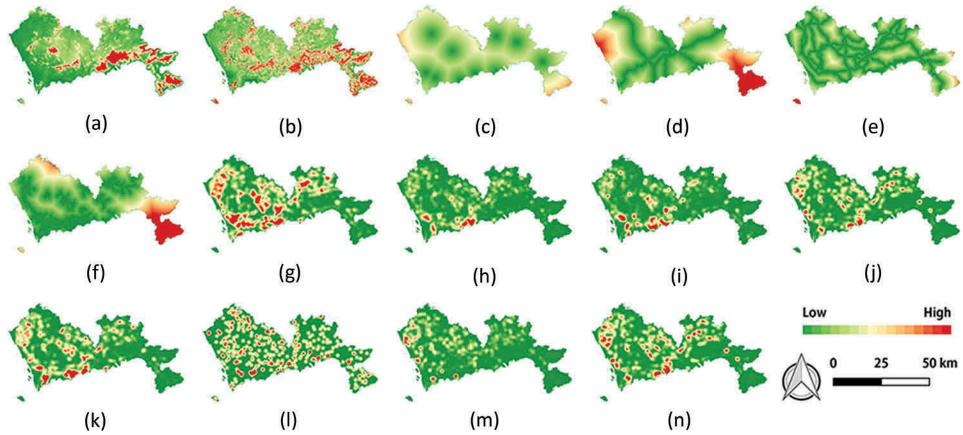

Figure 4. Auxiliary geospatial data sets: (a) DEM, (b) slope, (c) distance to district centers, (d) distance to railways, (e) distance to highways, (f) distance to roads, (g) density of bus stations, (h) density of medical facilities, (i) density of entertainment facilities, (j) density of shopping malls, (k) density of restaurants, (l) density of parks, (m) density of factories and (n) density of wholesale markets.

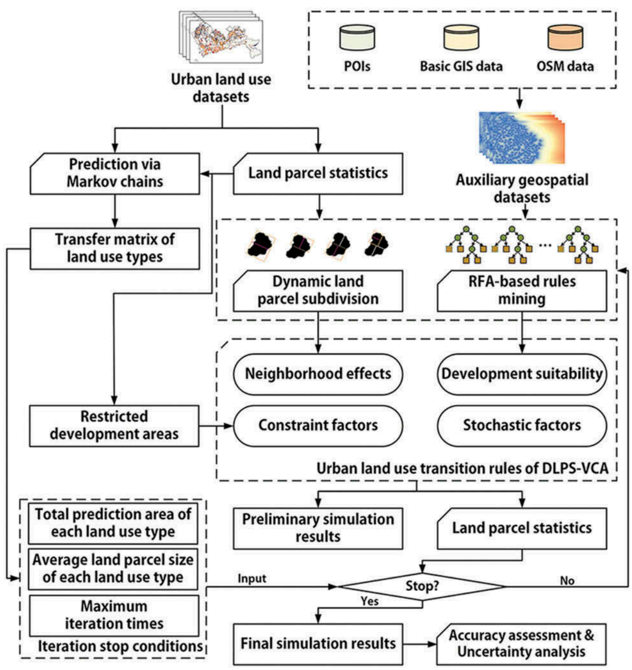

Figure 5. Flowchart to simulate urban land use with the proposed DLPS-VCA model.

sets and introduced an RFA model to explore the rules of urban expansion and land-use changes. (3) The proposed DLPS-VCA, which was validated by multi-period urban land-use data, was constructed to conduct urban land-use simulation and evaluate the performance of the simulation results by accuracy assessment and uncertainty analysis.

(4) We assumed a variety of urban development scenarios and used the proposed DLPS-VCA to forecast the urban land use of the study area.

3.1. *MABR-based DLPS*

Land-use change, which is mainly caused by urban expansion, is a random and highly fragmented process (Antrop 2004, Su *et al*. 2012). Based on a previous study by Vanegas *et al*. (2009), an iterative dichotomy strategy was used to split larger urban land parcels in this study to ensure that the entire direction of the block after division was extended along the road. Before each urban land-use change rule-mining process, we calculated the mean area μ_i and standard deviation σ_i of the i th land-use type block and split the i th land parcel $P_{i,j}$, which possessed an area greater than $(\mu_i + 2\sigma_i)$. As illustrated in Figure 6(a), the division process of land parcel $P_{i,j}$ included the following three steps:

(a) Convex hulls were acquired based on Cheng *et al*.'s (2008) model of arbitrary polygon-partitioning MABR (Cheng *et al*. 2008) by iterating the vertices of polygons to obtain an accurate MABR of land parcel $P_{i,j}$. Figure 6(a) shows that the directionality of the MABR was consistent with that of the original land parcel.
(b) The perpendicular bisector l was formed on the longer side of the MABR, which separated the land parcel $P_{i,j}$ into two new land parcels $P^1_{i,j}$ and $P^2_{i,j}$. Then, the algorithm calculated the areas of the newly obtained land parcels $P^1_{i,j}$ and $P^2_{i,j}$, and

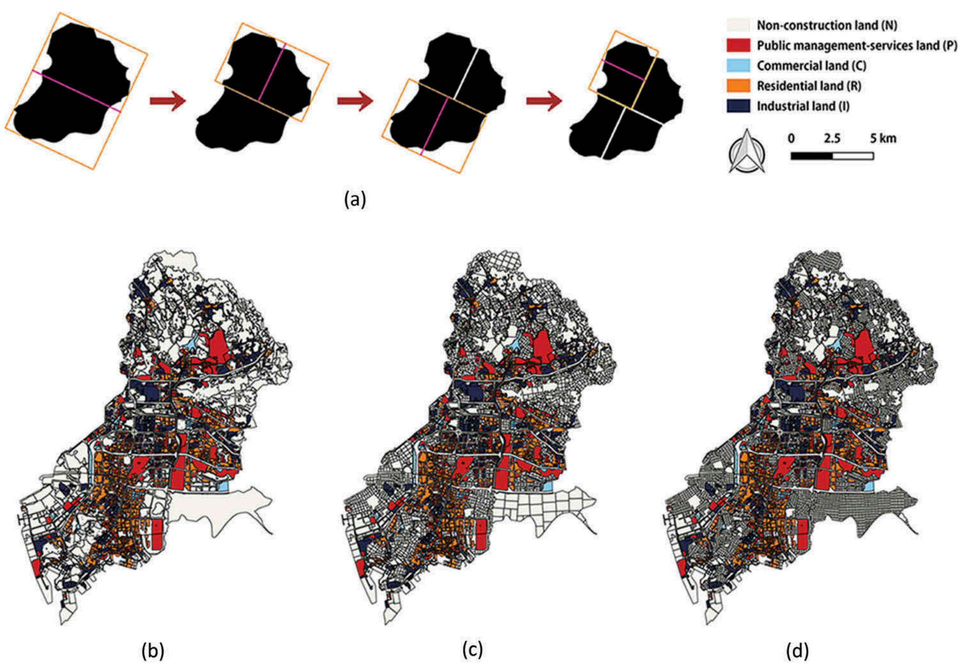

Figure 6. (a) Diagram of MABR-based land parcel subdivision. (b) Actual land use and parcels in Nanshan district. (c) Land parcel subdivision results in Nanshan district after the first iteration of land parcel subdivision. (d) Final result of land parcel subdivision in Nanshan district.

the MABR-based division continues to proceed for areas that were larger than (μ + 2σ).

(c) We repeated steps (a) and (b) until the areas of all the parcels were smaller than the current $(\mu_i + 2\sigma_i)$. After each urban land-use change simulation, the mean area μ_i and the standard deviation σ_i of each land parcel types were updated, thus realizing dynamic land-use parcel subdivision in the urban land-use simulation process.

Figure 6(b–d) shows the proposed MABR-based DLPS process and a case study in Shenzhen's Nanshan district, where the larger land parcels were gradually separated into basic block units that were closer to the mean areas and had a more fragmented distribution.

3.2. *RFA-based vector cellular automata (RFA-VCA)*

After the land parcel subdivision, we treated the land parcels as the elementary cells to conduct simulations that were based on the cellular automata model. Previous studies indicated that the development probability P of each cell consists of four factors, including the overall development suitability Pg, neighborhood effect Ω, constraint factor Pc and stochastic factor RA (Li and Yeh 2002, Chen *et al*. 2016).

