Spatial planning-constrained modeling of urban growth in the Yangtze River Delta considering the element flows

Pengshuo Li\textsuperscript{a,b}, Yongjiu Feng\textsuperscript{a,b}, Xiaohua Tong\textsuperscript{a,b}, Rong Wang\textsuperscript{a,b}, Shuting Zhai\textsuperscript{a,b}, Yan Tang\textsuperscript{a,b} and Weijia Liu\textsuperscript{a,b}

\textsuperscript{a}College of Surveying and Geo-Informatics, Tongji University, Shanghai, China; \textsuperscript{b}The Shanghai Key Laboratory of Space Mapping and Remote Sensing for Planetary Exploration, Tongji University, Shanghai, China

\textbf{ABSTRACT}

Urban growth models that consider spatial control and factor flows are important for the integrated management of urban agglomerations, and to date, no simulation method has been available to consider both thoroughly. We proposed a new cellular automaton (CA\textsubscript{GBDT}) model based on gradient boosting decision trees (GBDT) to simulate urban growth in the Yangtze River Delta (YRD) area of China. In this model, GBDT was used to retrieve land transition rules by integrating multiple weak learners and a feature expansion approach was employed to remove high correlations among urban growth drivers and then expand the factor features. We applied the final urban pattern as the dependent variable and selected nine drivers as the independent variables of urban growth in YRD. The simulation results show that the overall accuracies exceeded 89\% and the figure-of-merits (FOMs) exceeded 27\%, about 10\% higher than other similar simulations in YRD, indicating the strong ability of CA\textsubscript{GBDT} to simulate urban growth in YRD. Although the inclusion of inter-city material and information element flows in the model improves the accuracy by only 1\%, it reveals the different development patterns in Hangzhou Bay in the south of YRD and the Taihu Lake basin in the center of YRD. By considering urban scenarios under different strategies of spatial control, it shows that the simulated FOMs declined by 5\% with the stronger enforcement of spatial regulations, reflecting that the actual urban growth in fully non-developable areas and 20\% partly developable areas of YRD has already violated the regulations. The results can help urban planners and local authorities to develop solutions for the development of a high-quality YRD urban agglomeration. The proposed model is applicable to diagnose urban land-use change elsewhere, especially in rapidly developing cities that need balancing urban growth and ecological protection.

\textbf{1. Introduction}

The World Urbanization Prospects report shows that today 55\% of the world’s population lives in cities, and the proportion is expected to increase to 68\% by 2050 (UN-DESA 2018). An urban agglomeration, as a group of cities at many different levels, has been a major trend in urban development in recent decades and the future. Typically, an urban agglomeration is formed gradually during long-term socioeconomic and urban development, and is a highly urbanized spatial form with a high degree of integration and competition among member cities (Fang and Yu 2017). In addition to these characteristics, the formation of urban agglomerations in China is also attributed to administrative forces such as regional socioeconomic planning and territorial spatial planning (Ye et al. 2019). The cities link each other in a wide range of elements including economy, transportation, environment, science, and society (Storper and Scott 2016; Fang and Yu 2017; Zhen et al. 2017; Li and Phelps 2018, 2019) and have high densities of material, information, and population flows with each other, which also facilitate the rapid land-use change and urban development in the urban agglomeration.

The modeling and prediction of urban growth using earth observation and geographical information science (Im 2020) are of great significance to the sustainable development of urban agglomerations (Yu et al. 2021). Among many spatial methods, a wide range of cellular automata (CA) models have been applied to modeling land-use change and urban growth in many individual cities and urban agglomerations (Arsanjani et al. 2013; Li et al. 2017; Gemitzi 2021; Rahnama 2021; Saganeiti et al. 2021; Wu et al. 2022). Compared to an individual city, the member
cities in an urban agglomeration are more complex because they have different driving features of urban growth and strong information and migration flows among the cities (Lv et al. 2021). Spatial planning is an important influencing factor of urban growth, especially for cities in the Yangtze River Delta (YRD), the largest urban agglomeration in China. This has received unprecedented attention in the context of the “Integrated Development of YRD” policy (Liang et al. 2018; Duan et al. 2019). With the implementation of China’s national territory spatial planning, the spatially explicit simulation of urban growth not only promotes the application of CA simulation methods to large-scale areas but also possesses important values for the YRD’s integrated development.

The development of urban agglomerations is jointly affected by natural conditions, human activities and policies. The past development is the basis to select features of urban expansion, where features that well fit urban dynamics are often used as driving factors (Zhou et al. 2020). Based on past urban growth and the related features, modelers have proposed different methods to derive transition rules for better CA models (Omrani, Tavani, and Pijanowski 2017; Feng and Tong 2018; Saganeiti et al. 2021; Nuqraha et al. 2022). The representative approaches include spatial statistical methods, rough set theory, artificial intelligence, heuristics, machine learning, and system dynamics (Liu et al. 2014). For example, the representatives of heuristics applied in CA models are the gray wolf algorithm, bee colony algorithm and ant colony algorithm (Feng and Tong 2020; Yan et al. 2021), while the representatives of machine learning algorithms are convolutional neural networks, random forests, and support vector machines (Feng and Tong 2019; Zhou et al. 2020; Wu et al. 2022). The successful application of intelligent algorithms has greatly boosted CA models, and most of the above methods have been proved useful in modeling individual cities (Aburas et al. 2016). A few others such as genetic algorithm, generalized simulated annealing, and particle swarm optimization have also been successfully used in the YRD urban agglomeration (Feng, Liu, and Tong 2018). However, a single learner may not correctly determine the influences of many factors when modeling the rapid urban growth, leading to uncertainties in the simulation results. To well simulate and predict scenarios of mega-urban agglomeration like YRD, multiple learners are needed to accurately derive the development rules because such areas are complex in urban dynamics and policies.

It is noted that representative driving factors play an essential role in urban growth and spatial association (Basse, Charif, and Bodis 2016), thus factor selection is crucial in CA-based modeling. Biophysical and socio-economic conditions that substantially affect the development of an individual city have also been considered influential in urban agglomerations. For modeling development in urban agglomerations, interactions between two cities need to be carefully considered as critical external drivers (Li and Phelps 2018). The interactions can be reflected by multidimensional urban element flows such as internet information search and migration among member cities, and these should help improve the modeling ability when studying urban agglomerations (Ou et al. 2019; Xia et al. 2019). In the Pearl River Delta (PRD) area, modelers considered urban flows in their CA model (Lin and Li 2015); in the Jing-Jin-Tang urban agglomeration, modelers considered the effects of landscape dynamics in their CA model (He et al. 2013); in the same Jing-Jin-Tang area, modelers considered the spatial intensity in their urban growth modeling (Lv et al. 2021). In the urban agglomeration around Wuhan, a gravitational field-based CA model was proposed to simulate urban growth (Xia et al. 2019). Despite the progress, urban element flows are important components that can be revealed by increasingly abundant spatial big data, and can be used to address the interactions among internal cities in urban agglomerations.

