




Urban expansion simulation and scenario prediction using cellular automata: comparison between individual and multiple influencing factors

Yongjiu Feng  · Jiafeng Wang · Xiaohua Tong · Hossein Shafizadeh-Moghadam · Zongbo Cai · Shurui Chen · Zhenkun Lei · Chen Gao

Received: 14 November 2018 / Accepted: 4 April 2019 / Published online: 18 April 2019
© Springer Nature Switzerland AG 2019

Abstract Quantifying the contribution of driving factors is crucial to urban expansion modeling based on cellular automata (CA). The objective of this study is to compare individual-factor-based (IFB) models and multi-factor-based (MFB) models as well as examine the impacts of each factor on future urban scenarios. We quantified the contribution of driving factors using a generalized additive model (GAM), and calibrated six IFB-DE-CA models and fifteen MFB-DE-CA models using a differential evolution (DE) algorithm. The six IFB-DE-CA models and five MFB-DE-CA models were selected to simulate the 2005–2015 urban expansion of Hangzhou, China, and all IFB-DE-CA models were applied to project future urban scenarios out to the year 2030. Our results show that terrain (DEM) and population density (POP) are the two most influential

factors affecting urban expansion of Hangzhou, indicating the dominance of biophysical and demographic drivers. All DE-CA models produced defensible simulations for 2015, with overall accuracy exceeding 89%. The IFB-DE-CA models based on DEM and POP outperformed some MFB-DE-CA models, suggesting that multiple factors are not necessarily more effective than a single factor in simulating present urban patterns. The future scenarios produced by the IFB-DE-CA models are substantially shaped by the corresponding factors. These scenarios can inform urban modelers and policy-makers as to how Hangzhou city will evolve if the corresponding factors are individually focused. This study improves our understanding of the effects of driving factors on urban expansion and future scenarios when incorporating the factors separately.

Y. Feng · X. Tong (✉)
College of Surveying and Geo-Informatics, Tongji University,
Shanghai 200092, China
e-mail: xhtong@tongji.edu.cn

Y. Feng · J. Wang · Z. Cai · S. Chen · Z. Lei · C. Gao
College of Marine Sciences, Shanghai Ocean University,
Shanghai 201306, China

Y. Feng (✉)
School of Earth and Environmental Sciences, The University of
Queensland, Brisbane, QLD 4072, Australia
e-mail: yjfeng@tongji.edu.cn

H. Shafizadeh-Moghadam
Department of GIS and Remote Sensing, Tarbiat Modares
University, Tehran, Iran

Keywords Cellular automata · Differential evolution · Generalized additive model · Urban expansion · Scenario prediction · Hangzhou

Introduction

Urban expansion results from human alteration of the land surface (Lambin et al. 2001; Zadbagher et al. 2018) and is intensifying as a consequence of economic growth. Modeling urban expansion can reveal its underlying processes and spatiotemporal dynamics (Akin et al. 2015; Liu et al. 2017), providing planners and decision-makers with early warning of ecological and environmental consequences (Smidt et al. 2018; Sun

et al. 2018). Cellular automata (CA) models are effective tools to reconstruct past urban expansion and project future scenarios by analyzing land-use dynamics (Liu et al. 2018; Soares-Filho et al. 2013). Biophysical, socioeconomic, demographic, and meteorological prerequisites have been identified as factors that drive urban expansion (Ahmed et al. 2014; Osman et al. 2016). These factors are commonly applied to build transition rules in CA modeling. However, too many factors may yield multicollinearity that has undesirable impacts on the models (Feng and Tong 2017b; Von Thaden et al. 2018), leading to negative effects on their utility. Factor selection is therefore crucial in constructing CA models. As a consequence, the comparison of individual and multiple factors in CA-based urban modeling should improve our understanding of land-use dynamics.

Accurate analysis of the relationships between urban expansion and its drivers is fundamental in estimating land conversion probability, which guides CA models to control the distribution of new urban cells (Kamusoko and Gamba 2015). Wahyudi and Liu (2013) reported that, during the past two decades, an increasing number of driving factors of various categories have been applied in CA modeling. We have categorized these influencing factors as:

1. *Biophysical variables* that represent primary determinants of urban expansion, including meteorological and topographic conditions such as land surface temperature, normalized difference vegetation index, digital elevation model (DEM), and terrain slope (Dubovyk et al. 2011; Mahiny and Clarke 2012)
2. *Human-disturbance variables* that represent the proximity disturbance such as distance to city center, road networks, water bodies, ecological reserves, and farmland (Engelen 2002; Reilly et al. 2009), as well as the activity disturbance such as urban development activities and intensity (Li et al. 2011; Mitsova et al. 2011; Sang et al. 2011)
3. *Socioeconomic variables* that include social, economic, and demographic factors, such as employment potential, gross domestic product (GDP), land price, and population density (Dang and Kawasaki 2017; Haase et al. 2012; Poelmans and Rompaey 2009; Rienow and Goetzke 2015)
4. *Institutional variables* that represent macro-control policies on urban development, including urban

planning regulations, rural development policies, and land development regulations (Delden et al. 2010; Deng et al. 2015; Peña et al. 2005)

It is difficult to include all these factors in CA modeling, because they relate to hundreds of discrete variables that have been shown in many cases to be highly correlated. Factor correlation leads to variable multicollinearity, hence a negative impact on simulation accuracy. To reduce variable multicollinearity while minimizing information loss, modelers have identified dominant factors using different methods such as quantitative analysis, rough set theory, regression analysis, and spatial statistics (Mondal et al. 2015; Osman et al. 2016; Wang et al. 2011). To eliminate data redundancy, principal component analysis was applied to identify factors that dominate land development (Li and Yeh 2002). For example, Osman et al. (2016) applied an analytic hierarchy process to determine the weights of candidate factors in different regions of the Giza Governorate in Egypt. Rather than using all candidate factors, Wang et al. (2011) calibrated CA-based urban models using dominant factors identified by a rough set. A multivariate statistical model was applied to examine the influence of drivers on urban expansion at two different periods and reconstructed the urban expansion using CA models (Ahmed et al. 2014). Geographically weighted regression has been used to probe the spatial variation of driving factors for urban expansion modeling (Mondal et al. 2015; Shafizadeh-Moghadam and Helbich 2015). To assess the effects of driving factors on CA-based simulation results, González et al. (2015) applied a simplified global sensitivity analysis in modeling the urban expansion of Madrid.

Comprehensive analysis of the factors affecting urban development is generally beneficial for CA modeling. Feng and Tong (2017b) reported that CA model performance might decline when including more (e.g., larger than five) factors because multicollinearity among them may likely lead to inaccurate weight definition. With this knowledge, it is essential to ask: (1) Which factors are important to urban expansion in rapidly urbanizing areas, and is a specific factor useful in building a CA model? (2) Do multi-factor-based (MFB) CA models necessarily outperform individual-factor-based (IFB) CA models, if an individual factor can be used to build CA models? (3) How each factor affects the future

urban scenario prediction? As a result, the objective of this study is to address the above issues.

We applied a generalized additive model (GAM) that examines the ability of each factor to explain urban expansion and rank factors by their statistical significance. GAM is a nonparametric extension of a generalized linear model and uses an unspecified smoothing function to build the nonlinear relationship between response and explanatory variables (Larsen 2015). The model has been utilized to build transition rules of urban CA models (Brown and Goovaerts 2002; Feng and Tong 2017a). Here, we focus on the identification of explanatory ability and rank-order of each factor using GAM in constructing urban CA models. We then applied differential evolution (DE) to capture the land transition rules of CA models. Most recently, DE was integrated with CA to build a hybrid DE-CA model that successfully simulated rapid urban expansion in Kunming (Feng and Tong 2018a). DE automatically minimizes total residuals in the transition rules, resulting in improved accuracy over traditional CA methods. Since DE-CA can predict future scenarios, we applied this model to simulate urban expansion based on the GAM-identified individual and multiple factors. Hangzhou city that lies on the southeast coast of China was selected as our case study area. Modeling urban expansion and land-use change in Hangzhou is of great interest to researchers (Hou et al. 2019; Liu et al. 2018)

because it is a rapidly urbanizing and economically developed city in the Yangtze River Delta. A gradient CA model and a CA-Markov model were applied in these studies, but it is still unclear how urban scenarios of Hangzhou will be affected by each driving factors. For comparison, we calibrated DE-CA models based on individual (IFB-DE-CA) and multiple (MFB-DE-CA) factors using land-use change data from 2005 to 2015, and used these models to simulate the 2015 urban pattern of Hangzhou. We finally projected different urban scenarios for the year 2030 using the DE-CA models.

Study area and data

Study area

Hangzhou is the capital of Zhejiang province on the southeast coast of China (Fig. 1a). The city covers 16,596 km² and has jurisdiction over 13 sub-areas including 9 urban districts and 4 satellite cities (Fig. 1b, c).

Elevation gradually decreases from the southwest with a high slope to the northeast with a low slope. More than half of the study area is hilly, yielding the most extensive urban development in the northeast. In the *Urban Agglomeration Development Planning* of the Yangtze River Delta (NDRC 2016), Hangzhou was designated as a Type I large city with more than 5

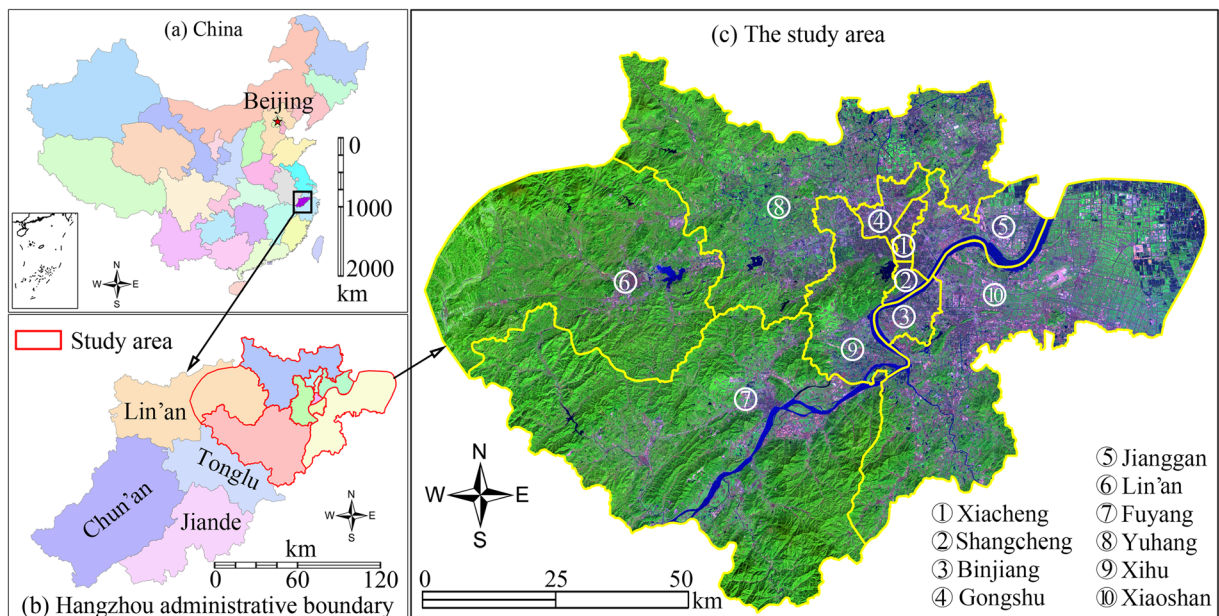


Fig. 1 The Hangzhou study area

million urban residents. According to the local Bureau of Statistics (tjj.hangzhou.gov.cn), Hangzhou had 9.2 million registered and short-term residents as of 2016, with a population density of 554 persons per km². Hangzhou's GDP increased substantially since China's opening up, rising from 284 million Chinese Yuan (about 45 million USD) in 1978 to 1130 billion Chinese Yuan (about 180 billion USD) in 2016. Like other big cities in China, the integrated effects of the population explosion and rapid economic growth have led to the fast expansion of the urban area. The built-up area of Hangzhou grew from 314 km² in 2005 to 506 km² in 2015 as reported by the local Bureau of Statistics. Comparison of land-use patterns in 2005 and 2015 classified from Landsat images shows that the newly built-up areas occur principally around the Hangzhou city center and two nearby suburban areas (Binjiang and Xiaoshan). In this study, we focused the Hangzhou city center and all the adjacent suburban areas that cover about 6300 km² (Fig. 1c).