This study adopted an RFA regression model to obtain the overall development suitability of each land parcel. Previous studies proved that RFAs are outstanding state-of-the-art machine-learning models that can obtain satisfactory results in many classification and regression tasks (Biau 2012, Fern and Ndez-Delgado *et al*. 2014). RFAs are constructed from a multitude of decision trees, which are built from each sub-data set that extracted random samples from the original training data set (Breiman 2001, Biau 2012). Simply, in the training process of RFA, sample data set X_i (i =1,2,…,$n$trees), which contains $n$trees items, is acquired by using Bootstrap sampling method to replacement sampling the original data set X in the first place, where data that does not appear in the X_i data set are called out-of-bag (OOB). Next, for each sample data set X_i, the following procedure is adopted to generate a non-pruning decision tree: let the dimension of the original data be N, given a positive integer n ($n \ll N$), for each node, n-dimensional features are randomly selected from the original N-dimensional features, and by calculating the amount of information contained in each feature, the feature most capable to classify is selected for node splitting. The above process continues to generate decision tree, and finally build M classification trees. Finally, according to these decision trees, classification results are decided according to the voting result of each record.

In particular, we can obtain an OOB-based estimation error report for each decision tree and further calculate the model's generalization error by averaging their errors. Tibshirani (1996), Wolpert and Macready (1999) and Breiman(1996) recommended to use the OOB error estimation as an integral part of the generalization error estimation. OOB error is an unbiased estimate with a result similar to the k-cross-validation that demands massive calculation, thus there is no need for further cross-validating or using an independent test set to get an unbiased estimate of the error. This RFA-based fitting model could overcome the multiple correlative problems among spatial variables, which is particularly useful in higher-dimensional fitting situations (Palczewska *et al*. 2014). We randomly screened samples based on the proportions of the land-use change areas to avoid the class imbalance problem (Wasikowski and Chen 2010) that is caused by disproportions in the number of land-use conversion samples. Thus, a balance between the input sample number of different land-use change types into the RFA and the conversion proportion of non-construction land into various types of urban land-use types was ensured. Hence, the overall development suitability of the i th non-construction land parcel when converted to the k th urban land-use type at time t is presented as follows:

$$Pg_i^{k,t} = \frac{\sum_{n=1}^{M} I(h_n(x) == Y_k)}{M} \tag{1}$$

In Equation (1), $I(\cdot)$ is the indicative function of the decision tree set; M is the total count of decision trees; x is a high-dimensional vector that consists of auxiliary spatial variables in the land parcel; and $h_n(x)$ is the prediction type of the nth decision tree for x, which is the land-use change type result of each decision tree for the i th non-construction land parcel transformation.

Neighborhood effects are one of the key considerations in the cellular automata modeling of complex geographic phenomena (Li and Yeh 2002, Dahal and Chow 2015). The basic units of VCA are irregular land parcels, and raster cells in the Moore neighborhood or Von Neumann neighborhood of traditional patch-based or raster-based CA cannot be obtained, so defining the rules of the VCA's neighborhood is very sensitive and difficult. A previous study indicated that the distance effect between cells in CA satisfied the law of exponential decay (Cohen and Kaplan 2007). Moreover, cells with larger areas, such as land parcels, had greater external effects and played pivotal roles in the transformation of the surrounding cellular states (Moreno *et al*. 2009). This study adopted a centroid-intercepted buffer rule that was based on land parcel area weighting to obtain the land parcel's neighborhood effects and thus improve the accuracy of multi-class land-use simulations (Abolhasani *et al*. 2016). If we assume that the j th land parcel is located in the buffer zone that was centered on the i th land parcel with a buffer range d and that no river barrier exists between the i th land parcel and the j th land parcel, then the neighborhood effect of the j th land parcel on the i th land parcel at time t is

$$\Omega_{i,j}^{t} = e^{-d_{ij}/d} \cdot \frac{S_j/S_i}{S_{\max}/S_{\min}} \tag{2}$$

In Equation (2), e is an exponential constant; d_{ij} is the center distance between the i th land parcel and the j th land parcel; S_i and S_j stand for the area of the i th land parcel and the j th land parcel, respectively; and $S_{\max}$ and $S_{\min}$ represent the maximum and minimum areas of the land parcels in the study area, respectively. Consequently, the neighborhood effect of the k th land-use type toward the i th land parcel at time t is presented as follows:

$$\Omega_i^{k,t} = \sum_j \Omega_{i,j}^{k,t}(\text{if } dis_{i,j} \leq \text{ buffer\_}d \text{ and No River between } i \text{ and } j) \tag{3}$$

The constraint factors are the particular land-use types that do not change to others during the simulation (Lin and Li 2016). In this study, we set water bodies (including rivers, lakes, sea areas, etc.) and roads as restricted development areas. The constraint factors of the i th land parcel could be calculated from Equation (4), where S_i is the development suitability state of the land parcels:

$$Pc_i^t = \begin{cases} 0 S_i = \text{restricted development area} \\ 1 S_i = \text{suitable development area} \end{cases} \tag{4}$$

To our knowledge, the factors that influence the transformation of land use are exceedingly complicated and highly stochastic (Li and Yeh 2002, Wu and Martin 2002). In this study, we introduced the stochastic factors $RA = 1 + (-\ln \gamma)^{\alpha}$, where γ is a random number between 0 and 1 and α is the parameter that controls the randomness with a constant value between 1 and 10 (Li and Yeh 2002, Wu and Martin 2002). In conclusion, the probability that the i th land parcel changes into the k th land-use type at time t is given as follows:

$$P_i^{k,t} = Pg_i^{k,t}\, \Omega_i^{k,t}\, Pc_i^t\, RA \tag{5}$$

During the simulation, we recomputed the land parcels' areas for each simulation result and performed dynamic land parcel subdivisions. By calculating the probabilities for each land parcel to be converted to the various land-use types respectively, we chose conversions that had the highest probabilities and exceeded the development thresholds to execute. In this study, the development thresholds of a certain land use class is set as the mean probability of all non-construction land parcel developing into this land use class. This ensures that if the i th type urban land use area does not meet the forecast requirements during the iteration process, the DLPS-VCA can reduce the threshold until all urban land area reach the predicted quantity.

To optimize the simulation results, the total amount of land-use change was controlled through the validation of multi-period land-use data during the simulation process by using Markov chains to predict changes in various land-use areas (Guan *et al.* 2011, Yang *et al.* 2012, 2014, Arsanjani *et al.* 2013). Markov chains is a predictive and optimized control method that assumes the state transitions depend on historical states. When performing a geographic simulation, Markov chains can not only be used to reveal the probability of conversion between different land types, but also to predict the ratio of various types of land use in the future stage, thus widely adopted in the prediction process of geographical research (Sang *et al.* 2011). In the Markov chains prediction process, $S(t+1) = P_{ij}\, S(t)$, where $S(t)$ and $S(t+1)$ represents the system status at time t and t + 1 respectively, which are used to represent the proportion of various land use types at time t and t + 1 respectively in this study; P_{ij} is a matrix representing the conversion probability of land uses. In this paper, the conversion matrix is calculated based on the ratio of land use in 2009–2014. In the DLPS-VCA model, through carrying out constraint control, the land use type is no longer developed for that where the area reaches the predicted value. Additionally, we predicted the future land-use patterns in the study area based on the different total simulation sizes of each land-use type in different future scenarios.

3.3. *Accuracy assessment and uncertainty analysis*

This study adopted the conventional cell-to-cell method, which calculates the overall accuracy (OA) and kappa coefficient, to evaluate the simulation accuracy by combining the actual and simulated urban land use into raster data to evaluate the final accuracy of our proposed DLPS-VCA model in simulating urban expansion and land-use change processes (Chen *et al*. 2014, 2016). Nevertheless, previous studies indicated that using the confusion matrix to evaluate the accuracy of the simulation results on a larger scale is unreasonable because the ratio of conversion to non-conversion is extremely low (Pontius *et al*. 2008, Du *et al*. 2012). Actually, we were interested in the relevance of actual urban land-use transformation and simulated urban land use. Consequently, this study adopted Figure-of-Merit (FoM) (Pontius *et al*. 2008) to evaluate the accuracy of the simulation results. FoM is an indicator to evaluate the consistency between the true (observed) transition pattern and the simulated (predicted) transition pattern, which equals the ratio between the intersection and union of the observed change and predicted change (Perica and Foufoula-Georgiou 1996). The formula is listed as follows (Pontius *et al*. 2008, Chen *et al*. 2016):

$$\text{FoM} = B/(A + B + C + D) \tag{6}$$

$$\text{Product's accuracy}(PA) = B/(A + B + C) \tag{7}$$

$$\text{User's accuracy}(UA) = B/(B + C + D) \tag{8}$$

In Equations (6)– (8), *A* indicates the error of the land-use change in reality and non-change in the simulation results; *B* represents the land where real transformations occurred and the simulation results correctly changed; *C* stands for the error that a transition had occurred and the simulation had actually changed but the type of the simulated transition did not match the truth; and *D* represents the error that the land-use change did not occur in reality but did occur in the simulated situation.

