J. Geogr. Sci. 2022, 32(3): 499-516
DOI: https://doi.org/10.1007/s11442-022-1958-9

Model construction of urban agglomeration expansion simulation considering urban flow and hierarchical characteristics

WANG Haijun[1,2], WU Yue[1], *DENG Yu[3], XU Shan[4]

1. School of Resource and Environmental Science, Wuhan University, Wuhan 430079, China;
2. Key Laboratory of Geographic Information Systems, Ministry of Education, Wuhan University, Wuhan 430079, China;
3. Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China;
4. College of Land Science and Technology, China Agricultural University, Beijing 100193, China

Abstract: Since the launch of China's reform and opening up policy, the process of urbanization in China has accelerated significantly. With the development of cities, inter-city interactions have become increasingly close, forming urban agglomerations that tend to be integrated. Urban agglomerations are regional spaces with network relationships and hierarchies, and have always been the main units for China to promote urbanization and regional coordinated development. In this paper, we comprehensively consider the network and hierarchical characteristics of an urban agglomeration, while using urban flow to describe the interactions of the inter-city networks and the hierarchical generalized linear model (HGLM) to reveal the hierarchical driving mechanism of the urban agglomeration. By coupling the HGLM with a cellular automata (CA) model, we introduced the HGLM-CA model for the simulation of the spatial expansion of an urban agglomeration, and compared the simulation results with those of the logistic-CA model and the biogeography-based optimization CA (BBO-CA) model. According to the results, we further analyzed the advantages and disadvantages of the proposed HGLM-CA model. We selected the middle reaches of the Yangtze River in China as the research area to conduct this empirical research, and simulated the spatial expansion of the urban agglomeration in 2017 on the basis of urban land-use data from 2007 and 2012. The results indicate that the spatial expansion of the urban agglomeration can be attributed to various driving factors. As a driving factor at the urban level, urban flow promotes the evolution of land use in the urban agglomeration, and also plays an important role in regulating cell-level factors, making the cell-level factors of different cities show different driving effects. The HGLM-CA model is able to obtain a higher simulation accuracy than the logistic-CA model, which indicates that the simulation results for urban agglomeration expansion considering urban flow and hierarchical characteristics are more accurate. When compared with the intelligent algorithm model, i.e., BBO-CA, the HGLM-CA model obtains a lower simulation

Received: 2021-08-17 **Accepted:** 2022-01-05
Foundation: National Natural Science Foundation of China, No.42171411; Youth Innovation Promotion Association, CAS, No.2019055
Author: Wang Haijun (1972–), PhD, specialized in geographic simulation, territorial spatial planning and land resource evaluation research. E-mail: landgiswhj@163.com
***Corresponding author:** Deng Yu (1985–), PhD, E-mail: dengy@igsnrr.ac.cn
This paper is initially published in *Acta Geographica Sinica* (Chinese edition), 2021, 76(12): 3012–3024.

accuracy, but it can analyze the interaction of the various driving factors from a hierarchical perspective. It also has a strong explanatory effect for the spatial expansion mechanism of urban agglomerations.

Keywords: urban flow; hierarchical characteristics; cellular automata; driving mechanism; spatial expansion; urban agglomeration; middle reaches of the Yangtze River

1 Introduction

Since the implementation of China's reform and opening up policy in 1978, the process of urbanization in China has accelerated significantly. With the continuous strengthening of economic, population, transportation, and information links, inter-city interactions are becoming closer (Wang *et al*., 2016). By taking cities as "nodes" and the inter-city interactions as "lines", it is possible to generate a network structure across urban spaces, representing an urban agglomeration that tends to be integrated (Fang *et al*., 2014). The Yangtze River Delta, Pearl River Delta, and Beijing-Tianjin-Hebei agglomerations, and 19 other urban agglomerations approved as part of the "13th Five-Year Plan" (2016–2020), contain 73.63% of China's population and create 90.87% of the country's GDP from only 32.67% of the land (Zhang, 2020). Urban agglomerations have thus become an important part of promoting new urbanization and leading regional development, and have become increasingly prominent in China's economic and social development pattern. The "14th Five-Year Plan" (2021−2025) clearly states that urban agglomerations should be the main body to promote coordinated regional development and new urbanization. Therefore, studying the spatial expansion of urban agglomerations is of great significance for promoting regional coordinated development and cultivating modern metropolitan areas.

Simulation and prediction are important components when studying the spatial expansion of urban agglomerations, and they are of both theoretical and practical value when identifying the development trends of urban agglomerations and planning the future patterns of urban agglomerations. The existing research has generally focused on the simulation and prediction of urban spatial expansion, for which the cellular automata (CA) model is an important method for simulating urban expansion (Wu *et al*., 2019). In the 1970s, Tobler (1970) discovered the advantages of the CA model in solving geographic problems, and used it to simulate the urban expansion of Detroit in the U.S. Through in-depth investigation of geographic CA theory (Couclelis, 1985; 1989), the CA model has been further developed and improved, and models such as CA-Markov (Luijten, 2003) and Logistic-CA (Wu *et al*., 1997) have been widely used. However, most of the existing studies have focused on the coupling of the CA model and intelligent algorithms, and have usually combined the CA model with other models such as multi-agent systems (MAS) (Yang *et al*., 2007), artificial neural networks (ANNs) (Li *et al*., 2002; Xie *et al*., 2020), the maximum entropy (MaxEnt) model (Wang *et al*., 2020; Zhang *et al*., 2020), and biogeography-based optimization (BBO) (Wang *et al*., 2017). This was done to make up for the defects of the CA model and make the simulation effect closer to real urban expansion. The above-mentioned studies have usually aimed to improve the existing methods, with the goal of mining cell transformation rules and improving the model simulation accuracy. Furthermore, most of these studies have focused on the expansion of a single city space, and have ignored the impact of inter-city interaction. In addition, although the combination with artificial intelligence algorithms can improve the simulation accuracy, the intelligent algorithms are mostly black box models, and they lack the ability to analyze the driving mechanism of urban expansion.