Data and variables

We used vector and raster data to produce both dependent and independent variables in our models. The Hangzhou administrative map was implemented to identify Hangzhou boundaries, city center, and county centers (Fig. 2a). We acquired traffic network datasets from OpenStreetMap, which classifies the roads into motorway, trunk, and primary based on their categories. Among them, we included national and provincial roads that have greater impacts on urban expansion, but excluded urban roads and rural lanes that have weaker influences on urban expansion. The raster datasets

include Landsat images acquired on March 7, 2005 and October 13, 2015, which were assembled from Geospatial Data Cloud (www.gscloud.cn). We then applied Mahalanobis distance classifier in ENVI 5.2 to classify these images to produce land-use patterns. As our focus is to model urban expansion, we only considered three types of land use (i.e., urban, non-urban, and water body) to produce the 2005–2015 urban changes (Fig. 2b). We also used a digital elevation model (DEM) from the ASTER Global Digital Elevation Map and the rasterized population density from Worldpop (worldpop.org.uk). Spatial resolution set at 30 m yielded 3011 × 4233 cells in each map.

The observed 2005–2015 urban expansion was applied to generate the dependent variable (y) with a value of 1 (urban expansion) or 0 (others). We extracted four distance-based independent variables for urban expansion drivers from vector maps using the Euclidean Distance tool in ArcGIS 10.1 (Table 1; Fig. 3a–d). We included the biophysical DEM factor to reflect the impact of terrain, and the demographic POP factor to reflect the influence of population density. All driving factors were normalized using the method of Feng and Tong (2018b) to reduce the influence of dimension on DE-CA parameterization.

Among the independent variables, CITY and COUN yielded the highest mean values, while DEM and POP yielded the smallest mean values (Table 1). There are no substantial differences in standard deviation between CITY and COUN, suggesting a similar distribution of these two factors across space. The standard deviation of RAIL is larger than that of ROAD, indicating the more discrete distribution of RAIL and a lower density of the railway networks. The DEM and POP factors have the

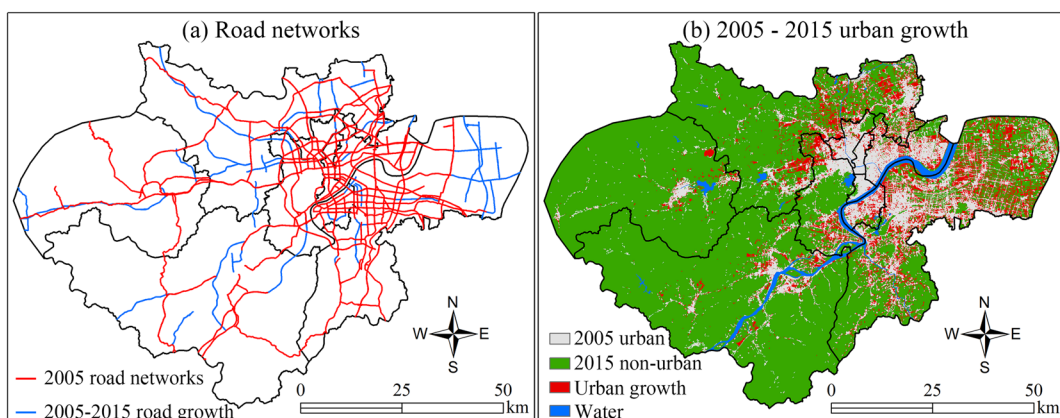


Fig. 2 Major road networks and urban expansion from 2005 to 2015

Table 1 Dependent [D] and independent [I] variables and summary statistics

Type	Variable	Meaning	Min	Max	Mean	Stdev
D	y	1 if non-urban cell is transformed to an urban cell, 0 if non-urban cell remains static	0.00	1.00	0.10	0.30
I	CITY	Effect of Hangzhou city center on urban expansion	0.00	1.00	0.46	0.22
I	COUN	Effect of county centers on urban expansion	0.00	1.00	0.41	0.21
I	ROAD	Effect of major roads on urban expansion	0.00	1.00	0.19	0.19
I	RAIL	Effect of railways on urban expansion	0.00	1.00	0.35	0.27
I	DEM	Effect of DEM on urban expansion	0.00	1.00	0.08	0.11
I	POP	Effect of population density on urban expansion	0.00	1.00	0.03	0.09

smallest standard deviation, suggesting that most areas are relatively plain and most regions have a low population density, sometimes with zero residents (Fig. 3e, f).

Methods

Workflow

Figure 4 shows our procedure of DE-CA modeling in four steps: (1) *data collection and preprocessing*

that include radiometric calibration, atmospheric and geometric rectification, and study area cropping; (2) *variable extraction* that includes classification of Landsat images to produce land-use patterns and spatial visualizations of driving factors; (3) *factor assessment using GAM* that selects samples from the land-use change map and driving factors using systematic sampling, and applies GAM to identify the rank-order of driving factor; and (4) *modeling and assessment* that construct CA transition rules using DE to build IFB-DE-CA and

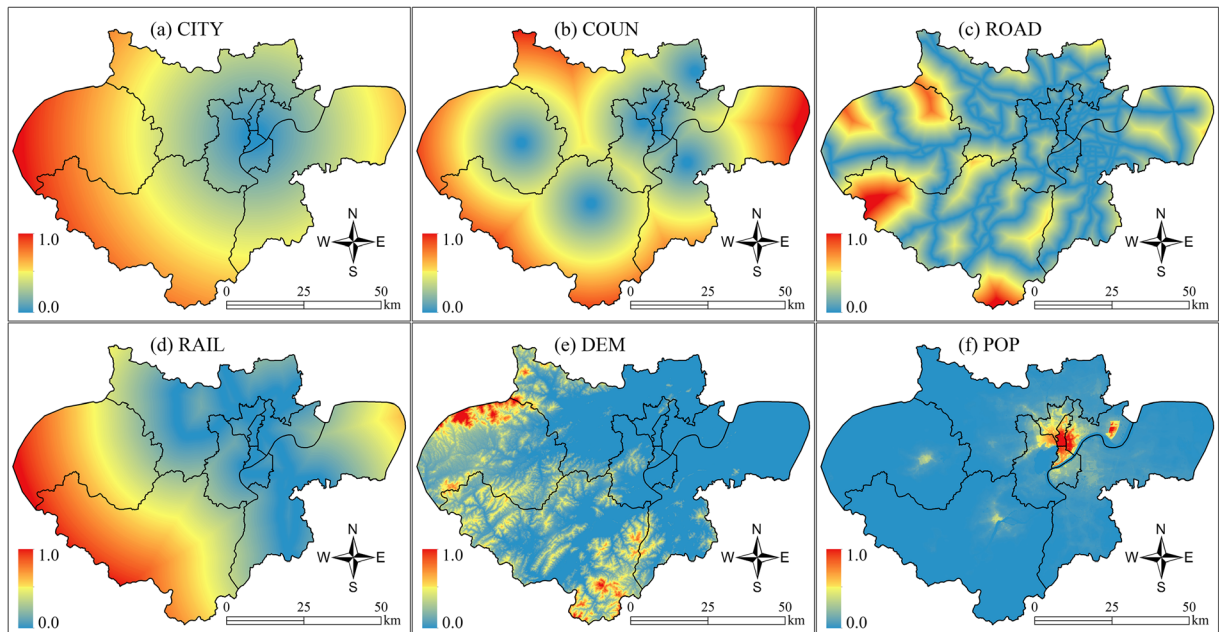


Fig. 3 Spatial visualization of independent variables in the DE-CA model

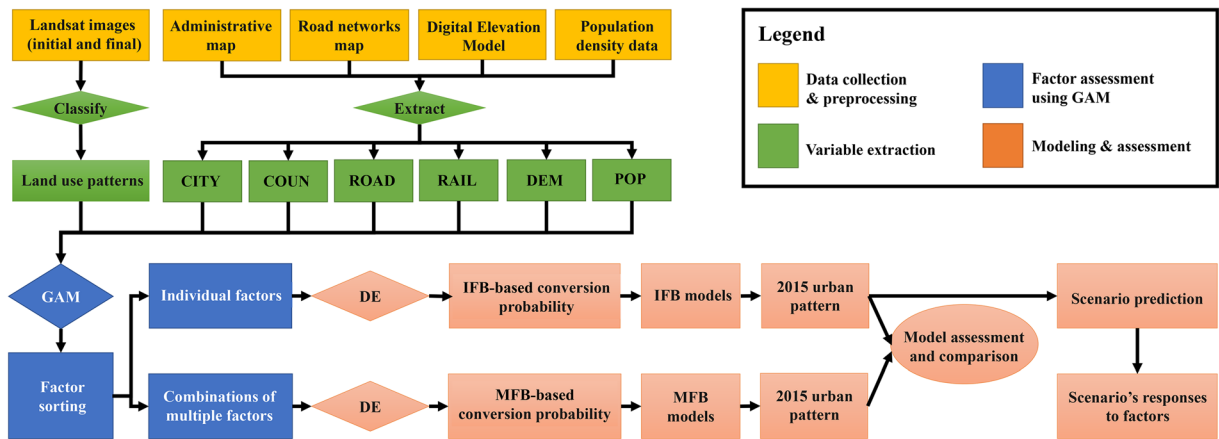


Fig. 4 Modeling procedure under different strategies for factor combination

MFB-DE-CA models, and simulate present urban pattern and future scenarios.

