

An extended patch-based cellular automaton to simulate horizontal and vertical urban growth under the shared socioeconomic pathways

Yimin Chen^{*}

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



ARTICLE INFO

Keywords:
 Cellular automaton
 Vertical urban growth
 Shared socioeconomic pathways

ABSTRACT

Most contemporary urban cellular automata (CA) models primarily focus on the simulation of urban land expansion, and cannot effectively simulate vertical urban growth. This study addresses this drawback by extending a patch-based urban CA model with a component that can predict the building volumes of an urban land expansion scenario. The proposed model is evaluated through a case study in the Guangzhou-Foshan metropolitan area, China. The horizontal urban growth simulations achieve a mean 'Figure-of-merit' value of 0.1406 at the cell level and an agreement of 97% at the pattern level. The building volume prediction made by the methods of random forest and k -nearest-neighbor has a testing R^2 of 0.90 and a mean percentage absolute error of 22%. The proposed model is applied to the urban growth projections under the shared socioeconomic pathways (SSPs). The results successfully reflect the influences that different SSPs have on vertical urban developments. These results also complement related research of urbanization projections under the SSPs, because most existing studies consider the impacts of horizontal urban growth only. As building volumes and heights are fundamental parameters to urban climate modeling, the ability of the proposed model to project future change in vertical urban developments can support the mitigation of climate change effects on human settlements.

1. Introduction

Today more than half of the global population are living in urban areas. Despite the small global land coverage, urban areas play a key role in driving global change. Urbanization causes anthropogenic heat release, greenhouse gas emission, and deforestation, thereby affecting global climate. On the other hand, urban inhabitants that lack basic infrastructure, services, and capacity to mitigate global change impacts are increasingly vulnerable. Therefore, it is important to increase the knowledge of future urbanization and its effects on the environment and human well-being as well.

The past two decades have witnessed the widespread application of cellular automata (CA) as a kind of urban simulation models to help advance the understanding of how urban systems evolve (Clarke, 2019; Liu, Batty, Wang, & Corcoran, 2021; Wu et al., 2019). Among the intensive studies of urban CA models, some have been devoted to the improvement of model performance (Ke, Zheng, Zhou, & Liu, 2017; Wang, Zhang, Xia, He, & Zhang, 2020) while others have focused more on integrating urban CA models into scenario analysis to evaluate the consequences of future urbanization (Guzman, Escobar, Peña, &

Cardona, 2020; He et al., 2021). In particular, many of the most recent studies appear to revolve around three topics:

- (1) To improve the performance of urban CA models by using machine learning algorithms. Because machine learning algorithms can efficiently handle the complicated relationships within data, they often yield better results than conventional methods such as logistic regression. Therefore, machine learning algorithms have been prevalently used to capture the effects of various driving factors (Shafizadeh-Moghadam, Minaei, Pontius Jr, Asghari, & Dadashpoor, 2021). Existing studies of urban simulations have evaluated the performance of a variety of machine learning methods, including but not limited to swarm-based optimization algorithms (Cao, Huang, Xu, Lü, & Chen, 2019; Tang, Liu, Song, & Chang, 2020), support vector machine (Karimi, Sultana, Babakan, & Suthaharan, 2019; Mustafa, Rienow, Saadi, Cools, & Teller, 2018), artificial neural network (ANN) (Zhai et al., 2020), evolutionary algorithms (Newland, van Delden, Zecchin, Newmann, & Maier, 2020), and random forest (RF) (Shafizadeh-Moghadam et al., 2021). A recent comprehensive comparison that

* Corresponding author at: School of Geography and Planning, Sun Yat-sen University, 135 West Xiangang Rd., Guangzhou, China.
 E-mail address: chenym49@mail.sysu.edu.cn.

involves 38 machine learning algorithms revealed that the RF algorithm outperforms most of the selected algorithms, and hence is a good initial choice for urban studies (Hagenauer, Omrani, & Helbich, 2019). Another case study also supports this finding (Rienow, Mustafa, Krelaus, & Lindner, 2021).

- (2) To improve the performance by modifying the traditional model structure. Empirical literature has revealed that conventional urban CA models cannot produce reliable simulations in some cases. For instance, the regular cell representation is not reasonable for fine-scale applications (Wang & Marceau, 2013). Hence, urban CA models with irregular representation of spatial units have been developed to address the drawbacks of the regular cell representation. There are two forms of irregular representation, i.e., vector cells (Abolhasani, Taleai, Karimi, & Rezaee Node, 2016; Chen, Liu, & Li, 2017a; Zhai et al., 2020) and raster patches (Chen, Li, Liu, & Ai, 2014; McGarigal et al., 2018; Meentemeyer et al., 2013; Wang & Marceau, 2013). For models implemented in a raster environment, patches are a better form to represent the homogeneous land units (e.g. a parcel) at local scales. Objects represented by fine-resolution raster data are usually composed of multiple cells, and it is not reasonable to assume that the change of objects' states is purely cell-by-cell. Patches are, therefore, a more natural form to represent these objects and their change (Chen, Li, Liu, Ai, & Li, 2016; Wang & Marceau, 2013). Moreover, patch-based models often produce more realistic simulations of urban growth. Models that accurately capture the dynamics of land change process can generate simulations with high aggregate similarity to the reference maps (Brown, Page, Riolo, Zellner, & Rand, 2005). In this sense, patch-based models seem better than the conventional cell-based models. Several recent studies have consistently reported that compared with the cell-based simulations, the patch-based simulations show higher agreements to the reference landscape patterns (Alaei Moghadam, Karimi, & Habibi, 2018; Li, Gong, Yu, & Hu, 2017; Xu & Brown, 2017).
- (3) To place urban CA models in a broad analysis framework of global change. Due to its far-reaching impacts, urbanization cannot be neglected in the projections of global change. A trend that has emerged recently in the applications of urban CA models is to explore future urbanization scenarios under the shared socioeconomic pathways (SSPs). Here the SSPs are narratives (see Section 3.4 for more details) that describe five alternative pathways for global society, demographics, and economics toward the 21st century (O'Neill et al., 2017). These five narratives are an important component of the ongoing IPCC assessment of global change. On the one hand, the qualitative narratives and quantitative elements of the SSPs provide a common framework and primary inputs for the projections of future urbanization. Therefore, many recent CA-based urbanization projections have been carried out under the SSPs at regional and global scales (Chen et al., 2020; Chen, Li, Liu, Zhang, & Huang, 2019; Gemitzi, 2021). On the other hand, it is also important to evaluate whether urbanization will induce extra environmental effects on top of climate change. To this end, urbanization projections for the SSPs can be conveniently linked with the Representative Concentration Pathways (RCPs) to further evaluate the joint impacts of urbanization and climate change on, for instances, flood risk (Chen, Zhang, Chen, & Huang, 2021), extreme heats (Rohat et al., 2019), air pollutions (Rao et al., 2017), and biodiversity (McManamy, Vernon, & Jager, 2021).

Despite the advances made, a major drawback of the contemporary urban CA models is the lack of an effective component to simulate vertical urban growth (Y. Liu et al., 2021). This problem also impedes

the understanding of the effects that different pathways have on vertical urban developments. As a result, most spatial projections of future urbanization for the SSPs only consider the impacts of horizontal urban growth (Chen et al., 2020; Chen, Li, Liu, Zhang, & Huang, 2019; Gemitzi, 2021). While horizontal growth takes place in the form of urban land expansion, vertical growth mainly refers to urban densification in the vertical dimension, including the increases in building volumes and population density. Hence, vertical urban growth is an important indicator to measure compact development, which has been considered effective to promote housing affordability (Goetz, 2013) and reduce energy consumption (Güneralp et al., 2017). Besides, the vertical attributes of built environments, such as building volumes and building heights, are fundamental parameters in modeling urban climate (Li, Schubert, Kropp, & Rybski, 2020). Accurate simulations of vertical urban growth are, therefore, critical to understanding the comprehensive effects of urban developments, especially in the context of global change.

A few methods have been proposed recently to enhance the simulations of vertical urban growth. Lin, Huang, Chen, and Huang (2014) developed an urban CA model to simulate vertical urban growth with a predefined set of 'IF-THEN' rules. He, Liu, Zeng, Chaohui, and Tan (2017) used a back propagation ANN to predict vertical urban growth based on the urban extents that were generated by an urban CA model. These two studies consistently represented vertical urban growth as the change in the types of building heights. Similarly, Huang et al. (2021) and Chen et al. (2021) used an ANN-based urban CA model to simulate the change of multiple urban land types, which consider the vertical characteristics of built environments. Koziatek and Dragičević (2017) proposed the 3D version of the conventional *iCity* model (Stevens, Dragicevic, & Rothley, 2007), which can simulate the increase of mid- and high-rise buildings at a cadastral lot resolution.

In short, existing models for simulating vertical urban growth suffer from several important issues: (1) Several recent studies have simplified the simulations of vertical urban growth into the simulations of change in the discrete types of, for example, building heights (e.g., 'low-rise', 'high-rise', etc.) (Chen, Xie, et al., 2021; He et al., 2017; Huang et al., 2021). Besides the modeling uncertainties, additional uncertainties could exist in both the definition and classification of the types, and the simulations may not be comparable from one case to another. (2) For some urban simulations that aim to simulate the continuous change in the vertical dimension (Lin et al., 2014), however, intensive expert knowledge is required for model verification due to the lack of automatic methods to capture the quantitative relationship between vertical urban growth and its driving factors. (3) Few have explicitly explored how urban form (in a vertical sense) would react to various development pathways (e.g., compact vs. sprawling forms), which is essential to help gain understanding of the impacts that urbanization have on environments.