As the highest spatial organization form of urban development in the mature stage, urban networks (Wu *et al.*, 2015) and urban hierarchies (Zhang *et al.*, 2020) coexist in urban agglomerations. Therefore, the simulation of the spatial expansion of an urban agglomeration differs from that for a single city. It is therefore necessary to consider the network interaction between cities and the hierarchical nature of urban agglomerations. In an urban spatial expansion model, urban network interaction usually refers to the flow of various factors such as economy and population between cities, i.e., the "urban flow", which is an important form of interaction and connection between cities. The existing studies have mainly taken the impact of urban interaction into account from the following two aspects. The first aspect is to improve the traditional CA model by characterizing the city flow. For example, He *et al*. (2016) quantified the city flow and embedded it as a conversion rule in the CA model to simulate the sprawl of urban agglomerations, and found that the simulation accuracy was improved after considering the urban flow. He *et al*. (2017) combined urban flow and a gravity model to simulate and predict the expansion of the Wuhan urban agglomeration, and concluded that urban flow has a significant impact on land-use changes in urban agglomerations. Xia *et al*. (2019) introduced bidirectional flow to improve the traditional gravity model, and constructed a model from the three perspectives of "macro", "meso", and "micro" (Xia *et al*., 2020), confirming that urban flow is an important driving factor of urban agglomeration, and that simulation results considering urban flow are usually more accurate. The second aspect is to combine an interactive model to reflect the inter-city long-distance connections and predict the urban land-use demand in different regions, while integrating the interactive model into the CA conversion rules. For example, Chen *et al*. (2019) incorporated the multi-region input-output (MRIO) model to describe the urban flow of land elements, and simulated nationwide urban expansion on the basis of estimating the relationship between urban land supply and demand, obtaining good simulation results. It can therefore be seen that, when simulating the spatial expansion of large-scale areas such as urban agglomerations, urban flow is one of the factors that cannot be ignored. The hierarchical characteristics of cities have also gradually attracted the attention of the academic community, but different people have different opinions on the understanding of urban hierarchies. The related research has also been based on different perspectives. For example, Sun *et al*. (2020) considered multi-level planning management and control, and built a multi-level vector CA model based on the hierarchical relationships in the urban planning system, which was used to simulate the land-use changes in the city of Jiangyin, China. Shu *et al*. (2020) considered the hierarchical characteristics of the land-use system and divided it into two levels–the cell and the township–and built a multi-level logistic–CA model to simulate the urban expansion of Tongshan District in the city of Xuzhou, China. The above-mentioned studies considering urban hierarchical characteristics have usually obtained superior simulation results. However, the existing studies are generally small in spatial scale, and most of them are limited to the simulation of spatial clusters within the city scale, without paying attention to the hierarchical problem of urban agglomeration expansion. The urban agglomeration system is formed by the agglomeration of multiple cities, with a wide range of space and frequent flow of internal elements. It is a complex system with multiple cores and multiple levels. Therefore, the role of urban flow and hierarchy at the scale of urban agglomerations is important, and they are important factors that cannot be ignored in the study of the spatial expansion of urban agglomerations.

In this paper, we consider the influence of urban network interaction when studying the spatial expansion of an urban agglomeration, and use urban flow to describe the interactions within the urban agglomeration. At the same time, taking into account the hierarchical characteristics of the urban agglomeration, the hierarchical generalized linear model (HGLM) is used to study the hierarchical driving mechanism of the urban agglomeration evolution, and the HGLM-CA model is introduced to simulate the spatial expansion of the urban agglomeration. Taking the urban agglomeration in the middle reaches of the Yangtze River in China as the study area, an empirical study was carried out to explain the driving mechanism of the spatial expansion of the urban agglomeration. The simulation results are then compared in this paper with the results of the logistic-CA model and BBO-CA model, to show the advantages and disadvantages of the HGLM-CA model.

2 Methodology

2.1 The hierarchical generalized linear model (HGLM)

The HGLM was developed from the hierarchical linear model (HLM). The dependent variable of the HLM is limited to continuous data. The HGLM improves on this basis and can handle binary dependent variables. The HLM model is a statistical analysis method that considers both overall factors and individual factors, and can process data with hierarchical characteristics. Traditional logistic regression can only consider factors at a single level, and usually ignores the level differences. The high-level factors are decomposed into lower levels for research, so that samples obtained from the same group are relevant and do not satisfy the independence assumption. The HLM overcomes the shortcomings of traditional logistic regression, and aggregates the parameters obtained after individual-level regression analysis, and then performs regression analysis again at the overall level, which is also called “regression of regression”, which can effectively process hierarchical feature data (Li *et al.*, 2006). However, in practical applications, the HLM models have certain requirements for the number of groups of samples (Hox, 1999), and too few groups may cause the model to fail to converge.

Multi-level structure data are ubiquitous (Lei *et al.*, 2002). In the problem of the spatial expansion of an urban agglomeration, the urban agglomeration has hierarchical characteristics. The land-use unit (cell) is embedded in the city, and the cell transition probability varies with the city where it is located. Therefore, the independent variables used to describe cellular characteristics are the individual variables at a lower level (level 1), while the independent variables used to describe urban characteristics are the group variables at a higher level (level 2), and group variables within the same city have the same value. In this paper, the urban agglomeration is divided into a cell level (level 1) and a city level (level 2). The formulas are as follows:

Level 1:

$$P_{Sij}\left(y_{ij}=1\right)=\varphi_{ij} \tag{1}$$

$$log\left[\varphi_{ij} / \left(1-\varphi_{ij}\right)\right] = \eta_{ij} \quad (2)$$

$$\eta_{ij} = \beta_{0j} + \sum_{n=1}^{k} \beta_{nj} X_{nij} \quad (3)$$

Level 2:

$$\beta_{0j} = \gamma_{00} + \sum_{m=1}^{l} \gamma_{0m} W_{mj} + \mu_{0j} \quad (4)$$

$$\beta_{nj} = \gamma_{n0} + \sum_{m=1}^{l} \gamma_{nm} W_{mj} + \mu_{nj} \quad (5)$$

where $P_{Sij}(y_{ij}=1)$ is the suitability of the cell with regard to being transformed into urban land, η_{ij} is the vector describing the state transition of the cell, and X_{nij} denotes the nth independent variable of cell i located in city j in the cell level (level 1). k is the number of independent variables in the cell level, β_{0j} is a random intercept, β_{nj} is the regression coefficient of X_{nij}, W_{mj} is the mth independent variable of city j in the city level (level 2), and l is the self-independent variable in the city level. The number of variables, γ_{00} and γ_{n0}, are the intercepts of β_{0j} and β_{nj}, respectively. γ_{0m} and γ_{nm} are the regression coefficients of W_{mj}, and μ_{0j} and μ_{nj} are residual items.

2.2 The hierarchical generalized linear model-cellular automata (HGLM-CA) model

2.2.1 Obtaining weight parameters based on the HGLM

Establishing a null model is the first step in applying the HGLM. The purpose is to determine whether the data are suitable for stratified research by calculating the intraclass correlation coefficients (ICCs). The ICC is the ratio of the variance between groups to the total variance. The larger the ICC value, the greater the influence of the difference between the groups on the dependent variable, i.e., the greater the probability of the cell state transition being affected by the city-level factors, and the more the city-level factors cannot be ignored. Generally speaking, ICC<0.059 is considered as low intra-group correlation, 0.059⩽ ICC<0.138 is moderate intra-group correlation, and ICC⩾0.138 is high intra-group correlation. The ICC can be calculated as follows:

$$ICC = \frac{\tau_{00}}{(\tau_{00} + \sigma^2)} \quad (6)$$

where τ_{00} is the variance between groups and σ^2 is the intraclass variance.

The cell-level and city-level variables are added into the empty model to obtain the random covariate model and the random intercept model, which can verify the significance of the influence of the cell-level and city-level variables on the dependent variable and the explanation degree of the variance. Two levels of variables are added into the model to build the full model, and the weight parameters of each variable and the relationship between the variables are then obtained (Figure 1).