The DE-CA model

Our DE-CA model is an improvement of a typical CA by automatic calibration using DE. In CA, four elements including the current cell state, contiguity cell effects, constraints, and land transition probability collaboratively determine the land transformation. The DE-CA transition rules can be expressed as

$$Stat_m^{t+1} = LandTrans(Stat_m^t, NeiEff, Con, P_t) \tag{1}$$

where

- *Land transition function (LandTrans)* denotes the transition rules that integrate the effects of the four current elements to determine the future cell state;
- *Future cell state (Stat_m^{t+1})* represents the state of cell *m* at time *t* + 1 and *current cell state (Stat_m^t)* represents the state of the cell *m* at time *t*;
- *Neighborhood effects (NeiEff)* denote the influence of neighboring urban cells on the state of the cell in processing, with 5 × 5 Moore square neighbors widely applied in DE-CA models (Pan et al. 2010; Wu 1998);
- *Constraints (Con)* denote prohibited areas resulting from unsuitable conditions or urban planning regulations, including broad water bodies, high-slope areas, and protected areas (Feng and Tong 2018b); and
- *Transition probability (P_t)* denotes the temporally stationary, but spatially non-stationary, land

conversion probability defined by driving factors. The probability is calculated by (Mustafa et al. 2018):

$$P_t = \frac{e^{(a_0+a_1 \times D_1+\dots+a_n \times D_n)}}{1 + e^{(a_0+a_1 \times D_1+\dots+a_n \times D_n)}} \tag{2}$$

where *a*₀ is the intercept, *n* is the number of factors, and *a*₁, ..., *a*_{*n*} are parameters for each factor representing its weight. Here, *n* ranges from 1 to 6, with *n* = 1 indicating only one factor included in the modeling and *n* = 6 indicating all six factors included in the modeling.

The parameters (*a_n*) are commonly retrieved using logistic regression (LR). Many studies have reported that LR alone cannot sufficiently address urban expansion dynamics (Feng and Tong 2018a; Li and Yeh 2002). In contrast, heuristics such as DE can search for parameters that represent the complex urban expansion and can minimize the residuals in fitting samples. Residuals are usually calculated as the root-mean-square error (RMSE) between the observed land conversion value and the predicted transition probability. A RMSE-based objective function can thus project the transition rule space into the DE space. A typical objective function is (Feng and Tong 2018a):

$$\begin{aligned} & \text{Min } Func(D_1, \dots, D_n) \\ & = \sqrt{\frac{\sum (P_v(D_1, \dots, D_n) - P_o)_m^2}{Q}} \end{aligned} \tag{3}$$

where *Func*(*D*₁, ..., *D*_{*n*}) is the objective function; for cell *m*, *P_v*(*D*₁, ..., *D*_{*n*}) is the predicted transition probability and *P_o* is the observed land conversion; and *Q* is the number of samples.

DE is a global heuristic similar to genetic algorithms (GAs) that are guided by an objective function as well as the population mutation, crossover, and selection operators. Compared with GA, DE searches for the global optimum using differentiation among populations, and does not easily fall into local optima (Das et al. 2009; Noroozi et al. 2011). DE starts with population initialization under predefined bound constraints. The process randomly selects three vectors from the initial population and then subtracts two of these to generate a difference vector that is added with the third vector to generate a new vector, which is compared with the third one. If the new individual's function value is smaller than that of the third one, the new one will replace the old one at the next generation. DE finds a near-optimal solution after the population undergoes mutation, crossover, and selection operations. Here, the solution is a set of optimal parameters for the CA transition rules.

Thresholds were defined according to the total land available for development and the maximum number of iterations from the initial to final year, where an iteration denotes 1 year. For modeling the urban pattern, the total land available was taken to be the actual urban area in the final year (2015) of the model. The threshold is preferred if it satisfies both the total land and the maximum iterations. We defined the thresholds for future predictions using the same method.

Factor combination using GAM

We adopted a GAM to assess the influence of each factor before building the transition rules. The model has the advantage of quantifying the factors' ability to explain urban expansion and ranking them accordingly. A standard GAM allows flexible estimation of the underlying nonlinear relationships between dependent and independent variables (Guisan Jr et al. 2002). An example of GAM can be represented as

$$\begin{aligned}
 g(E(y)) = & b_0 + s_1(\text{DEM}) + s_2(\text{POP}) + s_3(\text{CITY}) \\
 & + s_4(\text{RAIL}) + s_5(\text{ROAD}) \\
 & + s_6(\text{COUN}) \tag{4}
 \end{aligned}$$

where $g(E(y))$ is a function that links the expected value $E(y)$ to all independent variables, b_0 is the intercept, and $s_i(\cdot)$ is a nonparametric smoothing function that relates each variable to $g(E(y))$. The factors' positions in the above equation reflect their impacts, in descending

order, which may be different when using these factors to explain urban expansion at different locations and at different periods.

GAM is a rank-order sensitive method that introduces factors into the model one by one. An anterior factor has a stronger impact on urban expansion while a posterior factor has a weaker impact. Percentage of deviance explained (PDE) indicates the factor's ability to explain urban expansion and this contribution in GAM, and Akaike Information Criterion (AIC) indicates the model's comparative performance. We therefore applied PDE and AIC to define the rank order of each factor and combinations of factors.

Evaluation methods

Accuracy assessment

CA models are usually assessed using cell-by-cell comparison between the simulated patterns and the actual patterns (Feng and Tong 2018b; Liu et al. 2017; Musa et al. 2017). This method usually generates an error matrix that reports overall accuracy and Kappa coefficients. While pattern comparison indicates the global performance of CA models, its weakness is obvious when trying to identify the changes that are simulated correctly or falsely (Pontius and Millones 2011). CA models can also be assessed using an overlaid map of the simulated urban expansion and the actual urban expansion, which usually reports metrics such as hit, correct rejection, miss, and false alarm (Aldwaik and Pontius Jr 2012; Pontius et al. 2013). *Hit* means that the 2005–2015 reference urban expansion was correctly simulated as urban expansion, and *correct rejection* represents that the actual non-urban persistence during 2005–2015 was correctly identified. *Miss* indicates that the 2005–2015 urban expansion was incorrectly simulated as non-urban persistence, while *false alarm* indicates that non-urban persistence was mistakenly simulated as urban expansion.

Landscape metrics

To assess the agreement between the simulated landscape patterns and actual landscape patterns, researchers have applied a set of metrics that can quantify the landscape composition and structure (Chaudhuri and Clarke 2013; Feng and Tong 2017b; Whitsed and Smallbone 2017). The metrics include a number of

Table 2 The PDE, AIC, and rank order that show factor contribution for each IFB-DE-CA model

Model	Variable	Residual deviance	Deviance explained	Percentage of deviance explained (PDE; %)	Akaike information criterion (AIC)	Rank order
Model 0	Null	520.108			2722.681	
Model 1	DEM	453.707	66.401	12.77	2009.548	1
Model 2	POP	455.415	64.693	12.44	2030.229	2
Model 3	CITY	488.356	31.752	6.10	2401.780	3
Model 4	RAIL	496.085	24.023	4.62	2478.801	4
Model 5	ROAD	500.778	19.330	3.72	2530.451	5
Model 6	COUN	507.691	12.417	2.39	2608.308	6

landscape spatial characteristics at the landscape, class, and patch levels (Feng et al. 2018; Mcgarigal 2014). Among commonly used landscape metrics, area-edge metrics reflect the size and edges of patches, shape metrics reflect the complexity of patches, and aggregation metrics refer to the trend of spatial aggregation of urban patches (Mcgarigal 2014). In this study, we calculated ten class-level metrics to evaluate urban land-use patterns following Niesterowicz and Stepinski (2016):

- Percentage of landscape (PLAND), Largest Path Index (LPI), and total edge (TE) from the area-edge category
- Perimeter-area fractal dimension (PAFRAC) from the shape category
- Path density (PD), Interspersion Juxtaposition Index (IJI), Patch Cohesion Index (COHESION), Landscape Division Index (DIVISION), Aggregation Index (AI), and Splitting Index (SPLIT) from the aggregation category

Results

Factor effects

A null model with no factors (Model-0) and six IFB-DE-CA models (models 1–6) were constructed to evaluate each factor's impact on urban expansion (Table 2). The fitting statistics of GAMs show that the rank-order of the factors are DEM, POP, CITY, RAIL, ROAD, and COUN (Table 2). The terrain conditions (DEM) and population density (POP) have the highest influences with ~12.8% deviance explained while COUN

explained the least deviance ~2.4%. Models 1–2 yield the largest PDE and smallest AIC, indicating good performance of these two models. In contrast, models 3–6 yield smaller PDE and larger AIC, implying that these models were built using less influential factors.

For comparison, fifteen MFB-DE-CA models were constructed to assess the influences of multiple factors on urban expansion (Table 3). Of the two-factor models, model 7 (based on DEM and POP) explained ~15% deviance, while the other four models explained ~13% deviance. For the three-factor models, model 12 explained more than 16% deviance while the others explained deviance less than 16%. The models with four and five factors explained similar deviance. We selected the models that explained the highest deviance over those with the same number of factors, and selected model 21 that included all factors, to simulate the 2000–2015 urban expansion in Hangzhou. The deviance increased and the AIC decreased as more factors were put into the models, showing better fitting performance as more factors applied in the models, given that multicollinearity is not considered. While this suggests that more factors lead to more accurate simulation results, it may be violated by simulation practices.

Transition rules and land conversion probability maps

Controlling parameter assignment is critical for solving the objective function using DE because heuristics are sensitive to their controlling parameters. We followed Feng and Tong (2018a) to define the controlling parameters using the default values (Table 4) recommended in the package "DEoptim" in R-Gui. Among these parameters, the population size was assigned to 20 times the sum of the number of variables and an interpret (a_0). The lower and upper bounds are external DE parameters

Table 3 The PDE and AIC in GAMs that show factor contribution for the MFB-DE-CA models

Model	Number of variables	Variable	Residual deviance	Deviance explained	Percentage of total deviance explained (%)	AIC
Model 0	0	Null	520.108			2722.681
Model 7	2	DEM-POP	440.858	79.250	15.24	1869.766
Model 8	2	DEM-COUN	450.428	69.680	13.40	1984.731
Model 9	2	DEM-CITY	450.841	69.267	13.32	1990.716
Model 10	2	DEM-RAIL	451.557	68.551	13.18	1995.474
Model 11	2	DEM-ROAD	453.059	67.049	12.89	2009.631
Model 12	3	DEM-POP-COUN	435.437	84.671	16.28	1816.901
Model 13	3	DEM-POP-RAIL	436.966	83.142	15.99	1836.305
Model 14	3	DEM-POP-CITY	437.375	82.733	15.91	1840.479
Model 15	3	DEM-POP-ROAD	439.416	80.692	15.51	1861.715
Model 16	4	DEM-POP-COUN-RAIL	432.488	87.620	16.85	1793.699
Model 17	4	DEM-POP-COUN-CITY	433.582	86.526	16.64	1806.131
Model 18	4	DEM-POP-COUN-ROAD	434.436	85.672	16.47	1811.915
Model 19	5	DEM-POP-COUN-RAIL-CITY	431.099	89.009	17.11	1788.591
Model 20	5	DEM-POP-COUN-RAIL-ROAD	432.257	87.851	16.89	1791.312
Model 21	6	DEM-POP-COUN-RAIL-CITY-ROAD	430.910	89.198	17.15	1786.939

defined by the parameters calculated in LR. The lower bound of positive parameters and the upper bound of negative parameters are taken to be 0, while the upper bound of positive parameters and the lower bound parameters are defined as two times the calculated parameters by LR (Table 5).

Table 5 shows that the same factor may yield different parameters and even different signs in various models, suggesting the changing effects of factors across models. For factors except for POP, a negative parameter indicates a promotive effect on urban expansion while a positive parameter indicates a resistive effect. DEM has negative and the highest absolute parameters in models, suggesting that it strongly drives urban expansion. This confirms the rank order and the deviance explained for DEM (Table 3). The CITY’s parameters are all negative in models 3, 19, and 21, and there are no major differences between these parameters that show similar promoting effects. ROAD yield

negative parameters in models 5 and 21, where the larger absolute parameter in model 5 shows a stronger influence.