Moreover, landscape indices (LI) are used to assess the similarity of landscape patterns between the simulation results and the real scenario (McGarigal *et al*. 2012, Chen *et al*. 2014). This study adopted several LIs, including number of patches (NP), largest-patch index (LPI), mean Euclidean nearest-neighbor distance (ENN) and the mean perimeter–area ratio (PARA) to evaluate the similarity between the real and simulated block patterns. In this study, the LI were calculated by using Fragstats 4 (McGarigal *et al*. 2012). In addition, the similarity between the spatial patterns was measured by using the average difference in the LI between the simulation results and the real scenario. The calculation formula was as follows:

$$a_l = 1 - \frac{1}{8}\sum_i \Delta l_i \tag{9}$$

$$\Delta l_i = \begin{cases} \left| l_{i,s} - l_{i,o} \right| / l_{i,o}, l = \text{NP, ENN, PARA} \\ \left| l_{i,s} - l_{i,o} \right|, l = \text{LPI} \end{cases} \tag{10}$$

In Equations (9) and (10), $l_{i,s}$ and $l_{i,o}$ represent the *i* th LI of the simulated and real scenarios, respectively, and Δl_i is the normalized difference of the *i* th LI.

4. Results

Our research team built a software application and created the DLPS-VCA model that was proposed in Section 3 by using C++ on Windows Server 2008. Several open-source C/C++ libraries, such as CGAL, GDAL, OpenCV and Shark, were applied to this project to build the proposed DLPS-VCA model. The source codes of the proposed model were implemented in C++ with OpenMP and run on a multi-processor computation server, and the related applications will be released on the GeoSOS website (http://geosimulation.cn/).

4.1. *Implementation and results*

Our training data comprised selected samples that transformed from non-construction land into other land from 2009 to 2014 to increase the total number of samples to validate the proposed DLPS-VCA model. Additionally, the neighborhood distance was adjusted to 800 m. The relationship between the neighborhood size and simulation accuracy is analyzed in later sections. During RFA-based regression, we divided the training data set, including the original land-use change data and auxiliary geospatial data, into two components, namely, 60% training data and 40% validation data, to evaluate the fitting accuracy of this model. We set up 100 decision trees and 20% OOB data, cross-validated the result with boosted random sampling and iterated 100 epochs to obtain the average accuracy for the most reliable result.

In addition to using the proposed DLPS-VCA model, this study adopted two recently developed CA models, including RFA-Patch-CA and RFA-VCA, to simulate land use and evaluate the simulation accuracy of these results as a comparative experiment. RFA-Patch-CA adopts the same patch-based simulation strategy as Chen's Logistic-Patch-CA (Chen *et al*. 2014, 2016) but uses RFA-based regression to mine the land-use type change rules. As a comparative experiment of DLPS-VCA, RFA-VCA is a simulation of urban land-use changes with RFA-based regression based on existing urban parcels. RFA-VCA does not conduct dynamic land parcel subdivision on the land parcels of the initial simulation, which is the most common method that is used by most VCA studies for large-scale simulations (Barreira-González *et al*. 2015).

As described in the previous section, all three models were first validated at the cell level (Chen *et al*. 2016). Figure 7 demonstrates the actual and simulated urban land-use results in 2014 with the above three models. In addition to the local highlighted features in Figure 8, we provided the original image, which has a high resolution of 300 dpi and can be downloaded from http://pan.baidu.com/s/1jHMuuou, to offer more details.

Previous studies indicated that FoM is generally low when using CA models to simulate large-scale areas or cities with few land-use conversions (Pontius *et al*. 2008). In this study, only 0.60% of the land conversion occurred in the study area, and multiple land-use change simulations were proposed, which was more complex than the previous CA model rule for urban expansion (Du *et al*. 2012, Chen *et al*. 2016). Based on the RFA-based rule mining, the accuracy of the three land-use simulation models was relatively high and the overall FoM was significantly larger than 0.15. As illustrated in Table 2, the proposed DLPS-VCA model had the highest accuracy (FoM = 0.232), which was 18.97% and 32.57% higher than those of the RFA-VCA and Patch-CA models, respectively, where no land splitting was conducted. These results indicated two key

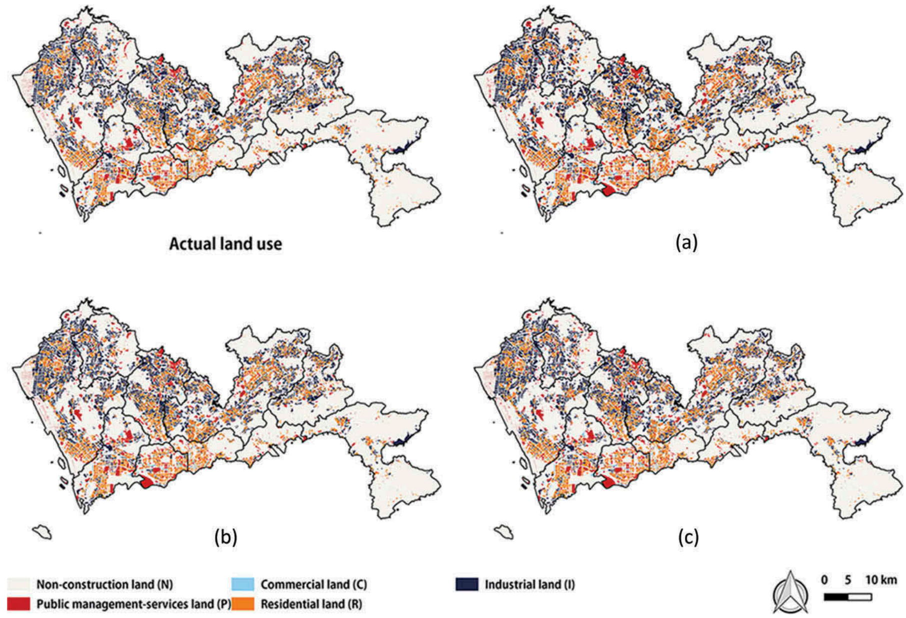

Figure 7. Actual and simulated urban land-use patterns in Shenzhen in 2014. Simulated urban land-use patterns in the study area via (a) Patch-CA, (b) RFA-VCA and (c) DLPS-VCA.

points. (1) Compared to patch-based CA, VCA models that are based on authentic land parcels can achieve much higher land-use simulation accuracy. (2) During urban development, larger urban land patches gradually split into smaller patches of land, so the proposed DLPS-VCA model could achieve higher simulation accuracy, which is a more realistic response to actual urban land-use development patterns.

Table 3 shows the results of the pattern-level similarities. Previous studies proved that Patch-CA could produce simulation results that are more similar to actual urban development patterns (Chen *et al*. 2014, 2016). Comparing the results of Patch-CA, RFA-VCA and DLPS-VCA with actual urban land-use patterns showed that the results for DLPS-VCA had the highest similarity (94.73%) with the actual land-use patterns in 2014. We controlled the size of the land parcels that altered their land-use types during dynamic land parcel subdivision, so the indices that were most relevant for the land parcels, such as LPI, ENN and PARA, also achieved the closest results compared to the other models. Figure 8 shows a comparison of the four typical areas in the study area. Obviously, the results of Patch-CA clearly connected adjacent land parcels, which induced a smaller NP and a salt-and-pepper noise.

The results of the RFA-VCA and proposed DLPS-VCA models were exceedingly similar in the downtown area. However, Figure 8(C1–C4) shows that the simulation results for DLPS-VCA were more precise than those of the other models in the newly developed regions. In conclusion, the proposed DLPS-VCA model more accurately matched the urban expansion simulation and reasonably mined the land-use development rules of various types of land. Moreover, the splitting results of the irregular plots were extremely similar to those of the actual urban land use based on the proposed DLPS method.