2.2.2 CA transition rules for urban agglomerations

Defining transition rules is the core of a CA model (Li *et al.*, 2007). In this paper, we com-

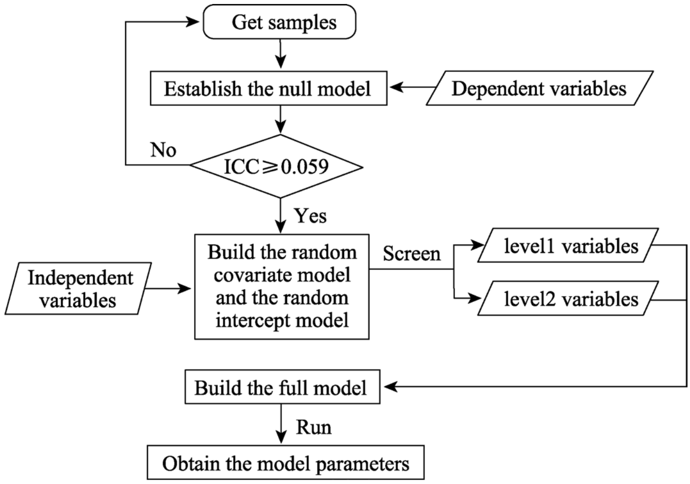


Figure 1 Flowchart of parameter acquisition based on the HGLM

prehensively consider the influence of the cell state transition suitability, the constraint conditions, and the neighborhood effects to construct the CA model transition rules. Since the follow-up experiments involved the accuracy comparison of different models, the conversion rule does not include stochastic perturbation, for the time being, which can be expressed as:

$$P_{ij} = P_{Sij} \times con\left(S_{ij}\right) \times \Omega_{ij} \tag{7}$$

where P_{ij} is the cell transition probability and P_{Sij} is the cell state transition suitability, as obtained by the HGLM:

$$P_{Sij} = \frac{1}{1+\exp(-\eta_{ij})} \tag{8}$$

$$\eta_{ij} = \gamma_{00} + \sum_{m=1}^{1}\gamma_{0m}W_{mj} + \sum_{n=1}^{k}\gamma_{n0}X_{nij} + \sum_{n=1}^{k}\sum_{m=1}^{1}\gamma_{nm}W_{mj}X_{nij} + \mu_{0j} + \sum_{n=1}^{k}\mu_{nj}X_{nij} \tag{9}$$

where $con(S_{ij})$ is the constraint condition. The areas that are not allowed to be developed (such as water areas) are assigned a value of 0, and the others are assigned a value of 1.

Ω_{ij} is the effect of the neighborhood. In this study, a 3×3 Moore neighborhood was selected for the research. The influence of the neighborhood on the cell transformation is as follows:

$$\Omega_{ij} = \frac{\sum con\left(S_{ij}=1\right)}{3\times 3-1} \tag{10}$$

A conversion threshold is set to determine whether the cell state will change. According to the number of cell transformations and the number of iterations, the number of cells that needs to be transformed in each iteration is obtained. We calculate the cell transition probability P, and arrange the cells that can be transformed according to the probability. The cell probability corresponding to the number of iterations is selected as the transition threshold $P_{threshold}$ to determine whether the cell state has changed. The state of the cell at the next moment is then:

$$S_{ij}^{t+1} = \begin{cases} 1, P \geqslant P_{threshold} \\ 0, P < P_{threshold} \end{cases} \tag{11}$$

In this paper, the intensity of the urban flow is used to describe the interaction between cities, and the HGLM is used to analyze the hierarchical driving mechanism of the spatial expansion of the urban agglomeration. The weights of the driving factors at different levels and their mutual relations are obtained as parameters of the CA model, and the HGLM-CA model is constructed to simulate the spatial expansion of the urban agglomeration (Figure 2).

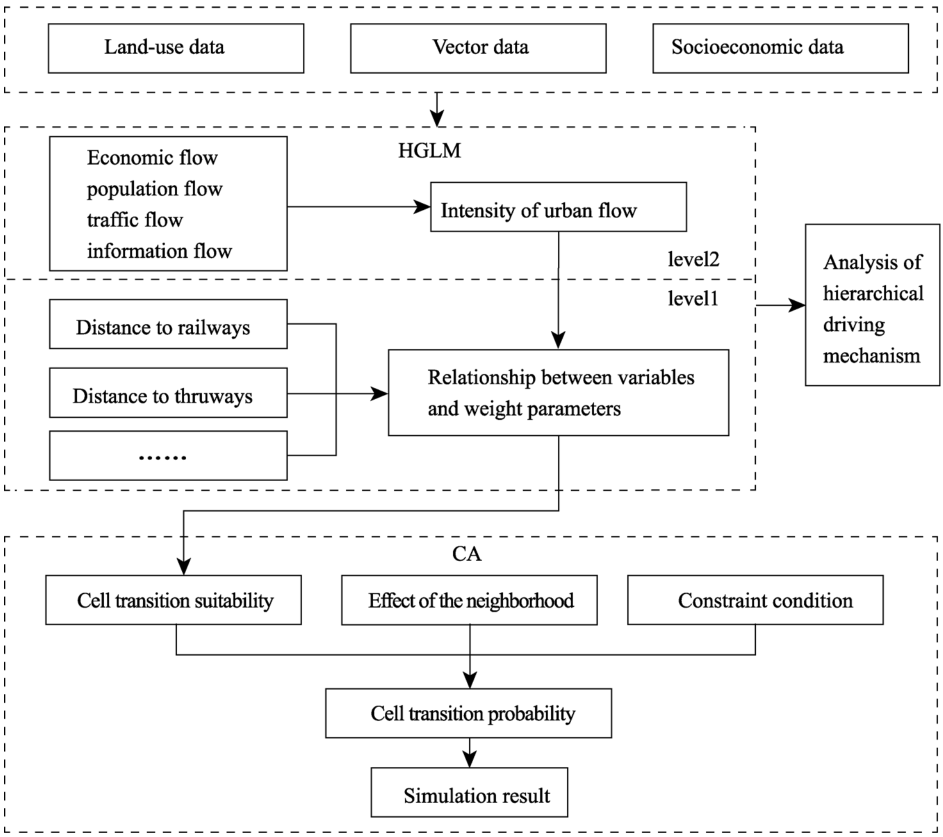


Figure 2 Framework of the HGLM-CA model

3 Implementation and results

3.1 Study area and data sources

In this study, the urban agglomeration in the middle reaches of the Yangtze River was selected as the study area, covering 31 cities in total, including Wuhan, Changsha, and Nanchang (Figure 3). The urban agglomeration in the middle reaches of the Yangtze River is a cross-regional national-level urban agglomeration that has been approved by the State Council of the People's Republic of China. It is a key area for coordinating regional development and promoting new urbanization. The urban agglomeration in the middle reaches of the Yangtze River covers a wide range of land, and there are differences in the economic strengths and urbanization levels of the cities. Wuhan had a permanent urban population of more than 9 million in 2019, with an urbanization rate of more than 80% (WSB, 2020), and has become a core city in central China. The permanent resident population and per capita GDP of Changsha and Nanchang are in the forefront, while the comprehensive strength of cities such as Huanggang and Yiyang is still weak. It can therefore be seen that the development situations of the cities in the middle reaches of the Yangtze River are significantly different.