Some factors yield changing impacts in different models. POP is positive in model 2 but negative in the other five MFB-DE-CA models, suggesting that POP promotes urban expansion in model 2 but inhibits it in the other models. COUN is negative in model 6 but positive in the other MFB-DE-CA models except for model 7, suggesting that COUN yields less strong influences when applying it with other factors in models. RAIL is negative and yields attractive effects in models 4 and 16, whereas it is positive and yields repulsive effects in models 19 and 21.

The conversion probabilities derived from the IFB-DE-CA models were stretched to range between 0 and 1, and used as input into the DE-CA models to simulate urban expansion. The maps in Fig. 5 show different spatial patterns that are characterized by their

Table 4 Definition of controlling parameters in DE for solving the objective function

Note: the crossover probability indicates how often crossover will be performed, and the scaling factor is a parameter to adjust mutation

Controlling setting	Parameter	Controlling setting	Parameter
Starting population	Null	Convergence tolerance	1e-10
Optimization strategy	Local-to-best/1/bin	Max iteration	5000
Crossover probability	0.5	Population size	20 × (N + 1)
Scaling factor [0, 2]	0.8		

Table 5 The lower and upper bounds for DE and the computed CA parameters of all the selected models

Model	Variable	Lower	Upper	Parameter	Model	Variable	Lower	Upper	Parameter
Model-1	Constant	-1.8	0	-0.4915	Model-12	Constant	-1.6	0	-0.5335
	DEM	-140	0	-134.6799		DEM	-150	0	-122.4532
	Best value			0.2238		POP	-2.7	0	-0.6996
Model-2	Constant	-4.3	0	-2.1301	Model-16	COUN	0	0.5	0.0765
	POP	0	3.5	1.0337		Best value			0.2237
	Best value			0.2415		Constant	-1.6	0	-0.5403
Model-3	Constant	-1.1	0	-0.8705	DEM	-150	0	-120.5924	
	CITY	-7.5	0	-2.8592	POP	-2.8	0	-0.7170	
	Best value			0.2364	COUN	0	0.5	0.1264	
Model-4	Constant	-2.4	0	-1.3045	RAIL	-0.6	0	-0.1103	
	RAIL	-6.3	0	-2.7965	Best value			0.2237	
	Best value			0.2364	Model-19	Constant	-1.2	0	-0.1266
Model-5	Constant	-2.8	0	-1.3765	DEM	-150	0	-132.3592	
	ROAD	-9.9	0	-5.3627	POP	-3.5	0	-1.2380	
	Best value			0.2374	COUN	0	0.9	0.7793	
Model-6	Constant	-3.3	0	-1.5288	RAIL	0	1.5	0.8958	
	COUN	-2.1	0	-1.4762	CITY	-2.5	0	-2.3986	
	Best value			0.2410	Best value			0.2233	
Model-7	Constant	-1.6	0	-0.4986	Model-21	Constant	-1.2	0	-0.1207
	DEM	-150	0	-122.9757	DEM	-150	0	-132.5431	
	POP	-2.7	0	-0.7665	POP	-3.6	0	-1.2540	
	Best value			0.2237	COUN	0	1	0.8026	
				RAIL	0	1	0.9023		
				CITY	-2.4	0	-2.3855		
				ROAD	-1.1	0	-0.1838		
				Best value			0.2233		

corresponding factors. In the DEM-based model 1 (Fig. 5a), conversion probability is low in the high-elevation areas but high in the low-lying areas. Model 2 yields high conversion probability in the city center with high population density, while it yields low conversion probability in other areas with low population density (Fig. 5b). In models 3–6 (Fig. 5c–f), high conversion probabilities occur principally in areas near the Hangzhou city center, railways, main roads, and county centers, respectively.

The multi-factor-based conversion probability maps in Fig. 6a–e are similar to each other, with minor differences. All maps based on multiple factors (Fig. 6a–e) are similar to the DEM-based map (Fig. 5a). This is probably due to the dominant effect of elevation and topography. Visual inspection shows different spatial patterns

for model 7 and models 12 and 16, but their differences are small and probably insignificant (Fig. 6f, g). While model 7 and models 19 and 21 (Fig. 6h, i) are quite similar, they differ widely in probability, ranging from 0.00 to 0.32. This suggests that model 7 may be quite different from models 19 and 21. All models yield the same min and max due to normalization, but most means and SDs differ between factors (Table 6), leading to different land conversion thresholds.

Simulation results

We simulated 11 urban patterns for 2015 using the DE-CA models but present only nine in Fig. 7 because models 12 and 16, and models 19 and 21, are very similar. Figure 7 shows that the selected DE-CA models

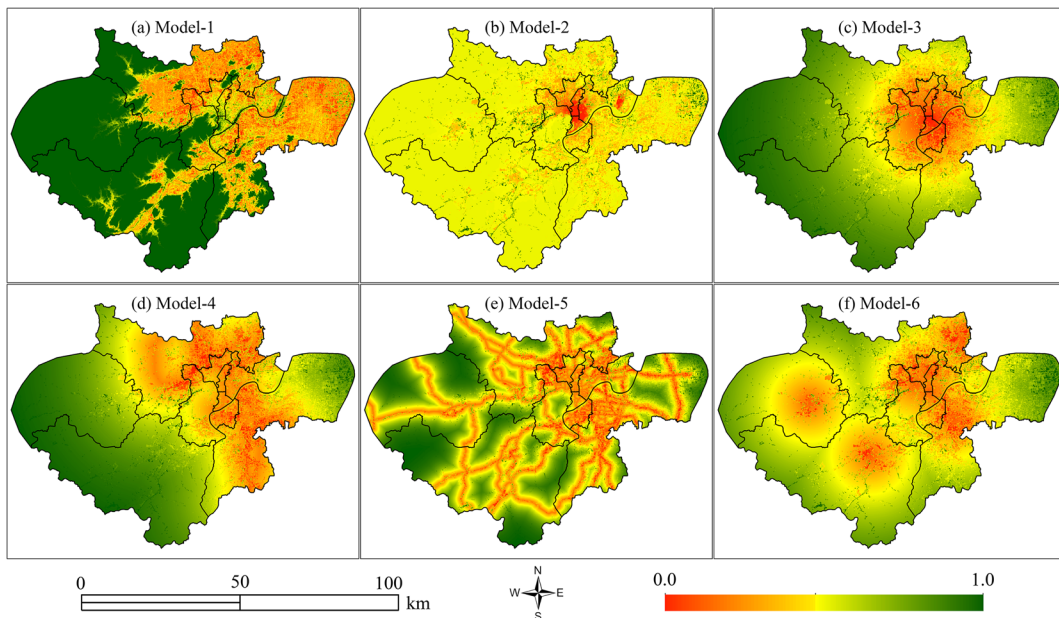


Fig. 5 Land conversion probability maps produced by all six IFB-DE-CA models

are capable of generating the 2015 urban patterns where newly built-up areas occur surrounding the existing built-up areas.

Models 1, 2, 7, 12, and 19 (Fig. 7a, b, g–i) yielded quite similar outcomes with minor differences. All models simulated fewer changes in the outer suburb Lin’an (region 3). Compared to the other models, model 2 simulated fewer changes in regions 1 and 2 (Fig. 7b) while model 19 simulated the fewest changes in region 1 but more changes in region 5 (Fig. 7i). Models 3 and 4 produced similar spatial patterns as they simulated fewer changes in the far suburbs for both Lin’an and Fuyang (Fig. 7c, d). The simulated urban expansion areas are mainly distributed around the Hangzhou city center. By comparison, model 3 allocated new urban areas in the east-west direction while model 4 allocated new urban areas in the north-south direction. The simulated changes in model 5 occur mainly along the road networks (Fig. 7e), confirming the significant influence of the major roads on urban expansion. Model 6 captured most of the urban expansion in the Hangzhou city center and two of its distant suburbs (Lin’an and Fuyang), but it missed many other urban areas in region 2 (Fig. 7f).

Accuracy analysis and model response

The overall accuracy for all the DE-CA models varies as the iteration number increases (Fig. 8). Models 1, 2, and

7 yielded the greatest accuracy (~91%) at the 10th iteration (Fig. 8a), models 3–6 yielded the greatest accuracy (~89%) at the 6th iteration (Fig. 8a), and the remaining four models yielded the greatest accuracy (~90%) at the 8th iteration (Fig. 8b). While the overall accuracy of models 1, 2, and 7 monotonously increased during iteration, their simulated quantity of urban expansion matched the total land available for development at the 10th iteration. The overall accuracy of models 3–6 first increased as the iteration number increased, then peaked at the 6th iteration. The new urban areas of these four models satisfied the total land available for development at the 10th iteration, where they have lower overall accuracy than at the 6th iteration. Models 12, 16, 19, and 21 yielded the greatest accuracy at the 8th iteration, similar to models 3–6. These results show that the DE-CA models containing different driving factors led to different simulation processes and outcomes.

To show the spatial patterns of accurate and erroneous simulations, we performed an overlay analysis for the 2005 reference map, the 2015 reference map, and the 2015 simulated map (Fig. 9). The overlay maps have five categories: urban agreement, non-urban agreement, miss, false alarm, and water body. The water body was excluded from the accuracy assessment. The maps show that the simulation results for models 1, 2, 7, 12, and 19 had fewer misses and false alarms (Fig. 9a, b, g–i), with

