TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 16/16 pp782-794
Volume 14, Number 6, December 2009

Beijing Urban Development Model: Urban Growth Analysis and Simulation*

LONG Ying (龙 瀛)[1,2,**], MAO Qizhi (毛其智)[1], DANG Anrong (党安荣)[2]

1. School of Architecture, Tsinghua University, Beijing 100084, China;
2. Beijing Institute of City Planning, Beijing 100045, China

Abstract: Urban growth analysis and simulation have been recently conducted by cellular automata (CA) models based on self-organizing theory which differs from system dynamics models. This paper describes the Beijing urban development model (BUDEM) which adopts the CA approach to support urban planning and policy evaluation. BUDEM, as a spatio-temporal dynamic model for simulating urban growth in the Beijing metropolitan area, is based on the urban growth theory and integrates logistic regression and MonoLoop to obtain the weights for the transition rule with multi-criteria evaluation configuration. Local sensitivity analysis for all the parameters of BUDEM is also carried out to assess the model's performances. The model is used to identify urban growth mechanisms in the various historical phases since 1986, to retrieve urban growth policies needed to implement the desired (planned) urban form in 2020, and to simulate urban growth scenarios until 2049 based on the urban form and parameter set in 2020. The model has been proved to be capable of analyzing historical urban growth mechanisms and predicting future urban growth for metropolitan areas in China.

Key words: cellular automata; policy simulation; logistic regression; planned urban form

Introduction

Along with the prosperity of Beijing macro-economy, especially as a result of the Olympics, the future form of the Beijing metropolitan area needs to be analyzed in the post-Olympic games period in 2009, at the end of the urban master planning in 2020, and for the 100th anniversary of the founding of the P. R. China's capital in 2049. Furthermore, long-term forecast of the urban form is essential for the next round of urban master planning. Comprehensive urban models have been developed to simulate urban systems for some major metropolitan areas, such as the London metropolitan area, the San Francisco Bay area, and the California area. However, there is no urban model for Beijing or most other cities in China.

Urban models, which were developed in the early years of the twentieth century, have progressed from structural models to static models to dynamic models. Traditional urban models, based on differential equations or quasi dynamic equations, usually simulate the urban system at a macro level, so they cannot accurately reflect the dynamic, self-organizing, or emerging characteristics of urban systems. The development of GIS and other complex adaptive models has led to urban models based on artificial life or discrete dynamics. In recent years, urban growth models have used the cellular automata (CA) approach which is based on self-organizing theory. The CA models are composed of a series of basic rules instead of strictly defined physics equations or functions. The discrete character

Received: 2008-10-16; revised: 2009-06-10
* Supported by the National Natural Science Foundation of China (No. 50678088) and the National Project of Scientific and Technical Supporting Programs Funded by the Ministry of Science & Technology of China (No. 2006BAJ14B08)
** To whom correspondence should be addressed.
E-mail: longying1980@gmail.com; Tel: 86-10-88073660

is a key characteristic of the time and space and status in CA. CA has been adapted to simulate the emergence, self-organizing, and chaos phenomena in urban systems.

CA is now a practical tool for simulating urban growth, which is the main field using CA. Tobler[1] initially simulated urban expansion in the Great Lakes region of the United States. Couclelis[2-5] claimed that CA with simple rules can be applied to generate complex urban forms in a virtual city, with great potential to simulate urban growth. White and Engelen[6] applied CA to urban planning and White and Engelen[7] simulated the land use patterns in Cincinnati, Ohio, United States. Clark and Gaydos[8] developed the SLEUTH model to simulate long-term urban growth in the San Francisco Bay area and the Washington-Baltimore area in the United States, like some of the earliest applications of CA which simulate urban growth in the real cities. Batty et al.[9-16] conducted several studies using fractal and CA to study urban formation and expansion. Xie[17] simulated the land use changes in Buffalo, NY, United States. Wu and Webster[18-20] used the multi-criteria evaluation (MCE) method to find the status transition rules in CA and applied the model for urban expansion simulations of Guangzhou, China.

In China, Li and Yeh[21-27] used various intelligent methods to identify the CA transition rules, with a constrained CA model to simulate sustainable urban growth in the Pearl River Delta and analyzed the uncertainty of CA. Others have used CA to simulate urban growth in Haikou[28], Wuhan[29], Fuzhou[30], Xi'an[31], Northern China[32], and part of the Beijing metropolitan area[33]. Thus, there have been many applications of CA to simulate urban growth in China. However, there are no studies using CA to simulate urban growth in the Beijing metropolitan area.

The Beijing urban development model (BUDEM) was developed to support urban planning and policy evaluation. This spatio-temporal dynamic model simulates urban growth for the Beijing metropolitan area using the CA approach. The model was developed for the Beijing metropolitan municipal government and planning committee with an area of 16 410 km^2 and with a spatial resolution of 500 m, including computer simulations using CA of the Beijing metropolitan area. The BUDEM urban form simulation platform was specially developed for Beijing urban planning in a mega-city. BUDEM integrates logistic regression and MonoLoop to identify the CA transition rule to realize the desired urban form. The model uses environmental constraints and urban planning conditions to reflect China's urban development characteristics. Then it is used to simulate the planned urban form of Beijing in 2020 and to predict the long-term unknown urban form of Beijing in 2049. Therefore, the paper presents a historical analysis of the different phases of Beijing's urban growth and model estimates for Beijing in 2020 (BEIJING2020), and then predicts the Beijing form in 2049 (BEIJING2049) based on the urban form and parameter set of Beijing in 2020.

1 BUDEM: CA-Based Urban Simulation Model

1.1 Spatial factors selection

Macro level urban growth research, which does not consider urban spatial distribution, regards the urban system as one whole entity. The driving forces for urban growth consist of population changes, economic changes, political structure, etc.[34]. However, we prefer the self-organizing process research within the urban system. Urban development is influenced by location and geographic conditions in a classical urban land use model. Alonso[35] pointed out that the distance to the urban center is the principal factor controlling the urban land use structure in his single center urban location theory. The optimum land type will change according to the distance to the urban center, which affects the accessibility and transportation cost. In addition, Doxiadis, who founded human settlement science, concluded that the distances to the present urban center, to the main road, or to the natural landscape are the main forces for human settlement[36]. The Hedonic model provides a clearer framework, which assumes that commodity prices are determined by the total utility of the different properties and depend on the number and composition of the commodity's properties[37]. For example, Butler[38] held that residential prices are affected by location, architectural structure, and neighborhood, and that the prices reflect the total preferences of the consumer. Urban development is related to residential prices and its probability reflects developer's preferences for the lot or block. Therefore, the spatial variables in the CA were chosen as shown in Table 1 based on Hedonic theory and the possibility of acquiring the data. Column "Dataset" stands for the corresponding dataset from which the spatial variable is obtained with detailed descriptions of the datasets in Section 1.4.