Figure 3 Location of the study area (the urban agglomeration in the middle reaches of the Yangtze River)

The data required for the research included land-use data from 2007, 2012, and 2017, road data, and urban flow data for the city cluster in the middle reaches of the Yangtze River (Table 1).

Table 1 The sources of data

Data name	Data specification	Data source
Land-use data	The impervious surface data of urban areas are obtained by using the reliable impervious surface mapping algorithms and GEE platform with a resolution of 30 m×30 m. The impervious surface is regarded as urban land, while the others are non-urban land, which is resampled to 90 m ×90 m	Published by Gong *et al*. (2019), Tsinghua University (http://data.ess.tsinghua.edu.cn/)
Road data	Shapefile data including railways, thruways and national highways	Resource and Environment Science and Data Center (http://www.resdc.cn/)
DEM	Based on the latest SRTM V4.1 data after collation and stitching, the resolution is 90 m×90 m	Resource and Environment Science and Data Center (http://www.resdc.cn/)
Urban flow data	According to statistical yearbook data and big data of spatio-temporal geography, models of economic flow, population flow, traffic flow, and information flow (Wang *et al*., 2018; Zhai, 2019) are constructed separately to obtain the intensity of each element flow, and the weighted average of the four element flows to obtain the urban flow intensity.	See references (Zhai, 2019)

3.2 Operation and simulation of the HGLM-CA model

The 2007, 2012, and 2017 land-use data were selected for the research. The first two phases of data (2007 and 2012) were used for the model parameter calibration, and the last phase of data (2017) was used for the model validation. In this study, nine spatial driving factors were selected in the experiments, i.e., urban flow intensity, elevation, slope, the Euclidean distance to city centers, the Euclidean distance to district and county centers, the Euclidean distance to railways, the Euclidean distance to national highways, the Euclidean distance to thruways, and the Euclidean distance to water (Figure 4). Under the ArcGIS platform, each driving factor was processed and standardized, and the area where the land-use status changed from non-urban land to urban land was extracted by overlay analysis. After random sampling, outliers were eliminated to obtain training samples. Due to the large research scope and calculation efficiency, 3,000 and 30,000 samples were selected from the transformed and untransformed areas, respectively, for the experiments conducted in this study. Variable parameters were obtained through the HGLM, which were then substituted into the CA model to obtain the simulation results.

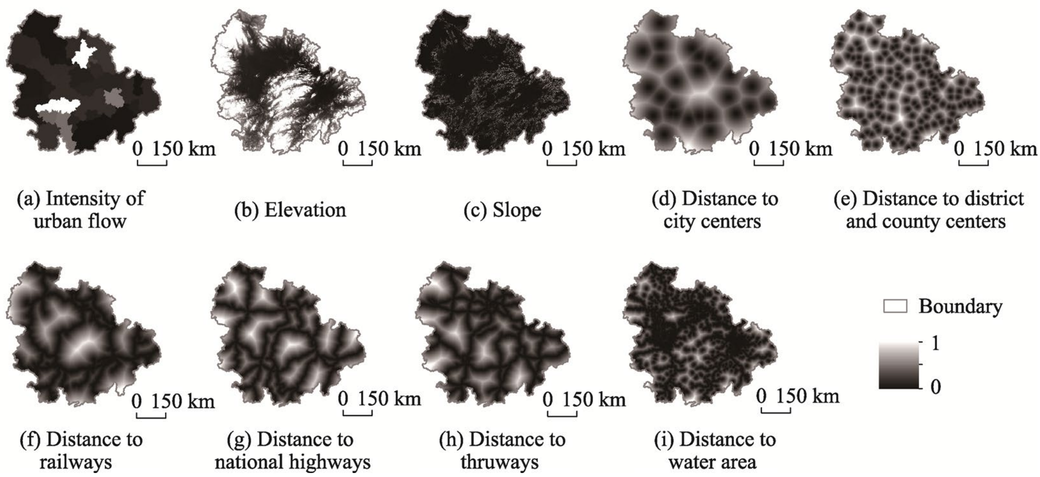


Figure 4 Driving factors of urban agglomeration spatial expansion in the middle reaches of the Yangtze River

3.3 Hierarchical driving mechanism of the spatial expansion of urban agglomerations

The spatial expansion of an urban agglomeration is the result of the combined effects of the different driving factors, and the analysis of the driving mechanism is of great significance to the development and planning of urban agglomerations (Wang *et al*., 2018). In this study, we built the HGLM using HLM 6.08 software, and analyzed the hierarchical driving mechanism of the urban agglomeration in the middle reaches of the Yangtze River. Firstly, an empty model was constructed for ANOVA, and the preliminary running results are listed in Table 2. According to Table 2, both the intercept mean (γ00) and slope mean (μ0) exist ($P <$ 0.01). The inter-group variance is 0.431. The variance of the logistic regression residual is usually $\pi^2/3$, and ICC = 0.116 could be obtained here, indicating that 11.6% of the variance is caused by the different cities where the cells were located. Therefore, the differences between the different cities could not be ignored, and a hierarchical analysis was necessary.

Table 2 ANOVA results of the HGLM

Fixed effect	Coefficient	Standard error	T-ratio	*P*-value	Random effect	Standard Deviation	Variance Component	Chi-square	*P*-value
γ00	–2.081	0.120	–17.337	0.000	μ0	0.657	0.431	1571.213	0.000

By constructing a random covariate model, a random intercept model, and a complete model, the independent variables were screened in several experiments, and were divided into explanatory variables and control variables.

The driving factors for the spatial expansion of the urban agglomeration were divided into the following levels:

Cell-level (level 1) independent variables, including distance to city centers, distance to district and county centers, distance to railways, distance to water, distance to national highways, distance to thruways, elevation, and slope. Among the variables, distance to city centers, distance to district and county centers, and distance to railways were used as explanatory variables. The other variables were considered to be control variables.

City-level (level 2) independent variables, referring to the intensity of the urban flow. The intensity of the urban flow varied between cities, and the intensity of the urban flow in the same city was the same. The dependent variable was a binary variable, where 1 meant that the cell was converted to urban land, and 0 meant that the state of the cell had not changed.

The final construction of the HGLM is shown in Figure 5. After running the HGLM, the parameter identification results were obtained (Table 3).