This study presents an integrated CA-based model that addresses the problems mentioned above. The proposed model is developed by extending a conventional patch-based urban CA model (Chen et al., 2014) with a component to predict the continuous change in building volumes. With the results of simulated building volumes, building heights can also be converted correspondingly. Compared with the conventional simulation models of vertical urban growth, the proposed model has three important features: (1) The patch-based urban CA model not only provides more reliable simulations of urban landscape change than the conventional cell-based models but also allows exploring the divergent outcomes under the assumptions of different pathways (e.g. SSPs). (2) The proposed model uses a machine learning algorithm that can automatically establish the quantitative relationship between building volumes and a series of factors. (3) The proposed model can produce multi-scenario simulations of vertical urban growth corresponding to different development pathways, thereby addressing the shortcoming of existing SSPs urban projections that consider horizontal urban growth only.

The proposed model is applied in the Guangzhou-Foshan metropolitan area, China. After calibrating the model with empirical data, scenario simulations of horizontal and vertical urban growth are carried out under the SSPs. Here the horizontal urban growth refers to urban land expansion, and the vertical urban growth refers to the increase in building volumes. The simulations of urban land expansion are at a 30-m resolution, while the corresponding predictions of building volumes are at a 150-m resolution. This resolution is chosen according to the settings in recent urban form studies, in which the adopted data resolution may range from 100 m (Chen, Xie, et al., 2021; Huang et al., 2021) to 250 m (Liang et al., 2021) or even coarser (Liu & Shi, 2020). Moreover, an open access urban land parcel dataset (Chen, Xu, & Gong, 2021) reveals that the average size of the land parcels in the Guangzhou-Foshan metropolitan area is 340,398 m², which corresponds to a grid cell of ~583 m. Therefore, the choice of 150 m is enough to suit the needs of typical urban form studies and can properly represent the local situations as well.

2. Case study area and data sources

The case study area is located in the Guangzhou-Foshan metropolitan area, China (Fig. 1(a)). Guangzhou and Foshan are two adjacent prefectures in the Pearl River Delta in South China. They share a similar history, culture, and complementary economic structures. With the implementation of a series of development agendas, Guangzhou and Foshan have become functionally integrated whole (Ye et al., 2018). In recent years, Guangzhou and Foshan's urban areas have merged into a large metropolitan area with more than 23 million inhabitants.

Table 1 summarizes all the data required for the case study, including land use maps, DEM, buildings polygons, road networks, point-of-interests (POIs), and SSP urban land area data. The land use maps for the years 2010, 2015, and 2018 are acquired from the Data Center for Resources and Environmental Sciences of the Chinese Academy of Sciences. The land-use categories include cropland, forest, grassland, water body (including rivers, lakes, fishponds, etc.), built-up land, and unused land. The spatial resolution is 30 m. According to the local development plans, forests, rivers, and lakes are restricted areas where no developments are allowed. Additionally, the 30-m DEM data is downloaded from the Land Processes Distributed Active Archive Center of NASA.

Table 1
Summary of data and their sources.

Data	Resolution	Sources
Land use maps (2010, 2015, 2018)	30 m	Data Center for Resources and Environmental Sciences of the Chinese Academy of Sciences
DEM	30 m	ASTER GDEM from NASA's Land Processes Distributed Active Archive Center
Polygons of individual buildings	—	Purchased
Road networks	—	OpenStreetMap
Point-of-interests data	—	Gaode Map
SSP urban land maps (2015–2050)	1 km	Chen et al. (2020)

The polygon data of individual buildings are purchased from a domestic data vendor. This data provides the spatial footprints and the heights of buildings in Guangzhou and Foshan (Fig. 1(b)-(c)). The building heights vary substantially across the case study area. More than half of the buildings' heights are below 10 m, and the mean height of all buildings approximates to only 12 m.

The data of road networks are collected from OpenStreetMap. Only the primary roads, secondary roads, tertiary roads, and highways (motorways) are used for subsequent analysis. Additionally, a dataset of (POIs) is collected through the API of Gaode Map. This dataset provides the locations of the major urban facilities and infrastructures, such as residential quarters, shopping centers, offices, metro stations, etc. Furthermore, the simulated SSPs global urban land maps developed by Chen et al. (2020) are acquired to localize the SSPs scenarios for Guangzhou-Foshan Metropolitan Area. The spatial resolution of these maps is 1 km. They are used to estimate the future demand of urban land for each SSP scenario.

The data mentioned above are processed into two sets of variables to estimate development potential and predict building volumes, respectively. Here, the development potential is used for the simulation of urban land expansion at a 30-m resolution. The estimation of development potential requires several spatial variables, including distance to the city center, distance to the nearest county center, distance to the nearest metro station, distance to the nearest highway entrance,

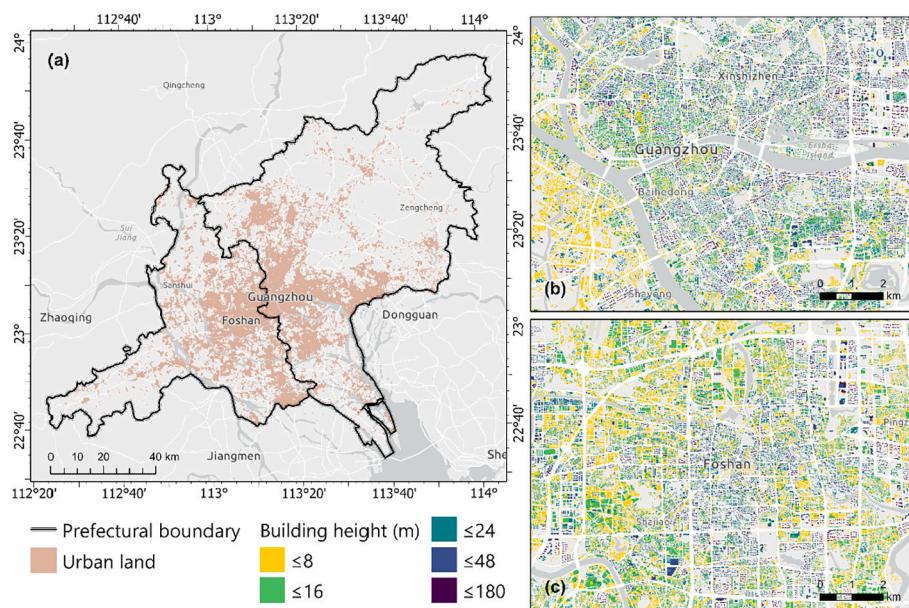


Fig. 1. (a) The geographic location of the Guangzhou-Foshan metropolitan area. (b)-(c) Polygon data of the individual buildings in central Guangzhou and Foshan. Supplementary Fig. S1 shows the spatial distribution of building polygons at full extent.

distance to the nearest railway station, distance to the road network, and slope (Fig. 2). They are selected according to the settings in empirical literature (Chen, Li, Liu, Huang, & Ma, 2019; Feng & Tong, 2020; Liang et al., 2020). The locations of city and county centers are determined through Gaode Map. The POIs of metro stations, highway entrances, and railway stations are used to generate the corresponding distance variables. The data of roads for generating the distance variable refer to the primary roads, secondary roads, and tertiary roads that are collected from OpenStreetMap. The variable of slope is produced using the DEM data and the surface analysis function in a GIS.

Because the change in building volumes is simulated at the resolution of 150 m, the building data is transformed and aggregated to the 150-m grid cells. Each grid cell i is assigned with the value of total building volume (denoted as V_i) in that cell (Fig. 2(h)):

$$V_i = \sum_j a_j h_j \quad (1)$$

where a_j and h_j are the polygon area and the height of building j , respectively. V_i is, therefore, the sum of volumes for all buildings that are within grid cell i . Additionally, the mean building height for each grid cell also can be obtained by aggregating the building data.

The spatial variables for predicting building volumes include the area of built-up land, the intensity of urban activities (Fig. 2), and all the variables used for modeling development potential as well. The intensity of urban activities is represented using the data of five POIs types, including shopping center, company, industrial, office, and residential. They are chosen because they strongly influence the spatial layout of buildings (Chen et al., 2017b; Li et al., 2017). These POIs data are transformed into five kernel density layers. The bandwidth of kernel density estimation is set automatically using ‘Silverman’s rule-of-thumb’ (Silverman, 1986). Similarly, the variable of built-up land area is generated by counting the number of urban land pixels (in year 2018) within each grid cell. All these variables are prepared at the spatial resolution of 150 m. Supplementary Table S1 summarizes the correlation coefficients between the spatial variables and building volumes at 150-m resolution, in which the kernel densities of the five POIs types consistently show significant positive correlations with building volumes. Their correlation coefficients are 0.5050, 0.4835, 0.2331, 0.5135, and 0.5383, correspondingly.

3. Methodology

The proposed methods for scenario simulations of horizontal and

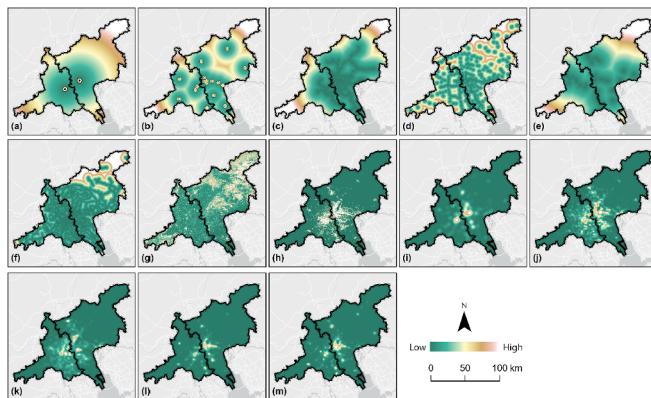


Fig. 2. Spatial variables for estimating development potential: (a) distance to the city center, (b) distance to the nearest county center, (c) distance to the nearest metro station, (d) distance to the nearest highway entrance, (e) distance to the nearest railway station, (f) distance to the road network, and (g) slope. Additional spatial data for predicting building volumes: (h) building volume distribution, (i-m) kernel density layers of the five POIs types, including shopping center, company, industrial, office, and residential.

vertical urban growth are implemented by the following procedures (Fig. 3): First, a model based on the random forest (RF) algorithm is built to produce urban development potential at a 30-m resolution (Section 3.1). Second, the patch-based urban CA model is calibrated to simulate urban land expansion with the development potential generated by RF (Section 3.2). Third, another RF-based model is built to predict building volumes at a 150-m resolution (Section 3.3). Finally, the estimated urban land areas of the five SSPs scenarios (Section 3.4) are used as the constraints for the scenario simulations of future urban land expansion. The results are further connected to the RF-based model to predict the building volumes for the five SSPs scenarios.