Table 1 Spatial variables in the CA

Type of variables	Name	Value	Description	Dataset
Self-status	isrural	0, 1	Whether the cell is rural built-up land in the previous iteration	LANDUSE, LANDi*
	isagri	0, 1	Whether the cell is agricultural land in the previous iteration	LANDUSE
Location	d_tam	⩾0	Minimum distance to Tian'anmen Square	LOCATION
	d_vcity	⩾0	Minimum distance to important new city	
	d_city	⩾0	Minimum distance to new city	
	d_vtown	⩾0	Minimum distance to important town	
	d_town	⩾0	Minimum distance to town	
	d_river	⩾0	Minimum distance to river	
	d_road	⩾0	Minimum distance to road	
	d_bdtown	⩾0	Minimum distance to town boundaries	BOUNDARY
Government	planning	0, 1	Whether planned as urban built-up	PLANNING
	con_f	0, 1	Whether in the forbidden zone	CONSTRAINT
	landresource	1-8 (integer)	Land suitability classified for agriculture	LANDRESOURCE
Neighbor	neighbor	0-0.125	Neighborhood development intensity	LANDi

* Dataset LANDi is the generated land use layer when CA iterating

1.2 Conceptual model

The model's premises are as follows: First, the urban system is a complex self-adaptive system that can be simulated with a bottom-up method. Second, the urban growth forces can be classified into promoting and restraining types. They can also be divided into market forces and government forces. Third, the historical development trends will continue in future development. Finally, various urban growth scenarios can be generated with variations of the baseline scenario.

With these assumptions, the conceptual CA model is shown in Eq. (1). The CA lattices based on the Beijing metropolitan area, which has an area of 16 410 km^2, are adjustable based on the simulation purpose. This study used cell sizes of 500 m×500 m for a total 65 628 cells in the lattices. The CA cell states are 0 or 1, where 1 stands for urban built-up land and 0 stands for non urban built-up land. MCE was used as the CA transition rule with a 3×3 Moore used as the CA neighborhood with 8 adjacent cells. The discrete time in the CA model was based on the total number of urban built-up cells.

$$V_{i,j}^{t+1} = f\{V_{i,j}^{t}, \text{Global variables}, \text{Local variable}\} = f\left\{\begin{array}{l} V_{i,j}^{t}, \text{isrural}_{i,j}, \text{isagri}_{i,j}, \text{d\_tam}_{i,j}, \text{d\_vcity}_{i,j}, \text{d\_city}_{i,j}, \\ \text{d\_vtown}_{i,j}, \text{d\_town}_{i,j}, \text{d\_river}_{i,j}, \text{r\_road}_{i,j}, \\ \text{d\_bdtown}_{i,j}, \text{planning}_{i,j}, \text{con\_f}_{i,j}, \text{landresource}_{i,j}, \\ \text{neighbor}_{i,j}^{t} \end{array}\right\} \quad (1)$$

where $V_{i,j}^{t}$ is the cell status at cell *ij* of time *t*, $V_{i,j}^{t+1}$ is of the cell status at *ij* of time t+1, *f* is the transition rule.

The cell status in each iteration is influenced by the self status variables in the previous iteration, global variables, and the local variable listed in Table 1. The self status variables include isrural and isagri, which indicate whether cell *i*, *j*'s status at time *t* is rural built-up land or agricultural land. Development control of the rural built-up land and the agricultural land is a sensitive question in China, so both of these two self status variables are important in the model. The global variables included location type and government type variables, which are static through all the iterations. Variable neighbor is the only local variable in the CA which changes during the iterations. The model simulates the transition from urban non built-up to urban built-up, but not the reverse process. Urban redevelopment is also not considered.

In addition to the CA model, a macro-restraining sub-model was also developed using the urban development time-serial data as described in Section 1.4 to predict the total amount of urban built-up each year. The total is the core parameter for controlling the number of CA iterations when the model terminates. The model also calculates the corresponding real time (Year) for each iteration as the CA discrete time.

Urban economists use regression methods, such as logistic regressions or multi-logit models to calculate the urban development probability. These methods can identify the relationships between land use change and locational characteristics. However, the self-organizing process of land development is not considered. Spontaneous growth and self-organizing growth are integrated in the current CA model, which is a key advantage of this model for simulating urban growth.

1.3 Status transition rule

The transition rule, as the core CA component, is an important topic in CA research. Various methods have been developed to determine the transition rule, such as MCE, grey theory, principal component analysis, artificial neural networks, genetic algorithms, rough sets, and case-based reasoning. This analysis uses MCE to establish the CA status transition rule.

Landis and Zhang[39-42] developed CUF and CUF-2 to predict urban forms as typical applications of MCE for urban growth modeling. CUF and CUF-2 use the developing land use (DLU) as the basic modeling unit, instead of a CA cell. Wu[20] determined the MCE-based transition rules for the CA urban growth simulation model based on

$$P_{\mathrm{c}}^t = P_{\mathrm{g}} \cdot \mathrm{con}(s_{ij}^t = \mathrm{suitable}) \cdot \Omega_{i,j}^t \qquad (2)$$

Yeh and Li's constrained CA model used a similar transition rule[25]. In Eq. (2), P_{g} is the urban growth suitability, which is the global probability calculated by the MCE method. Ω is the neighborhood effect, con is the environmental restraining effect, and P_{c}^t is the joint probability. Wu and Webster[18,19] used the analytic hierarchy process (AHP) method to obtain the weights of spatial variables in the MCE, while Wu[20] used logistic regression to obtain the weights based on historical development data. Wu found that the MCE method is a convenient way to obtain the CA transition rules, but that the weights for the spatial variables in the MCE are difficult to be accurately and comprehensively determined. The AHP method is not repeatable and overly subjective and it cannot identify historical urban development trends. In the AHP method for obtaining the MCE weights, the neighborhood and environmental effects are separately multiplied to get P_{g}, instead of being included in the logistic regression procedure. As a result, the logistic regression does not include all the relevant factors and the regression weights cannot entirely explain some historic urban growth development trends.

