GIScience & Remote Sensing

Taylor & Francis
Taylor & Francis Group

ISSN: 1548-1603 (Print) 1943-7226 (Online) Journal homepage: http://www.tandfonline.com/loi/tgrs20

Multiple intra-urban land use simulations and driving factors analysis: a case study in Huicheng, China

Dachuan Zhang, Xiaoping Liu, Xiaoyu Wu, Yao Yao, Xinxin Wu & Yimin Chen

To cite this article: Dachuan Zhang, Xiaoping Liu, Xiaoyu Wu, Yao Yao, Xinxin Wu & Yimin Chen (2018): Multiple intra-urban land use simulations and driving factors analysis: a case study in Huicheng, China, GIScience & Remote Sensing, DOI: 10.1080/15481603.2018.1507074

To link to this article: https://doi.org/10.1080/15481603.2018.1507074

Published online: 06 Aug 2018.

GIScience& Remote Sensing, 2018
https://doi.org/10.1080/15481603.2018.1507074

Multiple intra-urban land use simulations and driving factors analysis: a case study in Huicheng, China

Dachuan Zhang [a], Xiaoping Liu [a*], Xiaoyu Wu [b], Yao Yao [c,d], Xinxin Wu [a] and Yimin Chen [a]

[a]*School of Geography and Planning, Guangdong Key Laboratory for Urbanization and Geo-simulation, Sun Yat-sen University, Guangzhou, Guangdong province, China;* [b]*Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing, China;* [c]*School of Information Engineering, China University of Geosciences, Wuhan, China;* [d]*Department of Data Technology and Products, Alibaba Group, Hangzhou, China*

(*Received 14 December 2017; accepted 24 July 2018*)

Simulations of intra-urban land use changes have gradually attracted more attention as these approaches are extremely helpful in regard to decision making and policy formulation. While prior studies mostly focused on methods of developing intra-urban level simulations, very little research has been conducted explain the factors driving intra-urban land use change. Urban planners are highly concerned with how inner-city structures are formed and how they function. Here, to simulate multiple intra-urban land use changes and to identify the contribution of different driving factors, we developed a random forests (RF) algorithm-based cellular automata (CA) simulation model. In this study, the model applied diverse categories of spatial variables, including traffic location factors, environmental factors, public services, and population density, as the driving factors to enhance our understanding of the dynamics of internal urban land use. The CA model was tested using data from the Huicheng district of Huizhou city in the Guangdong province of China. The Model was validated using actual historical land use data from 2000 to 2010. By applying the validated model, multiple intra-urban land use maps were simulated for 2015. Simultaneously, spatial variable importance measures (VIMs) were calculated by using the out-of-bag (OOB) error estimation approach of the RF algorithm. Based on the calculation results, we assessed and analysed the significance of each intra-urban land use driver for this region. This study provides urban planners and relevant scholars with detailed and targeted information that can aid in the formulation of specific planning strategies for different intra-urban land uses and support the future evolution of this area.

Keywords: cellular automata; urban planning; multiple intra-urban land use simulations; driving factors analysis; random forests algorithm

1. Introduction

Urbanization is regarded as one of the most intensive human development activities and has become the dominant factor that exerting tremendous pressure on social, economic and urban environmental sustainability (Bihamta et al. 2015; Braimoh and Onishi 2007). Increasing urban populations (particularly in developing countries) and accelerating urban land use change have critical anthropogenic effects on the natural environment and have become a global concern (Kalnay and Cai 2003). At the city level, accelerating

*Corresponding author. Email: liuxp3@mail.sysu.edu.cn

urbanization poses challenges on the urban development and redevelopment processes (Zheng et al. 2015). Consequently, the planning of urban structure, urban renewal and urban transportation in growing urban areas has become more complex (Leao, Bishop, and Evans 2004; Wise, Crooks, and Batty 2016). A comprehensive understanding of land use change mechanisms and dynamics is required for future urban evolution.

The advent of urban growth models creates possibilities for comprehending urban growth mechanisms. Such models also provide educational tools for the scientists, politicians and city planners who can benefit from the visualization of different growth scenarios (Berling-Wolff and Jianguo 2004). Amongst these urban growth models lie systems based on cellular automata (CA) techniques. CA models simulate geographic space as an arrays of cells on which each cell assumes a particular state based on its previous state and that of neighbouring cells according to a set of transition rules (Santé et al. 2010; White and Engelen 2000). CA models offer advantages in flexibility, intuitiveness, their ability to generate spatiotemporal patterns, and capability to simulate complex dynamic systems (Aburas et al. 2016; Santé et al. 2010). Numerous recent studies have explored the potential use of CA models for urban growth simulation (Liu et al. 2017; Torrens and O'Sullivan 2001), and numerous studies have used urban CA models to simulate and predict urban growth phenomena (Abdullahi et al. 2015; Al-Shalabi et al. 2013; Azari et al. 2016; Chen et al. 2014; Feng et al. 2011; Feng and Tong 2018; Han et al. 2009; He et al. 2008; Arsanjani et al. 2013; Lagarias 2012; Leao, Bishop, and Evans 2004; Liu et al. 2010a, 2014, 2018; Mitsova, Shuster, and Wang 2011; Omrani, Tayyebi, and Pijanowski 2017; Hossein and Helbich 2013).

Although many of these prior studies have focused on land cover/land use simulations at the national scale and within city cluster areas, there is a great potential for research on land use change simulations and mechanisms at the local scale since few of these simulations have been scaled down to the intra-urban level. Macro-scale land cover/land use simulations view cities as homogeneous entities for monitoring urban sprawl or ecological impacts of land cover/land use change (Kuang 2012). However, the intra-urban simulations are not aimed at addressing these issues but rather at understanding the urban intrinsic evolution through intra-urban land use maps, urban socioeconomic attributes and spatial analysis tools. Decision making at the intra-urban level is also different from that at the macro scale and must be improved for sustainability because the intra-urban simulation aims not only to show the recent urban land use change processes but also to explain how this trend can provide essential information for urban planners to begin possible future spatial configuration and urban revitalization (Godoy and Soares-Filho 2008). In view of this, some studies have tried to scale down their analyses to the intra-urban level and apply machine learning algorithms to simulate intra-urban land use change (Almeida et al. 2008; Benenson, Omer, and Hatna 2002; Godoy and Soares-Filho 2008; Partanen 2016; Zheng et al. 2015). However, these relevant studies focused on simulating intra-urban land use technically, and few of them studied land use change mechanisms sufficiently. For example, Almeida's study in 2014 integrated the ANN algorithm and the CA model to simulate intra-urban land use dynamics in Midwest São Paulo. Consequentially, the employment of a supervised back-propagation neural network in the parameterization of several variables was quite meaningful. However, the proposed ANN-CA model used in Almeida's research has weaknesses. Since the internal training process that is used by the ANN algorithm is not highly comprehensive and since this model can be easily overfit (Hung and Hung 2014), it is difficult to use the ANN algorithm to measure the importance of each variable, which makes identifying and analysing the driving factors more challenging.

Factually, increasing our knowledge regarding the underlying determinants of driving factors is important because it can improve our understanding of local-scale urban development (Luo and Dennis Wei 2009). For urban planners and decision makers, the phenomena of multiple internal urban land use changes and dynamics (e.g., residential, industrial, commercial, etc.) and the drivers of these changes are some of the most important issues (Braimoh and Onishi 2007). For example, urban planners and decision makers are highly concerned with how inner-city structures are formed and which factors drive the intra-urban land use changes. Clearly, effective planning strategies and sound policies can be formulated only with a full understanding of intra-urban land use change mechanisms. To address this important knowledge gap, during intra-urban simulations, multiple driving factors should be selected. On this basis, reliable methods that can explore and identify the dominant driving factors (and all other factors) are required for improving our understanding of the dynamics of specific internal urban land uses.

Hence, in this study, we combined the random forests (RF) algorithm and the CA simulation model for the purpose of simulating multiple intra-urban land use changes in Huicheng district, which is a developing area located in the Pearl River Delta (PRD) city cluster of China. The RF algorithm was used to mine the CA transition rules. Seventeen categories of spatial variables are involved in the implementation as the driving factors of land use changes in this area. In recent years, mobile location-based service (LBS) and multisource geospatial Big Data gathering techniques have been quickly developed, and data for points of interest (POIs) can be obtained across the network (Liu et al. 2015; Yao et al. 2017a). Previous research has shown that POIs datasets are related to various human activities at the micro scale and effectively reflect the distribution of urban structures and land use (Ratti et al. 2006; Yao et al. 2016). Therefore, in addition to traditional categories and population density, this study also introduced a series of public services POIs in spatial variables datasets. More importantly, the contribution of each driving factor of each land use can be identified through variable importance measures (VIMs) using the OOB error estimation of the RF algorithm (see Section 4.2 for more details). We obtained the raw land use dataset of the study area for 2000, 2005, 2010 and 2015 from the official planning department of Huicheng district. After developing and validating the model, the model was deemed appropriate for simulating multiple land uses at the intra-urban scale. By applying the validated model, the multiple intra-urban land use maps were finally simulated from 2000 to 2015. Simultaneously, the identification of the factors driving land use dynamics was conducted. The results and analysis of this study can aid in the formulation of specific planning strategies for different intra-urban land uses and support the future evolution of this area.

2. Methodology

For intra-urban simulation needs, the proper strategy is to scale down the simulation (using a higher spatial resolution and more detailed categories of land use data) and construct a simulation model to analyse the mechanisms, such as the driving forces of intra-urban land use change. Since previous CA models focused only on methods of constructing intra-urban level simulations, they cannot provide much valuable assistance for urban planners and researchers from the endogenous perspective. As a result, we constructed an intra-urban simulation model in this study and simulated the intra-urban land use changes from 2000 to 2015 by combining the RF algorithm and the CA method. The proposed model was calibrated and validated by comparing the simulated land use to the actual land use for 2010. The internal urban datasets used in this study include intra-urban land use maps and spatial variables datasets that were obtained from the local planning department. Because the land use maps were updated every five years, data for 2000, 2005, 2010, and 2015 were adopted and used in the simulation. Other driving factors include diverse public services variables that were captured from Baidu (the world's largest Chinese web search engine) POIs. To avoid the oversimplification of the reality being modelled, all spatial data used in this research were processed at a high spatial resolution of 10 metres. The flowchart for the methodology is provided in Figure 1.