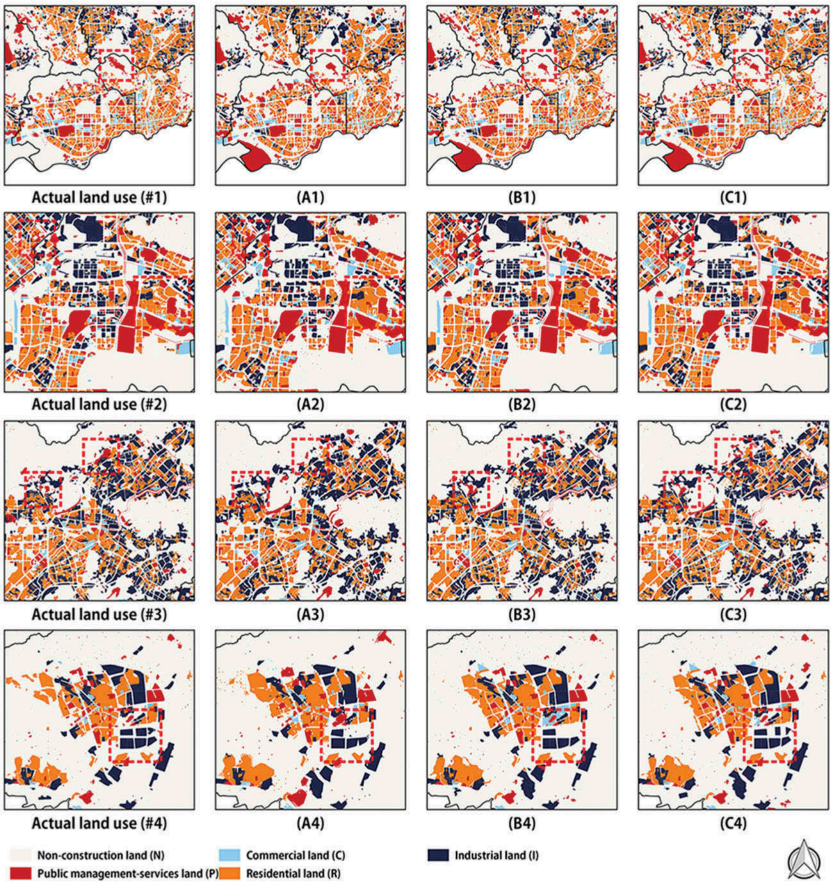

Figure 8. Details of the actual and simulated urban land-use patterns in (#1) the City center area, (#2) Nanshan district, (#3) Pingshan district and (#4) Dapeng district via different simulation models: (A1)–(A4) Patch-CA, (B1)–(B4) RFA-VCA and (C1)–(C4) DLPS-VCA.

Table 2. Overall FoM of the simulated results via different models.

Results	PA (%)	UA (%)	FoM
Patch-CA	29.83	29.05	0.175
RFA-VCA	32.74	31.91	0.195
DLPS-VCA	37.45	37.20	0.232

Table 3. Overall landscape indices of the simulated results via different models.

Results	NP	LPI	ENN	PARA	Similarity (%)
Actual land use	28,550	67.673	124.799	906.634	–
Patch-CA	26,826	66.793	121.022	923.85	87.63
RFA-VCA	27,961	67.001	121.589	911.377	90.95
DLPS-VCA	27,910	67.286	123.596	909.309	94.73

Additionally, we compared the simulated results for DLPS-VCA with the actual urban land use, as shown in Table 4 for FoM and Table 5 for the similarities in the different administrative districts. Compared to Baoan, Longhua, Pingshan and Dapeng, which are

Table 4. FoM of the simulated results via the DLPS-VCA model for different districts in the study area.

Districts	PA (%)	UA (%)	FoM
Futian	22.40	59.51	0.194
Luohu	17.13	26.31	0.116
Nanshan	21.24	42.77	0.167
Yantian	30.43	59.95	0.250
Baoan	41.57	34.67	0.233
Guangming	42.17	31.00	0.221
Longgang	24.70	35.30	0.181
Longhua	50.36	38.90	0.286
Pingshan	37.00	41.49	0.248
Dapeng	50.88	35.19	0.267
Shenzhen	37.45	37.20	0.232

Table 5. Landscape indices of the simulated results via the DLPS-VCA model in the study area.

Districts	Type	NP	LPI	ENN	PARA	Similarity (%)
Futian	Actual	1368	59.22	113.124	746.733	79.21
	Simulated	1366	57.566	112.871	750.805	
Luohu	Actual	1326	70.875	123.492	833.318	94.11
	Simulated	1317	71.296	118.485	831.682	
Nanshan	Actual	2199	63.139	161.124	794.401	89.64
	Simulated	2057	63.896	162.05	795.827	
Yantian	Actual	520	87.307	179.479	893.546	93.39
	Simulated	498	87.701	163.327	895.857	
Baoan	Actual	8250	60.291	111.796	970.609	87.44
	Simulated	8473	59.355	108.565	982.972	
Guangming	Actual	2168	69.718	140.569	936.472	71.70
	Simulated	2071	67.562	132.773	944.271	
Longgang	Actual	6230	59.392	119.576	893.631	95.27
	Simulated	5907	59.711	119.415	887.44	
Longhua	Actual	3531	50.376	117.156	936.245	79.91
	Simulated	3451	48.844	111.98	944.451	
Pingshan	Actual	1995	73.881	131.569	883.653	96.25
	Simulated	1924	73.641	128.66	885.43	
Dapeng	Actual	1242	92.088	205.705	913.509	97.36
	Simulated	1237	91.919	199.651	921.619	
Shenzhen	Actual	28,550	67.673	124.799	906.634	94.73
	Simulated	27,910	67.286	123.596	909.309	

not as economically developed, Futian, Luohu and Nanshan, which are the most prosperous downtown areas in Shenzhen, had less precise urban land-use change simulation results. These results illustrate the complexity of urban development: the urban development of economically developed regions is dominated by the internal transformation of land use, which is called urban renewal, while less developed regions are mainly led by urban expansion (Dai *et al*. 2010, Wang *et al*. 2013, Zheng *et al*. 2014). The results also essentially suggests that the proposed DLPS-CA model, which is considered as a 'bottom-up' local model, can be well used to simulate the non-stationary urban development. Moreover, Guangming and Longgang are newly developed areas from the planning of the Shenzhen government, and government policies occupied a dominant position in their development. The proposed DLPS-VCA model did not consider the factors of government decision-making, so the FoM index and the differences in the NP LI in the simulation results for Guangming and Longgang were both large, resulting in a relatively low accuracy. Therefore, adding government decision-making and residential activities as consideration factors is a key issue to improve the precision of DLPS-VCA simulations in future study.

4.2. Contribution weights of different land-use types

This study obtained the contribution weights of various auxiliary spatial variables when non-constructed land was converted to other land-use types based on the sensitivity analysis of input features by a RFA-based regression model (Breiman 2001, Biau 2012, Palczewska *et al*. 2014), as shown in Table 6. The development of public management-service land was mainly relevant to district centers (10.66%), parks (7.29%) and traffic facilities, including railways (10.26%), roads (9.59%), highways (8.71%) and bus stations (7.58%). On the one hand, this factor determined by the service attributes of public management-service land. On the other hand, the development of public management-service land and improvement of transportation facilities were mutually reinforcing.

Residential land development was closely related to district centers (9.41%), public transportation facilities (bus stations, roads and highways) and livelihood infrastructures (medical facilities and parks). Developed urban transport infrastructures can reduce the cost of living in urban areas, so the construction of district centers and livelihood infrastructures was strongly correlated to the living convenience of residents. At the same time, commercial land had a strong correlation with all manner of spatial variables but shared a stronger correlation with spatial variables that were related to location, transportation and logistics factors, such as district centers (9.05%), restaurants (8.77%), and whole markets (8.42%).

Industrial land development often requires the support of a certain infrastructure and the formation of a production chain, which is prone to spatial agglomeration and spatial clustering (Schweitzer and Steinbink 1997, Gordon and McCann 2000, Rauch 2013). Newly developed industrial areas had a strong correlation with existing factories (9.63%) because of government planning. Additionally, the distance to railways

Table 6. Contribution weights of different spatial variables when non-construction land (N) was changed to public management-service land (P), residential land (R), commercial land (C) and industrial land (I). Background color ranging from green to red indicated that the increasing influence of the spatial variable on the current land use.

Spatial variables	P	R	C	I
DEM	6.08%	3.00%	5.59%	5.95%
Slope	2.98%	6.95%	3.75%	3.82%
Distance to district centers	10.66%	9.41%	9.05%	10.48%
Distance to railways	10.26%	6.12%	6.76%	13.84%
Distance to highways	8.71%	8.51%	7.14%	7.11%
Distance to roads	9.59%	8.35%	8.31%	8.01%
Density of bus stations	7.58%	9.28%	7.29%	7.12%
Distance to medical facilities	5.64%	8.62%	7.02%	5.30%
Distance to entertainment facilities	5.93%	6.27%	7.15%	5.61%
Density of shopping malls	6.73%	6.53%	7.36%	5.59%
Density of restaurants	6.26%	6.88%	8.77%	6.40%
Density of parks	7.29%	7.41%	6.63%	5.20%
Density of factories	6.49%	5.68%	6.76%	9.63%
Density of whole markets	5.81%	6.99%	8.42%	5.95%

(13.84%), roads (8.01%) and highways (7.11%) was highly correlated with industrial land because plant locations must consider the transportation costs of the goods. On the contrary, the effect of industrial infrastructures on industrial land was minimal.