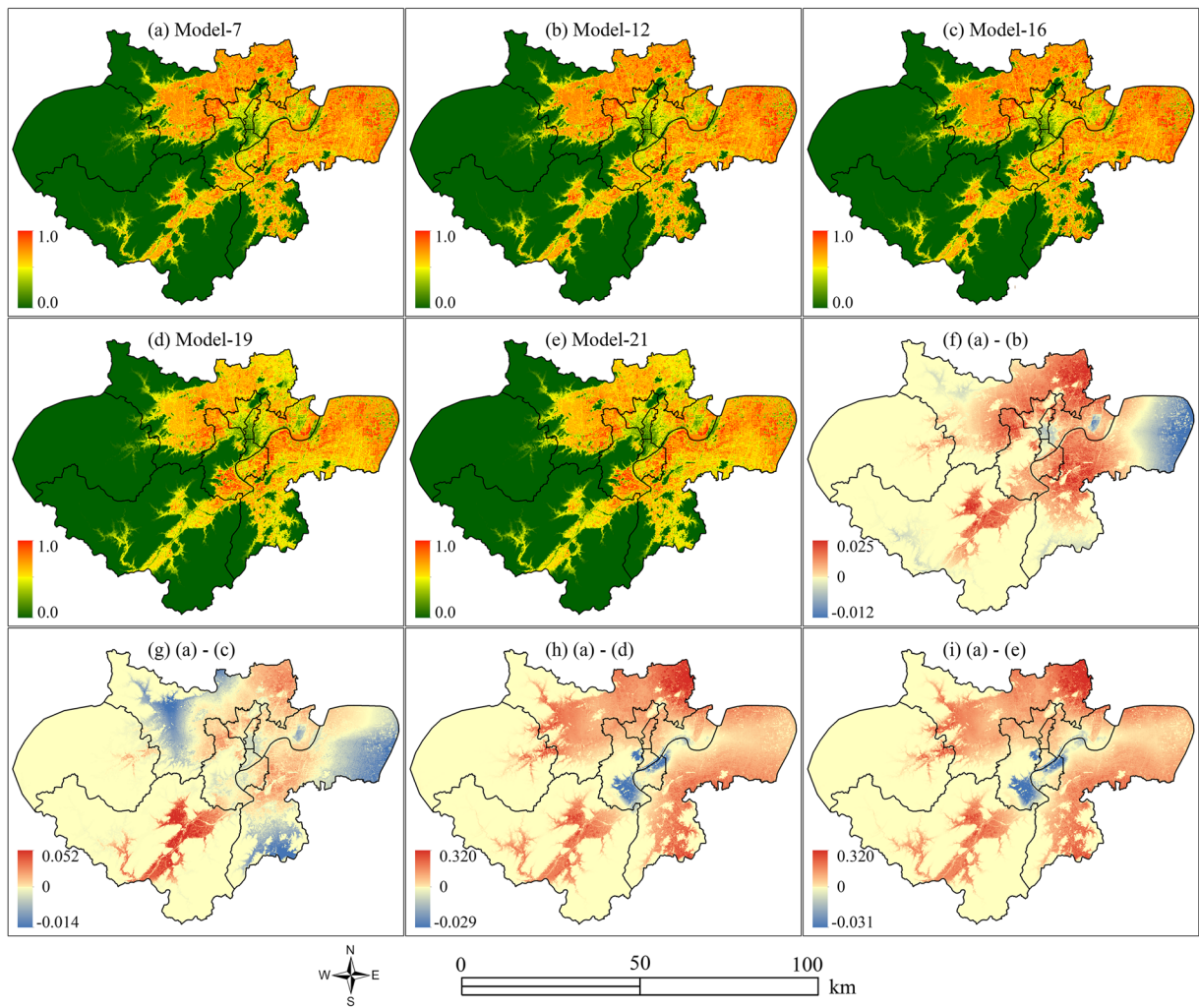


Fig. 6 Land conversion probability maps and their differences for the MFB-DE-CA models

Table 6 Summary statistics of land conversion probability from different models

Model	Variable	Mean	SD	Threshold
Model 1	DEM	0.24	0.28	0.496
Model 2	POP	0.40	0.06	0.400
Model 3	CITY	0.32	0.18	0.430
Model 4	RAIL	0.41	0.24	0.530
Model 5	ROAD	0.42	0.25	0.530
Model 6	COUN	0.49	0.15	0.490
Model 7	DEM-POP	0.24	0.28	0.496
Model 12	DEM-POP-COUN	0.23	0.28	0.490
Model 16	DEM-POP-COUN-RAIL	0.24	0.28	0.495
Model 19	DEM-POP-COUN-RAIL-CITY	0.20	0.24	0.450
Model 21	DEM-POP-COUN-RAIL-CITY-ROAD	0.20	0.24	0.448

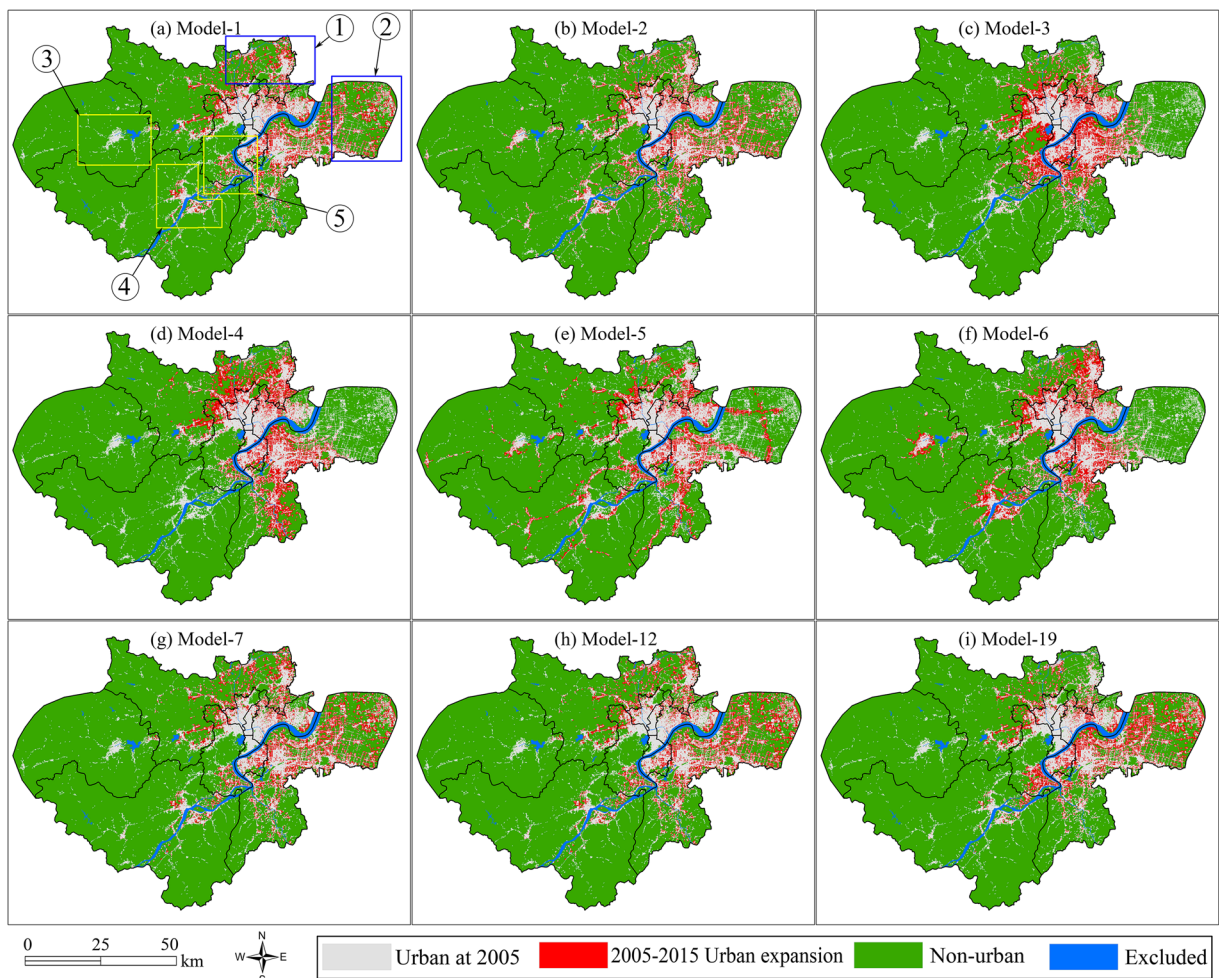


Fig. 7 Simulated 2015 Hangzhou urban patterns from DE-CA models with different factors

simulation errors principally occurring in the east and north of Hangzhou. In contrast, models 3–6 yielded more misses in eastern Hangzhou, and most of the false alarms occurred around the existing built-up area (Fig. 9c). Model-4 yielded more false expansion in the

southern and northern parts of Hangzhou (Fig. 9d) while model 5 yielded more false expansion along the road networks (Fig. 9e). For model 6, false alarms occurred in Hangzhou city center and north of Hangzhou, as well as in two satellite cities (Lin’an and Fuyang).

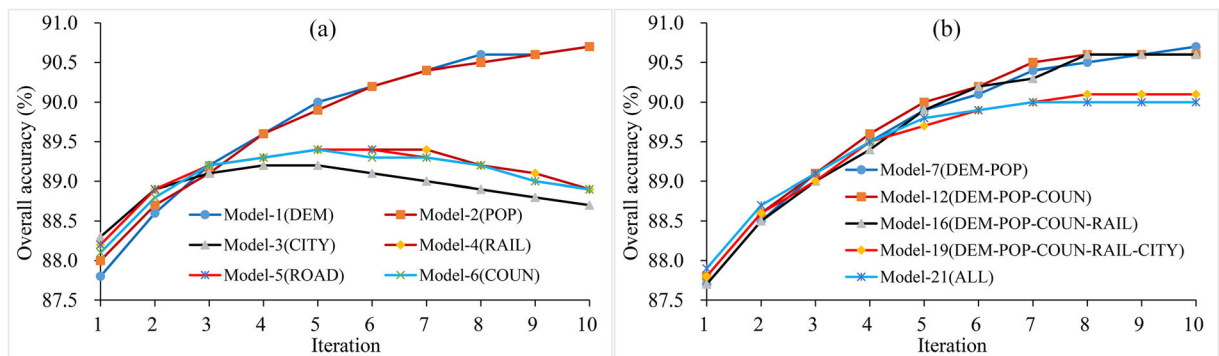


Fig. 8 Overall accuracy versus iteration number for 2015 urban expansion simulations