YRD is the leading urban agglomeration in China, thus its high-quality development and land-use optimization are of significance to its integrated development as issued by the Chinese central government. Indeed, uncontrolled urban growth could likely lead to various social and environmental problems such as encroachment on prime agricultural land and non-compact urban patterns (Teitz 2012; Moghadam et al. 2019). In China, the issued spatial planning usually includes three types: fully developable area (FDA), partly developable area (PDA) and fully non-developable area (FNA). For individual cities, Tong and Feng (2019) proposed a new method that can simulate and predict future urban scenarios with a case study in Ningbo of YRD. Specifically, this earlier work examined Ningbo’s future urban patterns in response to the implementation stringency of spatial
planning. However, when modeling urban growth of agglomerations such as YRD, it should be more complicated because of the complex interactions and element flows, simultaneously considering spatial planning. To date, this has not well been addressed by the literature on spatial modeling in urban agglomerations. Thus, to model urban patterns in urban agglomerations such as YRD, several questions need to be addressed: 1) how to integrate multiple machine learners to construct suitable CA models considering essential features (driving factors), 2) how to derive urban element flows that can improve the credibility of simulation results, and 3) how to integrate spatial planning when testing the implementation stringency of spatial planning on urban growth.

The objective of this work is to answer the three key questions mentioned above, with a case study of the YRD urban growth from 2008 to 2018. The answers to these questions are the novelty of this work, namely the use of an artificial neural network approach based on feature expansion to improve the quality of the drivers, where the factors include elemental flows such as information and migration. The novelty also lies in the construction of a new CA$_{\text{GBDT}}$ model based on the gradient boosting decision tree algorithm (GBDT). The government-issued spatial planning was also considered to examine the illegal land-use and the integrated degree of the YRD urban patterns. The method and practice should be useful to the goal of the “Integrated Development of YRD,” and the sustainable development of urban agglomeration elsewhere.

2. Study area and datasets

2.1 The YRD urban agglomeration

The YRD urban agglomeration, located on the southeast coast of China (Figure 1a), is China’s leading urban group with a large economic volume, high urbanization level and high population density. This region covers Shanghai, Jiangsu, Zhejiang and Anhui provinces, including 26 cities (Figure 1b) with a total area of 211,700 km$^2$, accounting for about 2.2% of China’s area. The density of towns in the YRD is more than 80 per 10,000 km$^2$, exceeding the national level by 4 times (SC-PRC 2019). During 2008–2018, the urbanized areas increased significantly by 2.26 times (Figure 1c), and its GDP increased from 7.6 trillion RMB to 22.1 trillion RMB (Figure 1d). As accelerated by the “Integrated Development of YRD” policy, this region has formed a networked and convergent urban agglomeration pattern, with Shanghai as the leading

![Figure 1. The study area of the YRD urban agglomeration. (a) The 26 cities of YRD, (b) the GDP growth during 2008–2018, and (c) the population growth during 2008–2018.](image-url)
city and Nanjing, Hangzhou, Hefei, Ningbo and Su-Xi-Chang as the node cities. Under the frame of “Integrated Development of YRD,” the YRD’s spatial planning is the most important influencing factor that may promote high-quality sustainable development and lead to different urban scenarios. We therefore selected YRD as the study area and aimed to address its issues considering spatial planning.

YRD has constructed high-level transport networks consisting of automobiles, trains, ships and planes in the last 10 years. The highways were about 4,194 km, and the railways reached 9,996 km including 3,357 km of high-speed railways and 6,639 km of intercity railways. There are 24 cities within a 2-hour rail distance from Shanghai and 13 airports in this region, suggesting that every two cities have an airport. As the leading city in YRD, Shanghai is the Asia-Pacific region’s major communication hub, and near Shanghai, there are several national-level mode cities with good communication facilities (SC-PRC 2019). The high-level transport networks facilitate the migration flow, and the well-developed internet networks facilitate the information flow, which are both the critical drivers of the urban growth of YRD. In recent years, the highly integrated transport and internet networks have greatly improved the socio-economic situation and the integrated development of YRD, finally driving the urban land-use change.

2.2 Raw datasets and input factors

The YRD urban growth is jointly influenced by various drivers including socio-economic conditions, material & information flows, and physical situations. The driver selection was based on the mechanism of urban growth, the interaction between individual cities within the urban agglomeration, the pre-analysis of the development of the YRD area, and the YRD spatial planning. We performed a preliminary screening of the extracted drivers and removed the factors not statistically significant. Based on the datasets in Table 1, we defined the final urban pattern as the dependent variable and 9 drivers as the independent variables of urban growth in YRD (Table 2).

The Landsat images provided by Geospatial Data Cloud Platform (www.gscloud.cn) were applied to produce the land-use maps, which served as the independent variable. The images acquired on 11 June 2008 and 17 June 2018 were classified using the Mahalanobis distance method (Table 1), where the overall accuracies for 2008 and 2018 both exceed 98%. The YRD's spatial planning published in 2016 was provided by the State Council of the People’s Republic of China.

To extract urban growth drivers, we collected spatial and nonspatial datasets including transport networks, information networks, terrain maps, population density, points of interest (POI), and population migration. The transport networks include highways, arterial ways, railways and subways that were extracted using OpenStreetMap (www.openstreetmap.org). The information networks and flows are represented by the frequency of internet searches from residents in YRD. We then collected the information flow between June 2018 and December 2018 recorded by the Baidu Index Website (index.baidu.com). The terrain map is the DEM derived using the 30 m GDEM data. The gridded maps of population density were collected from Worldpop (www.worldpop.org.uk). The POIs mainly include the city-center vector data from OpenStreetMap. The population migration data between January 2018 and December 2018 was crawled from the Tencent Location Big Data Platform (heat.qq.com/bigdata).

Transportation facilities have long been acknowledged as the substantial drivers of urban growth, which means a high probability of urban land-use transition if a land parcel is close to the transport networks. We produced 4 transport drivers including the proximity to highways, arterial ways, railways, and subways. Such proximity is negatively related to urban growth, which means smaller proximity could lead to a higher probability of a land parcel transforming into an urban state. The proximity to city-centers (V-cc) substantially affects land-use change in China because the administration centers usually promote urban growth. The population density (V-pop) is considered the most direct driver of urban growth. For our study area, the influence of migration (V-mf) and information (V-if) flows among the cities is important because they promote urban growth and the socio-economic aspects in YRD. We also considered physical conditions such as DEM and elevation to represent the impact of terrain on urban growth. All variables (Figure 3) were acquired in the boundary year (2018) of the consequent modeling, and have been normalized to eliminate the effects caused by inconsistencies in the dimensionality.
3. The new $CA_{\text{GBDT}}$ model

3.1 The $CA_{\text{GBDT}}$ modeling procedure and software

Figure 4 shows the proposed $CA_{\text{GBDT}}$ model of urban growth simulation that integrates CA, artificial neural network (ANN) and GBDT. The procedure includes three steps: 1) feature expansion-based sampling, 2) GBDT-based transition probability computation, and 3) the $CA_{\text{GBDT}}$ modeling. We first used Landsat imagery to generate an urban land-use map at the initial and final time points, and extracted urban growth driving factors from the social-economic, facility and elevation datasets. In the first step, we applied systematic sampling to select training samples from the land-use and driving factor maps, then used principal components analysis (PCA) to merge the high-correlation features. The original features were applied in the ANN training to expand high-dimension features, resulting in the final combined samples. In the second step, we applied GBDT to train the urban growth rules based on the combined features. In the third step, we defined the CA components, land-use planning constraints and thresholds to calibrate the $CA_{\text{GBDT}}$ model in YRD. In the urban growth modeling, a planning implementation parameter (PIP) was defined to reflect the spatial planning constraints in the model.

The $CA_{\text{GBDT}}$ model was trained and implemented using CityAIModel software. The CityAIModel software provides transition probability map modeling based on CA, ANN and GBDT, and supports users to independently select various model parameters for modeling urban land-use change and urban growth. The readers are referred to Feng and Tong (2020) for details.
Figure 2. The YRD’s important transport facilities that include highways, railways and airports.