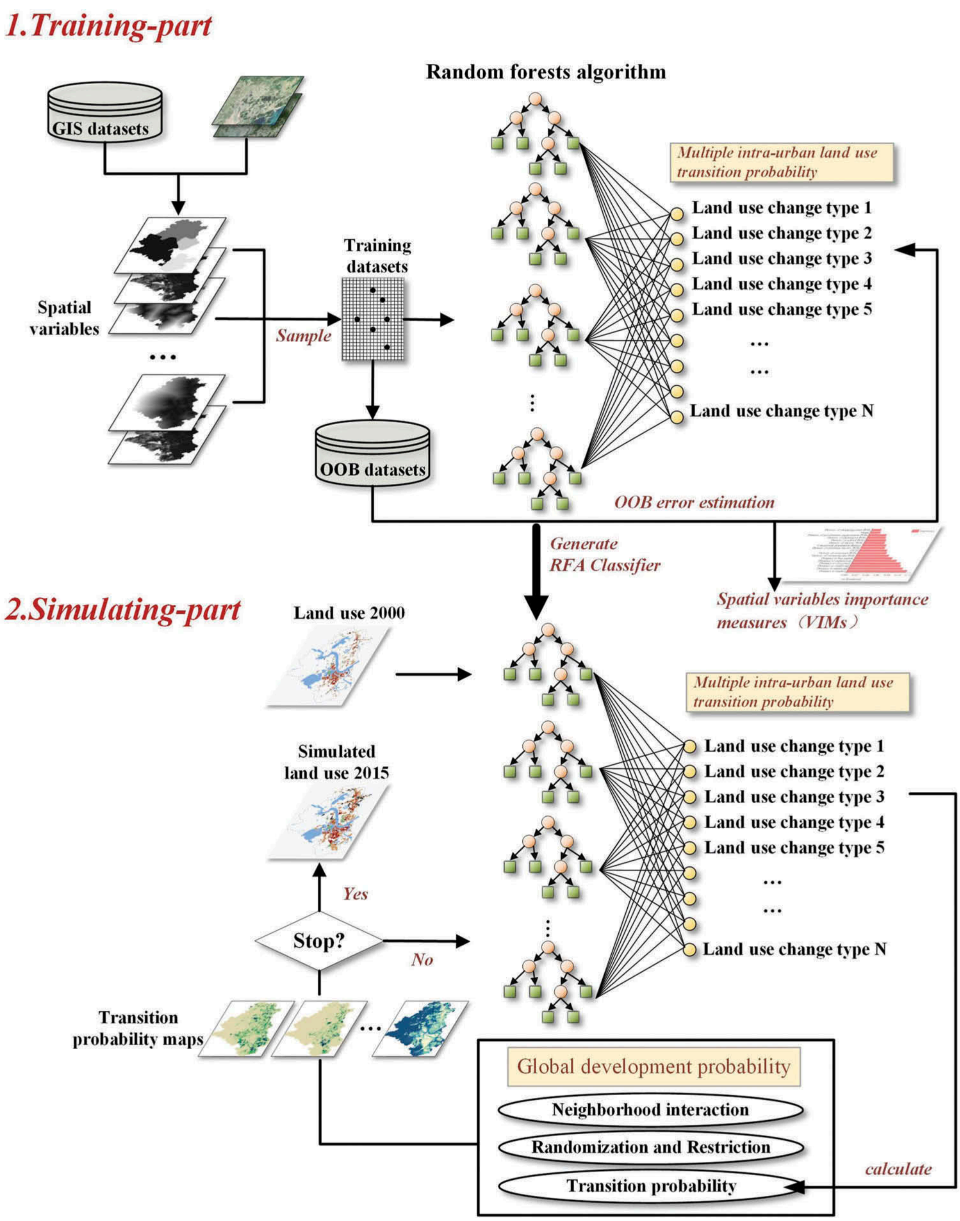

Figure 1. Structure of the RF-CA model.

2.1 RF-CA model

When using CA models to simulate intra-urban land use change, we must develop a nonlinear precise, less time consuming method that is difficult to overfit, includes a high tolerance for random variables and is able to reveal the mechanisms of multiple land use changes to mine the rules of intra-urban development and redevelopment. However, former applications combining machine learning algorithms with CA models have their respective shortcomings in simulating intra-urban land use change issues. More importantly, few methods can easily analyse the factors driving land use change. For instance, the spatial variables in traditional logistic regression are required to be linearly independent (Knol et al. 2012), while generally, most spatial variables used in a logistic-based CA model have a certain relationship (e.g., cells that are close to the urban centre are prone to be close to roads). Neural networks can solve the above problem, but the internal mechanism of the training process is not clear and is prone to overfitting, and the algorithm is time consuming and difficult to parallelize (Hung and Hung 2014). RF algorithms are some of the most suitable machine learning methods for land use/land cover issues (Gislason, Benediktsson, and Sveinsson 2006). Scholars have indicated that RF algorithms are capable of handling high data multicolinearity and dimensionality and are fast and intensive to overfitting (Belgiu and Lucian 2016, Wyner et al. 2017). In addition, recently, RF algorithms have been proven to be the most effective classifiers (Ndez-Delgado et al. 2014). Certain scholars have applied RF algorithms in CA models to simulate urban land use change (Du et al. 2017; Kamusoko and Gamba 2015; Yao et al. 2017b; Zhang et al. 2016). The RF algorithm is a decision tree-based ensemble classifier that can effectively analyse multi-class issues and is suitable for addressing categorical and numerical features, missing values, and redundant features (Reif et al. 2006). Therefore, this study employs an RF algorithm-based intra-urban CA model to simulate intra-urban land use changes. To summarize, we compared our model with other intra-urban CA models and other RF-CA models to show the improvement of our CA model (Table 1).

The RF-CA model in this study was divided into two components, namely, a training component and a simulation component. For the training component, an RF algorithm classifier is trained and calibrated using a training dataset that includes land use datasets and auxiliary geospatial variables datasets, which is created by the model. For the simulation component, the trained and calibrated classifier is used to

Table 1. Comparisons between our model and former CA models.

Research	Objective	Neighbourhood size	Cell space	VIMs
(Zheng et al. 2015)	MILU	Moore neighbourhood (3 × 3)	Not explicit	None
(Almeida et al. 2008)	MILU	Moore neighbourhood (3 × 3) and varying neighbourhood size	50 m	None
(Godoy and Soares-Filho 2008)	MILU	Not explicit	10 m	None
(Kamusoko&Gamba 2015)	DLC	Not explicit	30 m	Done
(Du et al. 2017)	MLC	Moore neighbourhood (3 × 3) and varying neighbourhood size	100 m	Done
Our research	MILU	7 × 7 neighbourhood and varying neighbourhood size	10 m	Done

Note: DLC represents dual state land cover issues (urban/non-urban), MLC represents multiple land cover issues, MILU represents multiple intra-urban land use issues calculate potential intra-urban multiple land use transitions and completes the simulation.

The RF algorithm is a multi-classifier combination algorithm that includes a large number of decision trees. Each decision tree can calculate an independent classification result sample, and then, the RF algorithm generates different training sub-datasets to increase the diversity between classifiers through feature selection in each sample. Therefore, the RF algorithm is considered an effective machine learning algorithm for minimizing the potential overfitting during the training process, which improves the predictive ability. When the multi-classifier RF algorithm has been generated using M training decision trees, Equation (1) measures the probability that unclassified data θ falls into each category, and Equation (2) generates the final classification results (Biau 2010; Breiman 2001).

$$P_i(x) = \frac{I(h_i(x) = Y_i)}{M} \tag{1}$$

$$H(x) = arg_{Y_i} max \sum_{i=1}^{M} I(h_i(x) = Y_i) \tag{2}$$

Specifically, $H(x)$ is the result of the multi-classifier combination algorithm, $h(x)$ is the result of a single decision tree, Y_i is the classification result of a single tree, and $I(\bullet)$ denotes the indicator function. The classification result of an RF algorithm is based upon a majority voting rule.

The global development probability P of each cell in this model consists of four factors (including the multi-class transition probability, neighbourhood interactions, stochastic factor and restriction), which is similar to ordinary CA models. In our study, the proposed model takes the advantage of the spectacular ability of the RF algorithm in multi-class classification to calculate each cell's transition probability. In Equation (1), $P_i(x)$ expresses the probability that unclassified data θ will be classified into category i. As a result, the probability that the N-class ($N > 2$) transition of cell k will be classified into category l at time t can be expressed by Equation (3).

$$P_g(k, t, l) = \frac{I(h_k(\theta) = Y_l)}{M} \quad l = (1, 2, 3, \ldots, N) \tag{3}$$

Neighbourhood interactions are some of the primary elements in land utilization simulation models that are based on CA (Shafizadehmoghadam et al. 2017; Verburg et al. 2004a). Generally, traditional urban CA models adopt a 3 × 3 Moore neighbourhood, which is more appropriate for dual-state issues (urban/non-urban). For a multi-classes issue, a larger neighbourhood should be used to better reflect the influence of the surroundings on the central cells. This research adopted a larger Moore neighbourhood to fully consider the possibility that the central cell shifts to all other land use categories. However, using an excessively large neighbourhood may cause fewer differences between the amounts of all land use categories in this neighbourhood, which reduces the sensitivity of the CA model's neighbourhood interaction (and thus reduces the simulation's accuracy). In addition, since human activities are influenced by wider spaces, an appropriate definition of the neighbourhood is essential (Li& Wenkai 2015; Verburg et al. 2004b; White and Engelen 2000). Therefore, this study conducted constant experimentation in the model calibration and validation procedures (mentioned in Section 3.2) to determine the effects of neighbourhood size on the model performance. Thus, we define an expression $\Omega_{k,l}^{t}$ to represent the neighbourhood interaction for cell k transforming into category l at time t. As Equation (4) shows, C_k represents the land use category of cell k, and $S_{k,l}^{t}$ is a discriminant function through assignment to judge whether C_k is l.

$$S_{k,l}^{t} = \begin{cases} 1, ifC_k = l \\ 0, ifC_k \neq l \end{cases} \quad \Omega_{k,l}^{t} = \frac{\sum_{n \times n} S_{k,l}^{t-1}}{n \times n - 1} l = (1, 2, 3, \ldots, N) \tag{4}$$

In this equation, the function of $\sum_{n \times n} S_{k,l}^{t-1}$ is to sum up the values of $S_{k,l}^{t-1}$ to count the number of cells that belong to the l-th land use category in a $n \times n$ neighbourhood after the last iteration. On the basis of knowing the current category of the central cell, Ω_k^t has N values of $\Omega_{k,l}^{t}$ that correspond to N cases of variable l.

Scholars have shown that land use transformation is influenced by some complex and stochastic factors (Wu and Martin 2002). We introduced the stochastic factors (R_a) to control the effect of model stochastic perturbation in our research. The stochastic factors are expressed as $1 + (-\ln\gamma)^{\alpha}$, where γ is a uniform random variable that ranges from 0 to 1 and α ranges from 0 to 10. CA models often use restriction (the factor R_s is used in this model) or constraints to represent the particular land use types that do not change to others during the simulation (Lin and Xia 2016). In this study, water bodies and urban road systems are selected as restricted development areas.