In conclusion, the urban land-use development rule that was based on the proposed GLCM-VCA mining had an interpretable rationality and could act as a guide for urban land-use planning.

4.3. *Parameter sensitivity analysis*

The definition of neighborhoods in CA rule mining is very important, while that in VCA is more complex (Dahal and Chow 2015, Abolhasani *et al*. 2016). This study adopted area-weighted centroid-intercepted buffer neighborhoods to define the neighborhood effects of land parcels (Wu *et al*. 2012, Dahal and Chow 2015). As illustrated in Figure 9, the proposed GLPS-VCA model was extremely sensitive to the distance configuration of the neighborhood. The highest simulation accuracy was achieved when the buffer distance was 800 m. When the buffer distance was less than 800 m, the simulation's precision increased with increasing buffer distance, while the simulation's precision gradually decreased when the buffer distance was greater than 800 m. This result mainly occurred because the number of land parcels within the buffer zone was exceedingly small when the buffer distance was small because of the area of the center cell itself, which produced small neighborhood effects and no significant differences between different cells.

Certainly, the number of neighbor cells in the buffer increased and the neighborhood effects of various land-use types could be distinguished with the expansion of the buffer range, thus improving the simulation's accuracy. When the buffer distance was too large, however, the neighborhood area of the center cell was relatively large, and distant cells excessively affected the central cell, resulting in simulation errors. Previous studies indicated that the internal structure of the urban functional area generally presents an aggregated distribution in space (Schweitzer and Steinbink 1997, Rauch 2013); in other words, urban functional areas from the same categories generally clustered together, and the urban functional structure was affected much more by the concentration center than by the outer urban functional area. In this study, we conducted the above experiment by selecting 800 m as the simulation parameter.

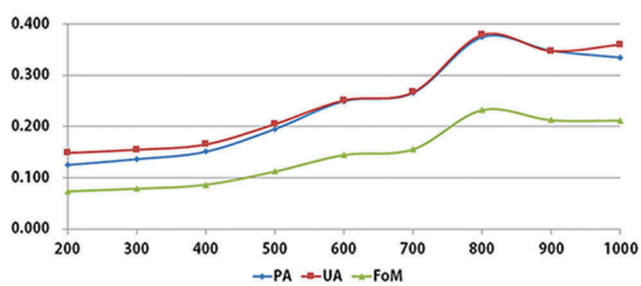

Figure 9. Changes in the accuracy indices (*y* axis) in the urban land-use simulation via DLPS-VCA by different neighborhood radii (*x* axis, unit: meters).

4.4. Future scenario simulation

The proposed DLPS-VCA model could also be used to simulate future developments under different scenarios. This study adopted three different future scenarios: (1) urban disordered development without any restrictions (UDDS), (2) urban sustainable development with ecology control (USDE) and (3) urban sustainable development with ecology control and 'job-housing balance' (USDB) (Frank 1994, Peng 1997, Zhao *et al*. 2011).

In the first two urban development scenarios (UDDS and USDE), the total conversion area for all the land-use types in 2020 and 2030 was based on an unrestricted Markov chain prediction. In USDE, however, we considered an official ecology control plan, in which the government participated in ecological protection interventions to protect the type of land within the ecological control line and did not allow this land to develop (Liu and Diamond 2005, Lu *et al*. 2013).

Furthermore, 'job-housing balance' refers to the spatial relationship between the number of jobs and housing units within a given geographical area (Peng 1997). Previous studies indicated that an area is considered balanced when the resident workers can obtain a job within a reasonable travel distance and when the available housing types can complement a variety of employees' housing demands (Zhao *et al*. 2011). Frank (1994) defined the job-housing balance within census tracts as a job-household ratio between 0.8 and 1.2 (Frank 1994). Hence, this study's urban development model for USDB was based on USDE, where the area ratio of residential land and land for working (including commercial land and industrial land) was controlled to approximately 1:1 during urban development. The conversion area of various land uses for USDB in 2020 and 2030 was predicted by restricted Markov chains.

Figure 10 and Table 7 show the simulation results and total quantities of different land-use types under the three proposed future scenarios. Figure 11 shows the detailed

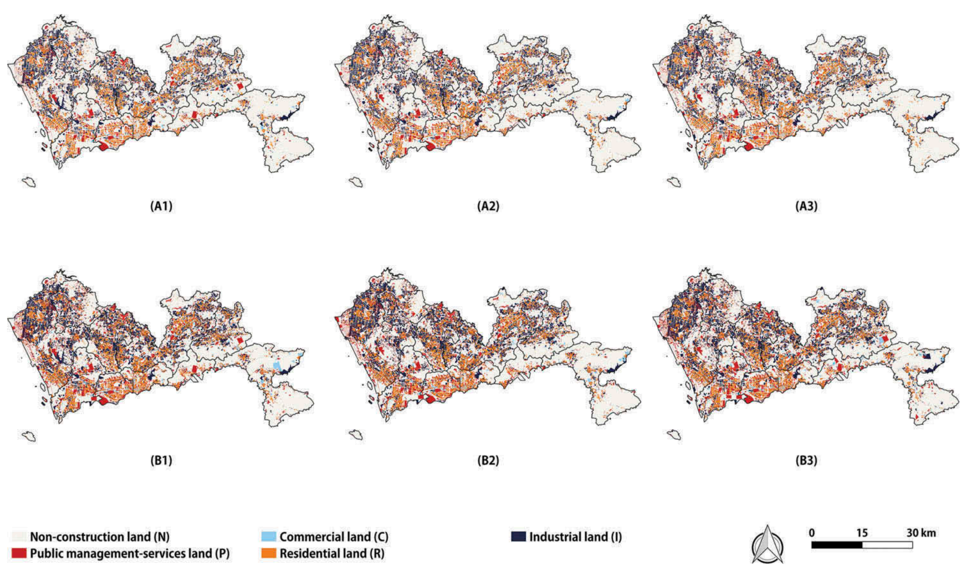

Figure 10. Simulated urban land-use patterns of Shenzhen under different urban development scenarios: (A1) UDDS, (A2) USDE and (A3) USDB in 2020; and (B1) UDDS, (B2) USDE and (B3) USDB in 2030.

Table 7. Total simulation area (unit: km^2) of each land-use type under different future scenarios.

Scenarios	UDDS		USDE		USDB	
Year	2020	2030	2020	2030	2020	2030
N	1307.076	1228.304	1312.581	1249.088	1322.504	1258.378
P	124.05	153.053	118.566	144.213	117.998	143.508
R	211.237	221.033	211.223	220.12	214.721	220.12
C	38.812	50.106	38.803	50.105	35.547	43.891
I	306.736	335.415	306.736	324.383	297.14	322.014

Land-use types: non-construction land (N), public management-service land (P), residential land (R), commercial land (C) and industrial land (I).

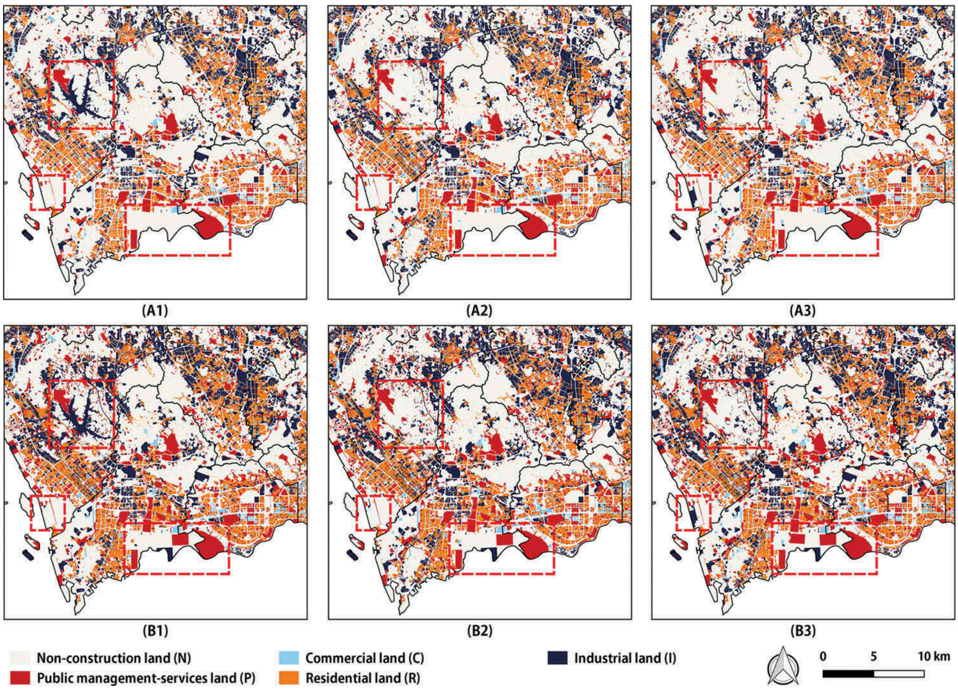

Figure 11. Details of the simulated urban land-use patterns in southwestern Shenzhen (Futian, Nanshan and Baoan districts) under different urban development scenarios: (A1) UDDS, (A2) USDE and (A3) USDB in 2020; and (B1) UDDS, (B2) USDE and (B3) USDB in 2030.