Clark and Gaydos[8] presented a rigorous calibration method. This method first generates simulation results with different parameter sets (nested loops). Then, each simulated result is matched with the observed form to calculate matching indexes (the r-squared fit between the actual and predicted number of urban pixels, edges in the images, separate clusters, and a modified Lee-Sallee shape index). The parameter set with the best matching index is then used to predict the future urban form. Five parameters are considered in this model, with 6, 6, 6, 5, 7 values for each parameter, respectively. Consequently, 7560 parameter sets are generated for the model calibration with 252 h of calculation time needed for the rigorous calibration process. As the number of parameters increased, the calibration time would increase exponentially. Our current model has 14 parameters. If each parameter had 6 values for testing with the same time cost for every loop as in the Clark and Gaydos' model, the total calibration time would be about 300 000 years, even with a limitation of only 6 choices for each parameter. Even though the Clark and Gaydos' method can identify the best parameter set for the simulation, the time cost is not acceptable even by the most advanced workstation. Therefore, the methods of Wu[20] and Clark and Gaydos[8] were integrated to obtain the weights for the MCE formatted transition rule as Eq. (3).

$$\begin{aligned} s_{ij}^t = {} & w_0 + w_1 \cdot \mathrm{isrural}_{ij} + w_2 \cdot \mathrm{isagri}_{ij} + w_3 \cdot \mathrm{d\_tam}_{ij} + \\ & w_4 \cdot \mathrm{d\_vcity}_{ij} + w_5 \cdot \mathrm{d\_city}_{ij} + w_6 \cdot \mathrm{d\_vtown}_{ij} + \\ & w_7 \cdot \mathrm{d\_town}_{ij} + w_8 \cdot \mathrm{d\_river}_{ij} + w_9 \cdot \mathrm{d\_road}_{ij} + \\ & w_{10} \cdot \mathrm{d\_bdtown}_{ij} + w_{11} \cdot \mathrm{planning}_{ij} + w_{12} \cdot \mathrm{con\_f}_{ij} + \\ & w_{13} \cdot \mathrm{landresource}_{ij} + w_{\mathrm{n}} \cdot \mathrm{neighbor}_{i,j}^t , \end{aligned}$$

$$p_g^t = \frac{1}{1+e^{-s_{ij}^t}},$$
$$p_{ij}^t = \exp\left[\alpha \cdot \left(\frac{p_g^t}{p_{g,\max}^t} - 1\right) \cdot \mathrm{RI}_{ij}^t\right],$$
$$\mathrm{RI}_{ij}^t = 1 + (\gamma_{ij}^t - 0.5)/k,$$
$$\text{if } p_{ij}^t > p_{\text{threshold}}, \text{ then } V_{ij}^{t+1} = 1 \quad (3)$$

In Eq. (3), s_{ij}^t is the development suitability, w is the logistic regression coefficient, p_g^t is the initial transition probability, $p_{g,\max}^t$ is the maximum p_g^t in iteration t, p_{ij}^t is the final transition probability, $p_{\text{threshold}}$ is the urban growth threshold, RI is the random item, γ is the random value varying from 0 to 1, k is the random index used to regulate RI, and α is the dispersion parameter ranging from 1 to 10, which indicates the rigid level of development. The larger α indicates stricter development control and lower development probability with the same suitability. Hence, the parameter α greatly impacts the simulated urban form.

All the spatial variables, except neighbor, are included in the logistic regression equation, and the corresponding coefficients, weights w_1-w_{13} in the MCE, can be obtained. In the regression analysis, the dependent variable is either 0 or 1, since the land use change status is either "developed" or "undeveloped". There are 13 spatial variables in the regression analysis (all except neighbor). The dependent variable is obtained by algebra operation on LANDUSE datasets at the start and the end of the year. The sample tool in the ESRI ArcGIS was used to sample the dependant variables and independent variables for the logistic regression into a table, which were then analyzed in SPSS to obtain the w_1-w_{13} coefficients.

Then, the weight for the variable neighbor (w_n^*) was calculated using the sole parameter looping method (MonoLoop), instead of looping through all the parameter weights as in Clark and Gaydos[8]. Various w_n were calculated to find w_n^* with the best matching index while keeping w_1-w_{13} constant. Then, w_n^* was obtained with w_1-w_{13} input to the transition rule to simulate the urban growth. The goodness-of-fit (G), accuracy of the point-to-point comparisons, was used to assess the degree of matching between the simulated and observed urban forms. G has a maximum of 100%. This method combining logistic regression and MonoLoop greatly reduces model calibration time and is still able to identify historical urban growth trends.

The RI in the transition rule allows development beyond what can be explained by the selected spatial variables, such as the leapfrog type development. The use of RI in the transition rule makes the simulated result more rational. A constant threshold $p_{\text{threshold}}$ is used instead of a static or random threshold to guarantee the same development standards for urban development in different phases. Variables isagri and isrural stand for the transition from agricultural land and rural built-up areas, which is a concern of the Beijing Municipal Government.

1.4 Dataset

The input spatial data is classified into seven types, LANDUSE, PLANNING, CONSTRAINT, LANDRESOURCE, LOCATION, BOUNDARY, as well as UrbanInfoSeries. All the spatial data was converted into the ESRI single band GRID format, using the same coordinate and projection system. The data descriptive statistics is listed in Table 2.

(1) LANDUSE: The most complete dataset includes landuse for 1947, 1964, 1976, 1981, 1986, 1991, 1996, 2001, and 2006. The data for 1947 was digitalized from a relief map, the data for 1964 was interpreted from a DISP aerial image, the data for 1976 and 1981 was interpreted from MSS images, and the others were interpreted from TM images. LANDUSE is classified into six land use types, including urban built-up area, rural built-up area, agriculture area, forest area, wetland area, and vacant area. The variable landuse was derived from the LANDUSE dataset with 1 for urban built-up areas and 0 for other areas.

(2) PLANNING: Five urban master plans for the Beijing metropolitan area were included for 1958, 1973, 1982, 1993, and 2004, with the areas classified into urban built-up areas and non urban built-up areas[43]. The variable planning is derived from the PLANNING dataset.

(3) CONSTRAINT: The data reflects the urban development constraints derived from 110 spatial layers of natural resource protection and hazard prevention according to current laws, legislations, and standards of China. CONSTRAINT is zoned into the forbidden built-up areas, constrained built-up areas, and suitable built-up areas. The variable con_f stands for the forbidden built-up areas.