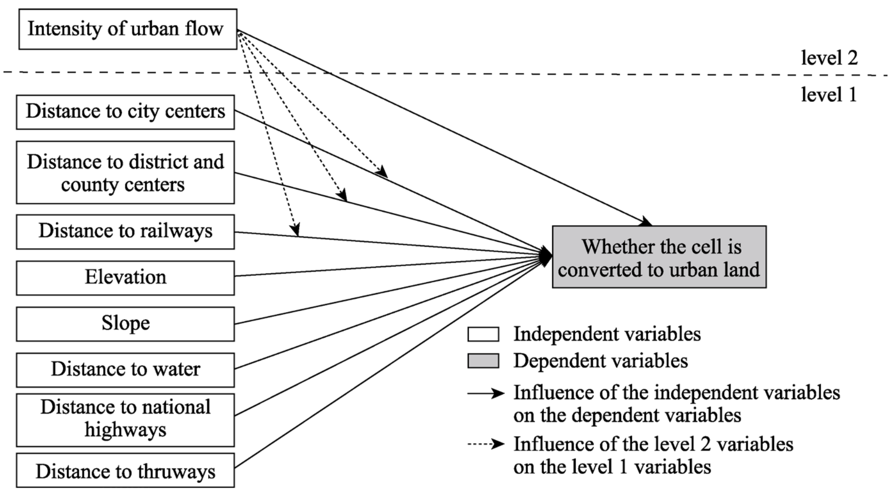


Figure 5 Schematic diagram of the HGLM

According to the parameter identification results, urban flow is positively correlated with the cell state transition probability (γ01 > 0), indicating that the urban flow has a significant impact on the spatial expansion of the urban agglomeration. The higher the intensity of the urban flow in a city, the easier it is for the cells in the city to transform into urban land. Among the independent variables, the coefficient β1 of distance to city centers has the largest weight, which reflects that city centers have a greater influence on the cell transition probability. The distance to city centers γ10 and the effect of city flow on the distance to city centers γ11 are negatively correlated with the cell state transition probability, indicating that

Table 3 Variable parameter identification results of the HGLM

	Coefficient	P		Coefficient	P
Intercept:			Control variables:		
Level 1 intercept β0			Elevation		
Intercept γ00	–1.429	0.000	Intercept γ40	–0.592	0.000
Urban flow γ01	0.947	0.000	Slope		
Independent variables：			Intercept γ50	–1.764	0.000
Distance to city centers β1			Distance to water		
Intercept γ10	–3.261	0.000	Intercept γ60	0.313	0.315
Urban flow γ11	–8.176	0.000	Distance to national highways		
Distance to district and county centers β2			Intercept γ70	–0.486	0.111
Intercept γ20	–1.899	0.001	Distance to thruways		
Urban flow γ21	–1.790	0.206	Intercept γ80	–0.297	0.391
Distance to railways β3					
Intercept γ30	–1.338	0.048			
Urban flow γ31	4.752	0.021			

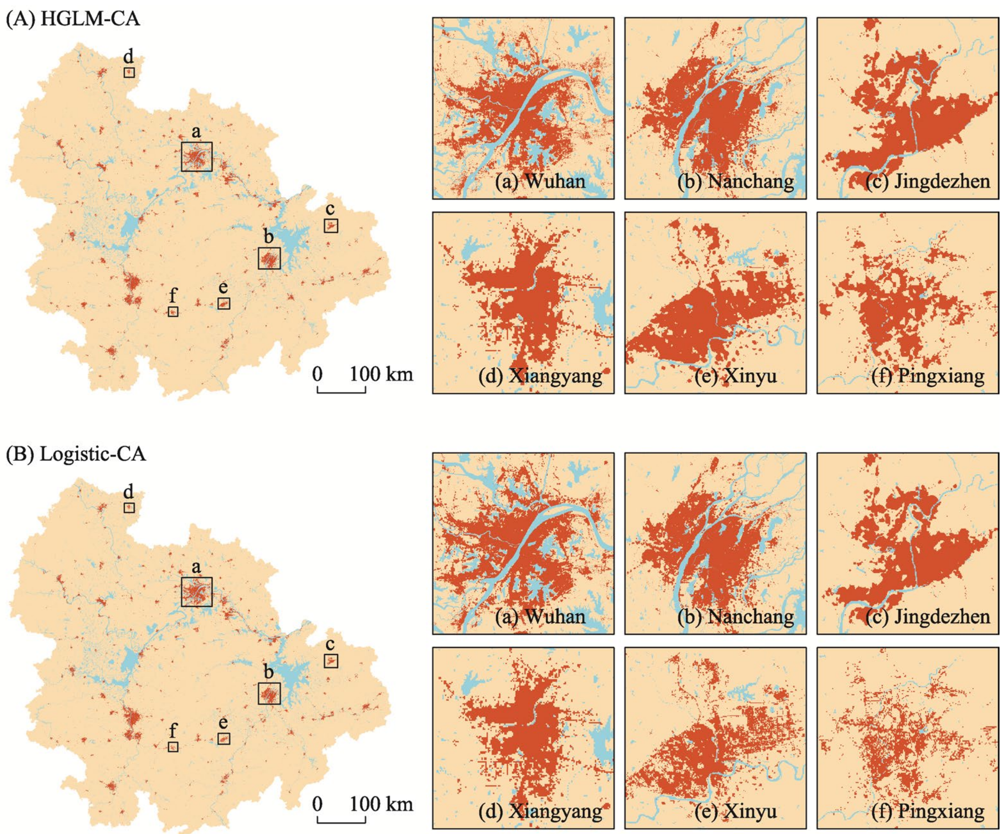
the existence of urban flow aggravates the impact of the distance to city centers variable, making it easier for land units closer to the city centers to be transformed into urban land, while making it more difficult for land units farther away from city centers to be transformed into urban land. This phenomenon is more obvious in areas with a higher intensity of urban flow. Distance to the district and county centers has an impact on the spatial expansion of the urban agglomeration that is second only to that of city centers, but the effect of urban flow on the district and county centers is not significant ($\gamma 21 = -1.790$, $P > 0.1$). Among the road factors, railways make a greater contribution to urban expansion, and urban flow has an obvious regulating effect on railways, which means that the influence of railways on the cell transition probability presents different effects in different cities. Since the cell transition probability is negatively correlated with the distance to railways γ30, and positively correlated with the effect of urban flow on the distance to railways γ31, in cities with a low intensity of urban flow, the effect of the urban flow is relatively weak, which weakens the influence of railways on the cell transition probability. In areas with a high intensity of urban flow, the influence of urban flow on the railway factors gradually appears with the increase of the urban flow, and it is very likely that the coefficient β3 of the distance to railways will change from negative to positive. Therefore, in cities with a higher level of economic and social development, the more likely it is that land units farther away from railways are converted to urban land. Among the control variables, elevation γ40 and slope γ50 have a greater impact on the spatial expansion of the urban agglomeration, reflecting that topography is still one of the important factors restricting urban development, and topography and landform play a certain role in shaping urban form.

From the HGLM analysis results, it can be seen that the urban flow, as a high-level (city-level) factor, not only has a significant positive impact on the cell transition probability, but also adjusts the degree of influence of the low-level (cell-level) independent variables on the cell transition probability. Furthermore, the same driving factor presents different effects in different cities.

3.4 Analysis of the HGLM-CA simulation results

The logistic model and the BBO model were used to obtain nine driving factor parameters, including the intensity of urban flow, and the logistic-CA model and the BBO-CA model were constructed and simulated. The simulation results of the HGLM-CA, logistic-CA, and BBO-CA models are shown in Figure 6. In this paper, the three indices of overall accuracy (OA), Kappa, and figure of merit (FoM) are used to evaluate the simulation results. The HGLM-CA model was compared with the logistic-CA and BBO-CA models, and the comparison results are shown in Table 4.