Within each iteration, every land use class will typically lose (or gain) some of its land to one or more of the other classes, and this process is decided by the global development probability. The global development probability is multivalued and is a concrete reflection of the complexity of simulating multiple land use changes. Equation (5) expresses the global intra-urban development probability of cell k at time t transforming into category l, which multiplies the multi-class transition probabilities, neighbourhood interactions, randomization and restrictions in the CA model. If there are N categories of land use that are included in the simulation with no restriction limits, then category l will include N values and $P(k, t, l)$ also corresponds to N values. Each cell k at time t can only transform into one land use category. Therefore, the global development probability for cell k is the maximum value in $NP(k, t, l)$ and can be expressed as shown in Equation (5).

$$\begin{aligned} P(k,t) &= \max(P(k,t,l)) = \max\left(P_g(k,t,l) \times \Omega_{k,l}^{t} \times R_a \times R_s\right) l = (1, 2, 3, \ldots, N) \\ &= \max\left(\frac{I(h_k(\theta) = Y_l)}{M} \times \frac{\sum_{n \times n} S_{k,l}^{t-1}}{n \times n - 1} \times (1 + (-\ln\gamma)^{\alpha}) \times R_s\right) \end{aligned} \tag{5}$$

2.2 Variable importance measures

The RF-based regression model is capable of calculating the contribution weights of the involved auxiliary variables contribution weights, namely, the VIMs (Biau 2010; Breiman 2001; Palczewska et al. 2013).

Specifically, VIMs are calculated in the training component of RF-CA model. The RF algorithm creates the training dataset Xi (i = 1,2,. . .,n trees), which includes n tree items, by using bootstrapping with replacement in the training component. *Xi* includes approximately 66% of the data from the initial dataset *X*, and the remaining 34% of the data create an out-of-bag (OOB) dataset. Specifically, we can obtain an OOB-based estimation error report for each decision tree and then measure the importance of the variables. The mechanism that conducts the OOB estimation to measure the importance of the variables in the RF algorithm can be generalized as follows (Breiman 2001). First, consider that *V* input variables exist for each OOB. After each decision tree in the RF is constructed, we calculate the OOB estimation errors (expressed as OOBerror1) using the corresponding OOB. Next, the process is repeated for $v = 1,2,\ldots, V$. Then, we calculate an additional OOB estimation error (expressed as OOBerror2) based on increasing the value of the *v*-th variable in each OOB. Finally, we compare and identify the difference between OOBerror1 and OOBerror2. The quantified result of this difference can be expressed as the importance of the *v*-th variable.

2.3 Model accuracy assessment

The RF-CA model adopts the “figure of merit” (FoM) indicator and the *kappa* statistic to estimate the simulation results of the model by comparing the model results to historical data. For the level of the cell, a “FoM” indicator has been commonly used by numerous scholars (He et al. 2015). The “FoM” indicator is a ratio that is included in Equation (6). The numerator (*B)* represents the number of samples that are observed and simulated correctly. The denominator (*A)* represents the error due to observed developments and is simulated as the persistence. *C* is defined as the error that is caused by the observed developments and is simulated as an incorrect gaining category. *D* represents the error caused by observed persistence and is simulated as developed (He et al. 2015; Pontius et al. 2008).

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

The *kappa* statistic and overall accuracy are other reliable assessment indices for land use simulation models (Vliet et al. 2016), particularly for multi-class simulations. When using the *kappa* statistic to assess the results of a multi-class simulation, a confusion matrix must be constructed, and then, Equation (7) is calculated using the values from the confusion matrix. For Equation (7), p_{ik} represents the rate that cells are simulated in category *j*, and p_{ki} represents the rate of cells for category *j* in the actual map. p_{ij} represents the proportion of cells that are simulated in category *i* but are actually in category *j* of the actual map. The assessment result depends on the value of Equation (7). Generally, the simulation results are good when the value is greater than 0.7, and a higher value implies a superior result (Sim and Wright 2005).

$$Kappa = \frac{\sum_{i=1}^{j} p_{ii} - \sum_{i=1}^{j} p_{ik} \times p_{ki}}{1 - \sum_{i=1}^{j} p_{ik} \times p_{ki}} \tag{7}$$

3. Implementation

3.1 Study area and data

3.1.1 Study area

The proposed model was tested in Huicheng district, which is in the centre of Huizhou city in the PRD of the Guangdong province, China. Huizhou, the 4th biggest city in the

PRD economic zone, is the regional economic centre of the PRD. Huizhou possesses large-scale electronic information estates and a number of manufacturing industries based on petroleum and chemical engineering. With the rapid economic development over the past decade, Huizhou is now a representative city of the rapid urbanization phenomenon in China. In addition, the Huicheng district, which is the downtown area of Huizhou, is the focus of local governmental regional planning efforts. Huicheng district has a population of approximately 1.1 million people and a land area of 27.8 km^2. The population grew at an annual average rate of 7% over the past 5 years according to the most recent official census. In addition, the intra-urban land use categories for this district include residential, administration and public services and industrial and commercial land use. Because of the potential urban development in Huicheng district, the land use types are continuously changing. After a long period of urbanization (Figure 2), the local

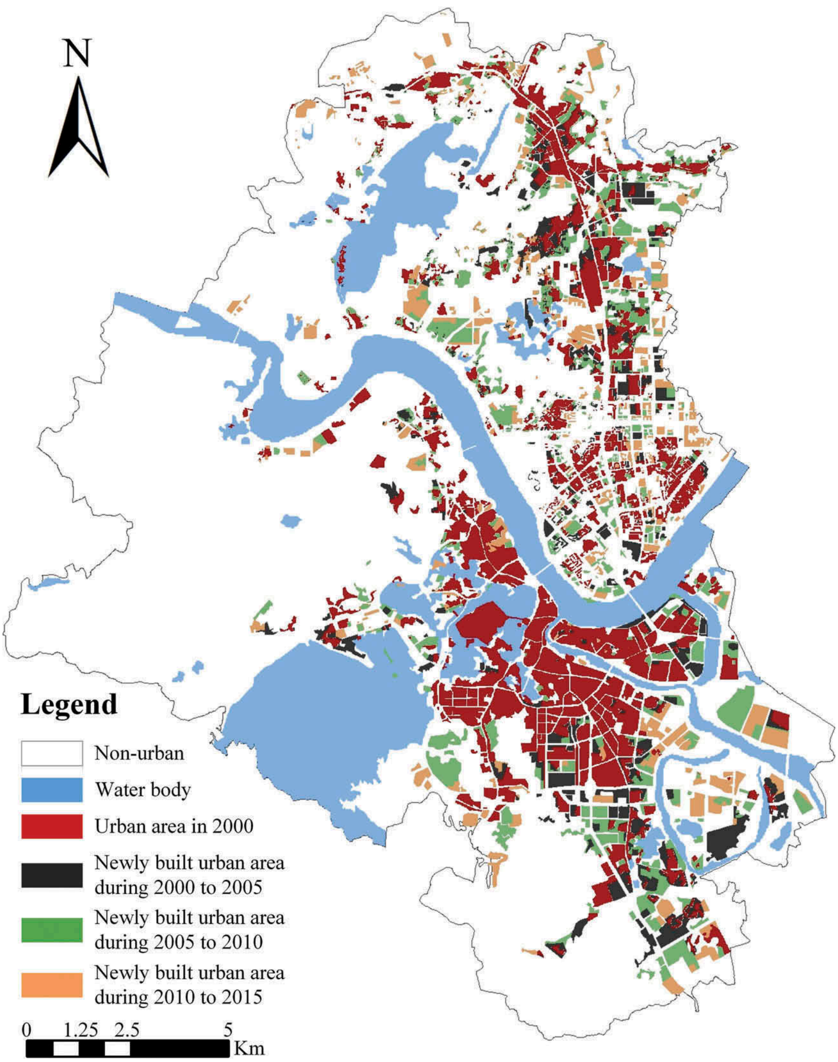

Figure 2. Trends of intra-urban land-use change from 2000 to 2015 in Huicheng district.

government is now facing the challenge as to how to formulate specific planning strategies for future urban evolution. Given this challenge, we decided to choose Huicheng district for testing our model as a specific case to assist urban planners in better comprehending the forces driving urban land use change.

3.1.2 Data and data source

The first step in intra-urban CA simulation is to develop internal urban datasets. Ensuring that the data are reliable is important due to a lack of fine-scale spatially consistent datasets, especially datasets with detailed intra-urban structure changes that reflect urban socioeconomic attributes (Kuang 2012; Xie et al. 2007). Similar studies have indicated that government departments and government research projects are common and reliable sources for data (Almeida et al. 2008; Godoy and Soares-Filho 2008; Zheng et al. 2015). This study obtained a raw land utilization dataset for the study area for 2000, 2005, 2010 and 2015 from the official planning department of Huizhou.

The initial land utilization dataset included 13 classes (Table 2). To simplify the initial dataset, the data were generalized and reclassified into 5 categories and similar land use types were reclassified into the same category. Water bodies were added, and the final experimental land use map included 6 classes. Every urban use map includes 4,016,732 cells with 1,798 rows and 2,234 columns. All the data layers applied in our study were registered to the Universal Transverse Mercator (UTM) coordinate system. All data were resampled at a 10 m × 10 m resolution.

CA models always repeat the selection of spatial variables even though their rules are quite different. For most CA models, the common spatial variables selected include road accessibility, distance to urban centres, environmental factors, planning factors, and population density (Santé et al. 2010). Previous studies indicated that these variables can reasonably reflect urban planning and human activities (Chen et al. 2017; Long and Liu 2015). Therefore, in this paper, we adopt various traffic-location factors, slope, elevation, and population density as part of the spatial variable datasets. To our knowledge, because of the limitations in data sources, certain public services-factors were not widely considered in previous CA simulations. With the growing

Table 2. Reclassification and generalization of land-use categories.

Initial land use categories	Spatial resolution	Generalized land use categories
Commercial land use	10 metres	Commercial land use (CL)
Industrial land	10 metres	Industrial land use (IL)
Residential land use Village residential land use High-rise residential land use	10 metres	Residential land use (RL)
Sports land use Cultural land use Sanitary land use Educational land use Land-use for public facilities Administration land use	10 metres	Administration and public services land use (APL)
Non-urban	10 metres	Non-urban (NU)
Water body	30 metres	Water body

popularity of mobile LBS and multisource geospatial Big Data gathering techniques, the data for POIs can be obtained across the network, which provides us with abundant information that reflects the distribution of public services (Liu et al. 2015; Sun et al. 2011; Yao et al. 2017a). Therefore, in addition to the basic selection of variables, several POIs datasets that represent public services factors were applied in our study. We retrieved Baidu's POIs datasets by using geospatial Big Data techniques and these POIs include public services such as hospitals, parks, bus stations, factories, and business establishments. In all, this study involved 17 spatial variables, including traffic-location factors, public services, natural factors, and population density as illustrated in Table 3. Figure 3 illustrates the raster layer for each auxiliary spatial variable that is used in this study.

3.2 Model calibration and validation

In this study, we built the RF-CA model by using C/C++ in Microsoft Visual Studio 2012. The RF algorithm was implemented by using C++ with the open-source C/C++ library Alglib 2.9 package (http://www.alglib.net/). The related applications will be released on the GeoSOS website (http://geosimulation.cn/).