3.1. Estimating development potential

Development potential is an essential input of urban CA for simulating urban growth. The development potential of a non-urban cell refers to the propensity of its state to become ‘urban’. As recommended by empirical literature (Hagenauer et al., 2019), this study employs the RF algorithm to estimate development potential.

The RF algorithm operates by training a set of randomized, unpruned decision trees with randomly selected samples and attributes, and the outputs of these decision trees are aggregated to generate a final prediction. Specifically, with the number of decision trees N_{tree} defined a priori, each of the trees is built by using a bootstrap sample that is randomly drawn with replacements from the input dataset. This sampling procedure creates a data subset for training a tree (approximately two thirds of the original dataset) and a testing data subset (approximately one third of the original dataset) for evaluating the performance of that tree. Here the prediction error evaluated with the testing data subset is called the out-of-bag error. Additionally, when training the decision trees, the features (i.e., attributes) are also a random subset of the full features of the original dataset, and the number of the randomly drawn features should be less than N_{feat} , which is defined a priori. By using these randomly drawn features in the decision trees, the best split of nodes is then created according to the Gini impurity criterion. Finally, after the training of decision trees, the prediction is made by averaging the class assignment probabilities that are yielded by the decision trees. Due to its reliable performance, RF has been applied in the simulations of urban growth in, for example, Tehran (Shafizadeh-Moghadam et al., 2021), Attica (Gounaris, Chorianopoulos, Symeonakis, & Koukoulas, 2019), Wuhan (Wu, Lin, Xing, Song, & Li, 2021), Shanghai (Zhou, Dang, Sun, & Wang, 2020), etc.

Before the training of the RF algorithm, a set of cells is randomly drawn from the land use maps of 2010 and 2015, including 10,000 cells changed from non-urban to urban and another 10,000 cells that are persistently non-urban from 2010 to 2015. They are randomly split into the training set and the testing set with the respective proportions of 67% and 33%. After training, the assignment probability of the ‘changed’ class yielded by the RF algorithm is regarded as the development potential value. The Kappa metric and the total operating characteristic (TOC) curve (Pontius Jr & Si, 2014) are used to evaluate

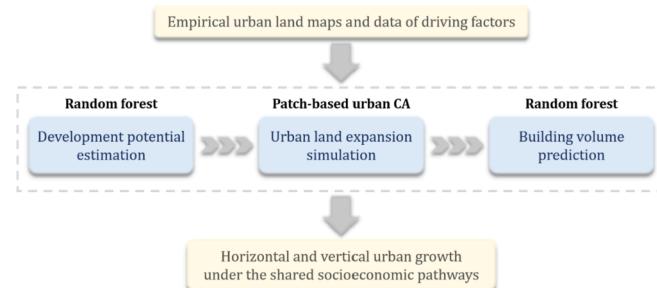


Fig. 3. The methodological framework of the extended patch-based urban CA model to simulate horizontal and vertical urban growth under the SSPs.

the performance of RF using the testing data.

3.2. Patch-based urban CA

A patch-based urban CA model is used to simulate horizontal urban growth (Chen et al., 2014). This model simulates the growth of one urban patch per iteration, and the entire simulation run is constrained by the pre-defined number of new urban cells. The simulation of the growth of an urban patch starts with the estimation of its size:

$$A_p = a_0(r_a)^{a_1} \quad (2)$$

where A_p is the estimated area of a simulated new urban patch; r_a is a randomly generated value between 0 and 1; a_0 and a_1 are parameters that can be calibrated by fitting the actual cumulative area distribution of patches of new urban land. Here the land use maps of 2010 and 2018 are used to obtain the patch areas of new urban land for calibrating a_0 and a_1 .

Once the area of a simulated new patch is set, its growth type, i.e., either spontaneous growth or organic growth, is determined by comparing a constant threshold T_{spont} with a randomly generated value between 0 and 1. Here the type of spontaneous growth refers to the urban land expansion that is not adjacent to existing urban areas. In contrast, organic growth refers to edge expansion, i.e., the urban land expansion at the edges of existing urban areas. If T_{spont} is greater or equal to the random value, then the type of the simulated new patch is set as spontaneous growth. Otherwise, it is organic growth.

The simulation of patch growth includes two procedures: determining the location of a seed and simulating subsequent developments that spread out from the seed. When allocating a seed for a patch that is the spontaneous growth type, a non-urban cell s is first randomly drawn from the space, and its development potential is compared with a randomly generated value between 0 and 1. If the development potential is greater or equal to the randomly generated value, the non-urban cell s is selected as the seed, and its state is converted into 'urban'. Otherwise, another non-urban cell is randomly drawn and so forth until the seed is set. The seed allocation procedure for a simulated patch of organic growth is almost identical, except that the product of the development potential and the neighborhood development density is used to compare against the randomly generated value. Here the neighborhood development density is calculated as:

$$\Omega = \frac{\sum_{n=1}^n con(c_n = \text{urban})}{n} \quad (3)$$

where Ω is the neighborhood development density; n is the number of cells in the neighborhood of the non-urban cell s ; $con(c_n = \text{urban})$ returns one if the state of the neighboring cell c_n is 'urban'. Here the 3×3 Moore neighborhood is used according to empirical literature (Chen et al., 2014).

The growth of a patch is then simulated using a 3×3 moving window. First, the window centers on the position of the seed. Second, within the window's scope, a non-urban cell with the greatest development probability (i.e., the product of the development potential and the neighborhood development density) is drawn and converted into 'urban'. Third, the window moves and centers on the newly converted cell, and again the non-urban cell with the greatest development probability is converted into 'urban'. Then the window moves on to convert the next non-urban cell until the total amount of simulated urban cells meets the estimated patch area or the non-urban cells within the window's scope are depleted.

The simulated horizontal urban growth is validated at both cell and landscape levels. At the cell level, the metric of 'Figure-of-Merit' (FoM) (Pontius et al., 2008) is used to evaluate the performance of the patch-based urban CA in predicting the locations of new urban cells. The FoM metric is calculated as follows:

$$F = \frac{B}{A + B + C + D} \quad (4)$$

where F is the value of FoM; A is the number of cells that are observed changed but simulated unchanged; B is the number of cells that are observed changed and simulated changed as well; C is the number of cells that are observed changed but simulated changed incorrectly; D is the number of cells that are observed unchanged but simulated changed. Here C is zero because the model only simulates the conversion from non-urban to urban.

At the landscape level, three landscape metrics are used to quantify the similarity between the observed and the simulated urban land patterns. These metrics include the number of patches (NP), mean Euclidean distance of nearest neighbors (ENN), and mean perimeter-to-area ratio (PARA). They are selected according to empirical literature (Li et al., 2017; Pickard et al., 2019). Based on these metrics, the indicator of pattern similarity a_l is calculated as follows:

$$a_l = 100\% - \frac{1}{3} \sum \Delta_k \quad (5)$$

$$\Delta_k = \frac{|l_{k,s} - l_{k,o}|}{l_{k,o}} \quad (6)$$

where Δ_k is the relative absolute difference between $l_{k,s}$ and $l_{k,o}$, which are the k th metrics of the simulated and observed urban land patterns, respectively.

3.3. Predicting building volumes

The RF algorithm is also applied to the prediction of building volumes at a 150-m resolution. The input variables include build-up land area, the five kernel density layers (for the POIs types of shopping center, company, industrial, office, and residential), the six distance variables (i.e., distances to city center, town center, road, railway stations, metro stations, and highway entrances), and slope. Additionally, as the spatial pattern of building volumes demonstrates strong spatial autocorrelation, including the neighborhood building volumes may help improve the prediction accuracy. Therefore, four additional variables, namely the mean building volume and the min-max range of building volumes in the 3×3 and 5×5 neighborhoods, are added to train the RF algorithm.

A set of randomly drawn grid cells (30% of total grid cells) are used to build the RF-based model to predict building volumes. The entire sample set is randomly split into the training set and the testing set with the proportions of 75% and 25%, respectively. Then the testing set is used to evaluate the performance of the RF algorithm with the metrics of squared-correlation-coefficient (R^2) and mean absolute percentage error (MAPE).

After training, the RF algorithm can be used to predict building volumes based on the simulated urban land expansion. Specifically, the grid cells where the simulated urban land expansion occurs (denoted as U_{hg}) are identified. The values of built-up land area are updated by aggregating the simulated urban land at 30-m resolution to U_{hg} . Afterward, the kernel density values of the five POIs types in U_{hg} are updated using the k -nearest-neighbor (k -nn) method with the variables of built-up land area, slope, and the distances to city center, town center, road, railway stations, metro stations, and highway entrances. These variables are used to quantify the similarity between an input instance and the existing cases that are stored in a library. Here the similarity is measured based on the Euclidean distance between the input variables with min-max normalization. The k existing cases with the shortest distance are extracted as the nearest neighbors of the input instance. Here k is set as 10 by balancing the accuracy achieved and the computation efficiency. The distances of the k nearest neighbors are further transformed into weights:

$$w_k = \frac{1/d_k^2}{\sum_k 1/d_k^2} \quad (7)$$

where w_k is the weight of the k th nearest neighbor, and d_k is the distance between the input instance and the k th nearest neighbor. By using these weights, the kernel density values of the input instance are calculated as follows:

$$\rho_f = \sum_k w_k \rho_{k,f} \quad (8)$$

where ρ_f is the estimated kernel density value of the f th POIs type for the input instance and $\rho_{k,f}$ is the kernel density value of the f th POIs type for the k th nearest neighbor. By this means, the kernel density values of the five POIs types in U_{hg} can be updated, thereby predicting the building volumes for U_{hg} using the trained RF.