Comparing the simulation results, it can be found that the overall simulation effect of the BBO-CA model is better than that of the HGLM-CA model, and the HGLM-CA model performs better than the logistic-CA model. Among the different models, the OA and Kappa values are relatively close, and the FoM value of the BBO-CA model is significantly higher than that of the other two models. Compared with the actual status of the land use in 2017, the simulation accuracies for the different cities are different. For example, the logistic-CA, HGLM-CA, and BBO-CA models show a poor simulation effect in Wuhan, especially in the marginal areas where a lot of urban land is not identified. However, the simulation effects of the three models in Nanchang and Jingdezhen were generally good, with high values of



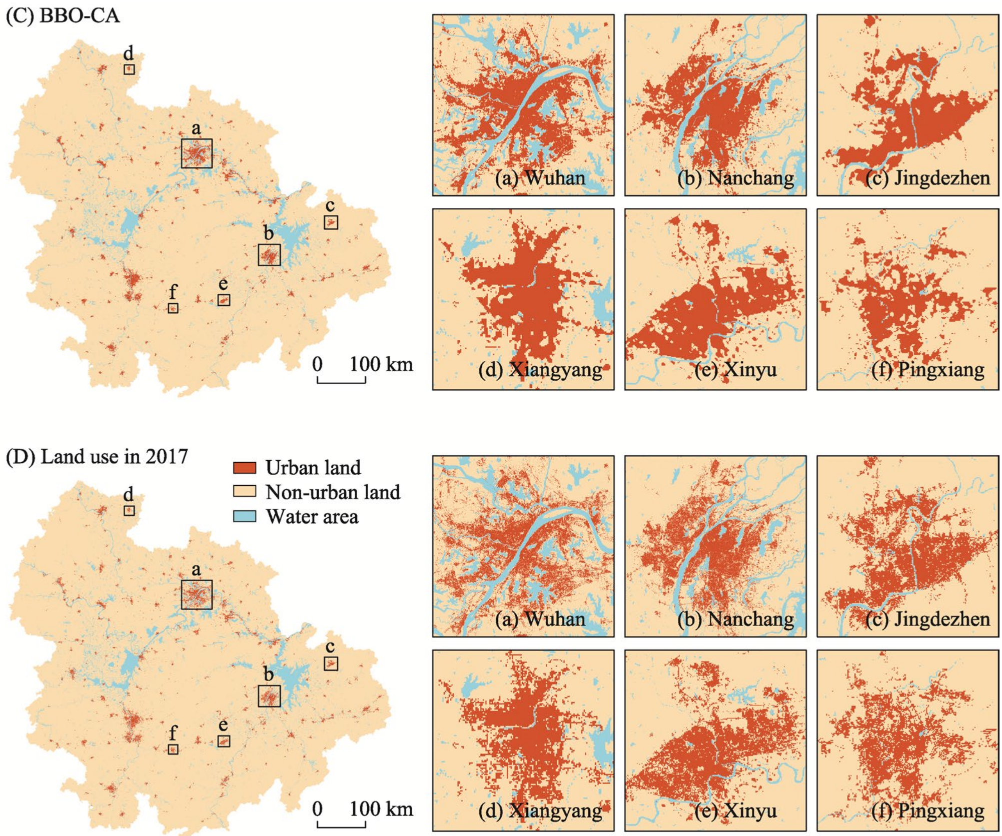


Figure 6 Comparison of the simulation results of the urban spatial expansion models in the middle reaches of the Yangtze River for 2017

OA,Kappa and FoM, In addition, the FoM value is significantly higher than that in other areas, at above 0.25. Furthermore, the simulation accuracies of the different models in the same city are also quite different. Due to the small difference between the OA and Kappa values of the three models, the simulation results are evaluated according to the FoM value. For FoM, the BBO-CA model shows the best simulation effect in most of the cities in the middle reaches of the Yangtze River, especially in Hunan and Hubei provinces, where the FoM values are generally higher than those of the other two models. In Xinyu and Pingxiang, the simulation accuracy of the HGLM-CA model is the best, with FoM values of 0.27223 and 0.17084, respectively. The FoM value of BBO-CA is slightly lower than that of HGLM-CA. In Nanchang, the logistic-CA model shows a good simulation effect, where the FoM value is 0.25614, which is slightly higher than that of the BBO-CA and HGLM-CA models. However, in cities such as Xiangyang and Xinyu, the FoM values of the logistic-CA model are all lower than 0.10, which is far from the simulation accuracy of the other two models.

HGLM-CA divides the driving factors into two levels, and the intensity of the urban flow is used as the city-level factor to adjust the cell-level factors. Overall, the OA, Kappa, and FoM values of HGLM-CA are slightly improved when compared to logistic-CA, and the simulation effect is better than that of logistic-CA in most cities, indicating that considering

Table 4 Comparison of the simulation accuracies of the urban spatial expansion models in the middle reaches of the Yangtze River for 2017