Figure 3. The urban growth drivers that reflect the YRD’s physical and socio-economic conditions (source: Table 1).

Figure 4. The workflow of the CA_{GBDT} model for simulating the YRD urban growth.
3.2 The proposed CAGBDT model

3.2.1 The fundamental CA modeling
In CA models, urban growth rules denote cell state transition possibility from non-urban to urban. The rules define the state \((STAT^{t+1})\) at the next time of the cell \((u,v)\) based on the integrated growth potential \((P_{\text{tran}})\). The integrated potential is defined by the cell’s current state \((STAT^t)\), the potential based on drivers \(P_{\text{GBDT}}(f_1 \ldots f_n)\), the neighborhood influences \((NI)\), the constraints \((CON)\), the time increase parameter \((TIP)\), and the local adjustment parameters \((LAP)\). The urban growth rules were given by (Feng and Tong 2020):

\[
\begin{align*}
P_{\text{tran}} &= \left( P_{\text{GBDT}}(f_1 \ldots f_n) \times \frac{1 + TIP \times 1 + NI \times L}{1 + TIP} \times CON \right) \\
STAT^{t+1} &= \text{Urban}, \text{ if } P_{\text{tran}} > P_{\text{thd}} \\
STAT^{t+1} &= \text{Nonurban}, \text{ if } P_{\text{tran}} \leq P_{\text{thd}}
\end{align*}
\]

where \(P_{\text{thd}}\) denotes the threshold of the transition potential. When \(P_{\text{tran}}\) is greater than or equal to the threshold, a nonurban cell can be converted to an urbanized cell; otherwise, the cell remains in its original state. The threshold is adjusted according to the difference in the urban areas between the predicted and simulated patterns at the end year. If the difference is smaller than 1%, then the threshold is considered optimal. \(P_{\text{GBDT}}(f_1 \ldots f_n)\) is the GBDT-based growth potential of cell \((u,v)\), where \((f_1 \ldots f_n)\) represent the urban growth drivers.

For a non-urbanized cell, more urban cells (neighbor cells) nearby in an \(m \times m\) square window would indicate a higher urban growth potential (Roodpohtsi, Hewitt, and Bryan 2020). The neighborhood influences \((NI)\) were given by (Ma et al. 2020):

\[
NI = \frac{\text{Count}(u,v,m \times m, STAT^t = \text{Urban})}{m \times m - 1}
\]

where \(\text{Count})\) calculates the number of urban cells in the square window.

\(CON\) is a parameter that prevents water bodies and natural reserves from being developed, where a value of 0 is assigned to denote the nonurban-to-urban constraint while 1 denotes the nonurban-to-urban availability. \(TIP\) is used to offset the transition potential and \(LAP\) is used to offset the increasing effect of neighborhood influences.

3.2.2 Element flows in the YRD urban agglomeration
Except for the seven drivers of proximity, population density and terrain (c.f. Table 1), the information and migration flows are typically important in the simulation of urban agglomeration because of the strong interaction among the cities. These two flow drivers (TotalInfor and TotalMigra) were produced using the Baidu search data and the Tencent immigration and emigration data (Figure 5).

The TotalInfor driver reflects the situation of communication facilities and the popularity of mobile phones among residents in YRD, as well as the willingness (as denoted by the Baidu search data) of traveling and migrating among the cities. The Baidu search data is based on massive internet user behaviors in the Baidu searching engine, denoting the concern intensity of residents in city \(A\) when searching another city \(B\). As such, the search data denotes information input \((\text{InforIn}, \text{see Figure 5a})\) for city \(A\) but information output \((\text{InforOut}, \text{see Figure 5b})\) for city \(B\). The information input and output of the 26 cities can be given by:

\[
\begin{align*}
\text{InforIn} &= \begin{bmatrix} \text{inforin}_1^1 & \cdots & \text{inforin}_{26}^1 \\ \vdots & \ddots & \vdots \\ \text{inforin}_{26}^1 & \cdots & \text{inforin}_{26}^{26} \end{bmatrix} \\
\text{InforOut} &= \begin{bmatrix} \text{inforout}_1^1 & \cdots & \text{inforout}_{26}^1 \\ \vdots & \ddots & \vdots \\ \text{inforout}_{26}^1 & \cdots & \text{inforout}_{26}^{26} \end{bmatrix}
\end{align*}
\]

where \(\text{inforin}_j^i\) is the information input for city \(A_i\) from city \(A_j\), reflecting the willingness of residents in city \(A_i\) traveling to city \(A_j\); \(\text{inforout}_j^i\) is the information output from city \(A_i\) to city \(A_j\), reflecting the attractiveness of city \(A_i\) to the residents in city \(A_j\). The combination of the information input and output produced the total information flow \((\text{TotalInfor})\) for all 26 cities:

\[
\begin{align*}
\text{TotalInfor} &= \begin{bmatrix} \text{City 1} : \sum_{j=1}^{26} (\text{inforin}_j^1 + \text{inforout}_j^1) \\ \vdots \\ \text{City } i : \sum_{j=1}^{26} (\text{inforin}_j^i + \text{inforout}_j^i) \\ \vdots \\ \text{City 26} : \sum_{j=1}^{26} (\text{inforin}_{26}^j + \text{inforout}_{26}^j) \end{bmatrix}
\end{align*}
\]

Another flow driver \((\text{TotalMigra})\) was generated considering the migration data crawled on the Tencent Location Big Data Platform. The time interval of the resident migration was from Jan 1 to 31
December 2018, where there are four migration indexes, namely immigration by train ($IBT$, see Figure 5c), emigration by train ($EBT$, see Figure 5d), immigration by bus ($IBB$, see Figure 5e) and emigration by bus ($EBB$, see Figure 5f). These can be given by:

\[
\begin{align*}
IBT &= \begin{bmatrix}
ibt_1^1 & \cdots & \ibt_{26}^1 \\
\vdots & \ddots & \vdots \\
\ibt_{26}^1 & \cdots & \ibt_{26}^{26}
\end{bmatrix} \\
IBB &= \begin{bmatrix}
ibb_1^1 & \cdots & \ibb_{26}^1 \\
\vdots & \ddots & \vdots \\
\ibb_{26}^1 & \cdots & \ibb_{26}^{26}
\end{bmatrix} \\
\text{EBT} &= \begin{bmatrix}
\text{ebt}_1^1 & \cdots & \text{ebt}_{26}^1 \\
\vdots & \ddots & \vdots \\
\text{ebt}_{26}^1 & \cdots & \text{ebt}_{26}^{26}
\end{bmatrix} \\
\text{EBB} &= \begin{bmatrix}
\text{ebb}_1^1 & \cdots & \text{ebb}_{26}^1 \\
\vdots & \ddots & \vdots \\
\text{ebb}_{26}^1 & \cdots & \text{ebb}_{26}^{26}
\end{bmatrix}
\end{align*}
\]

(5)

where $\ibt_j^i$ is the per-day immigrants by train from city $A_j$ to city $A_i$, $\ebt_j^i$ is the per-day emigrants by train from city $A_j$ to city $A_i$, $\ibb_j^i$ is the immigrants by bus from city $A_j$ to city $A_i$, and $\ebb_j^i$ is the per-day emigrants by bus from city $A_j$ to city $A_i$.