simulation results of Shenzhen's city center. Figures 10 and 11 show that a large number of ecological protection areas, such as Tiegang Lake (Figure 11(A1,B1)), were eroded by industrial land when the government did not implement ecology control. Moreover, approximately 7.9 km^2 of the non-construction land in the UDDS scenes was developed annually into urban land, especially in the Luohu District, Futian District and other economically developed downtown areas. In terms of ecology control, the development areas of the city were mainly dominated by internal urban renewal (Zheng *et al*. 2014) or development from public non-construction land, such as beaches (Figure 11(A2,B2). In terms of the 'job-housing balance', the growth of industrial land was weakened and more non-construction land was converted into residential land; however, by 2030, the residential land area was similar to those in the other two scenarios. Thus, the growth rates eventually stabilized under the current urban land-use pattern, even though the growth rate of the total area of residential land in Shenzhen changed.

5. Discussion and conclusions

This study improved the performance of CA simulations of fine-scale urban land-use changes at a large scale by integrating a vector-based simulation strategy and dynamic land parcel subdivision. First, we designed a dynamic MABR-based dynamic land-parcel subdivision method to split large urban land parcels. Second, a DLPS-VCA model was constructed by integrating a RFA-based regression model, which was used to mine the mutual conversion rules of different land use, and the land parcel subdivision method, and the correlations between the land-use conversion rules and several spatial variables were analyzed. Third, we constructed three different scenarios of urban land-use change, including UDDS, USDE and USDB, to simulate and analyze the land-use changes in the study area in 2020 and 2030.

Compared to the advanced Patch-CA (Chen *et al*. 2016) and RFA-VCA models, the proposed DLPS-VCA model achieved the highest simulation accuracy (FoM = 0.232) and was most similar to the real land-use scenario (similarity = 94.73%) at the pattern level. This result indicates that the reasonable division of land parcels during urban land-use change modeling can accurately simulate land-use change and urban land expansion processes simultaneously at a fine scale, which has rarely been addressed in previous studies. Additionally, we could feasibly analyze the driving factors of various land-use transformations and provide guidance for future urban development, based on the urban land-use conversion rules from the RFA-based regression model. Finally, this study obtained the land-use patterns of different development models in the study area by adjusting the constraint factors and the total amount of land-use types. Future studies should introduce more urban development scenarios based on the proposed DLPS-VCA model to analyze the changes in urban land use and functional structures.

Cities are increasingly considered complicated self-adaptive systems, so the evolution of urban land use can be depicted as intertwined processes from both top-down and bottom-up (Tian and Shen 2011, Long *et al*. 2012, Chen *et al*. 2017). In addition to self-organization, urban land-use changes are affected by urban planning, government decision-making and population activities, which are very complex and random processes that create difficulties in real land-use simulations, especially in China's developing cities. Hence, urban land-parcel division is a random and complex process (Antrop 2004, Su *et al*. 2012). Although DLPS-VCA achieved good simulation accuracy, this model still did not completely match reality. In future studies, we must consider the influence of government decision-making and human activities and obtain more comprehensive and accurate spatial variables through an agent-based system to improve the accuracy of urban land-use simulations.

Determining the effects of neighborhoods is a very important issue in CA, especially in VCA (Wu *et al*. 2012, Dahal and Chow 2015). This study adopted a rather simple neighborhood strategy to analyze the influence of the neighborhood distance determination, which obtained reasonable and precise simulation results. Future work should explore how the simulation performance of the proposed DLPS-VCA model changes if various extended neighborhood methods are applied. We must identify the most appropriate neighborhood configuration to achieve the best simulation results via DLPS-VCA and further generalize this approach for practical use.

Although the above questions have not yet been solved, CA models play an increasingly important role in government decision-making and urban planning (Lu *et al*. 2015, Chen *et al*. 2016). In fact, CA models can simulate complex dynamic urban spatial and temporal changes and can accurately predict and explain the development of cities. Overall, our proposed DLPS-VCA model can accurately simulate and predict complex urban expansion and land-use changes, which should provide a valuable reference for urban planners in the future.

Acknowledgment

We really appreciate Prof. May Yuan, Prof. David O'Sullivan and the three anonymous reviewers for their careful work, useful comments and suggestions. This study was supported by the National Key R&D Program of China [2017YFA0604402], the Key National Natural Science Foundation of China [Grant No. 41531176], the National Natural Science Foundation of China [Grant No. 41671398], National Natural Science Foundation of China [Grant No. 41601420] and the China Postdoctoral Science Foundation [Grant No. 2015M580753].

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study was supported by the National Key R&D Program of China [2017YFA0604402], the Key National Natural Science Foundation of China [Grant Numbers 41531176], [41671398] and [41601420], the China Postdoctoral Science Foundation [2015M580753].

ORCID

Yao Yao http://orcid.org/0000-0002-2830-0377
Xiaoping Liu http://orcid.org/0000-0003-4242-5392
Xia Li http://orcid.org/0000-0003-3050-8529
Penghua Liu http://orcid.org/0000-0002-8574-891X
Ye Hong http://orcid.org/0000-0002-8996-3748
Yatao Zhang http://orcid.org/0000-0001-5701-2836
Ke Mai http://orcid.org/0000-0002-3532-9872

References

Abolhasani, S., *et al*., 2016. Simulating urban growth under planning policies through parcel-based cellular automata (ParCA) model. *International Journal of Geographical Information Science*, 30 (11), 2276–2301. doi:10.1080/13658816.2016.1184271

Al-Ahmadi, K., *et al*., 2009. Calibration of a fuzzy cellular automata model of urban dynamics in Saudi Arabia. *Ecological Complexity*, 6 (2), 80–101. doi:10.1016/j.ecocom.2008.09.004

Antrop, M., 2004. Landscape change and the urbanization process in Europe. *Landscape and Urban Planning*, 67 (1), 9–26. doi:10.1016/S0169-2046(03)00026-4

Arsanjani, J.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. doi:10.1016/j.jag.2011.12.014

Ballestores Jr, F. and Qiu, Z., 2012. An integrated parcel-based land use change model using cellular automata and decision tree. *Proceedings of the International Academy of Ecology and Environmental Sciences*, 2 (2), 53.

Barredo, J.E.I., *et al*., 2003. Modelling dynamic spatial processes: simulation of urban future scenarios through cellular automata. *Landscape and Urban Planning*, 64 (3), 145–160. doi:10.1016/S0169-2046(02)00218-9

Barreira-González, P., Gómez-Delgado, M., and Aguilera-Benavente, F., 2015. From raster to vector cellular automata models: a new approach to simulate urban growth with the help of graph theory. *Computers Environment & Urban Systems*, 54, 119–131. doi:10.1016/j.compenvurbsys.2015.07.004

Basse, R.M., *et al*., 2014. Land use changes modelling using advanced methods: cellular automata and artificial neural networks. The spatial and explicit representation of land cover dynamics at the cross-border region scale. *Applied Geography*, 53, 160–171. doi:10.1016/j.apgeog.2014.06.016

Batty, M., Xie, Y., and Sun, Z., 1999. Modeling urban dynamics through GIS-based cellular automata. *Computers, Environment and Urban Systems*, 23 (3), 205–233. doi:10.1016/S0198-9715(99)00015-0

Benenson, I., Omer, I., and Hatna, E., 2002. Entity-based modeling of urban residential dynamics: the case of Yaffo, Tel Aviv. *Environment & Planning B Planning & Design*, 29 (29), 491–512. doi:10.1068/b1287

Biau, G.E.R., 2012. Analysis of a random forests model. *The Journal of Machine Learning Research*, 13 (1), 1063–1095.

Breiman, L., 1996. *Out-of-bag estimation*. Berkeley CA: Statistics Department, University of California Berkeley, Technical report, 1996b. 33, 34.