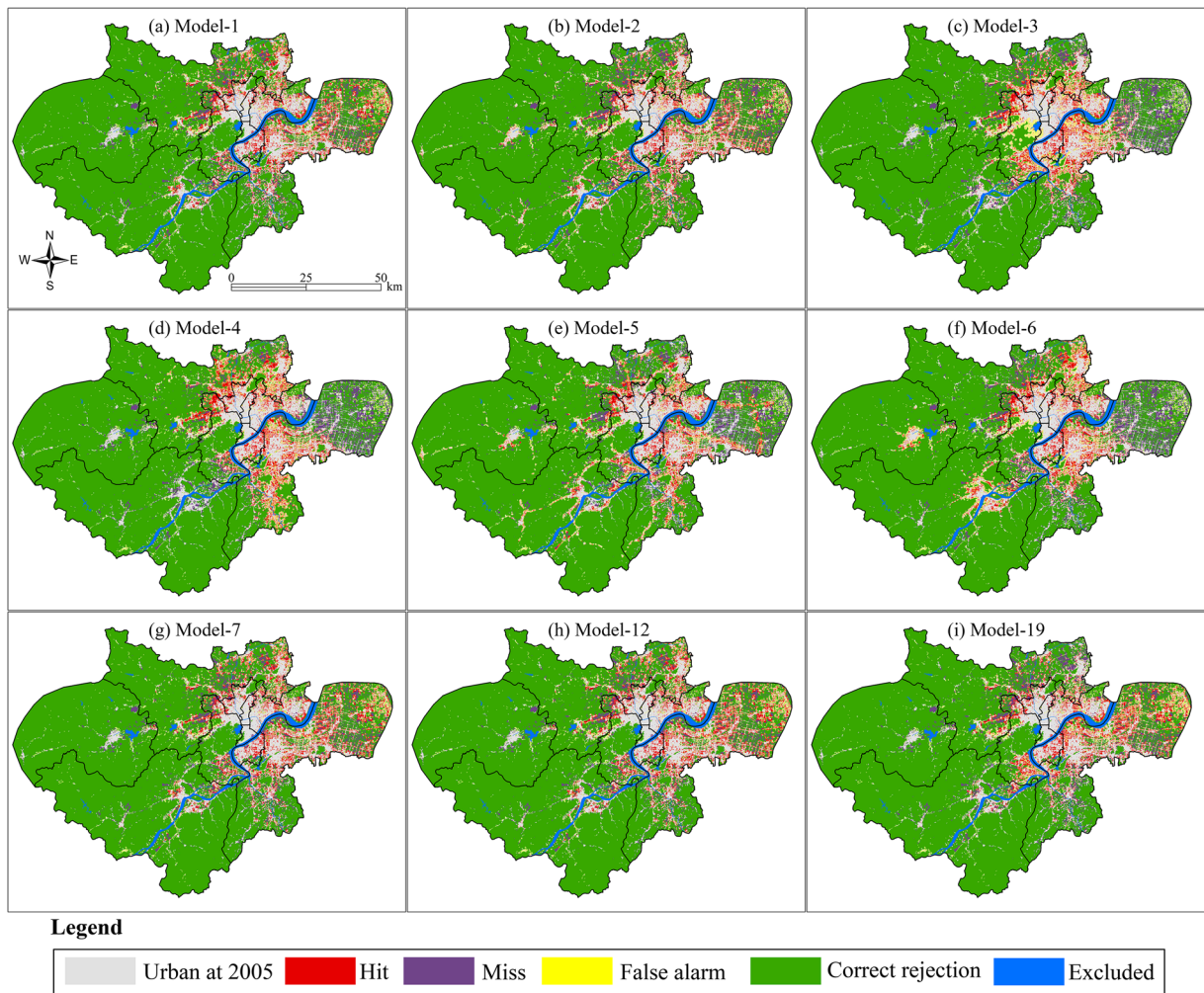


Fig. 9 Overlay maps showing 2015 simulation hits, misses, and false alarms

As each simulation progressed, hits increased monotonously, while misses decreased and false alarms increased (Fig. 10), indicating that more simulation errors were introduced when more urban expansion was accurately modeled. Quantitatively, models 1 and 2 hit 5.7% urban expansions correctly during 2005–2015, while models 3–6 hit 0.7% fewer urban expansions as compared with the former. Models 1 and 2 missed 4.9% actual urban expansions and generated 4.4% false alarms, whereas models 3–6 produced the largest misses (5.6%) and false alarms (5.5%). Models 7, 12, and 16 hit more than 5.7% of urban expansion, and there are no quantitative differences among them. The correctly captured urban expansions (~5.4%) of models 19 and 21 are slightly smaller than those of models 7, 12, and 16. Models

7 and 12 missed less than 5% of urban expansions while models 16, 19, and 21 missed more than 5% of urban expansions. Models 7 and 16 yielded similar false alarms ~4.4%, while models 12, 19, and 21 yielded slightly larger false alarms than the former models. Overall, the IFB-DE-CA models 1 and 2 and all MFB-DE-CA models are superior to the IFB-DE-CA models 3–6 as measured by all metrics including hit, miss, and false alarm, suggesting the latter four models are less preferred in modeling urban expansion at Hangzhou.

Predicted 2030 scenarios

To evaluate the effect of factors on urban scenarios, we predicted Hangzhou urban scenarios for 2030 using the

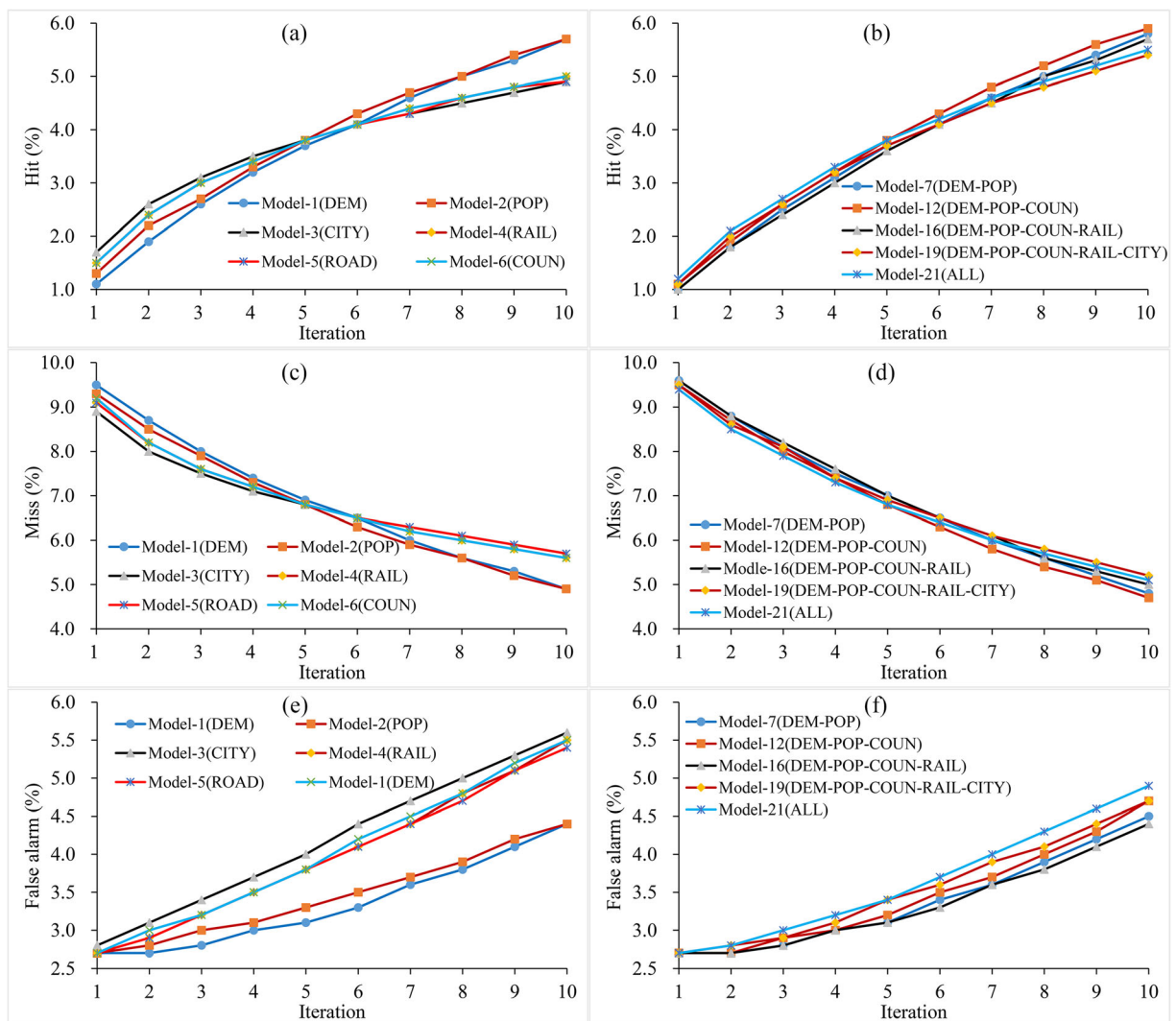


Fig. 10 Hit, miss, and false alarm percentages as a function of iteration number in simulating 2015 urban patterns

IFB-DE-CA models because each only includes only one factor. We applied a linear extrapolation to determine the amount of future urban expansion based on the 2005–2015 annual growth rate. Hangzhou’s urban area expanded by approximately 488 km² from 2005 to 2015, leading to a projected expansion of ~732 km² from 2015 to 2030 and a total urban area of ~2125 km² by 2030.

Figure 11 shows distinct differences in the spatial pattern among different scenarios, reflecting the effects of the corresponding factors on future urban expansion when applying them in DE-CA models. The model 1 scenario shows that the newly built-up areas will occur mainly in the flat and low-slope areas of eastern and northern Hangzhou (Fig. 11a). The model 2 scenario shows that urban expansion will primarily occur in highly populated areas

with less expansion in areas of low population density (Fig. 11b). For models 3 and 4, less expansion is observed in eastern Hangzhou and the outer suburbs Lin’an and Fuyang. Model 3 predicts greater urban expansion than Model 4 in southern and northwestern Hangzhou, while Model 4 predicts much more expansion in Xihu District. Model 5 suggests that new urban areas will form along existing major roads (Fig. 11e). Model 6 (Fig. 11f) projects expansion in the coming 15 years in the fringe areas of both Hangzhou city center and satellite cities Lin’an and Fuyang.

To examine differences in the predictions, ten class-level landscape metrics were used to characterize the spatial patterns of the scenarios (Table 7). PLAND shows the percentage of urban land use in 2030, where only ~0.7%

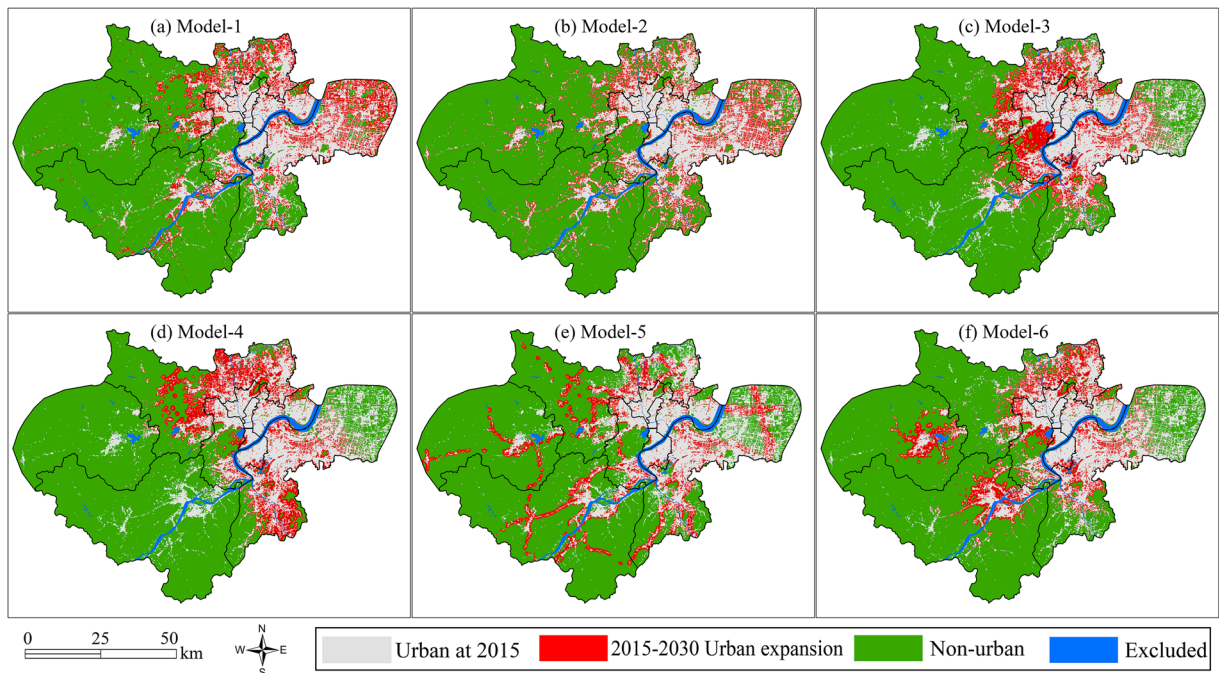


Fig. 11 The urban scenario prediction to the year 2030 using all six IFB-DE-CA models

differences exist between the models, indicating similar quantity control ability. Models 1–2 have greater LPI than the other models, suggesting that by 2030, these two models will fill more non-urban areas surrounded by the largest urban patch in 2015. These findings were confirmed by their relatively smaller values of TE and PD. PAFRAC suggests that the model 2 scenario yields a relatively less complex urban pattern than the other scenarios. As inferred by IJI, the urban patches of the model 3 scenario are more juxtapositioned while those of the model 5 scenario are less juxtapositioned compared with the other

models. This indicates that the CITY factor aggregates more new urban cells to the Hangzhou city center (Fig. 11c) while the ROAD factor attracts more urban cells along road networks (Fig. 11e). All scenarios have similar large values of the COHESION, DIVISION, and AI metrics, demonstrating the high connectivity among urban patches and the aggregation effect of the DE-CA models in predicting future urban expansion. Models 5–6 show the highest SPLIT, suggesting that urban landscapes in these simulations are more fragmented when compared to the other four scenarios. We applied the same samples to build