3.4. Localization of SSPs scenarios

There are five scenarios in the framework of SSPs (Jiang & O'Neill, 2017): SSP1 refers to a sustainability scenario that features fast urbanization with environmentally friendly living arrangements and compact urban development; SSP2 refers to a scenario of moderate urbanization that is the continue of historical development path (the business-as-usual scenario); SSP3 refers to a scenario that features low urbanization due to slow economic growth, limited mobility, and poor urban planning; SSP4 is a scenario of inequality, in which medium- and low-income countries will experience faster urbanization than high-income countries; and SSP5 (the ‘fossil-fueled development’ scenario) assumes fast urbanization in all countries and the advanced technologies that allows for large-scale developments of desirable housing.

Recently the SSPs narratives have been regionalized from the perspective of urban land expansion at the country or regional level (Chen et al., 2020; Xuecao Li, Zhou, Eom, Yu, & Asrar, 2019). However, urban land expansion scenarios that are consistent with the SSPs narratives are still not available at the level of metropolitan areas. To the best of the author’s knowledge, the finest dataset relevant to SSPs urban scenarios are the one developed by Chen et al. (2020) with a 1-km resolution, while others are as coarse as 0.25 degrees (Gao & O’Neill, 2020). Therefore, the 1-km SSPs urban land dataset Chen et al. (2020) is used for the localization of SSPs narratives with respect to urban land areas. Specifically, for each SSPs scenario, the growth rates of urban land from 2015 to 2050 are obtained from the 1-km projections. Then the growth rates are applied to the 30-m urban land map in 2015 to estimate the urban land area in 2050.

However, due to the differences of spatial resolution, the estimated urban land areas are unreasonably large compared to the historical ones (Table 2). Therefore, the estimated urban land areas for the five SSPs scenarios are further adapted according to the actual growth of urban land area from 2015 to 2018. First, the growth rates of the urban land area in the period of 2015–2018 are derived from the 30-m resolution maps and assumed to continue till 2050, thereby adjusting the urban land area in SSP2 to 3073 km². Second, for SSP2, a ratio is obtained between the original estimated increase and the adjusted increase of urban land areas from 2018 to 2050 (i.e., 1133 km² and 414 km²,

respectively), which is 2.7367. Third, the ratio is used to scale the estimated increases of urban land areas in the other four SSPs scenarios (see the last row of Table 2). For instance, the estimated increase of urban land areas is 1282 km² in SSP1 from 2018 to 2050. By applying the ratio, the estimated increase is scaled to 468 km², and hence the total urban land areas in 2050 for SSP1 is 3127 km². By this means, the increases of urban land areas are adjusted in the other three scenarios. The scaled urban land areas are then used to constrain the simulations of horizontal urban growth from 2018 to 2050 under the SSPs.

Moreover, the parameters of the patch-based urban CA, namely a_0 and T_{spon} , are also adjusted to reflect the features of different SSPs scenarios (Table 3). Here a_0 controls the size of a new urban patch. The higher the values of a_0 , the larger the size of a new urban patch is generated. The parameter T_{spon} controls the tendency of urban sprawl. Greater T_{spon} will generate more urban patches of the spontaneous growth type, and hence more fragmented patterns. According to the narratives of SSPs, SSP2 is the business-as-usual scenario, and hence the calibrated values of a_0 and T_{spon} are directly applied in this scenario. For the other four scenarios of SSPs, the narratives only provide qualitative descriptions of urban form with terms such as ‘compact’, ‘low’, ‘fast’, etc. Therefore, parameters a_0 and T_{spon} have to be subjectively determined to ensure that the resulting simulations fit the SSPs narratives. To this end, the parameters a_0 and T_{spon} in SSP2 are scaled proportionally and further applied in the other four SSPs.

The scaling factors for a_0 and T_{spon} are determined according to previous modeling experiences in the study area (Chen et al., 2014; Chen et al., 2016). Among the four remaining SSP scenarios, SSP1 and SSP5 represent two extreme cases of future urbanization, i.e., compact vs. sprawling developments. For SSP1, the parameter a_0 is set as five times of the a_0 in SSP2 while the parameter T_{spon} is set as 1/3 of the T_{spon} in SSP2 to ensure that the model can generate a compact urban form. Similarly, for SSP5, the settings of the parameters a_0 and T_{spon} should allow the simulations of sprawling urban form in an extreme case, and hence a_0 is set as 1/5 of the a_0 in SSP2 while T_{spon} is set as three times of the T_{spon} value in SSP2. For SSP3, a_0 is reduced into 80% of the SSP2 a_0 value and T_{spon} is enlarged by two times of the SSP2 T_{spon} value to match the assumption of low urbanization and relatively loose land control in SSP3. SSP4 is similar to SSP5, and the a_0 and T_{spon} are set respectively as 50% and 2.5 times of the corresponding SSP2 values.

4. Implementation and results

4.1. Model calibration and validation

The proposed methods are implemented using Python with additional packages such as ‘scikit-learn’, ‘rasterio’, and ‘pylandstats’. Model calibration using RF is carried out using ‘scikit-learn’ with the default parameter settings because they already yield satisfactory results. In order to obtain a more reliable result, the RF-algorithm is trained and tested for 30 rounds.

The testing accuracy, measured by the Kappa metric, is 0.8634 for the RF-based model for development potential estimation. This model is finalized using the complete set of data and used to generate the development potential at a 30-m resolution (Fig. 4(a)). The results of TOC show a curve above the random line and a high value of area-under-curve (AUC) for 0.9755 (Fig. 4(b)). These results suggest that RF performs much better than the random model (AUC = 0.5) in the estimation of urban development potential. Moreover, the parameters a_0 and a_1 are

Table 2

The scaled SSPs urban land areas for simulating urban growth between 2015 and 2050 (km²).

	2015	2018	SSP1	SSP2	SSP3	SSP4	SSP5
30-m map	2620	2659	–	–	–	–	–
1-km map	2513	–	3780	3637	3411	3577	3842
Est. area	–	–	3941	3792	3556	3729	4005
Scl. area	–	–	3127	3073	2987	3050	3151

Note: Est. area = Estimated area; Scl. area = Scaled area.

Table 3

The adjusted parameters for the simulation of horizontal urban growth between 2015 and 2050 under SSPs.

	Calibrated	SSP1	SSP2	SSP3	SSP4	SSP5
a_0	68.47	342.35	68.47	54.78	34.24	13.69
T_{spon}	0.2	0.06	0.2	0.4	0.5	0.6

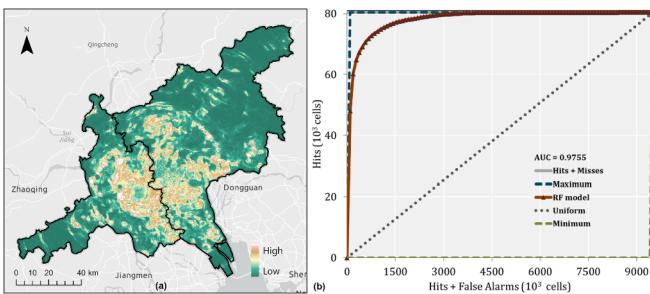


Fig. 4. (a) Development potential generated by the built RF model. (b) Validation of the development potential using the TOC approach.

calibrated using the empirical urban land maps in 2010 and 2018, with $a_0 = 68.47$ and $a_1 = -1.02$. The parameter T_{spn} is manually tuned to 0.2.

The performance of the patch-based urban CA model is tested by simulating the historical horizontal urban growth from 2010 to 2018 with the development potential and the calibrated parameters of a_0 , a_1 , and T_{spn} . The actual amount of urban land increase from 2010 to 2018 is used to constrain the simulations. As multiple runs can better understand the model behaviour, the patch-based urban CA model is run 100 times with the same parameter configuration. Then the 100 simulations are compared against the actual urban land pattern in 2018. The indicator of FoM and the three selected landscape metrics are used to quantify the agreements between the simulations and the actual urban land expansion at the cell and landscape levels, respectively (Fig. 5).