Table 2 Data descriptive statistics in the model (*N*=65 628)

Dataset	Variable	Minimum	Maximum	Mean	Std. deviation
LANDUSE	$landuse_{1986}$	1	6	4.32	1.02
	$landuse_{1991}$	1	6	4.31	1.03
	$landuse_{1996}$	1	6	4.27	1.10
	$landuse_{2001}$	1	6	4.24	1.14
	$landuse_{2006}$	1	6	4.19	1.20
PLANNING	$planning_{1958}$	0	1	0.04	0.20
	$planning_{1973}$	0	1	0.05	0.22
	$planning_{1982}$	0	1	0.03	0.16
	$planning_{1992}$	0	1	0.05	0.22
	$planning_{2004}$	0	1	0.14	0.35
CONSTRAINT	con_f	0	1	0.44	0.50
LANDRESOURCE	landresource	1	8	3.92	1.66
LOCATION	d_tam	0	129 711	57 737.08	26 952.19
	d_vcity	0	102 504	46 545.37	23 392.81
	d_city	0	78 824	24 801.90	14 730.19
	d_vtown	0	36 530	13 255.02	6968.38
	d_town	0	42 005	8286.05	5298.70
	d_river	0	14 230	3212.68	2416.59
	d_road_{1986}	0	29 681	2514.47	3675.36
	d_road_{1991}	0	29 954	2390.63	3577.05
	d_road_{1996}	0	29 820	2341.07	3635.95
	d_road_{2001}	0	24 000	1925.85	2494.29
	d_road_{2006}	0	29 820	2306.49	3613.74
BOUNDARY	d_bdtown	0	7762	1239.24	1173.94

(4) LANDRESOURCE: The data indicates the suitability for agricultural use classified into eight types ranging from 1 to 8[44]. The variable landresource is derived from LANDRESOURCE with the same values.

(5) LOCATION: The location indicates the minimum distance to urban or town centers for various administrative divisions, with d_road as the minimum distance to the road networks and d_river as the minimum distance to the rivers. The location spatial variables were derived from LOCATION using the Distance/Straight Line command in the ESRI ArcGIS.

(6) BOUNDARY: The boundaries include administrative, ring road, eco-zoning, and watershed boundaries, with various transition rules used in different sub-areas. The variable d_bdtown was derived from the dataset.

(7) UrbanInfoSeries: The data indicates population, environmental, economic, and social datasets in macro levels for the macro-restraint sub-model[45].

2 Parameter Estimates for Beijing

The urban growth driving forces in different historical phases were acquired by the logistic regression for 1947-1964, 1964-1976, 1976-1981, 1981-1986, 1986-1991, 1991-1996, 1996-2001, and 2001-2006. In the logistic regression, the variables, d_tam, d_vcity, d_city, d_vtown, d_town, d_bdtown, landresource, and con_f, do not change while the variables, planning, d_road, isrural, isagri, and the dependent variable vary with the historical phase. The variable neighbor is not considered in the logistic regression. The results can then be used to compare the urban growth mechanisms in various phases. The regression results for 2001-2006 are listed in Table 3, with an acceptable regression accuracy of 88.5%. As shown in Table 3, the coefficient of d_river is the largest, and accordingly the development probability will decrease by 0.024% with the increased distance of 1 m to the nearest river. The variable d_road is positively correlated to the development probability, but the coefficient is quite small, indicating that the urban growth along roads in this phase is not significant. The development probability of cells in the forbidden built-up zone is 65.9% lower than that of other areas, while a cell in the planned built-up area has a development probability of 36.7% higher than that of unplanned areas.

Table 3 Variables in the logistic regression equation for 2001-2006 (*B* is the regression coefficient.)

Name	*B*	Standard error	*p*	Exp(*B*)
isrural	3.284 309	0.321	0.000	26.691
isagri	4.376 267	0.157	0.000	79.541
d_tam	–0.000 042	0.000	0.000	1.000
d_vcity	–0.000 018	0.000	0.002	1.000
d_city	–0.000 032	0.000	0.000	1.000
d_vtown	–0.000 019	0.000	0.050	1.000
d_town	–0.000 036	0.000	0.320	1.000
d_river	–0.000 224	0.000	0.000	1.000
d_road_{2001}	0.000 061	0.000	0.578	1.000
d_bdtown	–0.000 057	0.000	0.510	1.000
$planning_{2004}$	0.312 422	0.139	0.025	1.367
con_f	–1.076 304	0.143	0.000	0.341
landresource	–0.023 686	0.040	0.556	0.977

The logistic regression coefficients for different historical phases are listed in Table 4. The LANDUSE data before 1986 is still incomplete so the logistic regressions were not conducted for data before 1986. Comparison of the dominant factors in the different phases shows that the urban growth mechanisms differ greatly with variations of the market and government roles in the different phases. 2001-2006 had more riverside development followed by development in the central city, with little development along the road. 1996-2001 had strong development along the roads followed by the new city development, with areas around the town growing more slowly. 1991-1996 and 1986-1991 both had significant development along the roads.

The results in Table 4 also show that the parameter of planning remained positive and reached a maximum in 1986-1991. The other phases all have the parameter of planning of about 0.4. Thus, in the first several years of the social market economy, urban planning played a leading role in urban growth. However, with the introduction of the market mechanism into China, its role was partially replaced by market factors. The parameter of con_f remained negative and decreased with time, indicating that the effect of ecological conservation and hazard prevention on the urban growth is increasing with time. The parameter of landresource remained positive but decreased with time, indicating less protection of cultivatable land. The parameter of d_bdtown was positive before 1996 and negative later, indicating that the restrictions of administrative boundaries are gradually decreasing in Beijing.