		HGLM-CA			Logistic-CA			BBO-CA		
		OA	Kappa	FoM	OA	Kappa	FoM	OA	Kappa	FoM
Overall accuracy	Urban agglomeration	0.99436	0.79872	0.18085	0.99427	0.79574	0.17374	0.99455	0.80567	0.19779
Local accuracy	Wuhan	0.95525	0.78844	0.14126	0.93924	0.74020	0.18817	0.94952	0.77246	0.18303
	Huangshi	0.97859	0.81788	0.17746	0.98138	0.83651	0.16789	0.98146	0.83816	0.18391
	Yichang	0.99445	0.78486	0.20519	0.99461	0.78646	0.18571	0.99448	0.78702	0.21576
	Xiangyang	0.99498	0.82902	0.15516	0.99514	0.82517	0.03963	0.99475	0.82610	0.19173
	Ezhou	0.94591	0.69046	0.14923	0.95070	0.70959	0.15293	0.96023	0.75093	0.16121
	Jingmen	0.99516	0.82205	0.11262	0.99563	0.83010	0.01952	0.99485	0.81586	0.14256
	Xiaogan	0.98711	0.77240	0.17225	0.98773	0.77889	0.16251	0.98763	0.77942	0.17561
	Jingzhou	0.99185	0.78990	0.15924	0.99196	0.79076	0.14808	0.99238	0.80328	0.18866
	Huanggang	0.99170	0.76696	0.15049	0.99208	0.77326	0.14080	0.99192	0.77603	0.18343
	Xianning	0.99392	0.82252	0.12214	0.99464	0.83704	0.08533	0.99347	0.81337	0.13292
	Xiantao	0.99208	0.83366	0.00041	0.99204	0.83445	0.02894	0.99065	0.82169	0.16428
	Qianjiang	0.98856	0.81205	0.04288	0.98857	0.81206	0.04001	0.98702	0.80615	0.18890
	Tianmen	0.99343	0.79276	0.01163	0.99346	0.79249	0.00210	0.99301	0.80157	0.19527
	Changsha	0.97542	0.79792	0.19991	0.97230	0.78174	0.21561	0.97574	0.80148	0.21485
	Zhuzhou	0.98999	0.77736	0.18150	0.99006	0.77920	0.18720	0.99081	0.79106	0.18815
	Xiangtan	0.97724	0.75437	0.21572	0.97718	0.75427	0.21746	0.98161	0.78615	0.20658
	Hengyang	0.99068	0.77667	0.20602	0.99089	0.78041	0.20796	0.99141	0.78814	0.20171
	Yueyang	0.99359	0.81224	0.14242	0.99422	0.82528	0.12560	0.99355	0.81314	0.16095
	Changde	0.99417	0.78845	0.17407	0.99461	0.79653	0.14288	0.99435	0.79537	0.19085
	Yiyang	0.99511	0.78612	0.18180	0.99544	0.79554	0.17560	0.99578	0.80790	0.18663
	Loudi	0.99284	0.80869	0.19307	0.99392	0.82569	0.13254	0.99350	0.82162	0.19003
	Nanchang	0.96873	0.79236	0.25293	0.96683	0.78406	0.25614	0.97187	0.80732	0.25144
	Jingdezhen	0.98913	0.79545	0.27337	0.98991	0.80425	0.26468	0.99082	0.81581	0.25770
	Pingxiang	0.98945	0.79041	0.17084	0.99075	0.79585	0.01749	0.98967	0.79303	0.16650
	Jiujiang	0.99296	0.80933	0.21376	0.99308	0.81093	0.20696	0.99303	0.81233	0.22926
	Xinyu	0.98812	0.85138	0.27223	0.99066	0.86902	0.09938	0.98876	0.85730	0.26929
	Yingtan	0.98118	0.73188	0.23667	0.98301	0.74787	0.23628	0.98613	0.77889	0.23856
	Ji’an	0.99532	0.77351	0.20956	0.99562	0.78092	0.19306	0.99591	0.79410	0.21761
	Yichun	0.99310	0.81064	0.12226	0.99302	0.80679	0.09893	0.99240	0.80345	0.18943
	Fuzhou	0.99558	0.80872	0.17760	0.99598	0.81430	0.09066	0.99566	0.81252	0.19129
	Shangrao	0.99397	0.79382	0.16224	0.99446	0.80489	0.14914	0.99459	0.81160	0.18193

the hierarchy of the urban agglomeration can improve the simulation accuracy of the CA model, to a certain extent. At the same time, it shows that there is a problem of underfitting when the logistic regression identifies the CA parameters, and the simulation effect is poor. After considering the hierarchical influence, the HGLM-CA model can mine the transition rules more deeply and obtain a better simulation effect. Compared with BBO-CA, the OA and Kappa values of the two other models are not much different, but the FoM value of the HGLM-CA model is significantly lower than that of the BBO-CA model, and the simulation accuracy in most cities is lower than that of BBO-CA. This shows that, compared with intelligent algorithms, the HGLM still has certain limitations in obtaining parameters. Although the BBO-CA model obtains the highest simulation accuracy, it is unable to analyze the driving mechanism.

4 Conclusions

Compared with the spatial expansion of a single city, the spatial expansion of an urban agglomeration is more complicated. In this study, by taking into account the urban flow and hierarchical characteristics, we constructed the HGLM-CA model to analyze the hierarchical driving mechanism of the spatial expansion of an urban agglomeration, and simulated the spatial expansion of the urban agglomeration in the middle reaches of the Yangtze River. We then compared the results with the results of the logistic-CA model and the BBO-CA model. The main conclusions are as follows:

(1) The urban agglomeration in the middle reaches of the Yangtze River is a large urban agglomeration in central China with close inter-city connections. Therefore, the influence of urban flow needs to be considered when analyzing the driving mechanism of its spatial expansion. In this study, we took urban flow as the driving factor at the urban level, and found that the intensity of the urban flow has a significant positive correlation with the spatial expansion of the urban agglomeration. Urban flow is therefore an important factor affecting the spatial expansion of the urban agglomeration in the middle reaches of the Yangtze River.

(2) Since the HGLM has high requirements on the number of sample groups, the use of the HGLM method to analyze the hierarchical driving mechanism is more suitable for areas with larger scales and more partitions. Therefore, in this study, we took the urban agglomeration in the middle reaches of the Yangtze River as the study area, which covers 31 cities, as this could effectively analyze the hierarchical driving mechanism of the urban agglomeration. Through experiments, it was concluded that the intensity of the urban flow of the urban factors plays an important role in regulating the driving factors at the cell level. Under the action of urban flow, the correlation between the cell-level factors and the cell transition probability changes, which can weaken, strengthen or even reverse strengthen the influence of the driving factors on the cell conversion probability, so that the driving effect of the cell-level factors in different cities is different, reflecting the inter-city heterogeneity of the driving factors.

(3) Compared with the traditional logistic-CA model, the OA, Kappa, and FoM values of the HGLM-CA model were improved, to a certain extent. Most cities in the local accuracy showed higher FoM values. Therefore, by considering the urban flow and hierarchical char-

acteristics of the urban agglomerations, the simulation results were more accurate. However, compared with the BBO-CA model, the accuracy of the HGLM-CA model was slightly lower. The artificial intelligence based BBO-CA model has certain advantages in parameter acquisition, and its simulation effect for the spatial expansion of urban agglomerations is better than that of linear models such as the HGLM-CA model. However, the aim of the BBO-CA model is to improve the simulation accuracy, and some parameters set in the optimization process can lack geographical meaning, which is not conducive to the analysis of the evolution mechanism of an urban agglomeration.

The HGLM-CA model explains the spatial expansion mechanism of the urban agglomeration by analyzing the correlation between the different levels of driving factors and the spatial evolution of the urban agglomeration, as well as the regulating effect of the high-level factors on the low-level factors, which reflects the complex interrelationships between the multi-level driving factors in the spatial expansion of urban agglomerations. The HGLM-CA model is therefore more suitable for establishing the internal mechanism of the spatial expansion of urban agglomerations on a scientific basis.

With the development of information technology, artificial intelligence has highlighted its unique advantages in remote sensing image interpretation, cartography, and other geographic fields. However, the traditional linear model is not completely without advantages in the face of the rapid rise of artificial intelligence in the relevant studies of the spatial expansion of urban agglomerations.

Urban agglomerations are complex evolutionary systems. The original intention of studying the spatial expansion of urban agglomeration was to simplify the complex urban systems, to facilitate understanding and research, instead of using more complicated methods for simulation and prediction in exchange for a slight improvement in simulation accuracy. From this perspective, the improvement of the traditional linear methods and the innovation of smart methods are equally valuable for research. How to retain the “simplicity” advantage of the traditional linear models, overcome the problem of underfitting, and endow the traditional linear models with artificial intelligence are issues worth discussing and studying in the spatial expansion of urban agglomerations. In addition, according to the experimental results obtained in this study for the urban agglomeration in the middle reaches of the Yangtze River, the simulation results of the HGLM-CA, logistic-CA, and BBO-CA models still had a large gap with the actual land-use status in 2017, and there were many examples of unidentified urban land. Urban agglomerations are complex, and the real spatial expansion of urban agglomerations tends to be both fragmented and sporadic. How to understand the driving mechanism of the spatial expansion of urban agglomerations and the law of spatial and temporal differentiation to effectively support model transition rules and overall optimization still needs to be explored in subsequent studies.