The 100 simulations are first compared cell-by-cell with the observed urban land map to detect errors and hits. A representative example is shown in Fig. 5(a) using the result of run #74. The corresponding amounts of hits and errors are used to calculate the indicator of FoM. The values of FoM for the 100 simulations range from 0.1222 to 0.1578, with a mean value of 0.1406 (Fig. 5(b)). Furthermore, the landscape metrics of ENN_MN, PARA_MN, and NP are computed for the 100 simulations (Fig. 5(c)). The relative errors between the simulated and the

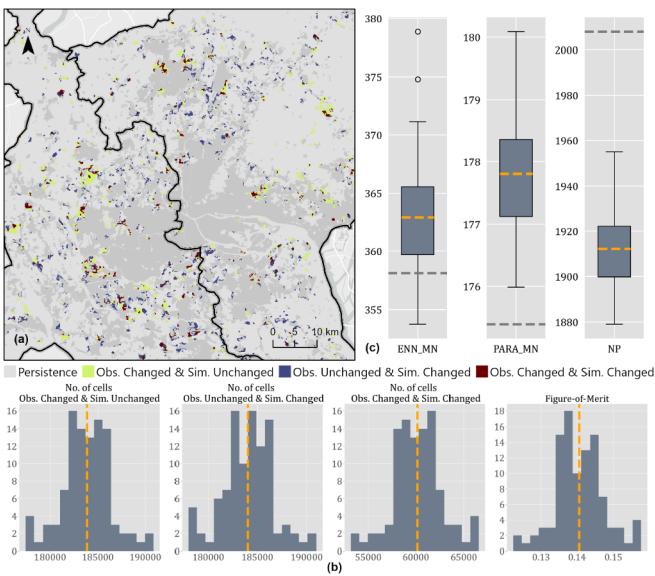


Fig. 5. (a) Cell-by-cell errors and hits of a representative run (#74) of horizontal urban growth simulations. (b) Histograms of the cell-by-cell errors and hits of the 100-run simulations. The orange dash lines are the mean values of errors and hits. The mean FoM is 0.1406. (c) Boxplots of the metrics ENN_MN, PARA_MN, and NP for the 100-run simulations. The orange dash lines are the mean values of these metrics, which are 362.92 for ENN_MN, 177.81 for PARA_MN, and 1912.25 for NP. The grey dash lines represent the observed values of ENN_MN (358.10), PARA_MN (175.40), and NP (2008).

observed metrics are 0.01%–5.80% for ENN_MN, 0.34%–2.67% for PARA_MN, and 2.64%–6.42% for NP. These errors are due to the overestimation of ENN_MN and PARA_MN, and the underestimation of NP. Nevertheless, the mean pattern similarity (Eq. (5)) is over 97%, suggesting that the patch-based urban CA can produce reliable simulations of horizontal urban growth.

The RF algorithm is also trained to predict building volumes based on the simulated horizontal urban growth. Fig. 6(a) shows the scatter plot of the actual building volumes derived from the testing data versus the predictions made by the RF algorithm. The testing R^2 is 0.85, and the MAPE is 37.89%. For the subsequent scenario simulations, this RF model is finalized by training it with the complete data set. Moreover, the k -nn approach ($k = 10$) for updating the kernel density values of the five POIs types is also evaluated using a sample dataset that is randomly drawn from the total grid cells by 30%. In this dataset, 75% of randomly chosen instances are used to form a case library, and the remaining instances are used to form the testing set for evaluating the performance of the k -nn approach. For a certain instance in the testing set, its kernel density values of the five POIs are replaced with the kernel density values estimated by k -nn, while holding the values of the other variables unchanged. Then this instance is put in the finalized RF model to predict the building volume. Finally, the MAPE of the predicted building volumes is calculated. The results show a testing MAPE of 22.37% and testing R^2 of 0.90 for the k -nn approach. In the subsequent scenario simulations, the complete data set is used as the case library to update the kernel density values for other grid cells in which horizontal urban growth occurs.

4.2. Projections of horizontal and vertical urban growth under SSPs

The calibrated patch-based urban CA model is applied to horizontal urban growth simulation under the SSPs. The scenario simulations are constrained by the urban land demands shown in Table 2. For each scenario, the model is run ten times to produce multiple simulations. Then the results are stacked to investigate the consistency of the multi-run simulations in each scenario (Fig. 7(a)).

The mean spatial consistencies of the multi-run simulations are all below 40% for the five SSPs scenarios. The low spatial consistencies are not surprising because the patch-based urban CA model is stochastic, and the simulations are affected by path dependence (Brown et al., 2005). These results also reflect that the patch-based urban CA model can simulate divergent development paths under the same process. The results also show that the proportions of cells where the model consistently predicts urban growth (in more than 50% of runs) vary largely across the five scenarios (Table 4). In SSP1, over 26% of the cells are consistently predicted as 'urban' in more than 50% of runs. In SSP3 and SSP4, however, the proportions are only 9.9% and 11.6%, respectively. The proportions in SSP2 and SSP5 are slightly higher, which are 17.6%

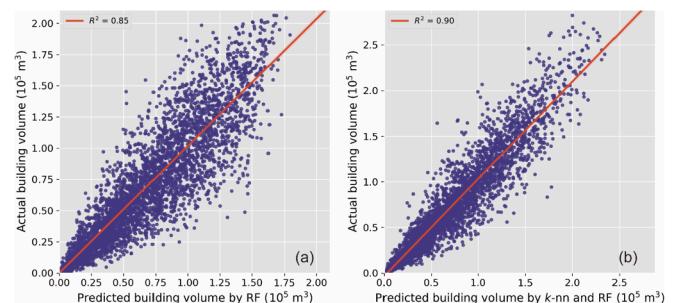


Fig. 6. (a) Comparing the actual building volumes and the predictions derived from RF with the testing data. (b) Comparing the actual building volumes and the predictions made by k -nn and RF with the testing data. This comparison evaluates the performance of the k -nn method in estimating the kernel density values for the predictions of building volumes under SSPs.

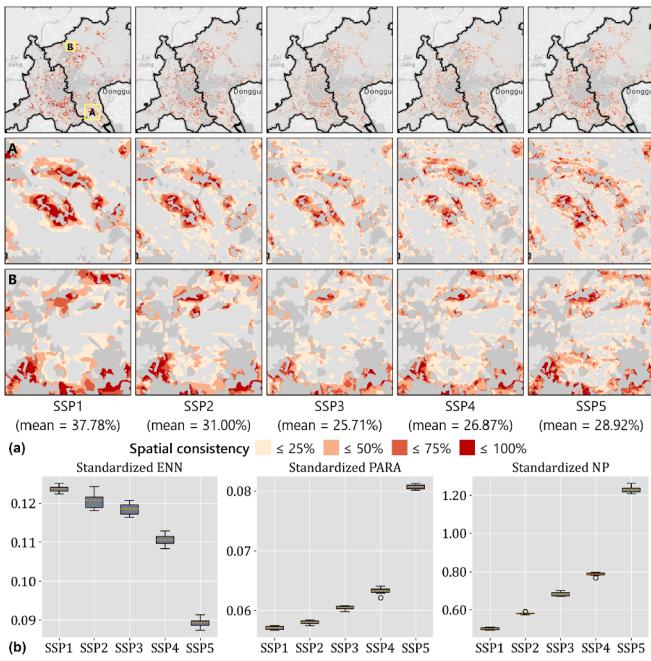


Fig. 7. (a) Spatial consistencies of the multi-run horizontal urban growth simulations, and (b) the corresponding results of landscape metrics in the five SSP scenarios.

Table 4
Spatial consistencies of the multi-run simulations and their proportions in the five SSP scenarios.

Consistency	SSP1	SSP2	SSP3	SSP4	SSP5
≤ 25%	42.2%	53.7%	61.9%	60.5%	56.7%
25%–50%	31.4%	28.7%	28.2%	27.8%	29.3%
50%–75%	14.5%	10.9%	7.2%	7.9%	9.0%
≥ 75%	11.9%	6.7%	2.7%	3.7%	5.0%

and 14%, respectively.

Such variations are mainly due to the configurations of each scenario, including the demand of urban land and the development tendency. SSP1 is the only scenario that features a large demand of urban land and a tendency to compact development. Urban growth in this scenario is simulated mainly in the forms of edge-expansion and infill developments. As a result, the model consistently predicts the non-urban cells that are adjacent to existing urban areas as the primary locations for urban developments. SSP3, SSP4, and SSP5 represent three scenarios of sprawl development. Therefore, the locations where the model predicts urban growth are more dispersed than those in SSP1 and SSP2 (the business-as-usual scenario) (Fig. 7(a)), and the spatial consistencies of the simulations are also relatively low in these three scenarios.

In addition to the assessment of spatial consistency, the scenario simulations are compared at landscape level using the metrics of ENN_MN, PARA_MN, and NP. By calculating the coefficient of variation for these metrics, it is found that ENN_MN and NP vary less than 2%, and PARA_MN varies less than 1% for the multi-run simulations in the same scenario. These results suggest that the patch-based urban CA model produces stable urban land patterns, although their cell-level consistencies are relatively low (Fig. 7(a)).

Because the demands of urban land vary from one scenario to another, the results of ENN_MN, PARA_MN, and NP cannot be compared directly between scenarios. Therefore, standardization is applied by dividing the metrics by urban land areas (Fig. 7(b)). As SSP1 represents a scenario of compact urban developments, the simulations in this scenario feature the highest ENN_MN and the lowest PARA_MN and NP.

SSP5 represents the other extreme, in which the simulated patterns are highly fragmented, dispersed, and complicated, resulting in the highest NP and PARA and the lowest ENN_MN. SSP3 and SSP4 also feature stronger tendencies to sprawl development than SSP2 (the business-as-usual scenario), and hence, they yield higher values of PARA and NP and lower values of ENN_MN.

The multi-run simulations of urban land expansion in each scenario are aggregated to the 150-m grid cells. Then they are used to predict building volumes using the calibrated RF model. For the grid-cells where urban land expansion occur, their kernel density values of the five urban POIs types are estimated using the *k*-nn method (see Sections 3.3 and 4.1). Fig. 8 and Table 5 show the simulated building volumes for each scenario.

The results of the simulated building volume increase vary largely across the SSPs scenarios. The red ellipses in Fig. 8 highlights the representative areas where the five scenarios produce different patterns of urban growth. The business-as-usual scenario (SSP2) predicts that urban land expansion occurs in 34,457 150-m grid cells on average and that leads to a mean building volume increase of 1214.56 million m³. If the compact development represented by SSP1 is implemented, the number of grid cells in which urban land expansion occurs is similar to those in SSP2, but the amount of building volume increase becomes 10.6% higher. The differences are more evident between the extreme cases of SSP1 and SSP5. SSP1 and SSP5 assume equivalent urban land demands (Table 2), and their projected increase in building volumes are also approximately equal. However, the number of 150-m grid cells that urban land expansion occurs in SSP5 is approximately 40% more than that in SSP1. The main reason is that sprawl development in SSP5 tends to expand the urban extent outward with dispersed, low-density developments. However, if compact development is promoted (SSP1), urban land expansion can be contained to a smaller extent, with the newly built areas being intensively developed and tightly connected with existing infrastructures. As a result, the building volume increase per grid cell is the lowest in SSP5 and the highest in SSP1 (Fig. 8). SSP4 is another scenario of urban sprawl. The projected increase in building volumes in SSP4 is less than that in SSP1 even though the number of grid cells that urban land expansion occurs is larger in SSP4. SSP3 yields the lowest projected increase in building volumes, largely related to the assumption of low urbanization. Nevertheless, Due to the stronger tendencies to sprawl development, the building volume increase per grid cell in SSP3 is lower than those in SSP1 and SSP2.