Table 4 Logistic regression coefficients for various historical phases

Name	*B*			
	2001-2006	1996-2001	1991-1996	1986-1991
isrural	3.284 309	–3.774 535	–3.949 259	–5.534 083
isagri	4.376 267	3.279 759	1.802 018	–0.156 322
d_tam	–0.000 042	–0.000 049	–0.000 054	–0.000 012
d_vcity	–0.000 018	–0.000 032	0.000 003	–0.000 047
d_city	–0.000 032	–0.000 094	–0.000 034	–0.000 028
d_vtown	–0.000 019	–0.000 029	–0.000 018	–0.000 014
d_town	–0.000 036	0.000 129	0.000 023	–0.000 021
d_river	–0.000 224	–0.000 078	–0.000 066	–0.000 021
d_road	0.000 061	–0.000 734	–0.000 365	–0.001 232
d_bdtown	–0.000 057	–0.000 463	0.000 001	0.000 182
planning	0.312 422	0.459 742	0.416 635	1.302 933
con_f	–1.076 304	–0.613 198	–0.449 983	–1.274 498
landresource	–0.023 686	–0.140 539	–0.131 834	–0.350 835

3 Identifying Parameters for the Planned Form

In 2004, the State Council of China approved the Beijing master plan from 2004 to 2020. The Beijing Municipal Planning Committee and the Beijing Municipal Government are now concerned about how to realize the planned urban form (represented by variable $planning_{2004}$ in dataset PLANNING) and how to predict the urban form from now until 2020. Accurate simulations of the future desired urban form cannot be based on the best parameter sets derived from differential equation or optimum theory methods. The nested-loop method cannot be used to determine the parameters. Therefore, the integrated logistic regression and MonoLoop method was used to obtain the needed parameter set for the transition rule of the desired urban form.

In the logistic regression for parameter identification, the dependent variable was calibrated via algebra operation on the $planning_{2004}$ and $landuse_{2006}$ with w_1-w_{13} then listed in Table 5 with the whole accuracy of 96.8%. Table 5 shows that each independent variable is significant at the acceptable level.

Table 5 Variables in the logistic regression equation for 2006-2020

Name	B	Standard error	p	Exp(B)
isrural	6.886 21	0.311	0.000	978.683
isagri	6.971 87	0.212	0.000	1066.213
d_tam	−0.000 10	0.000	0.000	1.000
d_vcity	−0.000 03	0.000	0.000	1.000
d_city	−0.000 10	0.000	0.000	1.000
d_vtown	−0.000 28	0.000	0.000	1.000
d_town	−0.000 11	0.000	0.000	1.000
d_river	−0.000 52	0.000	0.000	0.999
d_road	−0.000 96	0.000	0.000	1.001
d_bdtown	−0.000 27	0.000	0.001	1.000
planning	8.770 71	0.270	0.000	6442.743
con_f	−0.200 97	0.138	0.146	0.818
landresource	−0.093 55	0.039	0.016	0.911

The logistic regression was then followed by the MonoLoop procedure using 27 values for w_n which required 21.5 h for 2997 iterations. The relationship between G and w_n is shown in Fig. 1 where G remains steady at around 97.6% for w_n from 0 to 5, and decreases sharply to 93.0% for w_n from 5 to 35, and then rises for w_n larger than 35. For w_n bigger than 35, the number of developed cells in the first iteration was too large which indicates too many neighborhood interactions, exceeding the total final number of urban built-up cells, to complete the simulation. Therefore, w_n^* was set to 4.598 08 to simulate BEIJING2020, with G around 97.6%, which is near the ideal maximum of 98.9%. The ideal maximum was determined by the spatial relationship of $planning_{2004}$ which had 9376 urban built-up cells and the observed urban form of $landuse_{2006}$ which had 5297 urban built-up cells. The overlay analysis of $planning_{2004}$ and $landuse_{2006}$ shows that 712 developed cells in the observed urban form of $landuse_{2006}$ were not part of the planned form of $planning_{2004}$. Hence, the maximum G should be (65 628−712)/65 628=98.92%.

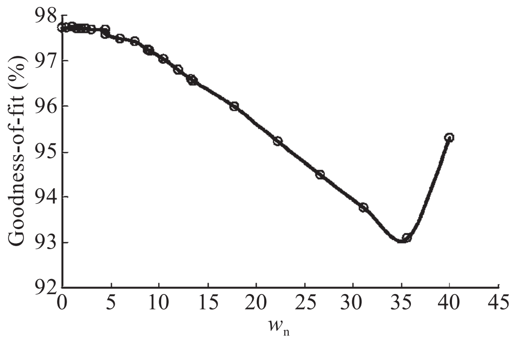

Fig. 1 Goodness-of-fit plot of the MonoLoop procedure for BEIJING2020

The weights w_1-w_{13} obtained by the logistic regression and w_n^* from MonoLoop were then input into the established transition rule to simulate the urban form of BEIJING2020 with 208 iterations and 6297 s of computation time. The simulated urban form shown in Fig. 2 has 10 104 developed cells. The current simulated urban form and the planned form are quite similar. The simulated urban form for 2020 and the forms in various phases from now to 2020 can be used by urban planning organizations.

Figure 3 with the simulations for BEIJING2020 shows that the maximum G appears at iteration 117 (G=97.75%), instead of at the last iteration, 208 (G=97.57%). Iteration 117 corresponds to the year 2019, which illustrates that the urban plan will be actually realized in 2019, rather than in 2020. Therefore, the Beijing Municipal Planning Committee should undertake a new round of urban planning in 2019, beginning far in advance, probably in 2015.

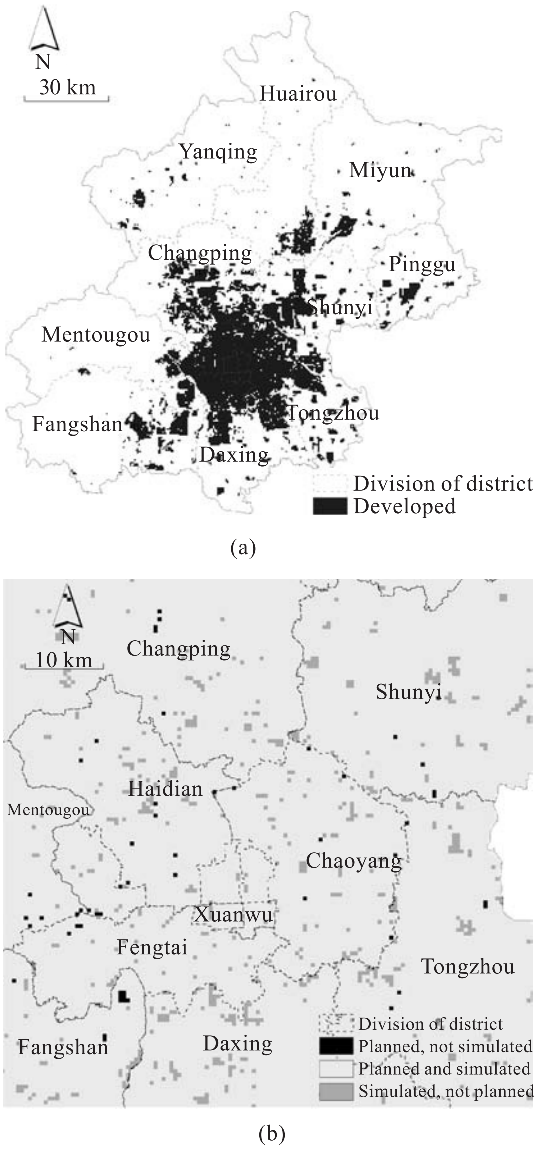

Fig. 2 Simulated urban form for BEIJING2020 (a) and its comparative result with planned form for the Beijing central metropolitan area (b)