Table 7 Landscape metrics of predicted 2030 urban land-use

Category	Metric	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Area-edge category	PLAND (%)	33.2774	33.3831	33.9916	33.8305	33.863	33.8677
	LPI (%)	13.7723	13.3585	11.7019	11.9556	11.415	10.8835
	TE (10 km)	558.7320	565.1070	590.9310	684.8370	722.6010	685.5840
Shape category	PAFRAC	1.3491	1.2671	1.3159	1.3453	1.3352	1.3014
Aggregation category	PD (n/100 ha)	0.1481	0.1429	0.1655	0.1676	0.1665	0.1593
	IJI (%)	33.7474	30.8198	35.1059	28.9423	25.2578	31.5347
	COHESION (%)	99.7655	99.7930	99.8014	99.8068	99.7972	99.7973
	DIVISION (%)	0.9729	0.9755	0.9726	0.9749	0.9794	0.9799
	AI (%)	97.9918	97.9836	97.9349	97.5782	97.4652	97.5883
	SPLIT (%)	36.9362	40.8499	36.5167	39.7660	48.6250	49.8690

the DE-CA models using DE, and the differences among the scenarios can be attributed to the effect of different driving factors.

Discussion

We used a stepwise GAM to examine the effects of driving factors and ranked these factors according to their statistical significance. Each factor was used to produce a land conversion probability map to create a DE-CA model. Such models based on individual factors have not been previously reported. We identified five combinations of land-use driving factors to generate conversion probability maps, which were in turn used to build five MFB-DE-CA models. The rapid urban expansion in Hangzhou during 2005–2015 was then simulated using all DE-CA models, yielding high overall accuracies exceeding 89%. Our accuracies are comparable to previous case studies in Hangzhou, which reported accuracies of 84–87% (Hou et al. 2019; Liu et al. 2018). To examine the impacts of individual factors on future urban expansion, we applied all six IFB-DE-CA models to predict future scenarios of Hangzhou to the year 2030.

Our models show that an individual factor can be applied to establish a CA model, which generated defensible simulation results. Among the factors, DEM and POP have the highest impact on urban expansion at Hangzhou from 2005 to 2015. Models incorporating two factors produced more accurate simulations when compared with models using other factors. While earlier work suggests including more significant driving factors into urban CA models (Engelen 2002; Wahyudi and Liu 2013), our findings show that MFB-DE-CA models are not necessarily superior to IFB-DE-CA models. The differences between the IFB-DE-CA and MFB-DE-CA models exist in both the simulated accuracy and spatial patterns. IFB-DE-CA models 1–2 performed better than the MFB-DE-CA models 19 and 21, which bettered the IFB-DE-CA models 3–6 by ~1%. This indicates that a CA model using only the most influential factor (e.g., DEM) performs better than a model using a less significant factor (e.g., RAIL) only and may perform better than models that include both the most influential factor and other less significant factors (e.g., DEM-POP-COUN-RAIL-CITY). Our findings confirmed the results of Feng and Tong (2017b), who showed that model performance can be reduced when too many factors are included.

CA models are built based on the land conversion probability defined by driving factors and the combined effects of other model elements. The definition of these elements and their processing may substantially affect the simulation results (Poelmans and Van Rompaey 2010; Sang et al. 2011). Possible issues include neighborhood configuration, constraints, spatial resolution, sampling methods, and particularly the algorithms applied to define land conversion probability (Feng and Tong 2017a). In this study, all DE-CA models applied the same settings except the factors included in the transition rules. As a result, the differences among the models in this research can be attributed mostly to the different impacts of the driving factors.

Driving factors are approximations of real geographic and socioeconomic elements, and the visualization of these elements may lead to the loss of some spatial detail (Goodchild et al. 1992). Different categories of factors have different effects on urban expansion. Factors belonging to the same category may generate different impacts on urban expansion, because they represent different geographical aspects. When more driving factors are included in DE-CA models, the interaction among factors does not necessarily lead to improvements in the simulation results and may even reduce the simulation accuracy. Applying more driving factors probably leads to multicollinearity (Feng and Tong 2017b), introducing more geographical data errors that propagate through data processing and model implementation and lead to negative impacts on modeling.

The selection of appropriate driving factors is crucial in CA modeling, which can produce more realistic simulation results by taking into account the most influential factors. GAM is suitable to quantify the impacts of factors using the fitting statistics and to select the appropriate combination of factors. Our study is a good example of how to select the most influential driving factors among the candidates for calibrating CA models. Future urban scenarios from various IFB-DE-CA modeling differ because each scenario emphasizes the different impacts of each factor. For a *comprehensive prediction* of urban expansion, IFB-DE-CA models may not be the best choice because cities are complex systems affected by numerous factors. However, IFB-DE-CA models can inform urban modelers and policy-makers about potential urban scenarios when considering each factor separately. These scenario predictions can be early warnings of the consequences of different urban development schemes and can help policy-makers to adjust and avoid unfavorable future urban scenarios.

Conclusions

We used a stepwise GAM to examine the rank orders of candidate driving factors and identify 21 combinations of these factors, which were applied to calibrate 21 DE-CA models. Among these models, six IFB-DE-CA models and five MFB-DE-CA models were applied to simulate urban expansion at Hangzhou from 2005 to 2015, and all these models produced defensible 2015 simulations with overall accuracies in excess of 89%. We finally applied all six IFB-DE-CA models to project future urban scenarios for the year 2030.

We concluded that (1) a CA model can be constructed using only one driving factor, (2) IFB-DE-CA models may outperform MFB-DE-CA models in simulating present urban patterns, and (3) the major differences between the two types of models are their projections of future urban scenarios. Each IFB-DE-CA model produces a very different future scenario that is shaped by the corresponding factor, informing the modelers and policy-makers about how cities will be formed if the corresponding factors are individually applied. This improves our understanding of the effects of driving factors on urban dynamics and their impacts on CA models when incorporating them into the models separately. In contrast, the MFB-DE-CA models take into account several factors simultaneously and allocate new urban cells where the combined effect of all factors is higher than the threshold. Because cities are complex systems, we suggest examining the significance and impact of each candidate factor using GAM to determine whether it should be included in the model.

Funding information This study was supported by the National Key R&D Program of China (2018YFB0505400 and 2018YFB0505402) and the National Natural Science Foundation of China (41771414 and 41631178).

References

- Ahmed, S. J., Bramley, G., & Verburg, P. H. (2014). Key driving factors influencing urban growth: spatial-statistical modelling with clue-s. In *Dhaka megacity* (pp. 123–145). Berlin: Springer.
- Akın, A., Sunar, F., & Berberoğlu, S. (2015). Urban change analysis and future growth of Istanbul. *Environmental Monitoring and Assessment*, 187(8), 506. <https://doi.org/10.1007/s10661-015-4721-1>.
- Aldwaik, S. Z., & Pontius, R., Jr. (2012). Intensity analysis to unify measurements of size and stationarity of land changes by interval, category, and transition. *Landscape and Urban Planning*, 106(1), 103–114. <https://doi.org/10.1016/j.landurbplan.2012.02.010>.
- Brown, D. G., & Goovaerts, P. (2002). Stochastic simulation of land-cover change using geostatistics and generalized additive models. *Photogrammetric Engineering & Remote Sensing*, 68(10), 1051–1062.
- Chaudhuri, G., & Clarke, K. C. (2013). How does land use policy modify urban growth? A case study of the Italo-Slovenian border. *Journal of Land Use Science*, 8(4), 443–465. <https://doi.org/10.1080/1747423X.2012.679748>.
- Dang, A. N., & Kawasaki, A. (2017). Integrating biophysical and socio-economic factors for land-use and land-cover change projection in agricultural economic regions. *Ecological Modelling*, 344, 29–37. <https://doi.org/10.1016/j.ecolmodel.2016.11.004>.
- Das, S., Abraham, A., Chakraborty, U. K., & Konar, A. (2009). Differential evolution using a neighborhood-based mutation operator. *IEEE Transactions on Evolutionary Computation*, 13(3), 526–553. <https://doi.org/10.1109/TEVC.2008.2009457>.
- Delden, H. V., Stuczynski, T., Ciaian, P., et al. (2010). Integrated assessment of agricultural policies with dynamic land use change modelling. *Ecological Modelling*, 221(18), 2153–2166. <https://doi.org/10.1016/j.ecolmodel.2010.03.023>.
- Deng, X., Huang, J., Rozelle, S., Zhang, J., & Li, Z. (2015). Impact of urbanization on cultivated land changes in China. *Land Use Policy*, 45, 1–7. <https://doi.org/10.1016/j.landusepol.2015.01.007>.
- Dubovyk, O., Sliuzas, R., & Flacke, J. (2011). Spatio-temporal modelling of informal settlement development in Sancaktepe district, Istanbul, Turkey. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(2), 235–246. <https://doi.org/10.1016/j.isprsjprs.2010.10.002>.
- Engelen, G. (2002). *Cellular automata for modelling land use change as driven by socio-economic, environmental and policy factors I*. Research Institute for Knowledge System (RIKS).
- Feng, Y., & Tong, X. (2017a). Calibrating nonparametric cellular automata with a generalized additive model to simulate dynamic urban growth. *Environmental Earth Sciences*, 76(14), 496. <https://doi.org/10.1007/s12665-017-6828-x>.
- Feng, Y., & Tong, X. (2017b). Using exploratory regression to identify optimal driving factors for cellular automaton modeling of land use change. *Environmental Monitoring and Assessment*, 189(10), 515. <https://doi.org/10.1007/s10661-017-6224-8>.
- Feng, Y., & Tong, X. (2018a). Calibration of cellular automata models using differential evolution to simulate present and future land use. *Transactions in GIS*, 22, 582–601. <https://doi.org/10.1111/tgis.12331>.
- Feng, Y., & Tong, X. (2018b). Dynamic land use change simulation using cellular automata with spatially nonstationary transition rules. *GIScience & Remote Sensing*, 55(5), 678–698. <https://doi.org/10.1080/15481603.2018.1426262>.
- Feng, Y., Liu, Y., & Tong, X. (2018). Spatiotemporal variation of landscape patterns and their spatial determinants in Shanghai, China. *Ecological Indicators*, 87, 22–32. <https://doi.org/10.1016/j.ecolind.2017.12.034>.
- González, P. B., Aguilera-Benavente, F., & Gómez-Delgado, M. (2015). Partial validation of cellular automata based model