To produce a cell-level map of the migration flow, we integrated these four migration indexes considering the distance of each cell $(u, v)$ to each city. As such, each cell $(u, v)$ has four migration indexes:

\[
\begin{align*}
\text{IBTcell}(u, v) &= \sum_{j=1}^{26} \frac{\ibt_j^i}{\text{dis}(u, v, \text{citycenter}_j)} \\
\text{EBTcell}(u, v) &= \sum_{j=1}^{26} \frac{\ebt_j^i}{\text{dis}(u, v, \text{citycenter}_j)} \\
\text{IBBcell}(u, v) &= \sum_{j=1}^{26} \frac{\ibb_j^i}{\text{dis}(u, v, \text{citycenter}_j)} \\
\text{EBBcell}(u, v) &= \sum_{j=1}^{26} \frac{\ebb_j^i}{\text{dis}(u, v, \text{citycenter}_j)}
\end{align*}
\]

(6)

where $\text{dis}(u, v, \text{citycenter}_j)$ denotes the Euclidean distance between the cell $(u, v)$ and the center of city $A_j$.

Therefore, the integrated migration index of cell $(u, v)$ can be given by:

\[
\text{TotalMigra}(u, v) = \text{IBTcell}(u, v) + \text{EBTcell}(u, v) + \text{IBBcell}(u, v) + \text{EBBcell}(u, v)
\]

(7)

Based on the methods above, we produced the maps of $\text{TotalInfor}$ (c.f. Figure 3h) and $\text{TotalMigra}$ (c.f. Figure 3j) as important inputs for the $\text{CA}_{\text{GBDT}}$ model. They were combined in deriving the urban growth potential for the YRD urban simulation.
3.2.3 Feature expansion-based sampling
To train CA models, samples are commonly selected from maps of the drivers and land-use using systematic sampling. Samples should sufficiently reflect most of the important drivers influencing urban growth; however, the selected drivers may likely be multicollinear while a few other important influencing factors may not be included in the modeling. We therefore applied a feature expansion approach to reduce the information redundancy in the initial samples, and simultaneously generated a few additional features from the initial samples. To reduce the information redundancy, we utilized PCA on the initial samples to extract linearly uncorrelated features (n1). We then applied ANN to these features to produce new features (n2) of high dimension, as extended in the penultimate layer in ANN. Although the number of the extended features and the number of the network layers can be adjusted, the penultimate layer is the cascade of the high-dimension features. As such, the final features include the initial n1 features and the extended n2 features, which were used in the GBDT-based transition potential training in Figure 4.

3.2.4 The GBDT-based transition probability
The urban growth rules are the center of CA models, which are highly related to the selected samples and the training method itself. The feature expansion-based sampling increased the sample quantity in the model training of the urban growth simulation. More features of low correlation can improve the urban simulation accuracy, but they also lead to difficulty in data fitting especially when studying urban agglomeration that covers many cities. While the decision trees have been successfully applied in CA modeling, a single weak learner is not enough when processing such vast samples. To overcome the shortcoming, we used GBDT to integrate several weak learners. By minimizing the information loss of weak learners, GBDT can output a strong learner for the CA rules training. We initialized a weak learner $P_0(X_i)$ as (Han, Zhang, and Shen 2018):

$$P_0(X_i) = \operatorname{Argmin}_{\text{pro}} \sum_{i=1}^{q} L(Y_i, \text{pro})$$  \hspace{1cm} (8)

where $L(Y_i, \text{pro})$ is the loss function measuring the distance between the cell-state $Y_i$ and the probability $\text{pro}$, which indicates the initial maximum likelihood probability of nonurban-to-urban change; $X_i$ is the $i^{th}$ sample vector that includes all initial and extended features; $q$ is the number of state change labels; $\operatorname{Argmin}$ calculates $\text{pro}$ as a constant by minimizing the sum of $L(Y_i, \text{pro})$.

We conducted $K$ iterations on the initial weak learner, and each iteration generated a new weak learner. In the $k^{th}$ iteration, we calculated the negative gradient $R_{kq}$ of $q$ samples as (Friedman 2001):

$$R_{kq} = [r_{k1}, r_{k2}, r_{k3}, \ldots, r_{kj} \ldots r_{kn}]^T$$

$$r_{ki} = -\left(\frac{\partial L(Y_i, P(X_i))}{\partial P(X_i)}\right)_{P(X_i) = P_{k-1}(X_i)}$$  \hspace{1cm} (9)

where $r_{ki}$ represents the negative gradient of $L(Y_i, P(X_i))$ on the changing probability $P(X_i)$. We then generated a training dataset $TD = \{(1, r_{k1}), \ldots, (i, r_{ki}) \ldots (n, r_{kn})\}$. In the $k^{th}$ iteration, the strong learner $P_{k}(X_i)$ can be given by (Friedman 2001):

$$P_k(X_i) = P_{k-1}(X_i) + \sum_{j=1}^{J} C_{kj} \times i, \text{where } X_i \in LNA_{kj}$$

$$C_{kj} = \operatorname{Argmin}_{\text{pro}} \sum_{X_i \in LNA_{kj}} L(Y_i, P_{k-1}(X_i) + \text{pro})$$  \hspace{1cm} (10)

where $P_k(X_i)$ is the urban growth potential in the $k^{th}$ iteration, $T_k$ is a CART regression tree trained using a training dataset $TD$ and $C_{kj}$ is the optimal output value in $LNA_{kj}$ that represents the leaf node areas of $T_k$. After $k$ iteration, the final strong ($k^{th}$) learner $P_k(X)$ subject to learn rate ($lr$) can be given by (Ke et al. 2017):

$$P_k(X_i) = P_0(X_i) + \sum_{k=1}^{K} \sum_{j=1}^{J} C_{kj} \times lr, \text{where } X_i \in LNA_{kj}$$  \hspace{1cm} (11)

3.2.5 Spatial planning constraints
Urban growth modeling considering spatial planning is an essential approach to promoting the integrated development of YRD. The YRD Urban Agglomeration Development Planning (2015–2035) issued by the Chinese central government indicates the spatial control for this area (Figure 6a). In the spatial control, FDA indicates the fully developable areas considering the relative integrity of urban infrastructure, PDA indicates the partly developable areas considering the balance between the production & living facilities and the environmental protection, FNA indicates the fully non-developable areas by focusing on the environmental and ecological protection. The classification of Landsat
images shows that the urbanization level from 2008 to 2013 was 7.1% to 11.5%, where the urbanized area highly concentrated in FDA (Figure 6b); from 2013 to 2018, the urbanized areas increased significantly by 4.5%, where new urban lands were found in both PDA and FNA (Figure 6c). This shows that the actual urban growth did not strictly comply with spatial planning. The CA-based prediction of future scenarios considering the spatial control should benefit the integrated development of YRD.

Here, we use a PIP to simulate the implementation of the spatial control for PDA (Tong and Feng 2019). The PIP ranges from 0 to 1, where 0 means that development is strictly prohibited in PDA and 1 means that development is fully allowed in PDA. Specifically, for example, a PIP of 0.7 means 70% of PDA can be developed. With the constraint of spatial planning, a land cell can be transformed into an urban state if it satisfies the following three rules:

\[
\begin{align*}
1. \quad P_{\text{tran}} &> P_{\text{thd}} \\
2. \quad R &\in [0, 1] \\
3. \quad \text{if } R \geq \text{PIP}, \text{STAT}^{t+1} &= \text{Urban}
\end{align*}
\]

(12)

3.3 The model assessment method

To assess the simulation accuracy, all land parcels were divided into four categories including true-positive (TP), false-positive (FP), true-negative (TN) and false-negative (FN). TP refers to urban growth parcels being correctly simulated as urban growth, FP refers to the nonurban persistent parcels being incorrectly simulated as urban growth, TN refers to urban and nonurban persistent parcels respectively being correctly simulated as urban and nonurban persistence, and FN refers to urban growth parcels being incorrectly simulated as nonurban persistence. Using these four metrics, we generated two categories of evaluation metrics including end-state and change-state (Table 3).