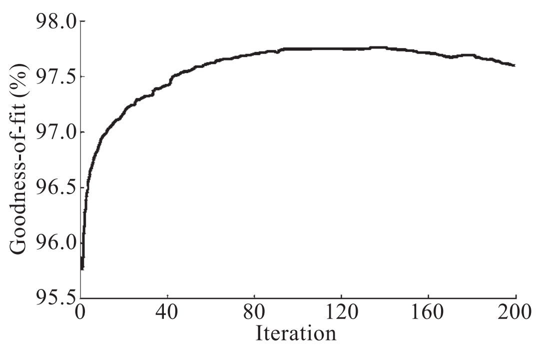

Fig. 3 Goodness-of-fit plot for the BEIJING2020 simulation

The urban growth policies to achieve the planned urban form can be identified by comparing the logistic regression coefficients of BEIJING2020 with those of various historical phases. Comparison of the present policies with achieved results can show whether the present policies and historical ones are consistent. In condition of inconsistence, the achieved parameters of BEIJING2020 can be regarded as new regulations. For instance, comparison of the 2001-2006 coefficients and the achieved parameters suggests that the realization of the planned urban form requires more urban planning implementing intensity, with an emphasis on the urban growth along main roads and increased constructions of important towns. The comparison also demonstrates that the built-up forbidden areas are well protected in the planned scheme and do not need more regulations. The model is capable of generating other urban forms in 2020 by adjusting the obtained policies for BEIJING2020 to evaluate various spatial development policies advanced by the government.

In contrast to the general CA model for urban growth, the MonoLoop process allows point to point validation of the CA, with the accuracy of the point to point comparison reaching 97.6%. G is the key constraint in the MonoLoop of BUDEM, which guarantees the highest model accuracy for the point to point comparison, illustrating the merits of the MonoLoop method. There are many other methods for validating the CA model, such as fractal indexes and spatial structure indexes. Here, the Moran I index (the degree of spatial autocorrelation) was also used to validate the BEIJING2020 simulation results. The Moran I was found to be 0.12 ($Z=31.1$) for the planned form and equal to 0.16 ($Z=43.6$) for the simulated form. Thus, both the simulated form and the planned form are comparatively concentrated, which is a characteristic of planned urban forms instead of natural growth, with the simulated form being more congregated than the planned one.

4 Future Urban Form Prediction Based on the BEIJING2020 Scenario

2020 is the final year of this round of the urban master plan drafted by the Beijing Municipal Planning Committee, while 2049 is the 100th anniversary of Beijing as the capital of P. R. China. Preparations for the next round of urban planning for Beijing will need predictions of the urban form from 2020 to 2049, especially 2049. The prediction should forecast the total amount of urban built-up area and the locations of these built-up areas.

The use of the planned urban form for 2020 will yield more accurate forecasts of the Beijing urban form in 2049 than the use of the observed urban form in 2006, based on the assumption that the urban plan of BEIJING2020 is more likely. In China, land development is strongly controlled by the government by means of the urban planning, so the urban planning can explain most of the urban development activities which will be located in areas planned for urban built-up. Therefore, long-term urban growth forecasts should use the planned urban form for an intermediate year to reduce the uncertainty of long-term forecasting. In some western countries, land use is more controlled by private owners and the urban development is less controlled by the government than in China. Therefore, the present forecast of the urban form for 2049 in the paper is based on the BEIJING2020 planned form.

4.1 Sensitivity analysis

To forecast urban form scenarios in 2049, the model parameters had to be adjusted, especially the weights of the spatial variables. Therefore, the parameter sensitivity must be analyzed to better understand the influence of each parameter on the model output. Generally, the two sensitivity analysis methods analyze the parameter sensitivity to the status and objectives, or analyze the sensitivity of the status to the objective. The first sensitivity analysis method is used here to emphasize the influence of the status transition rule on simulating the status variables and objectives.

The adjustable parameters in the model are w_1-w_{13}, w_n, α, k, and $p_{\text{threshold}}$. The base parameter set (BPS) will use the same w_1-w_{13} and w_n as in the BEIJING2020 scenario with α=3, k=20, $p_{\text{threshold}}$=0.99, and 10 iterations. The influence of each parameter will then be calculated by adjusting each parameter individually in the simulations. Five indicators are used to describe the influence, including x (the number of cells that change), S_{ave} (the average s_{ij} in all the cells), $P_{\text{g,ave}}$ (the average p_{g} in all the cells), P_{ave} (the average p_{ij} in all the cells), and G. The sensitivity analysis has three steps:

(1) Simulations with BPS are used to calculate each indicator's value $w_{m,\text{BPS}}$, where m=1,2,3,4,5 for the five indicators.

(2) Each parameter n in BPS is individually multiplied by 0.8 while holding the other parameters constant to get each indicator's value $w_{m,n}$, where n=1,…,17 for the 17 parameters.

(3) Calculate the sensitivity of parameter n to the indicator m: $U_{m,n}=\left|\frac{w_{m,n}}{w_{m,\text{BPS}}}-1\right|$. The sensitivity analysis results are listed in Table 6.