Through calibration, the RF-CA model was trained, and certain parameters were properly set. The size of the neighbourhood in this CA model was calibrated based on constant experiments (Table 4). As Table 4 shown, from 3 × 3 to 7 × 7, the overall accuracy and *kappa* statistic value increased slightly as the neighbourhood size increased. It is because the larger window size can better consider the possibility that the central cell shifts to the existing land use categories in the neighbourhood window. While using an excessively large neighbourhood (for example 9 × 9) may reduce the sensitivity of the CA model's neighbourhood interaction (and then reduces the simulation's accuracy). For all land use types considered the neighbourhood tends to become less specific with distance from the central cell. As a consequence, the size of the neighbourhood in this CA model was set to 7 × 7 based on the better model performance.

Additionally, some parameters and settings were also calibrated and optimized using the calibration procedure. We divided the training datasets into the training data and validation data, and the shares were set to 60% and 40%, respectively, to ensure the fitting accuracy and stability of this model. We also established 80 decision trees and 20% OOB data and cross-validated the model with boosted random sampling.

After the model calibration, the model was then validated by comparing the simulated land use map for 2010 and the actual land use map for 2010. The simulated map for 2010 was constructed based on the land use map in 2000 (See in Figure 4). Figure 4(a) is the actual land use map for 2000, Figure 4(b) is the actual land use map for 2010, and Figure 4(c) is the simulated land use map for 2010.

The model validation results indicate that the "FoM" indicator value is 24.31%, as indicated by Table 5 and the confusion matrix in Table 6. The *kappa* statistic is 0.7063, and the overall accuracy (OA) is 89.2%. The results of the "FoM" indicator and the *kappa* statistic indicate that the simulation accuracy of the model has met the simulation accuracy requirement.

Table 3. Details and generalization of each spatial variable.

Spatial Variables	Source	Initial Resolution	Definition	Categories
Distance to main road	Official planning department	10 m	Euclidean distance relation to main roads	Traffic-location factors
Distance to small road		10 m	Euclidean distance relation to small roads	
Distance to expressway		10 m	Euclidean distance relation to expressways	
Distance to city centre		10 m	Euclidean distance relation to the city centre	
Distance to market	Baidu Map POIs	10 m	Euclidean distance relation to markets	
Distance to bus station		10 m		
Density of factory POIs		10 m	The density of factory-points around each cell	Public services factors
Density of entertain facility POIs		10 m	The density of entertain facility-points around each cell	
Density of institutional facility POIs		10 m	The density of institutional facility-points around each cell	
Density of school POIs		10 m	The density of school-points around each cell	
Density of parking lot POIs		10 m	The density of parking lot-points around each cell	
Density of shopping centre POIs		10 m	The density of shopping centre-points around each cell	
Density of restaurant POIs		10 m	The density of restaurant-points around each cell	
Density of government organization POIs		10 m	The density of government organization-points around each cell	
Population density	National population census	Sub-district statistic data	Urban population density of year 2000,2010	Population factors
Elevation	DEM generated by China Geospatial Data Cloud	30 m	Terrain condition of elevation	Natural factors
Slope		30 m	Terrain condition of slope	

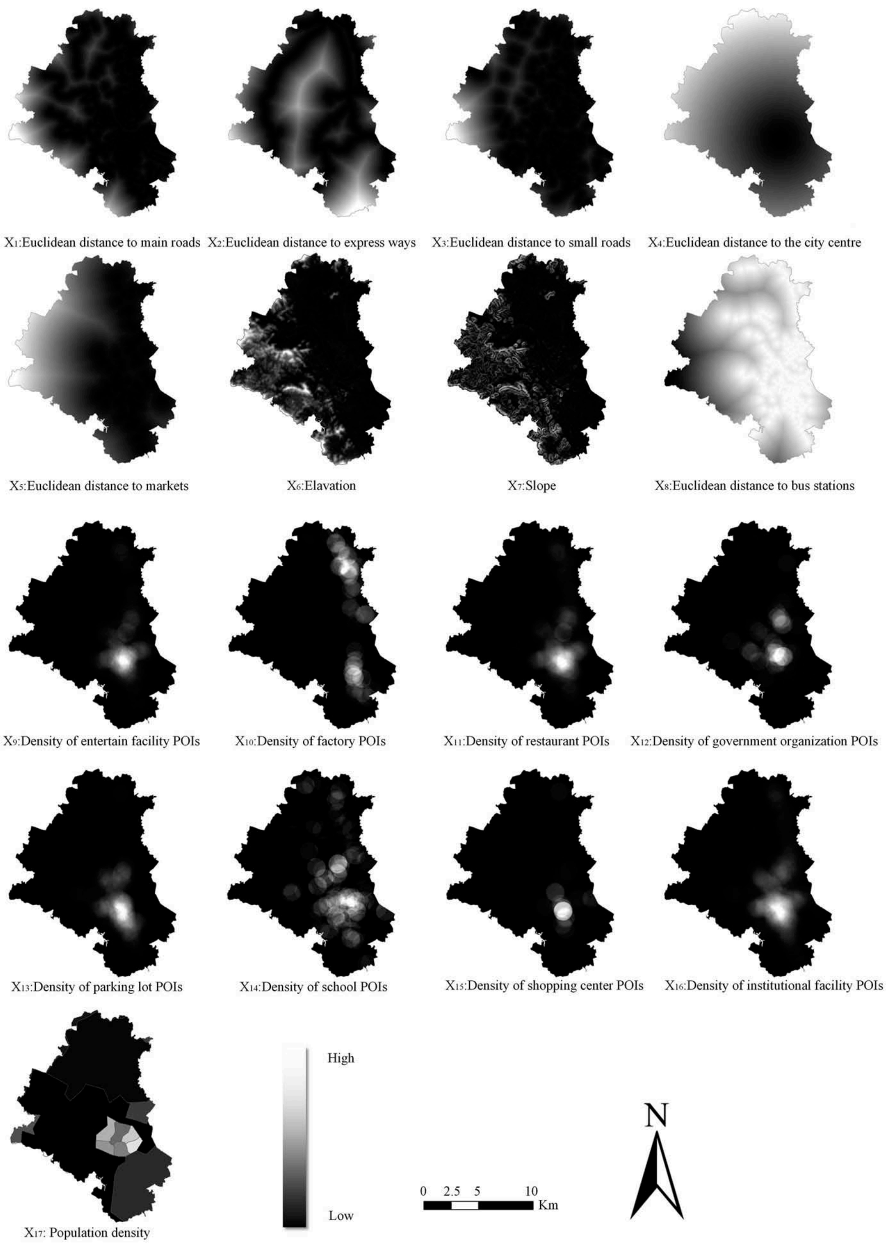

Figure 3. Raster layers of auxiliary spatial variables in this study.

4. Results

4.1 Simulation results and accuracy assessment

To obtain the simulation results for multiple intra-urban land uses after validating the model, we constructed 5 transition probability maps of multiple land uses using the RF algorithm of the simulation component of the RF-CA model (see Figure 5). These maps visualized the transition probability of 5 intra-urban land use types after the minimum-maximum

Table 4. Calibration of model neighbourhood's size.

Size of Neighbourhood	3 × 3	5 × 5	7 × 7	9 × 9	11 × 11
Overall Accuracy	88.16	88.35	89.15	84.53	83.38
Kappa Index Value	0.6997	0.6983	0.7039	0.6726	0.641

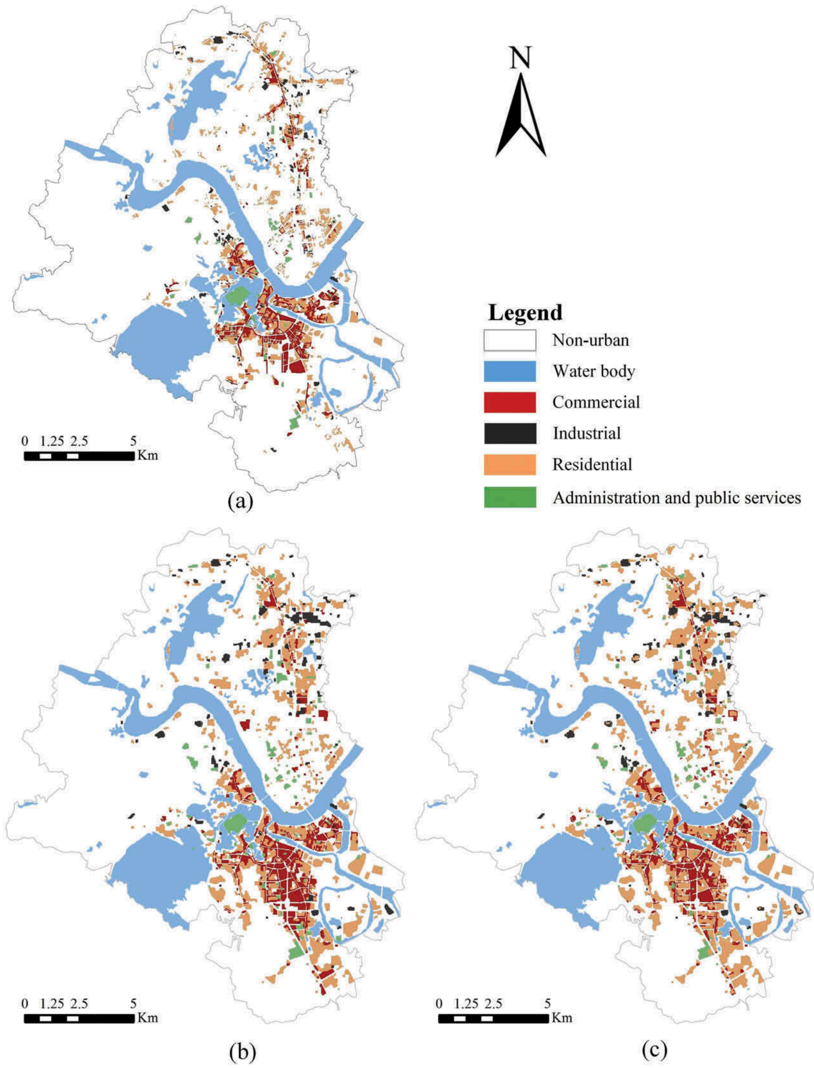

Figure 4. The real and simulated intra-urban land-use maps in 2010. (a) Real land-use map of year 2000. (b) Real land-use map of year 2010. (c) Simulated land-use map of year 2010.

normalization (min-max scaling) method. This process makes it easy for the simulation to accurately capture the distribution of each land use category and to compare the value of the probability for each land use in every raster of the probability map. Next, by continuously running the simulation component of the RF-CA model, multiple intra-urban land uses in

Table 5. The “FoM” indicator value of model validation.

Year:	A	B	C	D	“FoM” value
2000–2010	116,477	129,278	128,913	157,055	24.31%

Table 6. The confusion matrix of accuracy for model validation.

	Reality (Unit: percent)					
Simulation (Unit: percent)	0	1	2	3	4	Total
0	73.87	2.56	1.26	0.16	0.82	78.68
1	3.01	7.40	0.25	0.05	0.42	11.12
2	0.61	0.04	2.57	0.01	0.06	3.30
3	0.31	0.04	0.04	1.30	0.03	1.73
4	0.77	0.19	0.08	0.02	4.11	5.17
Total	78.58	10.23	4.20	1.55	5.44	100.00

Note: 0 represents non-urban land, 1 represents residential land use, 2 represents administration and public services land use, 3 represents commercial land use, 4 represents industrial land use

2015 were simulated (see Figure 6) based on multiple global development probabilities, which were calculated by integrating the transition probability, neighbourhood interactions, randomization and restrictions.