4. Results

4.1 The transition rules and model calibration

In the model training, we used the land-use maps of YRD in 2008 and 2018 as the initial and boundary datasets, respectively, and constructed one CA\textsubscript{GBDT} model excluding the element-flow factors and one CA\textsubscript{GBDT} model including the element-flow factors. The two models were constructed based on a sampling rate of 10%, 6 PCA-based features, 2 ANN’s hidden layers (a total of 8 features), and 200 iterations. Figure 7 shows the two maps of urban growth potential with low values located in the high-altitude areas, which are also shaped by the traffic road networks. Although the overall patterns of the two potential maps are highly similar, the regions with stronger information and migration flows yield significantly higher growth potential. This may lead to stronger urban expansion of larger-sized cities that yield stronger urban element exchanges.

![Figure 6](image_url)

Figure 6. The YRD’s spatial planning 2015–2030 (a; source: YRD’s spatial planning in Table 1), urban growth during 2008–2013 (b; source: Landsat-5 TM in Table 1), and urban growth during 2013–2018 (c; Landsat-8 OLI in Table 1).
Table 3. The evaluation metrics used to assess the CA\textsubscript{GBDT} simulation results.

| Category      | Metric       | Description                                                                 |
|---------------|--------------|-----------------------------------------------------------------------------|
| End-state     | Overall      | The percentage of all correctly simulated land parcels (TP and TN)          |
|               | accuracy     | to the total land parcels (TP, FP, TN and FN).                             |
| Precision     |              | The percentage of correctly simulated urban-development land parcels (TP)  |
|               |              | in all the land parcels simulated as a nonurban-to-urban transition (TP   |
|               |              | and FP).                                                                 |
| Recall        |              | The percentage of correctly simulated urban-development land parcels (TP)  |
|               |              | to the observed urban-development land parcels (TP and FN).                 |
| Kappa         |              | The consistency between the simulated and observed results excluding the    |
|               |              | consistency by chance.                                                     |
|                | F1-score     | The harmonic mean of precision and recall.                                 |
| Hit           |              | The percentage of correctly simulated urban-development land parcels (TP)  |
|               |              | to all land parcels (TP, FP, TN and FN).                                  |
| Miss          |              | The percentage of the missed urban growth parcels by the model (FN) to all |
|               |              | land parcels (TP, FP, TN and FN).                                          |
| False         |              | The percentage of the falsely simulated land parcels (FP and FN) to all    |
|               |              | land parcels (TP, FP, TN and FN).                                          |
|                | Figure-of-   | The percentage of correctly simulated urban-development land parcels (TP)  |
| merit         |              | to all observed and simulated urban-development land parcels.              |

4.2 The urban simulations without spatial planning

With the CityAIModel software we developed, we applied the GBDT approach to construct the CA\textsubscript{GBDT} model with the land-use map in 2008 as the initial state, the two maps of urban growth potentials as the input probability layers, and the urban area in 2018 as the total demand. Figure 8 shows the 2018 actual urban map (Figure 8a), the 2018 simulated urban map using the transition potential excluding the element flows (Figure 8b), and the 2018 simulated urban map including the element flows (Figure 8c). The two simulated results are both highly similar to the actual urban pattern, where the simulations seem to have more clustered-urban-patches while the actual has more dispersed-urban-patches. The urban patches are mainly located in Shanghai, Hangzhou Bay, the Taihu Lake basin, and the areas along the Yangtze River. Figure 8d shows that the accuracies and Kappa coefficients are over 0.9, the precisions are about 0.7, the recalls are over 0.72, and the FOMs are over 0.32. These indicate that the CA\textsubscript{GBDT} model has a strong ability to reproduce the urban pattern of YRD when excluding and including the element-flow factors. Including the information and migration flow factors into the CA\textsubscript{GBDT} model shows only slight improvements in accuracy, precision, Kappa and FOM (Figure 8d); however, the two urban simulations show differences in local areas.

The urban element flows directly reflect the gravitational attraction between any two cities and substantially affect urban expansion. Figure 9 shows that, after including the two element-flow factors, the statistics show that the simulated urban areas of Shanghai and Hang-Shao are larger (increased by 5.0% and 5.7%, respectively) while the urban areas of Su-Xi-Chang and Ningbo are smaller (decreased by 5.3% and 5.6%, respectively). Figure 9 also shows the quantitative comparison of four selected regions before and after

![Figure 7](image-url)  
Figure 7. The urban growth potential maps based on driving factors excluding the element flows (a) and including the element flows (b).
including the element flows, and the urban patch distribution in local areas in Shanghai (area-I), Su-Xi-Chang (area-II), Hang-Shao (area-III), and Ningbo (area-IV). The four local areas in Figure 9 show that, although the two types of simulations are more clustered than the actual, the simulated urban patterns are more...
similar to the actual urban patterns after including the urban element flows. This indicates the positive contribution of the information and migration flows to the urban growth simulation in YRD.

4.3 The urban simulations with spatial planning

To evaluate the effects of spatial planning on the urban simulations, we applied PIP to the CA\(_{GBDT}\) model to inform the specific urban growth in PDAs. We tested four PIPs (e.g., 0.0, 0.3, 0.5, and 0.7) in the modeling and produced eight urban patterns in 2018 (Figure 10), with an enlarged area to show the details for each simulation. There are many built-up areas in the strictly constrained area FNA, where could occur many new urban patches (2008–2018) when using CA models without spatial control. After implementing spatial control and PIP to the simulation, those new urban patches (2008–2018) that would be in FNA could appear in FDA and/or PDA. Compared to the observed urban area in 2018, the urban area in PDA decreased significantly from ~95% to ~75% as PIP decreased from 0.7 to 0.3 (Figure 10). This indicates that loose spatial planning was implemented in YRD during 2008–2018 and the actual urban growth in FNA and 20% PDA of YRD has already violated the regulations, but the implementation stringency did not meet the YRD Urban Agglomeration Development Planning (2015–2035).

For both excluding and including the element-flow factors, Table 4 shows that the overall accuracies of all simulation results exceed 89% for all PIPs, indicating high similarity between the simulated and actual urban patterns as well as the stability of the proposed CA\(_{GBDT}\) model. The precision, recall, Kappa, and F1-score metrics show a similar tendency as the overall accuracy, which is attributed to the implementation of spatial planning. The change-state metrics show that Hit and FOM decrease while Miss and False

![Figure 10](image-url). The simulated 2018 urban patterns under different PIPs before and after including the element-flow factors.
increase with the decreasing PIP. The 5% decline of the FOM suggests that looser regulations could lead to higher accurate results, implying that urban growth in YRD has already violated spatial planning.