Table 6 Sensitivity analysis results

Parameter	$U_{x,n}$	$U_{S_{\text{ave}},n}$	$U_{P_{\text{g,ave}},n}$	$U_{P_{\text{ave}},n}$	$U_{G,n}$	SUM
A	0.018	0.000	0.005	0.321	0.003	0.347
k	0.001	0.000	0.000	0.001	0.000	0.002
$p_{\text{threshold}}$	0.171	0.004	0.088	0.031	0.013	0.308
w_{isrural}	0.006	0.004	0.013	0.001	0.001	0.024
w_{isagri}	0.088	0.031	0.103	0.030	0.012	0.265
$w_{\text{d\_tam}}$	0.040	0.108	0.138	0.095	0.005	0.385
$w_{\text{d\_vcity}}$	0.005	0.022	0.029	0.022	0.001	0.078
$w_{\text{d\_city}}$	0.033	0.047	0.075	0.052	0.004	0.212
$w_{\text{d\_vtown}}$	0.054	0.067	0.163	0.141	0.007	0.433
$w_{\text{d\_town}}$	0.018	0.017	0.042	0.030	0.002	0.110
$w_{\text{d\_river}}$	0.023	0.031	0.049	0.036	0.003	0.141
$w_{\text{d\_road}}$	0.004	0.041	0.019	0.013	0.001	0.078
$w_{\text{d\_bdtown}}$	0.005	0.006	0.008	0.006	0.001	0.026
w_{planning}	0.146	0.019	0.210	0.166	0.020	0.562
$w_{\text{con\_f}}$	0.002	0.002	0.003	0.002	0.000	0.009
$w_{\text{landresource}}$	0.005	0.007	0.010	0.008	0.001	0.031
w_{neighbor}	0.013	0.011	0.112	0.100	0.002	0.238
SUM	0.633	0.418	1.068	1.055	0.076	

The results in Table 6 show that the weight for planning is the most sensitive, followed by the weight of d_vtown, with k being the least sensitive. Regarding the indicators, $P_{\text{g,ave}}$ is the most sensitive, while G is the least sensitive. For indicator x, $p_{\text{threshold}}$ is the most sensitive, followed by the weight for planning. For indicator G, planning is the most sensitive and then $p_{\text{threshold}}$. The sensitivity analysis then gives a basis for evaluating the model's uncertainties. The input and output relationships could be better understood by tracking the trends with the iteration number. For example $\mathrm{d}x/\mathrm{d}t$, $\mathrm{d}P_{\text{ave}}/\mathrm{d}t$, and $\mathrm{d}G/\mathrm{d}t$ could be plotted for each iteration for analysis.

4.2 BEIJING2049

The predicted total urban built-up in the Beijing metropolitan area in 2049 was estimated using Eq. (4).

$$y = a + b \cdot (x - c)^d \quad (4)$$

where a=2344 km^2 (total planned urban built-up area in 2020), b=30 km^2/a, c=2020, x is the forecast year, and y is the forecast urban built-up area (km^2). d is used to adjust the urban built-up developing speed.

After forecasting the total urban built-up area, the CA is used to simulate the layout of urban built-up areas. The transition rule indicates the impact of the corresponding urban growth policy (0 indicates the policy is not considered). The spatial variables also indicate urban growth policies, for example, the spatial distribution of con_f indicates for the extent of protected zones set by the policy. Thus, different urban growth scenarios for BEIJING2049 can be simulated by adjusting the weights or the variable spatial distributions. The possible scenarios can be urban sprawl, grape-clusters, sustainable development, center city focused, town focused, riverside development, roadside development, agricultural land protection, rural built-up control, etc. The Moron I index, agricultural land encroachment, forbidden built-up area encroachment, and rural built-up area can all be used to assess different scenarios to identify the effects of different policies. In this paper only the urban growth scenario with the same trends as BEIJING2020 is considered (d=1), with the same parameter set as in BEIJING2020. The simulated result is shown in Fig. 4, which shows that the urban growth in 2049 includes some forbidden areas with urban sprawl. Thus, the urban growth policies from 2020 to 2049 must be adjusted to generate a more sustainable urban form to avoid such problems.

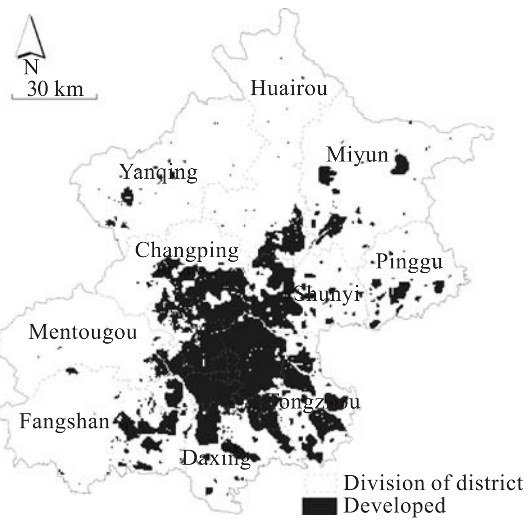

Fig. 4 Simulated urban form for BEIJING2049

5 Conclusions

This paper presents a CA urban growth analysis and simulation model, BUDEM, for the Beijing metropolitan area to analyze and simulate future urban form. The urban growth analysis and simulation platform based on an extensive database can identify urban growth mechanisms in various historical phases. A sensitivity analysis is used to evaluate the effects of the various model parameters on future urban growth predictions. Moreover, logistic regression was integrated with MonoLoop to derive the transition rule for the desired form of BEIJING2020. The weights were compared with historical weights to evaluate the effect of urban planning policies. The urban form until 2049 was also simulated to show that the CA can be applied to evaluate urban planning influences on urban growth and predict future urban growth.

Further work will use the model to analyze other related urban planning management policies by simulating urban growth. Updated BUDEM will be developed to simulate urban density effects and competing land use in parcel scale. Meanwhile, an agent-based module employing the discrete choice model will be added to simulate urban agents for various activities based on different urban policy scenarios.

References

[1] Tobler W R. A computer movie simulating population growth in the Detroit region. *Economic Geography*, 1970, **42**: 234-240.

[2] Couclelis H. Cellular worlds: A framework for modeling micro-macro dynamics. *Environment and Planning A*, 1985, **17**: 585-596.

[3] Couclelis H. Of mice and men: What rodent populations can teach us about complex spatial dynamics. *Environment and Planning A*, 1988, **29**: 99-109.

[4] Couclelis H. Macrostructure and microbehavior in a metropolitan area. *Environment and Planning B*: *Planning and Design*, 1989, **16**: 141-154.

[5] Couclelis H. From cellular automata models to urban models: New principles for model development and implementation. *Environment and Planning B*: *Planning and Design*, 1997, **24**: 165-174.

[6] White R W, Engelen G. Cellular automata and fractal urban form: A cellular modeling approach to the evolution of

urban land use patterns. *Environment and Planning A*, 1993, **25**: 1175-1193.

[7] White R W, Engelen G. Cellular automaton as the basis of integrated dynamic regional modeling. *Environment and Planning B*: *Planning and Design*, 1997, **24**: 235-246.