Each simulated land use displayed on the simulation map (Figure 6(b)) presented various distinguishing features. Residential land use was the most significant land use in the Huicheng district and was modelled acceptably. In the southern region of Huicheng district, residential land use had more vitality and ductility, and the residential areas of this region occupied a relatively large space. The shape and location of administration and public services land uses were simulated well and these regions was scattered around Huicheng district. Industrial land use was simulated properly at the industrial estate in the northeastern part of Huicheng district. Factories are clustered and industrial land use was more likely to be located in this area. In addition, the commercial land use was displayed in detail, particularly in the southern region of Huicheng district, which was the oldest and largest business area. Furthermore, the simulated commercial land use had prominent morphological similarities to the distribution of the actual map. (Figure 6(a)).

To assess the simulation results obtained by the RF-CA model, this study used the FoM indicator and the *kappa* statistic value. The simulation results indicate that the FoM indicator is 26.89% (Table 7), the *kappa* statistic value is 0.7306, and the OA is approximately 89.56% (Table 8). Therefore, when considering that this study simulated multiple land uses at an intra-urban scale rather than simple dual-state issues (urban/non-urban) at a macro scale, the simulation results and accuracies are quite acceptable.

4.2 Spatial variable importance measures and the identification of driving factors

When using CA models to simulate land use change, the spatial variable datasets in CA models are equivalent to driving factors for land use transformations. In addition, to implement the CA model, several relevant auxiliary spatial variables that represent the driving factors for land-use conversions are required (Conway and Wellen 2011; Lin and

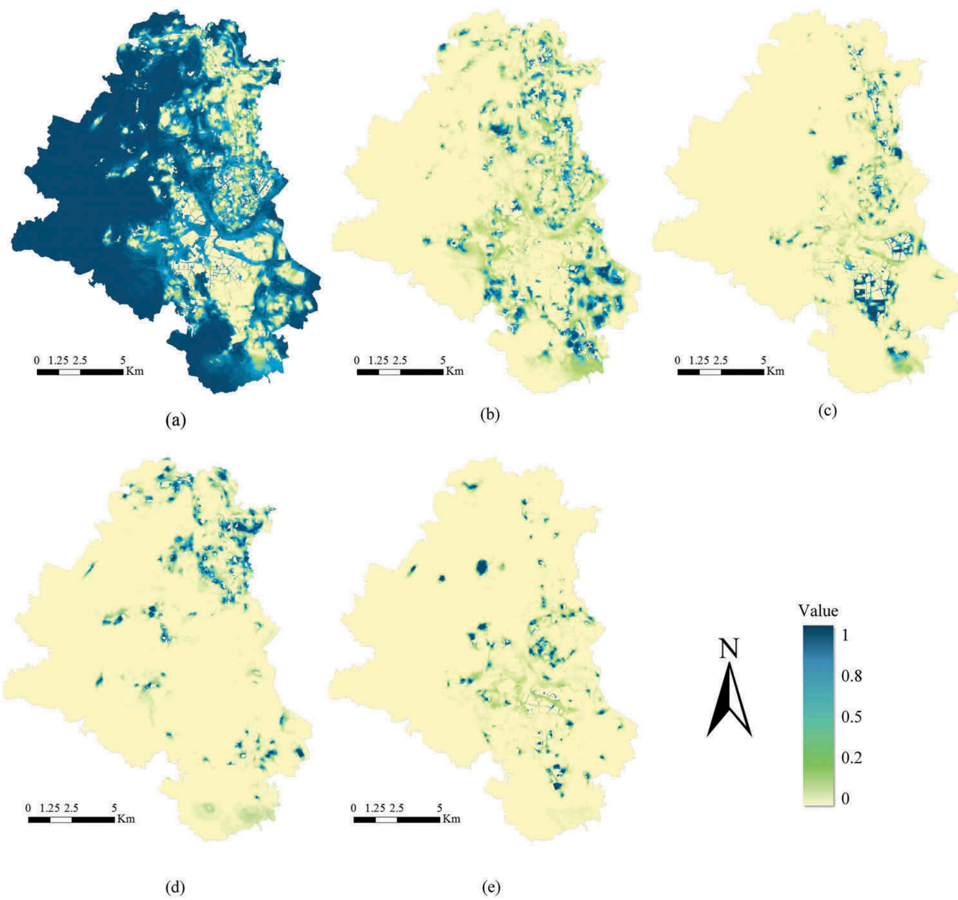

Figure 5. Transition probability maps of each land-use categories in 2015. (a) Probability of non-urban land-use. (b) Probability of residential land-use. (c) Probability of commercial land-use. (d) Probability of industrial land-use. (e) Probability of administration and public services land-use.

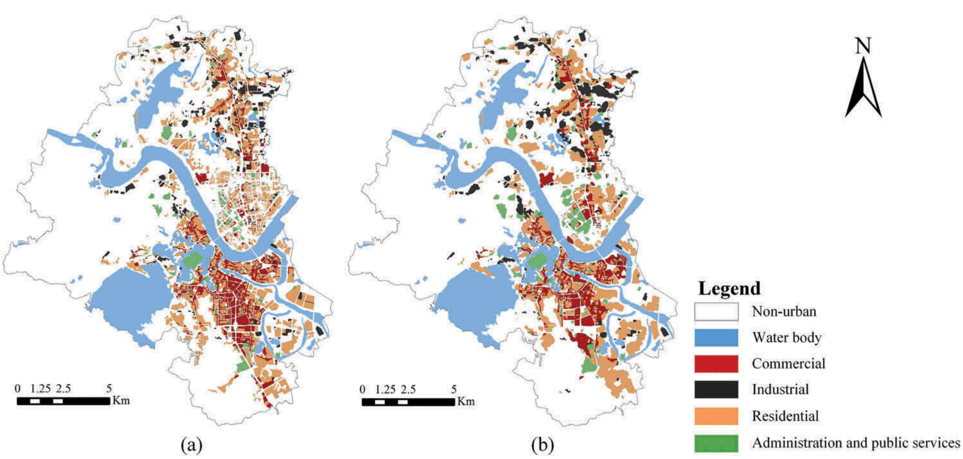

Figure 6. Simulation results of intra-urban land-use in 2015. (a) Real land-use map of year 2015. (b) Simulated land-use map of year 2015

Table 7. The “FoM” indicator value for simulation of 2010 to 2015.

Year:	A	B	C	D	“FoM” Value
2010–2015	122,261	135,552	139,191	107,056	26.89%

Table 8. The confusion matrices of accuracy for simulations of 2010 to 2015.

	Reality (unit: percent)					
Simulation (unit: percent)	0	1	2	3	4	Total
0	72.12	2.62	0.79	0.91	1.09	77.54
1	2.45	10.04	0.02	0.31	0.02	12.83
2	0.46	0.07	1.48	0.00	0.00	2.01
3	0.60	0.16	0.02	4.15	0.01	4.94
4	0.75	0.13	0.01	0.00	1.78	2.68
Total	76.39	13.02	2.32	5.37	2.90	100.00

Note: 0 represents non-urban land, 1 represents residential land use, 2 represents administration and public services land use, 3 represents commercial land use, 4 represents industrial land use

Xia 2016; Wu and Martin 2002). Hence, the measurements of spatial variable importance in CA models is coincident with the identification of driving factors acting on land use changes, and RF algorithms can accomplish both of these processes simultaneously.

From the perspective that the simulation was scaled down to the internal urban level, intra-urban spatial variables should be identified and analysed because they are viewed as driving factors that reflect land use dynamics and urban spatiotemporal patterns. We did not doubt that the land use types would transform quickly considering that the city was experiencing a period of rapid urbanization. However, these transformations were influenced by many relevant driving factors, particularly a series of urban spatial variables. Among these driving factors, certain factors were dominant, and others were inconsequential to intra-urban land use changes. As previously mentioned, the RF algorithm can obtain the OOB error to estimate the accuracy of the RF algorithm classifier and measure the importance of the spatial variables. For each intra-urban land use category, by using the trained RF algorithm classifier, the importance ranking of each spatial variable that was used in the simulation was calculated (see Figure 7).

Figure 7(a) illustrates that the distance to city life-related structures (markets, infrastructure POIs, restaurant POIs, etc.) and the distance to certain traffic location factors are more important for residential land use changes. Areas that are closer to downtown areas or key roads provide exceptional advantages for travelling to work or recreation and enjoying the convenience of city life-related facilities, which makes these spaces attractive for citizens to live in.

According to Figure 7(b), with regard to commercial land use, the distance to the city centre has the most influence on commercial land use changes. The city centre has the advantage of land value and attracts numerous citizens to purchase items in this area, which increases the attractiveness of business investments here. This phenomenon is evidenced by the categorical population density, which is the secondary factor in the measurement and constitutes 30 percent of the importance when considered with the distance to the city centre.

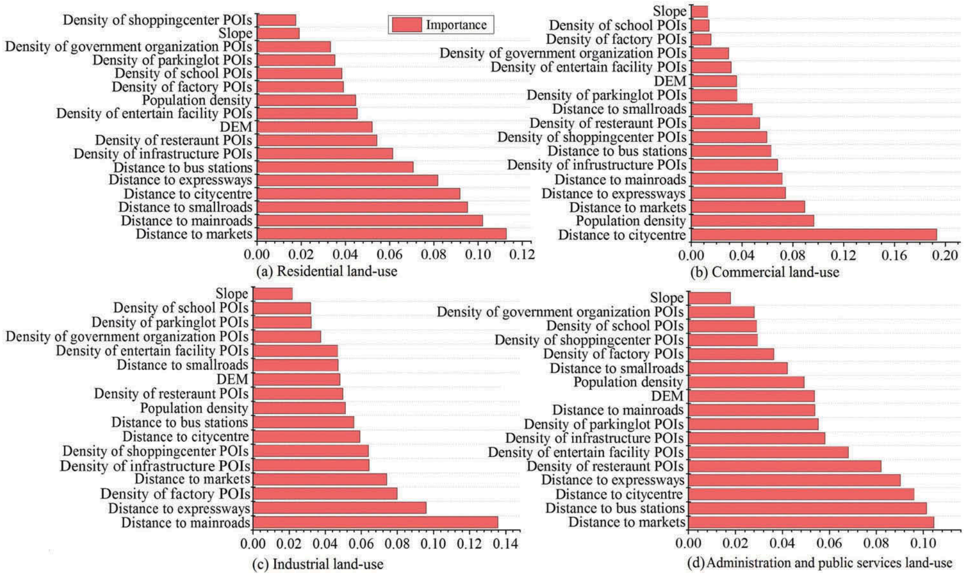

Figure 7. Spatial variables importance measures for each land-use category.