5. Discussion

5.1 The performance of the CA\textsubscript{GBDT} model

We developed a new urban growth simulation model (CA\textsubscript{GBDT}) by combining CA and GBDT, and performed it in the YRD urban agglomeration. Our CA\textsubscript{GBDT} model can reduce the correlation of huge samples while expanding features from low-dimensional features. In the model, we considered the urban element flows and spatial control to explore urban growth in YRD. The CA\textsubscript{GBDT} model yielded FOMs of about 30%, which are about 10% higher than urban simulations considering the element flows in other urban agglomerations, where the socio-economic context and urban growth stage are different from YRD (Lin and Li 2015; Xia et al. 2019; Lv et al. 2021). In this study, although the inclusion of urban element flows improved FOMs by only ~1%, the model has a strong ability to simulate urban growth in large-scale and highly developed urban agglomerations and can better address urban dynamics in YRD. In addition, considering the spatial control in the CA\textsubscript{GBDT} model helped us understand the effects of territory spatial planning on YRD’s urban growth.

To construct the CA\textsubscript{GBDT} model, GBDT used multiple decision trees to calculate the minimum negative gradient for defining the near-optimal land-use transition rules (Ke et al. 2017; Rao et al. 2019; Ma et al. 2020). The method expanded the features of the selected samples to derive more accurate transition rules, with acceptable increases in the demand for computational time (Du et al. 2018). In the YRD area, Feng, Liu, and Tong (2018) applied three heuristic CA models to simulate urban growth, and the Hits were about 4.6%, Misses were about 6.1% and False Alarms were about 5.9%. Using the CA\textsubscript{GBDT} model, our simulation accuracy was improved to 5.1% for the Hits, 4.98% for the Misses, and 5.66% for the False Alarms. Compared with the experimental results in the Seoul area using GBDT (Jun 2021), our accuracy (90.1%) is lower by 1.6% but our Kappa (90.1%) is higher by 32.7%. Because Kappa can reflect the difference between the urban and non-urban areas within the agglomeration, our GBDT model can maintain a high simulation accuracy as proved by Kappa, which indicates the strong ability to simulate urban growth in large-scale and highly developed areas. In addition, our average recall is 4.4% higher and the average precision is 5.6% lower than the simulations in Seoul, indicating the possible over-simulation of urban growth by our CA\textsubscript{GBDT} model, which confirms the properties of GBDT (Opitz and Maclin 1999). In fact, the simulation accuracy is not only affected by the model, but also by other factors such as regional characteristics, urban growth rate, and urbanization rate (Han et al. 2009). Urban growth in linked areas between different cities within an urban agglomeration is more difficult to capture because these land parcels are more influenced by regional characteristics and stochastic conditions.

5.2 Effects of element flows on urban growth modeling

For studies in urban agglomerations, a few modelers have considered urban flows to simulate urban growth, where the FOMs in Wuhan were reported about 0.18 (Xia et al. 2019), the highest FOM in Jing-Jin-Ji was about 0.204 (Lv et al. 2021), and the FOMs in PRD were about 0.11 (Lin and Li 2015). These valuable
works suggest that the FOMs of urban growth simulations considering the urban flows in large-scale areas are mainly between 0.1 ~ 0.2 in cities of China. Our experiments considering the element flows increased the FOMs to about 0.27 in YRD, which suggests that our FOMs are at high levels and indicates the superiority of the proposed model even though the element flows included in our $CA_{GBDT}$ model have improved the OA by only 1%.

After considering the urban element flows, the urban areas increased in several cities (e.g. Shanghai, Hangzhou and Shaoxing) but decreased in a few other cities (e.g. Changzhou and Ningbo), which suggests that stronger connections to Shanghai are in a superior position to gather population and information to promote urban growth. This indicates that a larger-size city having stronger urban element flows with Shanghai expand faster because information and migration are external, important factors of urban growth. It is speculated that the urban growth of smaller cities having weaker element flows with other bigger cities could less be promoted by external factors. In terms of the urban areas, the information and migration flows in YRD exacerbate the differences in urban growth among different cities rather than making balanced development. With the implementation of the “Integrated Development of YRD” policy, the economically developed cities should focus on urban quality rather than extent expansion. The local governments are suggested to guide new residents to small and medium-sized cities to promote high-quality development.

### 5.3 Spatial planning impacts and policy implications

As the “Integrated Development of YRD” policy has become a national strategy, spatial planning should be strictly enforced in YRD. Tong and Feng (2019) initially included spatial planning into CA models to detect illegal land development and predict future illegal land development in Ningbo, an important coastal city in YRD. In the whole YRD, our modeling results indicate that when the spatial planning regulations became more stringent, the urbanized area in PDAs decreased significantly, and the simulation accuracy decreased accordingly. However, because illegal development mostly occurs in PDAs, it does not mean that the spatial planning-based simulation is not consistent with the actual urban growth. The decreasing simulation accuracy with stricter spatial control indicated that the actual urban growth somewhat has already violated the policy. It is very difficult to strictly adhere to land development planning in urban growth despite that the planning is beneficial to long-term regional development.

In the planning constraint model, when the spatial control policy is implemented, illegal urban development in the PDA is controlled and illegal urban development in the FDA is eliminated, hence the urban scenario becomes more compact under stricter spatial control. In PDA, the locations where new urban patches appear are related to the flexibility of policy implementation, with strict implementation leading to new urban patches occurring only in the allowed sub-regions, and loose implementation leading to new urban patches appearing in more different sub-regions. In China, while there is a strict land development approval process, there also exists a regulatory blindside; for some planning authorities, CA simulation models that consider planning policies can help identify possible future illegal urban development. The constrained urban development in the Yangtze River Delta also suggests that the “Integrated Development of the Yangtze River Delta” policy provides guidelines for optimal land development, while the implementation of urban planning serves to restrain illegal urban development. Before the spatial control policy, the real estate industry was an important vehicle for the government to stimulate the economy (Fung, Jeng, and Liu 2010; Han, Zhang, and Zhao 2021), and the rapid urban growth has brought many environmental problems to YRD from 2000 to 2015 (C. Yang et al. 2021 and K. Yang et al. 2021). Fortunately, the implementation of the “Integrated Development of YRD” policy can help urban planners and local governments to formulate more livable cities and high-quality development in this urban agglomeration (Li et al. 2022).

### 5.4 Limitations and future works

In addressing material and information flows inter-cities, we only collected data on population and information flows due to the limitation of big data availability. Since urban agglomeration is a complex mega-system (Fang et al. 2017; Cui et al. 2020), more socioeconomic drivers need to be considered in the modeling of the gravitational attraction among cities. In future studies, the drivers also need to include
flows of people, goods, capital and information that are related to agriculture, manufacturing and services to quantify inter-city gravitational forces and thus maximize the advantages of the \( \text{CA}_{\text{GBDT}} \) model.

6. Conclusion

We developed a new \( \text{CA}_{\text{GBDT}} \) model combining the feature expansion strategy and the gradient boosting decision tree to simulate dynamic urban growth in large-scale areas, considering the element flows among cities simultaneously. The feature expansion eliminated the high correlation among driving factors using PCA while expanding the factor features using ANN. The GBDT algorithm of transition rule retrieval integrated multiple weak learners. The application in YRD showed that the overall simulation accuracies exceeded 89% and the FOMs were higher than 27%, showing the strong ability of the \( \text{CA}_{\text{GBDT}} \) model to simulate urban growth in YRD. Apart from 7 common driving factors, we collected big data to discover the effects of urban element flows on dynamic urban growth. While the inclusion of urban element flows into the model improved the results by only 1%, it revealed the spatially varying development modes in Hangzhou Bay and the Taihu basin.