The mean building height can be estimated by dividing the building volume by the built-up land area for each 150-m grid cell, and the results are shown in Figs. 9 and 10. According to the assumptions of SSPs, the simulations of SSP1 are expected to be the highest, following by those of

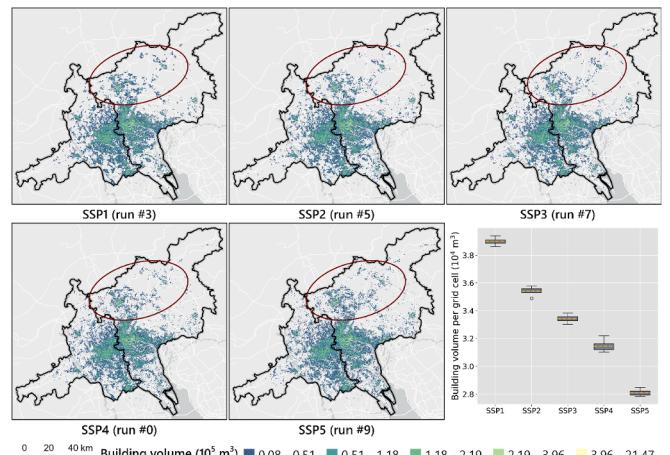


Fig. 8. Representative runs of building volume predictions and the growth of building volumes per 150-m grid cell in each SSP scenario.

Table 5

Summary of the 150-m resolution multi-run simulations of vertical urban growth with 95% confidence intervals for each SSP scenario. The total building volumes in 2018 is 4938 million m³.

	Increase in building volumes (10 ⁶ m ³)	Grid cells of urban growth
SSP1	1343.69 ± 20.18	34,457 ± 184
SSP2	1214.56 ± 13.46	34,302 ± 157
SSP3	947.27 ± 12.16	28,344 ± 272
SSP4	1101.63 ± 12.68	35,042 ± 337
SSP5	1355.25 ± 13.34	48,210 ± 315

SSP2, SSP3, SSP4, and SSP5. The profiles of the estimated building heights with respect to distance to the city center confirm this expectation (Fig. 9). According to the results of the binned distance to city center, all simulated urban growth is observed in the places that are over 10 km away from the city center. For the grid cells that are 10–20 km away from the city center, the estimated mean building heights are the greatest in SSP1 and SSP2, and the smallest in SSP5. When the distance to the city center gets longer, the differences in the estimated building heights become larger except for SSP2 and SSP3, of which the mean height curves starts overlapping after 30 km. In particular, the differences of the estimated building heights between SSP1 and the other four scenarios (especially SSP5) become the largest in the distance range of 30–50 km. Another notable feature is that for each distance bin, the grid

cells in SSP1 always have the smallest binned average distance to city center, following by SSP2, SSP3, SSP4, and SSP5. This result is echo with the assumptions of the five scenarios (i.e., compact vs. sprawling developments).

5. Discussion

Conventional studies of urban growth projections often consider the horizontal dimension (i.e., urban land expansion) only. While the environmental impacts of urban form are complicated, overlooking the change in vertical characteristics may lead to an incomplete understanding of future urbanization and its consequences. The proposed model can alleviate this issue by explicitly simulating vertical urban growth. In this case study, the scenario simulations reveal the varying distributions of building volumes and heights due to the different development tendencies in the SSPs. The results of the scenario simulations are critical to understanding the effects of various urban forms on, for instance, energy consumption (Güneralp et al., 2017), urban heat island effects (Zhang, Middel, & Turner, 2019), and ecosystem services (Chen, Li, Liu, Zhang, & Huang, 2019). Additionally, urban areas are vulnerable to many hazards (Güneralp, Güneralp, & Liu, 2015). As the spatial layout of building volumes can indicate how urban population distributes, the presented scenario simulations can help assess hazard risk and vulnerability in the context of climate change. While the SSPs

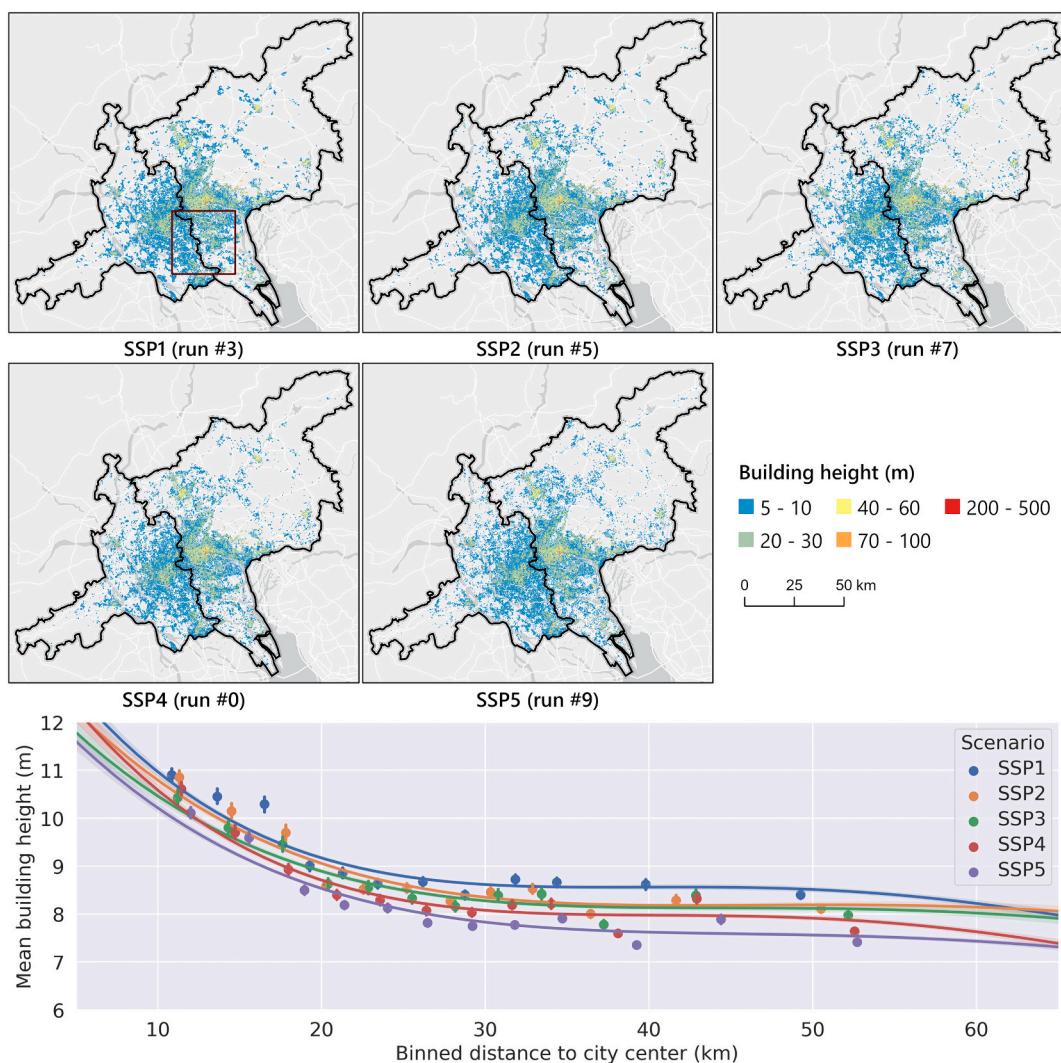


Fig. 9. Representative runs of building height predictions in each scenario and the profiles of the estimated mean building height with respect to distance to city center.

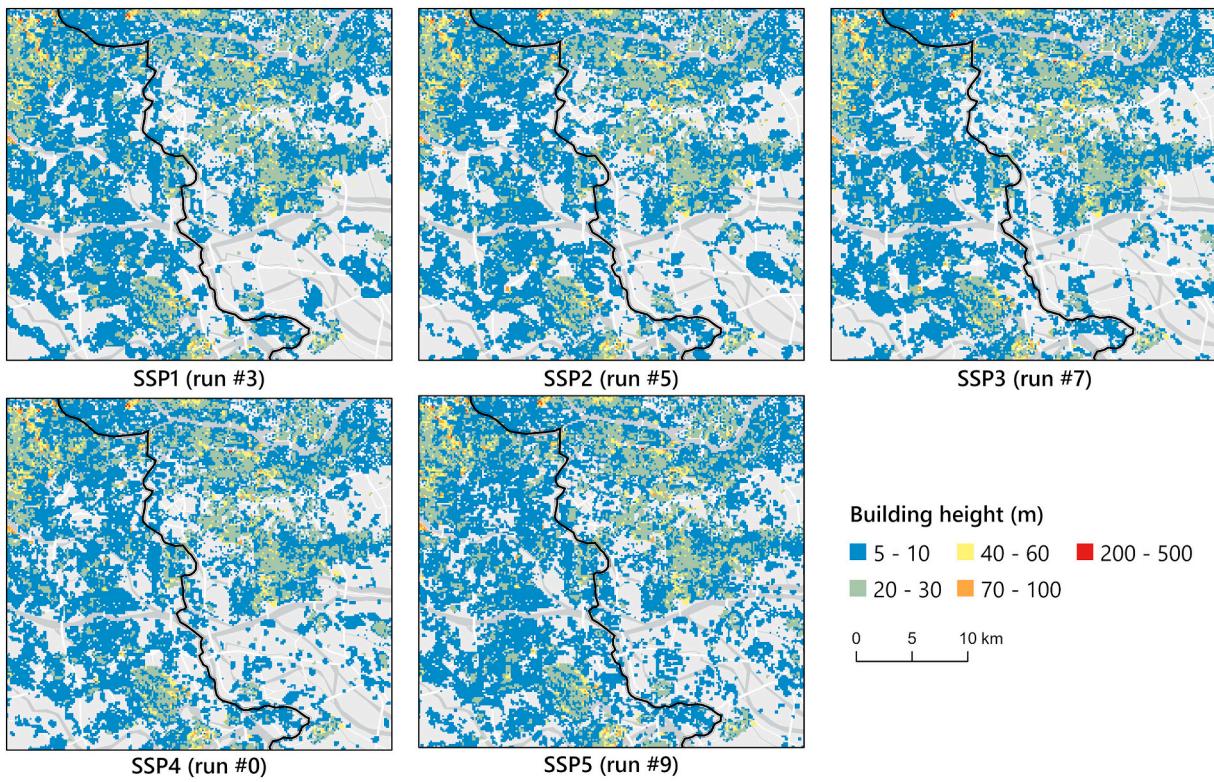


Fig. 10. Zoom-in view of the representative area highlighted (by a red box) in Fig. 9. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