Figure 7(c) indicates that transportation factors (including the distance to the expressway and the main road) and the density of factory POIs contribute nearly 30 percent of the importance for industrial land use changes. In the Huicheng district, industry primarily focuses on manufacturing and is generally located in scattered industrial areas. The distribution and density of factory POIs is relatively important for industrial land use. In addition, improved transportation factors can promote the flow of productive elements, reduce transportation costs, and accelerate the spatial agglomeration of manufacturing.

Because of the distance decay, if the immobile public goods are distributed at a farther distance, the residents will be less likely to use their (potential) benefit (Gregory et al. 2009). Some particular public facilities (hospitals, stadiums, libraries, etc.) and institutions have their own geographic service areas (GSAs) with circular buffers surrounding each facility for better coverage (Donnelly 2013; Kong et al. 2017; Xiong and Luo 2017). Therefore, the administration and public services land use is located near where the residents live or where there is high transportation accessibility (Abdullahi et al. 2015) to conveniently provide public services for residents. As Figure 7(d) indicates, transportation factors (including the distance to bus stations and to the express-way) and location factors (including the distance to city centre and the distance to market) are relatively important factors driving the public service land use transformation.

Urban planners may not pay special attention to relatively unimportant factors, but rather care more about how the different dominant driving factors act on each land use. Particularly, as the distance decreases, urban planners also consider whether the influence of dominant factors driving the land use change is spatially varying. Therefore, we adopt two indices here to analyse and explain this issue quantitatively. We use buffer analysis to build different buffers around the target driving factor features in the GIS analysis software. The buffers are zones or rings with specific distances around the target driving factor (Liu et al. 2010b). The first index (expressed as R_1) can be understood as the ratio of land change pixels in each buffer zone (A_i) to the area of each buffer zone (S_i). The second index (R_2) expresses the ratio of the land change pixels in each buffer zone (A_i) to the total of land change pixels in all buffer zones. The expressions are illustrated as follows:

$$R_1 = \frac{A_i}{S_i};\ R_2 = \frac{A_i}{\sum_{i=1}^{n} A_i} \qquad i = (1, 2, \ldots, n) \tag{8}$$

For each dominant driving factor, we calculated both R_1 and R_2 to quantitatively analyse the influences that these factors have on each land use category.

Road network – residential land use

We analysed the relationship between the distance from the Huicheng road network and residential land use change by calculating the R_1 and R_2 indices (Figure 8(a)). The results show that the region approximately 100 metres away from the road network is most attractive for residential estates and investments. This attraction decreases as the distance increases and the two indices vary accordingly.

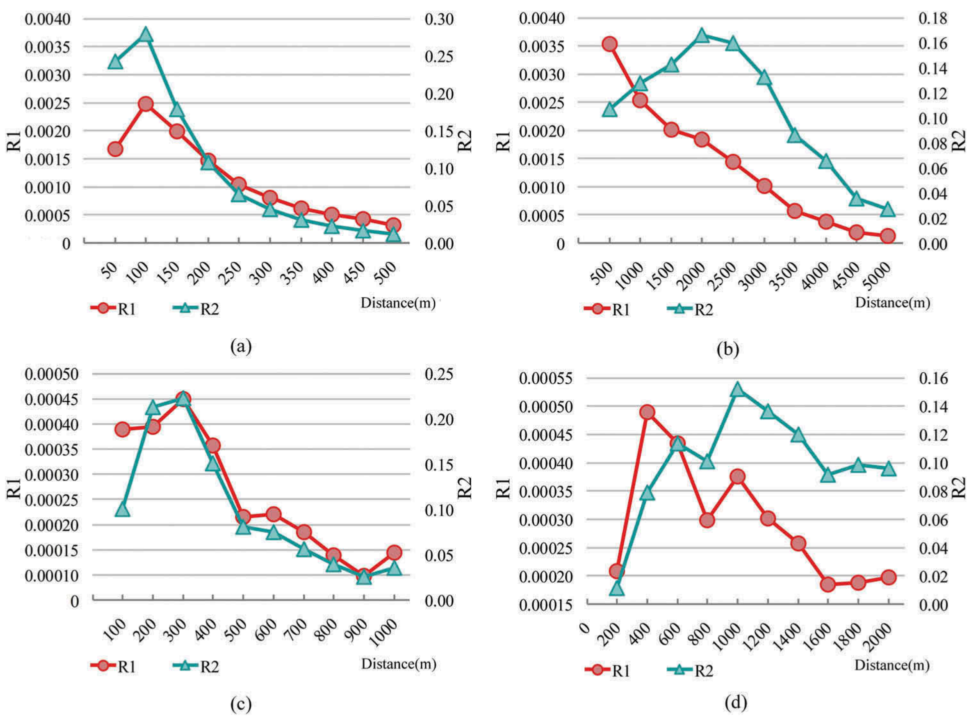

Figure 8. Quantitative analysis on the relationship between particular factors and land use change: (a) distance from road network – residential land use change, (b) distance from city center – commercial land use change, (c) distance from bus stations and administration – public services land use change, (d) distance from industrial estates – industrial land use change.

City centre – commercial land use

Figure 8(b) shows the relationship between the distance from the Huicheng city centre and commercial land use change. As the R_1 index shows, the relative amount of commercial land use change decreases slightly as the distance from road network increases. Thus, the distance to the city centre is an important determinant of commercial land change. Moreover, at distances as far as 1500 metres to 3000 metres from the city centre, from the R_2 index, we can see that the amount of commercial land use change has the largest proportion of all, which illustrates that this distance range is a hot spot of formally developing commercial estates.

Bus stations – administration and public services land use

Previous studies have indicated that the "'distance to bus stations'" does affect neighbouring land use conversions (Cervero and Kang 2011; Estupiñán and Rodríguez 2008). In this paper, we perform similar implementations for analysing the relationship between public services land use and the "distance to bus stations" through calculating the R_1 and R_2 indices. As Figure 8(c) shows, administration and public services land use change is less likely to occur in the vicinity of bus stations. We believe this could be due to the effect of the nuisance of heavy traffic and busy bus stations. Administration and public services land change is more likely to occur at the distances beyond 100 metres from bus stations. Beyond a distance of approximately 300 metres, administration and public services land use change is drastically reduced.

Industries – industrial land use

The analysis on the relationship between industrial land use change and distances from industrial district is supposed to be clear. However, the two indices actually show an unclear trend (Figure 8(d)): Industrial land use changed little at closer (within 300 metres) and farther (beyond 1500 metres) distances from industrial estates, the pixels for industrial land use change rise and fall in the range of 300 metres to 1500 metres to R_1 or R_2, and the proportion of industrial land change sharply declines at a distance of approximately 800 metres. We inferred that this decline may be caused by the occupancy of other land use at 800 metres distance away from the industrial cluster.

5. Discussion and conclusions

The purpose of this study was to propose an RF algorithm-based CA model (RF-CA model) to simulate multiple intra-urban land use changes and identify the contributions of each driving factor. The RF-CA model was successfully applied to the medium-sized Huicheng district located in Guangdong province in southern China. The internal urban datasets used in this study included intra-urban land use maps of the study area for 2000, 2005, 2010 and 2015 and multiple spatial variables. This study calibrated and validated the RF-CA model by comparing the simulated land use map for 2010 with the actual land use map for 2010. By applying the validated model, the multiple intra-urban land use map was simulated for 2015 with acceptable simulation results ("FoM" = 26.89%, *kappa* = 0.7305 and OA = 89.56%). Based on obtaining the intra-urban simulation results, this study focused on identifying the contribution of each driving factor (spatial variable) for each intra-urban land use using VIMs to understand how each driving factor influences each specific land use type.

We compared our method with former relevant studies (Almeida et al. 2008; Benenson, Omer, and Hatna 2002; Godoy and Soares-Filho 2008; Partanen 2016; Zheng et al. 2015) and summarized the superiorities of this paper's method. First, this study involved diverse categories of spatial variables, including traffic-location factors, natural factors, public services POIs, and population density as the driving factors to better reflect urban planning and human activity characteristics. Notably, public services POIs, which have rarely been considered in previous intra-urban CA models, were introduced in this study. These POIs were introduced because in intra-urban areas, particularly in areas such as Huicheng district in China, which are undergoing rapid urbanization, intra-urban POIs are important nodes for human activity and are very valuable in studying intra-urban land use change (Yao et al. 2017b). Generally, the distribution of public services POIs were closely related to citizens' behaviours (such as investing and purchasing houses) and influenced the distribution and transformation of intra-urban land uses. For example, urban citizens in China prefer to purchase residential homes in areas that are located near primary or secondary schools and to take advantage of the government's welfare policy regarding school district houses. In addition, urban citizens prefer to live near municipal hospitals. Because of these characteristic behaviours, internal urban POIs are essential factors that should be introduced in this study to enhance simulations at the intra-urban scale.

More importantly, we assessed and analysed the significance of each intra-urban land use driver for this region. With regard to the RF algorithm's VIMs, each category of intra-urban land use is sufficiently distinctive to be identified with the characteristics of different importance ranks in this paper (Table 9). After the importance ranks were aligned to each land use, the sequence of importance for each spatial variable are able to explain the reasons for land use changes. Furthermore, the importance sequences illustrate the types of driving factors that are crucial or dominant for a specific land use change (see in Section 3.4). Therefore, we can provide evidence and advice regarding development and

Table 9. The ranked contribution of each driving factors for four intra-urban land uses.

Driving factors (spatial variables)	RL	IL	APL	CL
Distance to small road	9.56%	4.73%	4.24%	4.86%
Distance to market	11.32%	7.44%	10.48%	8.98%
Distance to main road	10.25%	13.61%	5.40%	7.21%
Distance to expressway	8.22%	9.62%	9.05%	7.47%
Distance to city centre	9.21%	5.94%	9.64%	19.40%
Distance to bus station	7.09%	5.62%	10.17%	6.33%
Density of shopping centre POIs	1.78%	6.43%	2.96%	6.00%
Density of school POIs	3.87%	3.22%	2.92%	1.45%
Density of restaurant POIs	5.45%	4.99%	8.24%	5.42%
Density of parking lot POIs	3.56%	3.24%	5.56%	3.65%
Density of infrastructure POIs	6.17%	6.45%	5.84%	6.83%
Density of government organization POIs	3.35%	3.78%	2.83%	2.97%
Density of factory POIs	3.94%	8.02%	3.65%	1.58%
Density of entertain facility POIs	4.56%	4.71%	6.84%	3.18%
DEM	5.23%	4.84%	5.39%	3.61%
Slope	1.94%	2.21%	1.82%	1.33%
Population density	4.49%	5.13%	4.95%	9.72%

Note: RL represents residential land use, IL represents industrial land use, APL represents administration and public services land use, CL represents commercial land use. Background colour ranging from green to red indicated that the increasing influence of the driving factors acting on the current land use.

redevelopment for different intra-urban land use changes to assist decision-makers and urban planners with helpful and targeted information on future urban planning.