This study also examined the responses of urban scenarios to different strategies of spatial control. The results show that the simulation accuracy decreased with the enforcement of development planning getting stronger, reflecting that the actual urban growth in YRD has already violated the spatial regulations to a certain extent. Our results considering both the urban element flows and spatial planning constraints have displayed the spatially varying development of cities in the YRD urban agglomeration. The results are also helpful to urban planners and local authorities to formulate solutions for developing a high-quality YRD urban agglomeration. The proposed CA model is also applicable to diagnose urban land-use change elsewhere, especially in rapidly growing cities that need balancing urban growth and ecological protection. Future research will consider more socio-economic drivers, including population, goods, capital and information flows, to test the robustness of the proposed \( \text{CA}_{\text{GBDT}} \) model across more drivers and different cities.

Disclosure statement

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

Funding

Supported by the National Key R&D Program of China (2021YFB3900105-2) and the National Natural Science Foundation of China (42071371).

ORCID

Yongjiu Feng http://orcid.org/0000-0001-8772-7218

Data availability statement

https://figshare.com/s/8b260d3d4a9996a0e8ac

References

Aburas, M., Y. Ho, M. Ramli, and Z. 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 and Geoinformation 52: 380–389. doi:10.1016/j.ijag.2016.07.007.
Arsanjani, J., M. Helbich, W. Kainz, and A. 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.ijag.2011.12.014.
Basse, R. M., O. Charif, and K. Bodis. 2016. “Spatial and Temporal Dimensions of Land Use Change in Cross Border Region of Luxembourg. Development of a Hybrid Approach Integrating GIS, Cellular Automata and Decision Learning Tree Models.” Applied Geography 67: 94–108. doi:10.1016/j.apgeog.2015.12.001.
Cui, X., C. Fang, H. Liu, X. Liu, and Y. Li. 2020. “Dynamic Simulation of Urbanization and eco-environment Coupling: Current Knowledge and Future Prospects.” Journal of Geographical Sciences 30 (2): 333–352. doi:10.1007/s11442-020-1731-x.
Duan, D. Z., Y. Zhang, Y. Chen, and D. B. Du. 2019. “Regional Integration in the inter-city Technology Transfer System of the Yangtze River Delta, China.” Sustainability 11 (10): 19. doi:10.3390/su11102941.
Du, G., K. Shin, L. Yuan, and S. Managi. 2018. “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 32 (4): 757–782. doi:10.1080/13689998.2017.1410550.
Fang, C., and D. Yu. 2017. “Urban Agglomeration: An Evolving Concept of an Emerging Phenomenon.” Landscape and Urban Planning 162: 126–136. doi:10.1016/j.landurbplan.2017.02.014.
Fang, C., C. Zhou, C. Gu, L. Chen, and S. Li. 2017. “A Proposal for the Theoretical Analysis of the Interactive Coupled Effects between Urbanization and the eco-environment in...
mega-urban Agglomerations.” *Journal of Geographical Sciences* 27 (12): 1431–1449. doi:10.1007/s11442-017-1445-x.

Feng, Y., Y. Liu, and X. Tong. 2018. “Comparison of Metaheuristic Cellular Automata Models: A Case Study of Dynamic Land Use Simulation in the Yangtze River Delta.” *Computers Environment and Urban Systems* 70: 138–150. doi:10.1016/j.compenvurbsys.2018.03.003.

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. doi:10.1080/15481603.2018.1426262.

Feng, Y., and X. Tong. 2019. “Incorporation of Spatial heterogeneity-weighted Neighborhood into Cellular Automata for Dynamic Urban Growth Simulation.” *GiScience & Remote Sensing* 56 (7): 1024–1045. doi:10.1080/15481603.2019.1603187.

Feng, Y., and X. Tong. 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. doi:10.1080/13658816.2019.1648813.

Friedman, J. 2001. “Greedy Function Approximation: A Gradient Boosting Machine.” *Annals of Statistics* 29 (5): 1189–1232. doi:10.1214/aos/1013203451.

Fung, H.-G., J.-L. Jeng, and Q. W. Liu. 2010. “Development of China’s Real Estate Market.” *The Chinese Economy* 43 (1): 71–92. doi:10.2753/CES1097-1475430104.

Gemitz, A. 2021. “Predicting Land Cover Changes Using a CA Markov Model under Different Shared Socioeconomic Pathways in Greece.” *GiScience & Remote Sensing* 58 (3): 425–441. doi:10.1080/15481603.2021.1885235.

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.

Han, Q., X. Zhang, and W. Shen. 2018. “Lithology Identification Technology Based on Gradient Boosting Decision Tree (GBDT) Algorithm.” *Bull. Mineral. Petrol. Geochim* 37: 1173–1180.

Han, Y., H. Zhang, and Y. Zhao. 2021. “Structural Evolution of Real Estate Industry in China: 2002-2017.” *Structural Change and Economic Dynamics* 57: 45–56. doi:10.1016/j.strueco.2021.01.010.

He, C., Y. Zhao, J. Tian, and P. Shi. 2013. “Modeling the Urban Landscape Dynamics in a Megalopolitan Cluster Area by Incorporating a Gravitational Field Model with Cellular Automata.” *Landscape and Urban Planning* 113: 78–89. doi:10.1016/j.landurbplan.2013.01.004.

Im, J. 2020. “Earth Observations and Geographic Information Science for Sustainable Development Goals.” *GiScience & Remote Sensing* 57 (5): 591–592. doi:10.1080/15481603.2020.1763041.

Jun, M.-J. 2021. “A Comparison of A Gradient Boosting Decision Tree, Random Forests, and Artificial Neural Networks to Model Urban Land Use Changes: The Case of the Seoul Metropolitan Area.” *International Journal of Geographical Information Science* 35 (11): 2149–2167. doi:10.1080/13658816.2021.1887490.

Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu. 2017. “Lightgbm: A Highly Efficient Gradient Boosting Decision Tree.” *The proceeding of the 31st International Conference on Neural Information Processing Systems* 30: 3149–3157. doi:10.5555/3294996.3295074.

Liang, X., X. Liu, D. Li, H. Zhao, and G. Chen. 2018. “Urban Growth Simulation by Incorporating Planning Policies into a CA-based Future land-use Simulation Model.” *International Journal of Geographical Information Science* 32 (11): 2294–2316. doi:10.1080/13658816.2018.1502441.

Li, X., G. Chen, X. Liu, X. Liang, S. Wang, Y. Chen, F. Pei, and X. Xu. 2017. “A New Global land-use and land-cover Change Product at A 1-km Resolution for 2010 to 2100 Based on human-environment Interactions.” *Annals of the American Association of Geographers* 107 (5): 1040–1059. doi:10.1080/24694452.2017.1303357.

Li, L., S. Ma, Y. Zheng, and X. Xiao. 2022. “Integrated Regional Development: Comparison of Urban Agglomeration Policies in China.” *Land Use Policy* 114: 105939. doi:10.1016/j.jlandusepol.2021.105939.

Lin, J., and X. Li. 2015. “Simulating Urban Growth in a Metropolitan Area Based on Weighted Urban Flows by Using Web Search Engine.” *International Journal of Geographical Information Science* 29 (10): 1721–1736. doi:10.1080/13658816.2015.1034721.

Li, Y., and N. Phelps. 2018. “Megalopolis Unbound: Knowledge Collaboration and Functional Polycentricity within and beyond the Yangtze River Delta region in China, 2014.” *Urban Studies* 55 (2): 443–460. doi:10.1177/0042098016656971.

Li, Y., and N. A. Phelps. 2019. “Megalopolitan Glocalization: The Evolving Relational Economic Geography of Intercity Knowledge Linkages within and beyond China’s Yangtze River Delta region, 2004-2014.” *Urban Geography* 40 (9): 1310–1334. doi:10.1080/02723638.2019.11585140.