focus on the socioeconomic dimension of global change, the possible futures of global climate can be represented with the RCPs. Hence, it is feasible to further integrate the SSPs simulations of horizontal and vertical urban growth with the RCPs to explore the interacting processes of urbanization and climate change.

The proposed extended patch-based urban CA model can be easily adapted to applications in other geographic regions. The methods used to calibrate the model can be replaced with others. For instance, besides the RF algorithm, alternative machine learning algorithms, such as artificial neural networks and support vector machine, also are feasible methods to estimate development potential and predict building volumes. Additionally, the localized SSPs parameters in this study is one of the probable parameter configurations for SSPs. The model parameters can be flexibly determined for other geographic regions according to local conditions and development tendencies. The SSPs urban land demand can also be acquired using alternative data sources (Gao & O'Neill, 2020) or explicitly estimated based on socioeconomic models (Xuecao Li et al., 2019). Finally, several open access datasets, such as the Microsoft Building Footprint Data, have provided the information of individual buildings for some countries and regions in the world (e.g. USA, Canada and Australia), which can be used to estimate and predict vertical urban growth.

This case study has several limitations. First, the proposed model simulates horizontal and vertical urban growth separately, which is not fully aligned with actual urban planning activities. More techniques such as Building Information Modeling (BIM) can be incorporated to produce more realistic simulations at fine scales. Second, due to the issue of data availability, the building data and POIs data are only for a single year and the conditions of urban infrastructures (e.g., the road network) are held constant in the scenario simulations. Solving this problem requires additional efforts in future work to collect and build a multi-temporal dataset of urban buildings and POIs. Third, this case study focuses on urban growth, and the vertical change induced by urban renewal is not explicitly modeled. Adding a new component to

assess the probability of urban renewal may enhance the performance of the proposed model.

6. Conclusion

Accurate simulations of vertical urban growth are fundamental to understanding the social and environmental effects of future urbanization. While urban CA models have been widely used to simulate urban expansion, many urban CA models cannot effectively simulate vertical urban growth. This study addresses this drawback by extending a patch-based urban CA model with a component that can predict the associated building volumes of an urban land expansion scenario. Compared with conventional models that only represents the change in the types of vertical characteristics (e.g., from 'low-rise' to 'mid-rise'), the proposed model can explicitly simulate the continuous vertical growth, thereby providing more quantitative details than conventional models to explore the effects of different urban forms. This study also contributes methodologically to literature of urban consequences under the SSPs, because most existing models seldom address the impacts of different pathways on vertical urban developments.

The proposed model is evaluated through the historical urban growth simulations in the Guangzhou-Foshan metropolitan area, China. The simulated urban land expansion yields a mean FoM value of 0.1406 at the cell level and an agreement of 97% at the pattern level. The error of building volume prediction is approximately 22%, as estimated with the testing data. The proposed model is further applied to the simulations of urban growth under the SSPs. The scenario simulations successfully reflect the effects of different development tendencies on the change in building volumes and building heights. As the simulated building volumes and heights are fundamental parameters to urban climate modeling, the proposed model can support the mitigation of climate change effects on human settlements.

The major limitation of this study is the lack of multi-temporal data for urban buildings and facilities. As a result, the temporal relationships

between vertical urban growth and its driving factors cannot be fully captured. Another limitation is that the proposed model cannot represent the urban renewal process that may also induces change in building volumes and heights. Therefore, future work will focus on addressing these issues by acquiring more data and explicitly modeling the urban renewal process.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Credit author statement

Yimin Chen: Conceptualization, Methodology, Data curation, Validation, Visualization, Investigation, Writing.

Acknowledgements

The author sincerely thanks the reviewers for their helpful comments and suggestions. The research was supported by the National Key R&D Program of China (2019YFA0607201, 2017YFA0604401), the National Natural Science Foundation of China (Grant No. 41871306), the Guangdong Natural Science Funds for Distinguished Young Scholar (Grant No. 2021B1515020104), the Fundamental Research Funds for the Central Universities (20lgzd09), and the Open Fund of Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University (Grant No. KLGIS2020A02).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.compenvurbsys.2021.101727>.