- simulations of urban growth: an approach to assessing factor influence using spatial methods. *Environmental Modelling & Software*, 69, 77–89. <https://doi.org/10.1016/j.envsoft.2015.03.008>.
- Goodchild, M. F., Guoqing, S., & Shiren, Y. (1992). Development and test of an error model for categorical data. *International Journal of Geographical Information Systems*, 6(2), 87–103. <https://doi.org/10.1080/02693799208901898>.
- Guisan, A., Jr., E, T. C., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, 157(2), 89–100. [https://doi.org/10.1016/S0304-3800\(02\)00204-1](https://doi.org/10.1016/S0304-3800(02)00204-1).
- Haase, D., Haase, A., Kabisch, N., Kabisch, S., & Rink, D. (2012). Actors and factors in land-use simulation: The challenge of urban shrinkage. *Environmental Modelling & Software*, 35, 92–103. <https://doi.org/10.1016/j.envsoft.2012.02.012>.
- Hou, H., Wang, R., & Murayama, Y. (2019). Scenario-based modelling for urban sustainability focusing on changes in cropland under rapid urbanization: a case study of Hangzhou from 1990 to 2035. *Science of the Total Environment*, 661, 422–431. <https://doi.org/10.1016/j.scitotenv.2019.01.208>.
- Kamusoko, C., & Gamba, J. (2015). Simulating urban growth using a random forest-cellular automata (RF-CA) model. *ISPRS International Journal of Geo-Information*, 4(2), 447–470. <https://doi.org/10.3390/ijgi4020447>.
- Lambin, E. F., Turner, B. L., Geist, H. J., Agbola, S. B., Angelsen, A., Bruce, J. W., Coomes, O. T., Dirzo, R., Fischer, G., Folke, C., George, P. S., Homewood, K., Imbernon, J., Leemans, R., Li, X., Moran, E. F., Mortimore, M., Ramakrishnan, P. S., Richards, J. F., Skánes, H., Steffen, W., Stone, G. D., Svedin, U., Veldkamp, T. A., Vogel, C., & Xu, J. (2001). The causes of land-use and land-cover change: moving beyond the myths. *Global Environmental Change*, 11(4), 261–269. [https://doi.org/10.1016/S0959-3780\(01\)00007-3](https://doi.org/10.1016/S0959-3780(01)00007-3).
- Larsen, K. (2015). *GAM: the predictive modeling silver bullet. Multithreaded*. San Francisco: Stitch Fix 30.
- Li, X., & Yeh, A. G. O. (2002). Urban simulation using principal components analysis and cellular automata for land-use planning. *Photogrammetric Engineering & Remote Sensing*, 68(4), 341–352.
- Li, X., Lao, C., Liu, X., & Chen, Y. (2011). Coupling urban cellular automata with ant colony optimization for zoning protected natural areas under a changing landscape. *International Journal of Geographical Information Science*, 25(4), 575–593. <https://doi.org/10.1080/13658816.2010.481262>.
- Liu, X., Liang, X., Li, X., Xu, X., Ou, J., Chen, Y., Li, S., Wang, S., & Pei, F. (2017). A future land use simulation model (FLUS) for simulating multiple land use scenarios by coupling human and natural effects. *Landscape and Urban Planning*, 168, 94–116. <https://doi.org/10.1016/j.landurbplan.2017.09.019>.
- Liu, X., Hu, G., Ai, B., Li, X., Tian, G., Chen, Y., & Li, S. (2018). Simulating urban dynamics in China using a gradient cellular automata model based on S-shaped curve evolution characteristics. *International Journal of Geographical Information Science*, 32(1), 73–101. <https://doi.org/10.1080/13658816.2017.1376065>.
- Mahiny, A. S., & Clarke, K. C. (2012). Guiding SLEUTH land-use/land-cover change modeling using multicriteria evaluation: towards dynamic sustainable land-use planning. *Environment and Planning B: Planning & Design*, 39(5), 925–944. <https://doi.org/10.1068/b37092>.
- Mcgarigal, K. (2014). *Fragstats v4: spatial pattern analysis program for categorical and continuous maps-help manual*. Amhers: University of Massachusetts <http://www.umass.edu/landeco/research/fragstats/fragstats.html>. Accessed 21 May 2018.
- Mitsova, D., Shuster, W., & Wang, X. (2011). A cellular automata model of land cover change to integrate urban growth with open space conservation. *Landscape and Urban Planning*, 99(2), 141–153. <https://doi.org/10.1016/j.landurbplan.2010.10.001>.
- Mondal, B., Das, D. N., & Dolui, G. (2015). Modeling spatial variation of explanatory factors of urban expansion of Kolkata: a geographically weighted regression approach. *Modeling Earth Systems and Environment*, 1(4), 29. <https://doi.org/10.1007/s40808-015-0026-1>.
- Musa, S. I., Hashim, M., & Reba, M. N. M. (2017). A review of geospatial-based urban growth models and modelling initiatives. *Geocarto International*, 32(8), 813–833. <https://doi.org/10.1080/10106049.2016.1213891>.
- Mustafa, A., Heppenstall, A., Omrani, H., Saadi, I., Cools, M., & Teller, J. (2018). Modelling built-up expansion and densification with multinomial logistic regression, cellular automata and genetic algorithm. *Computers, Environment and Urban Systems*, 67, 147–156. <https://doi.org/10.1016/j.compenvurbsys.2017.09.009>.
- NDRC. (2016). *Urban agglomeration development planning of the Yangtze River Delta*, National Development and Reform Commission. <http://www.ndrc.gov.cn/fzgggz/fzgh/zcfg/201606/W020160715547746056307.pdf>. Accessed 21 May 2018.
- Niesterowicz, J., & Stepinski, T. F. (2016). On using landscape metrics for landscape similarity search. *Ecological Indicators*, 64, 20–30. <https://doi.org/10.1016/j.ecolind.2015.12.027>.
- Noroozi, V., Hashemi, A. B., & Meybodi, M. R. (2011). CellularDE: a cellular based differential evolution for dynamic optimization problems. In *International Conference on Adaptive and Natural Computing Algorithms* (p. 340–349).
- Osman, T., Divigalpitiya, P., & Arima, T. (2016). Driving factors of urban sprawl in Giza governorate of greater Cairo metropolitan region using ahp method. *Land Use Policy*, 58, 21–31. <https://doi.org/10.1016/j.landusepol.2016.07.013>.
- Pan, Y., Roth, A., Yu, Z., & Doluschitz, R. (2010). The impact of variation in scale on the behavior of a cellular automata used for land use change modeling. *Computers, Environment and Urban Systems*, 34(5), 400–408. <https://doi.org/10.1016/j.compenvurbsys.2010.03.003>.
- Peña, J., Bonet, A., Bellot, J., & Sánchez, J. R. (2005). Trends and driving factors in land use changes (1956–2000) in Marina Baixa, SE Spain.
- Poelmans, L., & Rompaey, A. V. (2009). Detecting and modelling spatial patterns of urban sprawl in highly fragmented areas: a case study in the Flanders–Brussels region. *Landscape and Urban Planning*, 93(1), 10–19. <https://doi.org/10.1016/j.landurbplan.2009.05.018>.
- Poelmans, L., & Van Rompaey, A. (2010). Complexity and performance of urban expansion models. *Computers, Environment and Urban Systems*, 34(1), 17–27. <https://doi.org/10.1016/j.compenvurbsys.2009.06.001>.

- Pontius, R. G., & Millones, M. (2011). Death to kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, 32(15), 4407–4429. <https://doi.org/10.1080/01431161.2011.552923>.
- Pontius, R. G., Gao, Y., Giner, N. M., Kohyama, T., Osaki, M., & Hirose, K. (2013). Design and interpretation of intensity analysis illustrated by land change in Central Kalimantan, Indonesia. *Land*, 2(3), 351–369. <https://doi.org/10.3390/land2030351>.
- Reilly, M. K., O'Mara, M. P., & Seto, K. C. (2009). From Bangalore to the Bay Area: comparing transportation and activity accessibility as drivers of urban growth. *Landscape and Urban Planning*, 92(1), 24–33. <https://doi.org/10.1016/j.landurbplan.2009.02.001>.
- Rienow, A., & Goetzke, R. (2015). Supporting SLEUTH—enhancing a cellular automaton with support vector machines for urban growth modeling. *Computers, Environment and Urban Systems*, 49, 66–81. <https://doi.org/10.1016/j.compenvurbsys.2014.05.001>.
- Sang, L., Zhang, C., Yang, J., Zhu, D., & Yun, W. (2011). Simulation of land use spatial pattern of towns and villages based on CA-Markov model. *Mathematical and Computer Modelling*, 54(3–4), 938–943. <https://doi.org/10.1016/j.mcm.2010.11.019>.
- Shafizadeh-Moghadam, H., & Helbich, M. (2015). Spatiotemporal variability of urban growth factors: A global and local perspective on the megacity of Mumbai. *International Journal of Applied Earth Observation and Geoinformation*, 35, 187–198. <https://doi.org/10.1016/j.jag.2014.08.013>.
- Smidt, S. J., Tayyebi, A., Kendall, A. D., Pijanowski, B. C., & Hyndman, D. W. (2018). Agricultural implications of providing soil-based constraints on urban expansion: land use forecasts to 2050. *Journal of Environmental Management*, 217, 677–689. <https://doi.org/10.1016/j.jenvman.2018.03.042>.
- Soares-Filho, B., Rodrigues, H., & Follador, M. (2013). A hybrid analytical-heuristic method for calibrating land-use change models. *Environmental Modelling & Software*, 43, 80–87. <https://doi.org/10.1016/j.envsoft.2013.01.010>.
- Sun, X., Crittenden, J. C., Li, F., Lu, Z., & Dou, X. (2018). Urban expansion simulation and the spatio-temporal changes of ecosystem services, a case study in Atlanta metropolitan area, USA. *Science of the Total Environment*, 622, 974–987. <https://doi.org/10.1016/j.scitotenv.2017.12.062>.
- Von Thaden, J. J., Laborde, J., Guevara, S., & Venegas-Barrera, C. S. (2018). Forest cover change in the Los Tuxtlas biosphere reserve and its future: the contribution of the 1998 protected natural area decree. *Land Use Policy*, 72, 443–450. <https://doi.org/10.1016/j.landusepol.2017.12.040>.
- Wahyudi, A., & Liu, Y. (2013). Cellular automata for urban growth modeling: a chronological review on factors in transition rules. In *13th International conference on computers in urban planning and urban Management (CUPUM 2013)*.
- Wang, F., Hasbani, J.-G., Wang, X., & Marceau, D. J. (2011). Identifying dominant factors for the calibration of a land-use cellular automata model using rough set theory. *Computers, Environment and Urban Systems*, 35(2), 116–125. <https://doi.org/10.1016/j.compenvurbsys.2010.10.003>.
- Whitsed, R., & Smallbone, L. T. (2017). A hybrid genetic algorithm with local optimiser improves calibration of a vegetation change cellular automata model. *International Journal of Geographical Information Science*, 31(4), 717–737. <https://doi.org/10.1080/13658816.2016.1231315>.
- Wu, F. (1998). SimLand: a prototype to simulate land conversion through the integrated GIS and CA with AHP-derived transition rules. *International Journal of Geographical Information Science*, 12(1), 63–82. <https://doi.org/10.1080/136588198242012>.
- Zadbagher, E., Becek, K., & Berberoglu, S. (2018). Modeling land use/land cover change using remote sensing and geographic information systems: case study of the Seyhan Basin, Turkey. *Environmental Monitoring and Assessment*, 190(8), 494. <https://doi.org/10.1007/s10661-018-6877-y>.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.