The integration of RF algorithms within CA models, which contributed to the application to Huicheng district, also has some limitations. One of the limitations is that we have not tested our model on other areas due to the lack of fine-scale datasets with sufficient intra-urban structures and details. In addition, due to the lack of various land use regulation datasets, including conservation areas, historical heritage sites, and green belts, this model considers only water bodies and urban road systems as the restriction area in the simulation. Consequently, the RF-CA model does not explicitly consider various scenarios of multiple intra-urban land uses dynamics. However, this limitation does not indicate that the model is lack of model pervasiveness and sustainability. By integrating the model with high-resolution remote sensing images and auxiliary spatiotemporal data, obtaining time series of urban spatial datasets at fine scales is feasible using GIS software (Treitz and Rogan 2004).Thus, simulations in other study areas under various land use scenarios (such as fast-development scenarios or scenarios under strict planning guidance) based on a more thorough preparation of the data must be introduced in our future studies.

To our knowledge, most CA models use common categories of spatial variables to simulate all land use types even though the factors driving each land use type may be unique. Whether we can apply more unique and typical spatial variables in the CA models to better represent the uniqueness of different urban land use types deserves additional considerations. In future studies, our RF-CA model has great potential in analysing the uniqueness of different land use types to assist in urban planning issues because the RF-based CA model is relatively robust compared to other machine learning algorithms-based CA models since RF algorithm is stable of overcoming the multiple correlative problems, nonstationarity issues, and noise among spatial variables.

Although the above questions have not yet been solved, our study can accurately identify the potential factors driving land use sprawl in cities. The mechanisms of multiple land use dynamics were explained to a large extent, and targeted information i that will be valuable for future urban planning is provided. Ongoing studies will analyse various simulation scenarios to better assist in urban planning decision making. In addition, finer categories of intra-urban land use and more targeted driving factors will be introduced into our method to improve our model's sustainability and applicability.

Highlights

- We proposed an RF-based CA model (RF-CA model) for simulating multiple intra-urban land-use change at a high resolution of 10 m with acceptable results ('FoM' = 26.89%, Kappa = 0.7306, OA = 89.56%).
- We identified the contributions of 17 driving factors for each intra-urban land-use through Variable Importance Measures (VIMs). These factors were categorized as traffic-location factors, natural factors, public services, and population density.
- We ranked the contribution of all driving factors for each land-use, identified the dominant factors driving each land use transformation, and quantitatively analysis the reasons for their formation. We believe that it would help urban planners and urban researchers to better understand how inner city structures are formed and how they function. This study would support the future evolution of this field.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was funded by the National Key R& D Program of China (2017YFA0604404) and the National Natural Science Foundation of China (Grant No. 41671398).

ORCID

Dachuan Zhang http://orcid.org/0000-0002-5586-9227

Xiaoping Liu http://orcid.org/0000-0003-4242-5392

Xiaoyu Wu http://orcid.org/0000-0002-4957-1643

Yao Yao http://orcid.org/0000-0002-2830-0377

Xinxin Wu http://orcid.org/0000-0002-4957-1643

References

Abdullahi, S., B. Pradhan, S. Mansor, and A. R. Mohamed Shariff. 2015. "GIS-based Modeling for the Spatial Measurement and Evaluation of Mixed Land Use Development for a Compact City." *GIScience& Remote Sensing* 52 (1): 18–39. doi:10.1080/15481603.2014.993854.

Aburas, M. M., Y. M. Ho, M. F. Ramli, and Z. H. Ash Aari. 2016. "The Simulation and Prediction of Spatio-Temporal Urban Growth Trends Using Cellular Automata Models: A Review." *International Journal of Applied Earth Observation& Geoinformation* 52: 380–389. doi:10.1016/j.jag.2016.07.007.

Almeida, C. M., J. M. Gleriani, E. F. Castejon, and B. S. Soares-Filho. 2008. "Using Neural Networks and Cellular Automata for Modelling Intra-Urban Land-Use Dynamics." *International Journal of Geographical Information Science* 22 (9): 943–963. doi:10.1080/13658810701731168.

Al-Shalabi, M., L. Billa, B. Pradhan, S. Mansor, and A. A. A. Al-Sharif. 2013. "Modelling Urban Growth Evolution and Land-Use Changes Using GIS Based Cellular Automata and SLEUTH Models: The Case of Sana'a Metropolitan City, Yemen." *Environmental Earth Sciences* 70 (1): 425–437. doi:10.1007/s12665-012-2137-6.

Arsanjani, J., M. H. Jamal, W. Kainz, and A. D. Boloorani. 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.

Azari, M., A. Tayyebi, M. Helbich, and M. A. Reveshty. 2016. "Integrating Cellular Automata, Artificial Neural Network, and Fuzzy Set Theory to Simulate Threatened Orchards: Application to Maragheh, Iran." *GISciences& Remote Sensing* 53 (2): 183–205. doi:10.1080/15481603.2015.1137111.

Belgiu, M., and D. Lucian. 2016. "Random Forest in Remote Sensing: A Review of Applications and Future Directions." *Isprs Journal of Photogrammetry& Remote Sensing* 114: 24–31. doi:10.1016/j.isprsjprs.2016.01.011.

Benenson, I., I. Omer, and E. Hatna. 2002. "Entity-Based Modeling of Urban Residential Dynamics: The Case of Yaffo, Tel Aviv." *Environment and Planning B: Planning and Design* 29 (4): 491–512. doi:10.1068/b1287.

Berling-Wolff, S., and W. Jianguo. 2004. "Modeling Urban Landscape Dynamics: A Case Study in Phoenix, USA." *Urban Ecosystems* 7 (3): 215–240. doi:10.1023/B:UECO.0000044037.23965.45.

Biau, G. 2010. "Analysis of a Random Forests Model." *Journal of Machine Learning Research* 13 (2): 1063–1095.

Bihamta, N., A. Soffianian, S. Fakheran, and M. Gholamalifard. 2015. "Using the SLEUTH Urban Growth Model to Simulate Future Urban Expansion of the Isfahan Metropolitan Area, Iran." *Journal of the Indian Society of Remote Sensing* 43 (2): 407–414. doi:10.1007/s12524-014-0402-8.

Braimoh, A. K., and T. Onishi. 2007. "Spatial Determinants of Urban Land Use Change in Lagos, Nigeria." *Land Use Policy* 24 (2): 502–515. doi:10.1016/j.landusepol.2006.09.001.

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

Cervero, R., and C. D. Kang. 2011. "Bus Rapid Transit Impacts on Land Uses and Land Values in Seoul, Korea." *Transport Policy* 18 (1): 102–116. doi:10.1016/j.tranpol.2010.06.005.

Chen, Y., L. Xia, X. Liu, and A. Bin. 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., X. Liu, L. Xia, X. Liu, Y. Yao, H. Guohua, X. Xiaocong, and F. Pei. 2017. "Delineating Urban Functional Areas with Building-Level Social Media Data: A Dynamic Time Warping (DTW) Distance Based K -Medoids Method." *Landscape& Urban Planning* 160: 48–60. doi:10.1016/j.landurbplan.2016.12.001.

Conway, T. M., and C. C. Wellen. 2011. "Not Developed Yet? Alternative Ways to Include Locations without Changes in Land Use Change Models." *International Journal of Geographical Information Science* 25 (10): 1613–1631. doi:10.1080/13658816.2010.534738.

Donnelly, F. P. 2013. "The Geographic Distribution of United States Public Libraries: An Analysis of Locations and Service Areas." *Journal of Librarianship& Information Science* 46 (2): 110–129. doi:10.1177/0961000612470276.

Du, G., J. S. Kong, L. Yuan, and S. Managi. 2017. "A Comparative Approach to Modelling Multiple Urban Land Use Changes Using Tree-Based Methods and Cellular Automata: The Case of Greater Tokyo Area." *International Journal of Geographical Information Science* 4: 1–26.

Estupiñán, N., and D. A. Rodríguez. 2008. "The Relationship between Urban Form and Station Boardings for Bogotá's BRT." *Transportation Research Part A Policy& Practice* 42 (2): 296–306. doi:10.1016/j.tra.2007.10.006.

Feng, Y., and X. Tong. 2018. "Dynamic Land Use Change Simulation Using Cellular Automata with Spatially Nonstationary Transition Rules." *GIScience& Remote Sensing. 55(5):678-698*.

Feng, Y., Y. Liu, X. Tong, M. Liu, and S. Deng. 2011. "Modeling Dynamic Urban Growth Using Cellular Automata and Particle Swarm Optimization Rules." *Landscape and Urban Planning 102 (3):* 188-196.

Gislason, P. O., J. A. Benediktsson, and J. R. Sveinsson. 2006. ""Random Forests for Land Cover Classification." *Pattern Recognition Letters*. doi:10.1016/j.patrec.2005.08.011.

Godoy, M., and B. S. Soares-Filho. 2008. *Modelling Intra-Urban Dynamics in the Savassi Neighbourhood, Belo Horizonte City, Brazil,* edited by Paegelow M. and Olmedo M.T.C: Springer, Berlin, Heidelberg. doi: https://doi.org/10.1007/978-3-540-68498-5_12.

Gregory, D., R. Johnston, G. Pratt, M. Watts, and S. Whatmore. 2009. *The Dictionary of Human Geography.* 5th ed. Chichester:Wiley-Blackwell.

Han, J., Y. Hayashi, X. Cao, and H. Imura. 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. doi:10.1016/j.landurbplan.2008.12.002.

He, C., N. Okada, Q. Zhang, P. Shi, and L. Jinggang. 2008. "Modelling Dynamic Urban Expansion Processes Incorporating a Potential Model with Cellular Automata." *Landscape& Urban Planning* 86 (1): 79–91. doi:10.1016/j.landurbplan.2007.12.010.

He, Y., B. Ai, Y. Yao, and F. Zhong. 2015. "Deriving Urban Dynamic Evolution Rules from Self-Adaptive Cellular Automata with Multi-Temporal Remote Sensing Images." *International Journal of Applied Earth Observation and Geoinformation*. doi:10.1016/j.jag.2014.12.014.

Hossein, S. M., and M. Helbich. 2013. "Spatiotemporal Urbanization Processes in the Megacity of Mumbai, India: A Markov Chains-Cellular Automata Urban Growth Model." *Applied Geography* 40: 140–149. doi:10.1016/j.apgeog.2013.01.009.

Hung, B. L., and H. A. Hung. 2014. "Using One-Dimensional Linear Interpolation Method to Check Over-Fitting in Neural Network with Multi-Dimensional Inputs." *Applied Mechanics& Materials* 496-500: 2228–2232. doi:10.4028/www.scientific.net/AMM.496-500.2228.

Kalnay, E., and M. Cai. 2003. "Impact of Urbanization and Land-Use Change on Climate." *Nature* 423 (6939): 528–531. doi:10.1038/nature01675.

Kamusoko, C., and J. Gamba. 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.

Knol, M. J., S. Le Cessie, A. Algra, J. P. Vandenbroucke, and R. H. Groenwold. 2012. “Overestimation of Risk Ratios by Odds Ratios in Trials and Cohort Studies: Alternatives to Logistic Regression.” *CMAJ* 184 (8): 895–899. doi:10.1503/cmaj.101715.