References

- Abolhasani, S., Taleai, M., Karimi, M., & Rezaee Node, A. (2016). Simulating urban growth under planning policies through parcel-based cellular automata (ParCA) model. *International Journal of Geographical Information Science*, 30(11), 2276–2301.
- Alaei Moghadam, S., Karimi, M., & Habibi, K. (2018). Simulating urban growth in a megalopolitan area using a patch-based cellular automata. *Transactions in GIS*, 22, 249–268.
- Brown, D. G., Page, S., Riolo, R., Zellner, M., & Rand, W. (2005). Path dependence and the validation of agent-based spatial models of land use. *International Journal of Geographical Information Science*, 19(2), 153–174.
- Cao, M., Huang, M., Xu, R., Lü, G., & Chen, M. (2019). A grey wolf optimizer–cellular automata integrated model for urban growth simulation and optimization. *Transactions in GIS*, 23(4), 672–687.
- Chen, B., Xu, B., & Gong, P. (2021). Mapping essential urban land use categories (EULUC) using geospatial big data: Progress, challenges, and opportunities. *Big Earth Data*, 5(3), 410–441.
- Chen, G., Li, X., Liu, X., Chen, Y., Liang, X., Leng, J., ... Huang, K. (2020). Global projections of future urban land expansion under shared socioeconomic pathways. *Nature Communications*, 11(1), 1–12.
- Chen, G., Xie, J., Li, W., Li, X., Chung, L. C. H., Ren, C., & Liu, X. (2021). Future “local climate zone” spatial change simulation in Greater Bay Area under the shared socioeconomic pathways and ecological control line. *Building and Environment*, 203, 108077.
- Chen, X., Zhang, H., Chen, W., & Huang, G. (2021). Urbanization and climate change impacts on future flood risk in the Pearl River Delta under shared socioeconomic pathways. *Science of the Total Environment*, 762, 143144.
- Chen, Y., Li, X., Liu, X., & Ai, B. (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.
- Chen, Y., Li, X., Liu, X., Ai, B., & Li, S. (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.
- Chen, Y., Li, X., Liu, X., Huang, H., & Ma, S. (2019). Simulating urban growth boundaries using a patch-based cellular automaton with economic and ecological constraints. *International Journal of Geographical Information Science*, 33(1), 55–80.
- Chen, Y., Li, X., Liu, X., Zhang, Y., & Huang, M. (2019). Tele-connecting China's future urban growth to impacts on ecosystem services under the shared socioeconomic pathways. *Science of the Total Environment*, 652, 765–779.
- Chen, Y., Liu, X., & Li, X. (2017a). Calibrating a land parcel cellular automaton (LP-CA) for urban growth simulation based on ensemble learning. *International Journal of Geographical Information Science*, 31(12), 2480–2504.
- Chen, Y., Liu, X., Li, X., Liu, X., Yao, Y., Hu, G., ... Pei, F. (2017b). 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.
- Clarke, K. C. (2019). Mathematical foundations of cellular automata and complexity theory. In L. D'Acci (Ed.), *The mathematics of urban morphology* (pp. 163–170). Cham, Switzerland: Springer.
- Feng, Y., & Tong, X. (2020). A new cellular automata framework of urban growth modeling by incorporating statistical and heuristic methods. *International Journal of Geographical Information Science*, 34(1), 74–97.
- Gao, J., & O'Neill, B. C. (2020). Mapping global urban land for the 21st century with data-driven simulations and shared socioeconomic pathways. *Nature Communications*, 11(1), 1–12.
- Gemitzi, A. (2021). Predicting land cover changes using a CA Markov model under different shared socioeconomic pathways in Greece. *GIScience & Remote Sensing*, 1–17.
- Goetz, A. (2013). Suburban sprawl or urban centres: Tensions and contradictions of smart growth approaches in Denver, Colorado. *Urban Studies*, 50(11), 2178–2195.
- Gounaris, D., Chorianopoulos, I., Symeonakis, E., & Koukoulas, S. (2019). A random forest-cellular automata modelling approach to explore future land use/cover change in Attica (Greece), under different socio-economic realities and scales. *Science of the Total Environment*, 646, 320–335.
- Güneralp, B., Güneralp, I., & Liu, Y. (2015). Changing global patterns of urban exposure to flood and drought hazards. *Global Environmental Change*, 31, 217–225.
- Güneralp, B., Zhou, Y., Ürge-Vorsatz, D., Gupta, M., Yu, S., Patel, P. L., ... Seto, K. C. (2017). Global scenarios of urban density and its impacts on building energy use through 2050. *Proceedings of the National Academy of Sciences*, 114(34), 8945–8950.
- Guzman, L. A., Escobar, F., Peña, J., & Cardona, R. J. L. U. P. (2020). A cellular automata-based land-use model as an integrated spatial decision support system for urban planning in developing cities: The case of the Bogotá region. 92 p. 104445.
- Hagenauer, J., Omrani, H., & Helbich, M. (2019). Assessing the performance of 38 machine learning models: the case of land consumption rates in Bavaria, Germany. *International Journal of Geographical Information Science*, 1–21.
- He, M., Chen, C., Zheng, F., Chen, Q., Zhang, J., Yan, H., & Lin, Y. (2021). An efficient dynamic route optimization for urban flooding evacuation based on cellular automata. *Computers, Environment and Urban Systems*, 87, 101622.
- He, Q., Liu, Y., Zeng, C., Chaohui, Y., & Tan, R. (2017). Simultaneously simulate vertical and horizontal expansions of a future urban landscape: A case study in Wuhan, Central China. *International Journal of Geographical Information Science*, 31(10), 1907–1928.
- Huang, K., Leng, J., Xu, Y., Li, X., Cai, M., Wang, R., & Ren, C. (2021). Facilitating urban climate forecasts in rapidly urbanizing regions with land-use change modeling. *Urban Climate*, 36, 100806.
- Jiang, L., & O'Neill, B. C. (2017). Global urbanization projections for the shared socioeconomic pathways. *Global Environmental Change*, 42, 193–199.
- Karimi, F., Sultana, S., Babakan, A. S., & Suthaharan, S. (2019). An enhanced support vector machine model for urban expansion prediction. *Computers, Environment and Urban Systems*, 75, 61–75.
- Ke, X., Zheng, W., Zhou, T., & Liu, X. (2017). A CA-based land system change model: LANDSCAPE. *International Journal of Geographical Information Science*, 31(9), 1798–1817.
- Koziatek, O., & Dragičević, S. (2017). iCity 3D: A geosimulation method and tool for three-dimensional modeling of vertical urban development. *Landscape and Urban Planning*, 167, 356–367.
- Li, X., Gong, P., Yu, L., & Hu, T. (2017). A segment derived patch-based logistic cellular automata for urban growth modeling with heuristic rules. *Computers, Environment and Urban Systems*, 65, 140–149.
- Li, X., Zhou, Y., Eom, J., Yu, S., & Asrar, G. R. (2019). Projecting global urban area growth through 2100 based on historical time series data and future shared socioeconomic pathways. *Earth's Future*, 7(4), 351–362.
- Li, Y., Schubert, S., Kropf, J. P., & Rybski, D. (2020). On the influence of density and morphology on the urban heat island intensity. *Nature Communications*, 11(1), 1–9.
- Liang, X., Guan, Q., Clarke, K. C., Chen, G., Guo, S., & Yao, Y. (2021). Mixed-cell cellular automata: A new approach for simulating the spatio-temporal dynamics of mixed land use structures. *Landscape and Urban Planning*, 205, 103960.
- Liang, X., Liu, X., Chen, G., Leng, J., Wen, Y., & Chen, G. (2020). Coupling fuzzy clustering and cellular automata based on local maxima of development potential to model urban emergence and expansion in economic development zones. *International Journal of Geographical Information Science*, 34(10), 1930–1952.
- Lin, J., Huang, B., Chen, M., & Huang, Z. (2014). Modeling urban vertical growth using cellular automata—Guangzhou as a case study. *Applied Geography*, 53, 172–186.
- Liu, S., & Shi, Q. (2020). Local climate zone mapping as remote sensing scene classification using deep learning: A case study of metropolitan China. *ISPRS Journal of Photogrammetry and Remote Sensing*, 164, 229–242.
- Liu, Y., Batty, M., Wang, S., & Corcoran, J. (2021). Modelling urban change with cellular automata: Contemporary issues and future research directions. *Progress in Human Geography*, 45(1), 3–24.
- McGarigal, K., Plunkett, E. B., Willey, L. L., Compton, B. W., DeLuca, W. V., & Grand, J. (2018). Modeling non-stationary urban growth: The SPRAWL model and the ecological impacts of development. *Landscape and Urban Planning*, 177, 178–190.
- McManamay, R. A., Vernon, C. R., & Jager, H. I. (2021). Global biodiversity implications of alternative electrification strategies under the shared socioeconomic pathways. *Biological Conservation*, 260, 109234.

- Meentemeyer, R. K., Tang, W., Dorning, M. A., Vogler, J. B., Cunniffe, N. J., & Shoemaker, D. A. (2013). FUTURES: Multilevel simulations of emerging urban–rural landscape structure using a stochastic patch-growing algorithm. *Annals of the Association of American Geographers*, 103(4), 785–807.
- Mustafa, A., Rienow, A., Saadi, I., Cools, M., & Teller, J. (2018). Comparing support vector machines with logistic regression for calibrating cellular automata land use change models. *European Journal of Remote Sensing*, 51(1), 391–401.
- Newland, C. P., van Delden, H., Zecchin, A. C., Newman, J. P., & Maier, H. R. (2020). A hybrid (semi) automatic calibration method for cellular automata land-use models: Combining evolutionary algorithms with process understanding. *Environmental Modelling & Software*, 134, 104830.
- O'Neill, B. C., Kriegler, E., Ebi, K. L., Kemp-Benedict, E., Riahi, K., Rothman, D. S., ... Kok, K. (2017). The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century. *Global Environmental Change*, 42, 169–180.
- Pickard, B. R., Meentemeyer, R. K. J. C., & Environment, & Systems, U. (2019). *Validating land change models based on configuration disagreement*. 77 p. 101366.
- Pontius, R. G., Jr., & Si, K. (2014). The total operating characteristic to measure diagnostic ability for multiple thresholds. *International Journal of Geographical Information Science*, 28(3), 570–583.
- Pontius, R., Boersma, W., Castella, J., Clarke, K., de Nijs, T., Dietzel, C., ... Kok, K. (2008). Comparing the input, output, and validation maps for several models of land change. *The Annals of Regional Science*, 42(1), 11–37.
- Rao, S., Klimont, Z., Smith, S. J., Van Dingenen, R., Dentener, F., Bouwman, L., ... van Vuuren, D. P. (2017). Future air pollution in the shared socio-economic pathways. *Global Environmental Change*, 42, 346–358.
- Rienow, A., Mustafa, A., Krelaus, L., & Lindner, C. (2021). Modeling urban regions: Comparing random forest and support vector machines for cellular automata. *Transactions in GIS*, 25(3), 1625–1645.
- Rohat, G., Wilhelm, O., Flacke, J., Monaghan, A., Gao, J., Dao, H., & van Maarseveen, M. (2019). Characterizing the role of socioeconomic pathways in shaping future urban heat-related challenges. *Science of the Total Environment*, 695, 133941.
- Shafizadeh-Moghadam, H., Minaei, M., Pontius, R. G., Jr., Asghari, A., & Dadashpoor, H. (2021). Integrating a forward feature selection algorithm, random forest, and cellular automata to extrapolate urban growth in the Tehran-Karaj region of Iran. *Computers, Environment and Urban Systems*, 87, 101595.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis* (vol. 26). CRC press.
- Stevens, D., Dragicevic, S., & Rothley, K. (2007). iCity: A GIS-CA modelling tool for urban planning and decision making. *Environmental Modelling & Software*, 22(6), 761–773.
- Tang, D., Liu, H., Song, E., & Chang, S. (2020). Urban expansion simulation from the perspective of land acquisition-based on bargaining model and ant colony optimization. *Computers, Environment and Urban Systems*, 82, 101504.
- Wang, F., & 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.
- Wang, H., Zhang, B., Xia, C., He, S., & Zhang, W. (2020). Using a maximum entropy model to optimize the stochastic component of urban cellular automata models. *International Journal of Geographical Information Science*, 34(5), 924–946.
- Wu, H., Li, Z., Clarke, K. C., Shi, W., Fang, L., Lin, A., & Zhou, J. (2019). Examining the sensitivity of spatial scale in cellular automata Markov chain simulation of land use change. *International Journal of Geographical Information Science*, 33(5), 1040–1061.
- Wu, H., Lin, A., Xing, X., Song, D., & Li, Y. (2021). Identifying core driving factors of urban land use change from global land cover products and POI data using the random forest method. *International Journal of Applied Earth Observation and Geoinformation*, 103, 102475.
- Xu, H., & Brown, D. G. (2017). Sensitivity of a stochastic land-cover change model to pixel versus polygonal land units. *International Journal of Geographical Information Science*, 31(4), 738–762.
- Ye, Y., Bryan, B. A., Connor, J. D., Chen, L., Qin, Z., & He, M. (2018). Changes in land-use and ecosystem services in the Guangzhou-Foshan metropolitan area, China from 1990 to 2010: Implications for sustainability under rapid urbanization. *Ecological Indicators*, 93, 930–941.
- Zhai, Y., Yao, Y., Guan, Q., Liang, X., Li, X., Pan, Y., ... Zhou, J. (2020). Simulating urban land use change by integrating a convolutional neural network with vector-based cellular automata. *International Journal of Geographical Information Science*, 34(7), 1475–1499.
- Zhang, Y., Middel, A., & Turner, B. (2019). Evaluating the effect of 3D urban form on neighborhood land surface temperature using Google street view and geographically weighted regression. *Landscape Ecology*, 34(3), 681–697.
- Zhou, L., Dang, X., Sun, Q., & Wang, S. (2020). Multi-scenario simulation of urban land change in Shanghai by random forest and CA-Markov model. *Sustainable Cities and Society*, 55, 102045.