Breiman, L., 2001. Random forests. *Machine Learning*, 45 (1), 5–32. doi:10.1023/A:1010933404324

Caruso, G., Rounsevell, M., and Cojocaru, G., 2005. Exploring a spatio-dynamic neighbourhood-based model of residential behaviour in the Brussels periurban area. *International Journal of Geographical Information Science*, 19 (2), 103–123. doi:10.1080/13658810410001713371

Chen, Q. and Mynett, A.E., 2003. Effects of cell size and configuration in cellular automata based prey-predator modelling. *Simulation Modelling Practice & Theory*, 11 (7–8), 609–625. doi:10.1016/j.simpat.2003.08.006

Chen, Y., *et al*., 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

Chen, Y., *et al*., 2016. Capturing the varying effects of driving forces over time for the simulation of urban growth by using survival analysis and cellular automata. *Landscape and Urban Planning*, 152, 59–71. doi:10.1016/j.landurbplan.2016.03.011

Chen, Y., *et al*., 2017. Delineating urban functional areas with building-level social media data: a dynamic time warping (DTW) distance based k-medoids method. *Landscape and Urban Planning*, 160, 48–60. doi:10.1016/j.landurbplan.2016.12.001

Cheng, P., Yan, H., and Han, Z., 2008. An algorithm for computing the minimum area bounding rectangle of an arbitrary polygon [J]. *Journal of Engineering Graphics*, 1 (1), 122–126.

Cohen, E. and Kaplan, H., 2007. Spatially-decaying aggregation over a network. *Journal of Computer and System Sciences*, 73 (3), 265–288. doi:10.1016/j.jcss.2006.10.016

Crooks, A.T., 2010. Constructing and implementing an agent-based model of residential segregation through vector GIS. *International Journal of Geographical Information Science*, 24 (5), 661–675. doi:10.1080/13658810903569572

Dahal, K.R. and Chow, T.E., 2014. A GIS toolset for automated partitioning of urban lands. *Environmental Modelling & Software*, 55 (55), 222–234. doi:10.1016/j.envsoft.2014.01.024

Dahal, K.R. and Chow, T.E., 2015. Characterization of neighborhood sensitivity of an irregular cellular automata model of urban growth. *International Journal of Geographical Information Science*, 29 (3), 475–497. doi:10.1080/13658816.2014.987779

Dai, X., *et al*., 2010. Spatio-temporal pattern of urban land cover evolvement with urban renewal and expansion in Shanghai based on mixed-pixel classification for remote sensing imagery. *International Journal of Remote Sensing*, 31 (23), 6095–6114. doi:10.1080/01431160903376407

Du, Y., *et al*., 2012. Comparison between CBR and CA methods for estimating land use change in Dongguan, China. *Journal of Geographical Sciences*, 22 (4), 716–736. doi:10.1007/s11442-012-0958-6

Feng, Y. and Liu, Y., 2013. A heuristic cellular automata approach for modelling urban land-use change based on simulated annealing. *International Journal of Geographical Information Science*, 27 (3), 449–466. doi:10.1080/13658816.2012.695377

Feng, Y., Liu, Y., and Batty, M., 2016. Modeling urban growth with GIS based cellular automata and least squares SVM rules: a case study in Qingpu–Songjiang area of Shanghai, China. *Stochastic Environmental Research and Risk Assessment*, 30 (5), 1387–1400. doi:10.1007/s00477-015-1128-z

Fern, A., *et al*., 2014. Do we need hundreds of classifiers to solve real world classification problems. *Journal Mach Learning Researcher*, 15 (1), 3133–3181.

Foley, J.A., *et al*. 2005. Global consequences of land use. *Science*, 309 (5734), 570–574. doi:10.1126/science.1111772

Frank, L.D., 1994. An analysis of relationships between urban form (density, mix, and jobs: housing balance) and travel behavior (mode choice, trip generation, trip length, and travel time). *Transportation Research Part A*, 1 (30), 76–77.

Gordon, I.R. and McCann, P., 2000. Industrial clusters: complexes, agglomeration and/or social networks? *Urban Studies*, 37 (3), 513–532. doi:10.1080/0042098002096

Guan, D., *et al*., 2011. Modeling urban land use change by the integration of cellular automaton and Markov model. *Ecological Modelling*, 222 (20), 3761–3772. doi:10.1016/j.ecolmodel.2011.09.009

Hu, S. and Li, D., 2004. Vector cellular automata based geographical entity. *Geoinformatics*.

Jiang, S., *et al*., 2015. Mining point-of-interest data from social networks for urban land use classification and disaggregation. *Computers, Environment and Urban Systems*, 53, 36–46. doi:10.1016/j.compenvurbsys.2014.12.001

Kamusoko, C. and Gamba, J., 2015. Simulating urban growth using a Random Forest-Cellular Automata (RF-CA) model. *ISPRS International Journal of Geo-Information*, 4 (2), 447–470. doi:10.3390/ijgi4020447

Kocabas, V. and Dragicevic, S., 2006. Assessing cellular automata model behaviour using a sensitivity analysis approach. *Computers, Environment and Urban Systems*, 30 (6), 921–953. doi:10.1016/j.compenvurbsys.2006.01.001

Lambin, E.F. and Geist, H.J., 2008. *Land-use and land-cover change: local processes and global impacts*. Berlin Heidelberg: Springer Science & Business Media.

Lau, K.H. and Kam, B.H., 2005. A cellular automata model for urban land-use simulation. *Environment and Planning B: Planning and Design*, 32 (2), 247–263. doi:10.1068/b31110

Lez, G.A., *et al*., 2015. Vector-based cellular automata: exploring new methods of urban growth simulation with cadastral parcels and graph theory. In: *Proceedings CUPUM*, Cambridge, MA USA.

Li, X., *et al*., 2017. A new global land-use and land-cover change product at a 1-km resolution for 2010 to 2100 based on human–environment interactions. *Annals of the American Association of Geographers*, 107 (5), 1040–1059.

Li, X. and Yeh, A.G., 2000. Modelling sustainable urban development by the integration of constrained cellular automata and GIS. *International Journal of Geographical Information Science*, 14 (2), 131–152. doi:10.1080/136588100240886

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. doi:10.1080/13658810210137004

Lin, J. and Li, X., 2016. Conflict resolution in the zoning of eco-protected areas in fast-growing regions based on game theory. *Journal of Environmental Management*, 170, 177–185. doi:10.1016/j.jenvman.2015.11.036

Liu, J. and Diamond, J., 2005. China's environment in a globalizing world. *Nature*, 435 (7046), 1179–1186. doi:10.1038/4351179a

Liu, X., *et al*., 2014. Simulating urban growth by integrating landscape expansion index (LEI) and cellular automata. *International Journal of Geographical Information Science*, 28 (1), 148–163. doi:10.1080/13658816.2013.831097

Liu, X., *et al*., in press. A Future Land Use Simulation Model (FLUS) for simulating multiple land use scenarios by coupling human and natural effects. Landscape and Urban Planning.

Long, Y., Gu, Y., and Han, H., 2012. Spatiotemporal heterogeneity of urban planning implementation effectiveness: evidence from five urban master plans of Beijing. *Landscape and Urban Planning*, 108 (2–4), 103–111. doi:10.1016/j.landurbplan.2012.08.005

Long, Y. and Liu, X., 2016. Automated identification and characterization of parcels (AICP) with OpenStreetMap and Points of interest. *Environment and Planning B: Planning and Design*, 43 (2), 341–360. doi:10.1177/0265813515604767

Lu, Y., *et al*., 2013. Redlines for the greening of China. *Environmental Science & Policy*, 33, 346–353. doi:10.1016/j.envsci.2013.05.007

Lu, Y., Cao, M., and Zhang, L., 2015. A vector-based Cellular Automata model for simulating urban land use change. *Chinese Geographical Science*, 25 (1), 74–84. doi:10.1007/s11769-014-0719-9

McGarigal, K., Cushman, S.A., and Ene, E., 2012. *FRAGSTATS v4: spatial pattern analysis program for categorical and continuous maps*. USA: University of Massachusetts, Amherst, Massachusetts. *goo. gl/aAEbMk*

Mitsova, D., Shuster, W., and Wang, X., 2011. A cellular automata model of land cover change to integrate urban growth with open space conservation. *Landscape and Urban Planning*, 99 (2), 141–153. doi:10.1016/j.landurbplan.2010.10.001

Moreno, N., Ménard, A., and Marceau, D.J., 2008. VecGCA: a vector-based geographic cellular automata model allowing geometric transformations of objects. *Environment & Planning B Planning & Design*, 35 (4), 647–665. doi:10.1068/b33093

Moreno, N., Wang, F., and Marceau, D.J., 2009. Implementation of a dynamic neighborhood in a land-use vector-based cellular automata model. *Computers Environment & Urban Systems*, 33 (1), 44–54. doi:10.1016/j.compenvurbsys.2008.09.008

O'Sullivan, D., 2001a. Exploring spatial process dynamics using irregular cellular automaton models. *Geographical Analysis*, 33 (1), 1–18. doi:10.1111/j.1538-4632.2001.tb00433.x

O'Sullivan, D., 2001b. Graph-cellular automata: a generalised discrete urban and regional model. *Environment and Planning B: Planning and Design*, 28 (5), 687–705. doi:10.1068/b2707

Palczewska, A., *et al*., 2014. Interpreting random forest classification models using a feature contribution method. In: *integration of Reusable Systems*. Springer International Publishing, 193–218.