[8] Clark K C, Gaydos L J. Loose-coupling a cellular automation model and GIS: Long-term urban growth prediction for San Francisco and Washington/Baltimore. *Geographical Information Sciences*, 1998, **12**(7): 699-714.

[9] Batty M. Cities as fractals: Simulating growth and form. In: Crilly T, Earnshaw R A, Jones H, eds. Fractals and Chaos. New York: Springer-Verlag, 1991: 41-69.

[10] Batty M, Xie Y. From cells to cities. *Environment and Planning B*: *Planning and Design*, 1994, **21**: 31-48.

[11] Batty M, Xie Y. Preliminary evidence for a theory of the fractal city. *Environment and Planning A*, 1996, **28**: 1745-1762.

[12] Batty M. Cellular automata and urban form: A primer. *Journal of the American Planning Association*, 1997, **63**: 266-274.

[13] Batty M, Xie Y. Possible urban automata. *Environment and Planning B*: *Planning and Design*, 1997, **24**: 175-192.

[14] Batty M, Couclelis H, Eichen M. Urban systems as cellular automata. *Environment and Planning B*: *Planning and Design*, 1997, **24**: 159-164.

[15] Batty M. Urban evolution on the desktop: Simulation using extended cellular automata. *Environment and Planning A*, 1998, **30**: 1943-1967.

[16] Batty M, Xie Y, Sun Z. Modeling urban dynamics through GIS-based cellular automata. *Computers*, *Environment and Urban Systems*, 1999, **23**: 205-233.

[17] Xie Y. Analytical models and algorithms for cellular urban dynamics [Dissertation]. State University of New York at Buffalo, NY, USA, 1994.

[18] Wu F. Simland: A prototype to simulate land conversion through the integrated GIS and CA with AHP-derived transition rules. *International Journal of Geographical Information Science*, 1998, **12**(1): 63-82.

[19] Wu F, Webster C J. Simulation of land development through the integration of cellular automata and multicriteria evaluation. *Environment and Planning B*: *Planning and Design*, 1998, **25**: 103-126.

[20] Wu F. Calibration of stochastic cellular automata: The application to rural-urban land conversions. *International Journal of Geographical Information Science*, 2002, **16**(8): 795-818.

[21] Li X, Yeh A G O. Principal component analysis of stacked multi-temporal images for monitoring of rapid urban expansion in the Pearl River delta. *International Journal of Remote Sensing*, 1998, **19**(8): 1501-1518.

[22] Li X, Yeh A G O. Modeling sustainable urban development by the integration of constrained cellular automata and GIS. *International Journal of Geographical Information Science*, 2000, **14**(2): 131-152.

[23] Li X, Yeh A G O. Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographical Information Science*, 2002, **16**(4): 323-343.

[24] Li X, Yeh A G O. Data mining of cellular automata's transition rules. *International Journal of Geographical Information Science*, 2004, **18**(8): 723-744.

[25] Yeh A G O, Li X. A constrained CA model for the simulation and planning of sustainable urban forms by using GIS. *Environment and Planning B*: *Planning and Design*, 2001, **28**: 733-753.

[26] Yeh A G O, Li X. A cellular automata model to simulate development density for urban planning. *Environment and Planning B*: *Planning and Design*, 2002, **29**: 431-450.

[27] Yeh A G O, Li X. Errors and uncertainties in urban cellular automata. *Computers*, *Environment and Urban Systems*, 2006, **30**(1): 10-28.

[28] Wu X B, Zhao J. Simulation urban growth using cellular automata: One case at Haikou. *Urban Planning*, 2002, **8**: 69-73. (in Chinese)

[29] Luo M H. Motivity of spatial urban expansion by cellular automata model: Taking analysis of Wuhan spatial urban expansion as an example. *Geomatics and Information Science of Wuhan University*, 2005, **30**(1): 51-55. (in Chinese)

[30] Li S L. Urban simulating model using CA: Case study at Fuzhou city [Dissertation]. Fuzhou University, China, 2005. (in Chinese)

[31] Guo P, Xue H F, Zhao N, et al. Complex adaptive system theory and cellular automata model in simulating urban growth. *Geography and Geo-information Science*, 2004, **20**(6): 69-73. (in Chinese)

[32] He C Y, Okada Norio, Zhang Q F, et al. Modeling urban expansion scenarios by coupling cellular automata model and system dynamic model in Beijing, China. *Applied Geography*, 2006, **26**: 323-345.

[33] Shen T Y, Wang W D, Hou M, et al. Study on

spatio-temporal system dynamic models of urban growth. *Theory and Practice of Systematic Engineering*, 2007, **1**: 10-17. (in Chinese)

[34] Stern P C, Young O R, Druckman D. Global environmental change: Understanding the human dimensions. National Research, UK, 1992.

[35] Alonso W. Location and Land Use: Towards a General Theory of Land Rent. Cambridge, MA, USA: Harvard University Press, 1964.

[36] Wu Liangyong. Introduction to Sciences of Human Settlements. Beijing: Architecture Industry Press of China, 2001. (in Chinese)

[37] Lancaster K. A new approach to consumer theory. *Journal of Political Economy*, 1966, **74**(2): 132-157.

[38] Butler R W H. A structural analysis of the Moine Thrust zone between Loch Eriboll and Foinaven, NW Scotland. *Journal of Structural Geology*, 1982, **4**: 19-29.

[39] Landis L D. The California urban future model: A new generation of metropolitan simulation models. *Environment and Planning*: *Planning and Design B*, 1994, **21**(4): 399-420.

[40] Landis L D. Imaging land use futures: Applying the California urban future model. *Journal of American Planning Association*, 1995, **61**(4): 438-457.

[41] Landis L D, Zhang M. The second generation of the California urban future model, Part 1: Model logic and theory. *Environment and Planning B*: *Planning and Design*, 1998, **25**(5): 657-666.

[42] Landis L D, Zhang M. The second generation of the California urban future model, Part 2: Specification and calibration results of the land-use change submodel. *Environment and Planning B*: *Planning and Design*, 1998, **25**(6): 795-824.

[43] Beijing Municipal Planning Committee, Beijing Institute of City Planning, Beijing Academy of Urban Planning. Beijing Urban Planning Atlas (1949-2005), 2006. (in Chinese)

[44] Beijing Planning Commission. Beijing Land Resource. Beijing: Beijing SciTech Press, 1988. (in Chinese)

[45] Beijing Statistical Bureau. Fifty Years of Beijing 1999. Beijing: Chinese Statistic Press, 1999. (in Chinese)