References

Chen Y, Li X, Liu X *et al*., 2019. Tele-connecting China’s future urban growth to impacts on ecosystem services under the shared socioeconomic pathways. *Science of the Total Environment*, 652: 765–779.

Couclelis H, 1985. Cellular worlds: A framework for modeling micro-macro dynamics. *Environment & Planning*

A, 17(5): 585–596.

Couclelis H, 1989. Macrostructure and microbehavior in a metropolitan area. *Environment & Planning B Planning & Design*, 16(2): 141–154.

Fang C, 2014. Progress and the future direction of research into urban agglomeration in China. *Acta Geographica Sinica*, 69(8): 1130–1144. (in Chinese)

Gong P, Li X, Zhang W, 2019. 40-year (1978–2017) human settlement changes in China reflected by impervious surfaces from satellite remote sensing. *Science Bulletin*, 64(11): 756–763.

He J, Shi X, Gong J *et al*., 2016. Modeling the spatial expansion of urban agglomeration considering their spatial interaction: A case study of Wuhan metropolitan area. *Geomatics and Information Science of Wuhan University*, 41(4): 462–467. (in Chinese)

He L, Liu Y, 2017. Simulating urban cooperative expansion in the metropolitan region based on improved CA model: A case study of Wuhan Urban Agglomeration, China. *Journal of Central China Normal University*, 51(2): 224–230. (in Chinese)

Hox J, 1999. Multilevel Modeling: When and Why. Berlin Heidelberg: Springer.

Lei L, Zhang L, 2002. Principle and application of the hierarchical linear model. *Journal of Capital Normal University (Social Sciences Edition)*, 30(2): 110–114. (in Chinese)

Li X, Xin T, 2006. Hierarchical Linear Model for binary data: Principle and application. *Psychological Development and Education*, 22(4): 97–102. (in Chinese)

Li X, Ye J, 2002. Neural-network-based cellular automata for realistic and idealized urban simulation. *Acta Geographica Sinica*, 57(2): 159–166. (in Chinese)

Li X, Ye J, 2007. Geographical Simulation System: Cellular Automata and Multi-Agent System. Beijing: Science Press, 43–46. (in Chinese)

Luijten J C, 2003. A systematic method for generating land use patterns using stochastic rules and basic landscape characteristics: Results for a Colombian hillside watershed. *Agriculture*, *Ecosystems & Environment*, 21(95): 427–441.

Shu B, Zhu S, Qu Y *et al*., 2020. Modeling multi-regional urban growth with multilevel logistic cellular automata. Computers, *Environment and Urban Systems*, 80: 101457. doi: 10.1016/j.compenvurbsys.2019.101457.

Sun Y, Yang J, Song S *et al*., 2020. Modeling of multi-level vector cellular automata and its simulation of land use change based on urban planning. *Acta Geographica Sinica*, 75(10): 2164–2179. (in Chinese)

Tobler W R, 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46 (Suppl.1): 234–240.

Wang H, Xia C, Liu X *et al*., 2016. Theoretical and methodological perspectives of fine-scale urban expansion cellular automata for the large regions. *Geography and Geo-Information Science*, 32(5): 1–8. (in Chinese)

Wang H, Xia C, Zhang A *et al*., 2017. Calibrating urban expansion cellular automata using biogeography-based optimization. *Geomatics and Information Science of Wuhan University*, 42(9): 1323–1329. (in Chinese)

Wang H, Zhai L, Liu Y, 2018. Urban connection and function in Wuhan urban agglomeration based on multi-dimensional urban factor flows. *Economic Geography*, 38(7): 50–58. (in Chinese)

Wang H, Zhang B, Liu Y *et al*., 2018. Urban expansion patterns and their driving forces based on the center of gravity-GTWR model: A case study of the Beijing-Tianjin-Hebei urban agglomeration. *Acta Geographica Sinica*, 73(6): 1076–1092. (in Chinese)

Wang H, Zhang B, Xia C *et al*., 2020. Using a maximum entropy model to optimize the stochastic component of urban cellular automata models. *International Journal of Geographical Information Science*, 34(5): 924–946.

Wu F, Yeh A G, 1997. Changing spatial distribution and determinants of land development in Chinese cities in the transition from a centrally planned economy to a socialist market economy: A case study of Guangzhou. *Urban Studies*, 34(11): 1851–1879.

Wu H, Li Z, Clarke KC *et al*., 2019. Examining the sensitivity of spatial scale in cellular automata Markov chain simulation of land use change. *International Journal of Geographical Information Science*, 33(5): 1040–1061.

Wu K, Fang C, Zhao M, 2015. The spatial organization and structure complexity of Chinese intercity networks. *Geographical Research*, 34(4): 711–728. (in Chinese)

Wuhan Statistical Bureau (WSB), 2020. 2019 Statistical bulletin of the national economic and social development of Wuhan Province. www.wuhan.gov.cn/zwgk/tzgg, 2020-03-29(006). (in Chinese)

Xia C, Zhang A, Wang H *et al*., 2019. Modeling urban growth in a metropolitan area based on bidirectional flows, an improved gravitational field model, and partitioned cellular automata. *International Journal of Geographical Information Science*, 33(5): 877–899. doi: 10.1016/j.landusepol.2019.104332.

Xia C, Zhang A, Wang H *et al*., 2020. Delineating early warning zones in rapidly growing metropolitan areas by integrating a multiscale urban growth model with biogeography-based optimization. *Land Use Policy*, 90: 104332.

Xie Z, Wang H, Zhang B *et al*., 2020. Urban expansion cellular automata model based on multi-structures convolutional neural networks. *Acta Geodaetica et Cartographica Sinica*, 49(3): 375–385. (in Chinese)

Yang Q, Li X, 2007. Integration of multi-agent systems with cellular automata for simulating urban land expansion. *Scientia Geographica Sinica*, 27 (4): 542–548. (in Chinese)

Zhai L, 2019. Urban spatial connection and function of urban agglomeration in the middle reaches of Yangtze River [D]. Wuhan: Wuhan University. (in Chinese)

Zhang B, Wang H, Xia C *et al*., 2020. Analyzing the effects of stochastic perturbation and fuzzy distance transformation on Wuhan urban growth simulation. *Transactions in GIS*, 24(6): 1779–1798.

Zhang K, Qian Q, Yang Q, 2020. An analysis of multilevel variables influencing China's land urbanization process. *Acta Geographica Sinica*, 75(1): 179–193. (in Chinese)

Zhang Y, 2020. Promoting China's green urbanization based on ecological civilization. *China Population, Resources and Environment*, 30(10): 19–27. (in Chinese)