Pang, M.Y.C. and Shi, W., 2002. Development of a process-based model for dynamic interaction in spatio-temporal GIS. *Geoinformatica*, 6 (4), 323–344. doi:10.1023/A:1020876609236

Parker, D.C., *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. doi:10.1111/1467-8306.9302004

Peng, Z., 1997. The jobs-housing balance and urban commuting. *Urban Studies*, 34 (8), 1215–1235. doi:10.1080/0042098975600

Perica, S. and Foufoula-Georgiou, E., 1996. Model for multiscale disaggregation of spatial rainfall based on coupling meteorological and scaling. *Journal of Geophysical Research-Atmospheres*, 101 (D21), 26–347. doi:10.1029/96JD01870

Pinto, N.N. and Antunes, A.O.N.P., 2010. A cellular automata model based on irregular cells: application to small urban areas. *Environment and Planning B: Planning and Design*, 37 (6), 1095–1114. doi:10.1068/b36033

Pontius Jr., R.G., *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. doi:10.1007/s00168-007-0138-2

Rahimi, A., 2016. A methodological approach to urban land-use change modeling using infill development pattern-a case study in Tabriz, Iran. *Ecological Processes*, 5 (1), 1. doi:10.1186/s13717-016-0044-6

Rauch, F., 2013. Cities as spatial clusters. *Journal of Economic Geography, Journal of Economic Geography*, 14 (4), 759–773. doi:10.1093/jeg/lbt034

Rienow, A. and Goetzke, R., 2015. Supporting SLEUTH-Enhancing a cellular automaton with support vector machines for urban growth modeling. *Computers, Environment and Urban Systems*, 49, 66–81. doi:10.1016/j.compenvurbsys.2014.05.001

Sang, L., *et al*., 2011. Simulation of land use spatial pattern of towns and villages based on CA–Markov model. *Mathematical & Computer Modelling*, 54 (3), 938–943. doi:10.1016/j.mcm.2010.11.019

Santé, I., *et al*., 2010. Cellular automata models for the simulation of real-world urban processes: a review and analysis. *Landscape & Urban Planning*, 96 (2), 108–122. doi:10.1016/j.landurbplan.2010.03.001

Schweitzer, F. and Steinbink, J., 1997. Urban cluster growth: analysis and computer simulations of urban aggregations. *World*, 101, 102.

Semboloni, F., 2000. The growth of an urban cluster into a dynamic self-modifying spatial pattern. *Environment & Planning B Planning & Design*, 27 (4), 549–564. doi:10.1068/b2673

Shi, W. and Pang, M.Y.C., 2000. Development of voronoi-based cellular automata -an integrated dynamic model for geographical information systems. *International Journal of Geographical Information Science*, 14 (5), 455–474. doi:10.1080/13658810050057597

Stevens, D. and Dragićević, S., 2007. A GIS-based irregular cellular automata model of land-use change. *Environment and Planning B: Planning & Design*, 34 (4), 708–724.

Stevens, D., Dragicevic, S., and Rothley, K., 2007. iCity: a GIS-CA modelling tool for urban planning and decision making. *Environmental Modelling & Software*, 22 (6), 761–773. doi:10.1016/j.envsoft.2006.02.004

Su, S., *et al*., 2012. Characterizing landscape pattern and ecosystem service value changes for urbanization impacts at an eco-regional scale. *Applied Geography*, 34, 295–305. doi:10.1016/j.apgeog.2011.12.001

Tepe, E. and Guldmann, J., 2017. Spatial and temporal modeling of parcel-level land dynamics. *Computers, Environment and Urban Systems*, 64, 204–214. doi:10.1016/j.compenvurbsys.2017.02.005

Tian, L. and Shen, T., 2011. Evaluation of plan implementation in the transitional China: a case of Guangzhou city master plan. *Cities*, 28 (1), 11–27. doi:10.1016/j.cities.2010.07.002

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society.Series B (Methodological)*, 267–288.

Turner, B.L., Lambin, E.F., and Reenberg, A., 2007. The emergence of land change science for global environmental change and sustainability. *Proceedings of the National Academy of Sciences*, 104 (52), 20666–20671. doi:10.1073/pnas.0704119104

Vanegas, C.A., *et al*., 2009. Visualization of simulated urban spaces: inferring parameterized generation of streets, parcels, and aerial imagery. *IEEE Transactions on Visualization and Computer Graphics*, 15 (3), 424–435. doi:10.1109/TVCG.2008.193

Wand, M.P. and Jones, M.C., 1994. *Kernel smoothing*. Boca Raton, FL USA: Crc Press.

Wang, F. and Marceau, D.J., 2013. A patch-based cellular automaton for simulating land-use changes at fine spatial resolution. *Transactions in GIS*, 17 (6), 828–846. doi:10.1111/tgis.2013.17.issue-6

Wang, H., *et al*., 2013. An integrated approach to supporting land-use decisions in site redevelopment for urban renewal in Hong Kong. *Habitat International*, 38, 70–80. doi:10.1016/j.habitatint.2012.09.006

Wasikowski, M. and Chen, X., 2010. Combating the small sample class imbalance problem using feature selection. *IEEE Transactions on Knowledge and Data Engineering*, 22 (10), 1388–1400. doi:10.1109/TKDE.2009.187

White, R. and Engelen, G., 2000. High-resolution integrated modelling of the spatial dynamics of urban and regional systems. *Computers, Environment and Urban Systems*, 24 (5), 383–400. doi:10.1016/S0198-9715(00)00012-0

Wickramasuriya, R., *et al*., 2011. An automated land subdivision tool for urban and regional planning: concepts, implementation and testing. *Environmental Modelling & Software*, 26 (12), 1675–1684. doi:10.1016/j.envsoft.2011.06.003

Wolpert, D.H., and Macready, W.G., 1999. An efficient method to estimate bagging's generalization error. *Machine Learning*, 35 (1), 41–55. doi:10.1023/A:1007519102914

Wu, D., *et al*., 2010. Simulating urban expansion by coupling a stochastic cellular automata model and socioeconomic indicators. *Stochastic Environmental Research and Risk Assessment*, 24 (2), 235–245. doi:10.1007/s00477-009-0313-3

Wu, F. and Martin, D., 2002. Urban expansion simulation of Southeast England using population surface modelling and cellular automata. *Environment and Planning A*, 34 (10), 1855–1876. doi:10.1068/a3520

Wu, H., *et al*., 2012. Quantifying and analyzing neighborhood configuration characteristics to cellular automata for land use simulation considering data source error. *Earth Science Informatics*, 5 (2), 77–86. doi:10.1007/s12145-012-0097-8

Yang, Q., Li, X., and Shi, X., 2008. Cellular automata for simulating land use changes based on support vector machines. *Computers & Geosciences*, 34 (6), 592–602. doi:10.1016/j.cageo.2007.08.003

Yang, X., Zheng, X., and Chen, R., 2014. A land use change model: integrating landscape pattern indexes and Markov-CA. *Ecological Modelling*, 283, 1–7. doi:10.1016/j.ecolmodel.2014.03.011

Yang, X., Zheng, X., and Lv, L., 2012. A spatiotemporal model of land use change based on ant colony optimization, Markov chain and cellular automata. *Ecological Modelling*, 233, 11–19. doi:10.1016/j.ecolmodel.2012.03.011

Yao, Y., *et al*., 2016. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. *International Journal of Geographical Information Science*, 31 (4), 825–848. doi:10.1080/13658816.2016.1244608

Yuan, J., Zheng, Y., and Xie, X., 2012. Discovering regions of different functions in a city using human mobility and POIs. *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 186–194.

Zhao, P., Lu, B., and De Roo, G., 2011. Impact of the jobs-housing balance on urban commuting in Beijing in the transformation era. *Journal of Transport Geography*, 19 (1), 59–69. doi:10.1016/j.jtrangeo.2009.09.008

Zheng, H.W., Shen, G.Q., and Wang, H., 2014. A review of recent studies on sustainable urban renewal. *Habitat International*, 41, 272–279. doi:10.1016/j.habitatint.2013.08.006