Kong, X., Y. Liu, Y. Wang, D. Tong, and J. Zhang. 2017. “Investigating Public Facility Characteristics from A Spatial Interaction Perspective: A Case Study of Beijing Hospitals Using Taxi Data.” *ISPRS International Journal of Geo-Information* 6 (2): 38. doi:10.3390/ijgi6020038.

Kuang, W. H. 2012. “Spatio-Temporal Patterns of Intra-Urban Land Use Change in Beijing,China between 1984 and 2008.” *Chinese Geographical Science* 22 (2): 210–220. doi:10.1007/s11769-012-0529-x.

Lagarias, A. 2012. “Urban Sprawl Simulation Linking Macro-Scale Processes to Micro-Dynamics through Cellular Automata, an Application in Thessaloniki, Greece.” *Applied Geography* 34: 146–160. doi:10.1016/j.apgeog.2011.10.018.

Leao, S., I. Bishop, and D. Evans. 2004. “Simulating Urban Growth in a Developing Nation’s Region Using a Cellular Automata-Based Model.” *Journal of Urban Planning and Development*. doi:10.1061/(ASCE)0733-9488(2004)130:3(145).

Li, T., and L. Wenkai. 2015. “Multiple Land Use Change Simulation with Monte Carlo Approach and CA-ANN Model, a Case Study in Shenzhen, China.” *Environmental Systems Research* 4 (1): 1. doi:10.1186/s40068-014-0026-6.

Lin, J., and L. Xia. 2016. “Conflict Resolution in the Zoning of Eco-Protected Areas in Fast-Growing Regions Based on Game Theory.” *Journal of Environmental Management* 170: 177. doi:10.1016/j.jenvman.2015.11.036.

Liu, X., H. Guohua, and A Bin. 2018. “Simulating Urban Dynamics in China with a Gradient Cellular Automata Model Based on S-shape Curve Evolution Characteristics.” *International Journal Of Geographical Information Science* 32 (1): 73-101.

Liu, X., L. Xia, X. Shi, X. Zhang, and Y. Chen. 2010b. “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. doi:10.1080/13658810903270551.

Liu, X., L. Xun, L. Xia, X. Xiaocong, and O Jinpei. 2017. “A Future Land Use Simulation Model (Flus) for Simulating Multiple Land Use Scenarios by Coupling Human and Natural Effects.” *Landscape and Urban Planning* 168: 94-116.

Liu, X., M. Lei, L. Xia, A. Bin, L. Shaoying, and H. Zhijian. 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. P., L. Xia, Y. M. Chen, Z. Z. Tan, S. Y. Li, and B. Ai. 2010a. “A New Landscape Index for Quantifying Urban Expansion Using Multi-Temporal Remotely Sensed Data.” *Landscape Ecology* 25 (5): 671–682. doi:10.1007/s10980-010-9454-5.

Liu, Y., X. Liu, S. Gao, L. Gong, C. Kang, Y. Zhi, G. Chi, and L. Shi. 2015. “Social Sensing: A New Approach to Understanding Our Socioeconomic Environments.” *Annals of the Association of American Geographers* 105 (3): 512–530. doi:10.1080/00045608.2015.1018773.

Long, Y., and X. Liu. 2015. “Automated Identification and Characterization of Parcels (AICP) with OpenStreetMap and Points of Interest.” *Environment& Planning B* 43 (2): 498–510.

Luo, J., and Y. H. Dennis Wei. 2009. “Modeling Spatial Variations of Urban Growth Patterns in Chinese Cities: The Case of Nanjing.” *Landscape and Urban Planning* 91 (2): 51–64. doi:10.1016/j.landurbplan.2008.11.010.

Mitsova, D., W. Shuster, and X. Wang. 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.

Ndez-Delgado, M., E. Cernadas, S. Barro, and D. Amorim. 2014. “Do We Need Hundreds of Classifiers to Solve Real World Classification Problems?” *Journal of Machine Learning Research* 15 (1): 3133–3181.

Omrani, H., A. Tayyebi, and B. Pijanowski. 2017. “Integrating the Multi-Label Land-Use Concept and Cellular Automata with the Artificial Neural Network-Based Land Transformation Model: An Integrated ML-CA-LTM Modeling Framework.” *GIScience& Remote Sensing* 54 (3): 283–304. doi:10.1080/15481603.2016.1265706.

Palczewska, A., J. Palczewski, R. M. Robinson, and D. Neagu. 2013. “Interpreting Random Forest Classification Models Using a Feature Contribution Method.” *Integration of Reusable Systems* 263: 193–218.

Partanen, J. 2016. “An Urban Cellular Automata Model for Simulating Dynamic States on a Local Scale.” *Entropy* 19: 1. doi:10.3390/e19010012.

Pontius, R. G., W. Boersma, J.-C. Castella, K. Clarke, T. De Nijs, C. Dietzel, Z. Duan, 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.

Ratti, C., D. Frenchman, R. M. Pulselli, and S. Williams. 2006. "Mobile Landscapes: Using Location Data from Cell Phones for Urban Analysis." *Environment& Planning B Planning& Design* 33 (5): 727–748. doi:10.1068/b32047.

Reif, D. M., A. A. Motsinger, B. A. Mckinney, J. E. Crowe, and J. H. Moore. 2006. "Feature Selection Using a Random Forests Classifier for the Integrated Analysis of Multiple Data Types." in IEEE Symposium on Computational Intelligence in Bioinformatics& Computational Biology.

Santé, I., A. M. García, D. Miranda, and R. Crecente. 2010. "Cellular Automata Models for the Simulation of Real-World Urban Processes: A Review and Analysis." *Landscape and Urban Planning* 96 (2): 108–122. doi:10.1016/j.landurbplan.2010.03.001.

Shafizadehmoghadam, H., A. Asghari, M. Taleai, M. Helbich, and A. Tayyebi. 2017. "Sensitivity Analysis and Accuracy Assessment of the Land Transformation Model Using Cellular Automata." *Giscience& Remote Sensing* 54(5):639-656. doi: https://doi.org/10.1080/15481603.2017.1309125.

Sim, J., and C. C. Wright. 2005. "The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements." *Phys Ther* 85 (3): 257–268.

Sun, J. B., J. Yuan, Y. Wang, H. B. Si, and X. M. Shan. 2011. "Exploring Space–Time Structure of Human Mobility in Urban Space." *Physica A Statistical Mechanics& Its Applications* 390 (5): 929–942. doi:10.1016/j.physa.2010.10.033.

Torrens, P. M., and D. O'Sullivan. 2001. "Cellular Automata and Urban Simulation: Where Do We Go from Here?" *Environment and Planning B: Planning and Design* 28 (2): 163–168. doi:10.1068/b2802ed.

Treitz, P., and J. Rogan. 2004. "Remote Sensing for Mapping and Monitoring Land-Cover and Land-Use Change—An Introduction." *Progress in Planning* 61 (4): 269–279. doi:10.1016/S0305-9006(03)00064-3.

Verburg, P. H., P. P. Schot, M. J. Dijst, and A. Veldkamp. 2004a. "Land Use Change Modelling: Current Practice and Research Priorities." *GeoJournal* 61 (4): 309–324. doi:10.1007/s10708-004-4946-y.

Verburg, P. H., T. C. M. De Nijs, J. R. Van Eck, H. Visser, and K. D. Jong. 2004b. "A Method to Analyse Neighbourhood Characteristics of Land Use Patterns." *Computers Environment& Urban Systems* 28 (6): 667–690. doi:10.1016/j.compenvurbsys.2003.07.001.

Vliet, J., A. K. Bregt, D. G. Brown, H. Van Delden, S. Heckbert, and P. H. Verburg. 2016. "A Review of Current Calibration and Validation Practices in Land-Change Modeling." *Environmental Modelling& Software* 82: 174–182. doi:10.1016/j.envsoft.2016.04.017.

White, R., and G. Engelen. 2000. "High-Resolution Integrated Modelling of the Spatial Dynamics of Urban and Regional Systems." *Computers Environment& Urban Systems* 24 (5): 383–400. doi:10.1016/S0198-9715(00)00012-0.

Wise, S., A. Crooks, and M. Batty. 2016. "Transportation in Agent-Based Urban Modelling." In *Agent Based Modelling of Urban Systems*. ABMUS 2016. Lecture Notes in Computer Science, edited by Namazi-Rad MR., Padgham L., Perez P., Nagel K., Bazzan A., 10051:129-148. Springer, Cham. doi: https://doi.org/10.1007/978-3-319-51957-9_8 129–148.

Wu, F., and D. Martin. 2002. "Urban Expansion Simulation of Southeast England Using Population Surface Modelling and Cellular Automata." *Environment& Planning A* 34 (10): 1855–1876. doi:10.1068/a3520.

Wyner, A. J., M. Olson, J. Bleich, and D. Mease. 2017. "Explaining the Success of AdaBoost and Random Forests as Interpolating Classifiers." *Journal of Machine Learning Research* 18(48):1−33.

Xie, Y., G. C. Chuanglin Fang, S. Lin, H. Gong, and B. Qiao. 2007. "Tempo-Spatial Patterns of Land Use Changes and Urban Development in Globalizing China: A Study of Beijing." *Sensors* 7 (11): 2881–2906. doi:10.3390/S7112881.

Xiong, X., and L. Luo. 2017. "Use of Geographical Information Systems for Delimiting Health Service Areas in China." *Geospatial Health* 12 (1): 486. doi:10.4081/gh.2017.486.

Yao, Y., L. Xia, X. Liu, P. Liu, Z. Liang, J. Zhang, and K. Mai. 2016. "Sensing Spatial Distribution of Urban Land Use by Integrating Points-Of-Interest and Google Word2Vec Model." *International Journal of Geographical Information Systems* 31 (4): 825–848. doi:10.1080/13658816.2016.1244608.

Yao, Y., X. Liu, L. Xia, J. Zhang, Z. Liang, K. Mai, and Y. Zhang. 2017a. "Mapping Fine-Scale Population Distributions at the Building Level by Integrating Multisource Geospatial Big Data." *International Journal of Geographical Information Systems* 31 (6): 1220–1244.

Yao, Y., X. Liu, L. Xia, P. Liu, Y. Hong, Y. Zhang, and K. Mai. 2017b. "Simulating Urban Land-Use Changes at a Large Scale by Integrating Dynamic Land Parcel Subdivision and Vector-Based Cellular Automata." *International Journal of Geographical Information Science* 31(12): 2452-2479.

Zhang, D. C., X. P. Liu, Y. Yao, and J. B. Zhang. 2016. "Simulating Spatiotemporal Change of Multiple Land Use Types in Dongguan by Using Random Forest Based on Cellular Automata." *Geography and Geo-Information Science. 32(5):* 29-36. In Chinese. doi:10.3969/j.issn.1672-0504.2016.05.005.

Zheng, H. W., G. Q. Shen, H. Wang, and J. Hong. 2015. "Simulating Land Use Change in Urban Renewal Areas: A Case Study in Hong Kong." *Habitat International* 46: 23–34. doi:10.1016/j.habitatint.2014.10.008.