Liu, X., L. Ma, X. Li, B. Ai, S. Li, and Z. He. 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.

Lv, J., Y. Wang, X. Liang, Y. Yao, T. Ma, and Q. Guan. 2021. “Simulating Urban Expansion by Incorporating an Integrated Gravitational Field Model into a demand-driven Random forest-cellular Automata Model.” *Cities* 109: 16. doi:10.1016/j.cities.2020.103044.

Ma, J., J. C. Cheng, F. Jiang, W. Chen, and J. Zhang. 2020. “Analyzing Driving Factors of Land Values in Urban Scale Based on Big Data and non-linear Machine Learning Techniques.” *Land Use Policy* 94: 104537. doi:10.1016/j.jlandusepol.2020.104537.

Moghadam, H., M. Minaei, Y. Feng, and R. Pontius. 2019. “GlobeLand30 Maps Show Four Times Larger Gross than Net Land Change from 2000 to 2010 in Asia.” *International Journal of Applied Earth Observation and Geoinformation* 78: 240–248. doi:10.1016/j.jag.2019.01.003.
Nugraha, A. T., B. J. Waterson, S. P. Blainey, and F. J. Nash. 2022. “Unravelling the Dynamics behind the Urban Morphology of port-cities Using a LUTI Model Based on Cellular Automata.” *Computers, Environment and Urban Systems* 92: 101733. doi:10.1016/j.compenvurbsys.2021.101733.

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.

Opitz, D., and R. Maclin. 1999. “Popular Ensemble Methods: An Empirical Study.” *Journal of Artificial Intelligence Research* 11: 169–198. doi:10.1613/jair.614.

Ou, C., J. Yang, Z. Du, X. Zhang, and D. Zhu. 2019. “Integrating Cellular Automata with Unsupervised deep-Learning Algorithms: A Case Study of urban-sprawl Simulation in the Jingjintang Urban Agglomeration, China.” *Sustainability* 11 (9): 20. doi:10.3390/su11092464.

Rahnama, M. R. 2021. “Forecasting land-use Changes in Mashhad Metropolitan Area Using Cellular Automata and Markov Chain Model for 2016-2030.” *Sustainable Cities and Society* 64: 102548. doi:10.1016/j.scs.2020.102548.

Rao, H., X. Shi, A. K. Rodrigue, J. Feng, Y. Xia, M. Elhoseny, X. Yuan, and L. Gu. 2019. “Feature Selection Based on Artificial Bee Colony and Gradient Boosting Decision Tree.” *Applied Soft Computing* 74: 634–642. doi:10.1016/j.asoc.2018.10.036.

Roodposhti, M. S., R. J. Hewitt, and B. A. Bryan. 2020. “Towards Automatic Calibration of Neighbourhood Influence in Cellular Automata land-use Models.” *Computers, Environment and Urban Systems* 79: 101416. doi:10.1016/j.compenvurbsys.2019.101416.

Saganeiti, L., A. Mustafa, J. Teller, and B. Murgante. 2021. “Modeling Urban Sprinkling with Cellular Automata.” *Sustainable Cities and Society* 65: 102586. doi:10.1016/j.scs.2020.102586.

SC-PRC (The State Council of the People’s Republic of China). 2019. “Development Planning of the Yangtze River Delta Urban Agglomerations.” https://www.ndrc.gov.cn/xxgk/zcfb/ghwb/201606/W020190905497826154295.pdf.

Storper, M., and A. J. Scott. 2016. “Current Debates in Urban Theory: A Critical Assessment.” *Urban Studies* 53 (6): 1114–1136. doi:10.1177/0042098016634002.

Teitz, M. 2012. “Megapolitan America; A New Vision for Understanding America’s Metropolitan Geography.” *Journal of the American Planning Association* 78 (4): 488. doi:10.1080/01944363.2012.722892.

Tong, X., and Y. Feng. 2019. “How Current and Future Urban Patterns Respond to Urban Planning? an Integrated Cellular Automata Modeling Approach.” *Cities* 92: 247–260. doi:10.1016/j.cities.2019.04.004.

UN-DESA (The United Nations Department of Economic and Social Affairs). 2018. “World Urbanization Prospects: The 2018 RevisionWorld Urbanization Prospects: The 2018 Revision.” https://www.un.org/development/desa/publications/2018-revision-of-world-urbanization-prospects.html.

Wu, X., X. Liu, D. Zhang, J. Zhang, J. He, and X. Xu. 2022. “Simulating Mixed land-use Change under multi-label Concept by Integrating a Convolutional Neural Network and Cellular Automata: A Case Study of Huizhou, China.” *GiScience & Remote Sensing* 59 (1): 609–632. doi:10.1080/15481603.2022.2049493.

Xia, C., A. Zhang, H. Wang, B. Zhang, and Y. Zhang. 2019. “Bidirectional Urban Flows in Rapidly Urbanizing Metropolitan Areas and Their Macro and Micro Impacts on Urban Growth: A Case Study of the Yangtze River Middle Reaches Megalopolis, China.” *Land Use Policy* 82: 158–168. doi:10.1016/j.landusepol.2018.12.007.

Yan, X., Y. Feng, X. Tong, P. Li, Y. Zhou, P. Wu, H. Xie, et al. 2021. “Reducing Spatial Autocorrelation in the Dynamic Simulation of Urban Growth Using Eigenvector Spatial Filtering.” *International Journal of Applied Earth Observation and Geoinformation* 102: 15. doi:10.1016/j.jag.2021.102434.

Yang, K., W. Sun, Y. Luo, and L. Zhao. 2021b. “Impact of Urban Expansion on Vegetation: The Case of China (2000–2018).” *Journal of Environmental Management* 291: 112598. doi:10.1016/j.jenvman.2021.112598.

Yang, C., R. Xia, Q. Li, H. Liu, T. Shi, and G. Wu. 2021a. “Comparing Hillside Urbanizations of Beijing-Tianjin-Hebei, Yangtze River Delta and Guangdong–Hong Kong–Macau Greater Bay Area Urban Agglomerations in China.” *International Journal of Applied Earth Observation and Geoinformation* 102: 102460. doi:10.1016/j.jag.2021.102460.

Ye, C., Z. Liu, W. Cai, R. Chen, L. Liu, and Y. Cai. 2019. “Spatial Production and Governance of Urban Agglomeration in China 2000–2015: Yangtze River Delta as a Case.” *Sustainability* 11 (5): 1343. doi:10.3390/su11051343.

Yu, J., A. Hagen-Zanker, N. Santitisadeekorn, and S. Hughes. 2021. “Calibration of Cellular Automata Urban Growth Models from Urban Genesis onwards-a Novel Application of Markov Chain Monte Carlo Approximate Bayesian Computation.” *Computers, Environment and Urban Systems* 90: 101689. doi:10.1016/j.compenvurbsys.2021.101689.

Zhen, X., F. Cao, X. Qin, and B. Wang. 2017. “Delineation of an Urban Agglomeration Boundary Based on Sina Weibo Microblog ‘check-in’ Data: A Case Study of the Yangtze River Delta.” *Cities* 60: 180–191. doi:10.1016/j.cities.2016.08.014.

Zhou, L., X. Dang, Q. Sun, and S. Wang. 2020. “Multi-scenario Simulation of Urban Land Change in Shanghai by Random Forest and CA-Markov Model.” *Sustainable Cities and Society* 55: 102045. doi:10.1016/j.scs.2020.102045.