Front. Earth Sci. 2018, 12(2): 299–310
https://doi.org/10.1007/s11707-017-0639-y

RESEARCH ARTICLE

Simulating land-use changes by incorporating spatial autocorrelation and self-organization in CLUE-S modeling: a case study in Zengcheng District, Guangzhou, China

Zhixiong MEI (✉), Hao WU, Shiyun LI

School of Geography, South China Normal University, Guangzhou 510631, China

Abstract The Conversion of Land Use and its Effects at Small regional extent (CLUE-S), which is a widely used model for land-use simulation, utilizes logistic regression to estimate the relationships between land use and its drivers, and thus, predict land-use change probabilities. However, logistic regression disregards possible spatial autocorrelation and self-organization in land-use data. Autologistic regression can depict spatial autocorrelation but cannot address self-organization, while logistic regression by considering only self-organization (NE-logistic regression) fails to capture spatial autocorrelation. Therefore, this study developed a regression (NE-autologistic regression) method, which incorporated both spatial autocorrelation and self-organization, to improve CLUE-S. The Zengcheng District of Guangzhou, China was selected as the study area. The land-use data of 2001, 2005, and 2009, as well as 10 typical driving factors, were used to validate the proposed regression method and the improved CLUE-S model. Then, three future land-use scenarios in 2020: the natural growth scenario, ecological protection scenario, and economic development scenario, were simulated using the improved model. Validation results showed that NE-autologistic regression performed better than logistic regression, autologistic regression, and NE-logistic regression in predicting land-use change probabilities. The spatial allocation accuracy and kappa values of NE-autologistic-CLUE-S were higher than those of logistic-CLUE-S, autologistic-CLUE-S, and NE-logistic-CLUE-S for the simulations of two periods, 2001–2009 and 2005–2009, which proved that the improved CLUE-S model achieved the best simulation and was thereby effective to a certain extent. The scenario simulation results indicated that under all three scenarios, traffic land and residential/industrial land would increase, whereas arable land and unused land would decrease during 2009–2020. Apparent differences also existed in the simulated change sizes and locations of each land-use type under different scenarios. The results not only demonstrate the validity of the improved model but also provide a valuable reference for relevant policy-makers.

Keywords CLUE-S, land-use change simulation, spatial autocorrelation, self-organization

1 Introduction

Land-use change is the result of multiple interacting processes driven by socioeconomic and biophysical factors and has significant effects on the eco-environment (Verburg et al., 2002; Jiang et al., 2011; Lin et al., 2014). Land-use change models play an important role in analyzing the driving forces, processes, and effects of land-use change, as well as in exploring future land-use changes under different scenarios (Verburg et al., 2004b; Liu and Deng, 2010; Lin et al., 2011; Liu et al., 2011). Hence, modeling land-use changes can help in understanding current land-use dynamics and in formulating future land-use policies (Lin et al., 2014; Luo et al., 2014). Currently, several models, such as cellular automata (CA), agent-based models, the Conversion of Land Use and its Effects (CLUE) model, and CLUE at Small regional extent (CLUE-S) model, have been widely used to detect land-use change.

CA provide an effective way of simulating complex geographical phenomena (Liu et al., 2008a, b) and have increasingly been used to simulate urban expansion and land-use change (Wu, 2002; Verburg et al., 2004a; Stevens and Dragicevic, 2007; Liu et al., 2010, 2014b). However, although appropriate for short-term projections, CA cannot

Received July 18, 2016; accepted February 3, 2017

E-mail: zhixiongmei76@126.com

predict changes when the demands for different land-use types change; moreover, CA can only simulate the conversion of one land-use type (Verburg et al., 2002; Zheng et al., 2015). Agent-based models simulate decision-making by individual agents of land-use change explicitly addressing interactions among individuals (Verburg et al., 2004b), and thus they can precisely reflect the role of human decision-making in land-use change (Zheng et al., 2015). However, the characterization of rules for different decision-making processes of land use and planning policies is difficult (Luo et al., 2014), and gathering sufficient data at the individual level is challenging to validate an agent-based model (Zheng et al., 2015).

The CLUE-S model, developed from the CLUE model proposed by Veldkamp and Fresco (1996), was specifically presented by Verburg et al. (2002) to simulate small-scale land-use change based on an empirical analysis of location suitability combined with dynamic modeling (Li et al., 2012; Lin et al., 2014). Unlike most empirical models, CLUE-S can simulate multiple land-use types simultaneously by addressing competition between land-use types (Verburg et al., 2002; Hu et al., 2013). It has been widely applied in simulating land-use changes at the regional level (Castella and Verburg, 2007; Overmars et al., 2007; Liu et al., 2009; Luo et al., 2010; Park et al., 2011; Zheng et al., 2012; Zhang et al., 2013; Yang et al., 2014; Wu et al., 2015; Zheng et al., 2015).

CLUE-S includes a non-spatial demand module and a spatially explicit allocation procedure. The former calculates the total area demands for all land-uses, while the latter translates these demands into land-use changes at various locations based on the quantified relationships between land use and its drivers and pre-specified decision rules (Verburg et al., 2002). In CLUE-S, logistic regression is often used to fit the relationships between land use and its drivers, and thereby calculate the probability of land-use change for each grid cell prior to land allocation. However, logistic regression disregards both spatial autocorrelation and self-organization effects that may occur in land-use data (Wear and Bolstad, 1998; Wu, 2002; Overmars et al., 2003; Verburg, 2006; Syartinilia and Tsuyuki, 2008); hence, the CLUE-S model based on logistic regression has difficulties in simulating land-use changes with significant spatial autocorrelation and self-organization.

The autologistic regression proposed by Besag (1972) has been recently integrated into CLUE-S to address the spatial autocorrelation effect (Wu et al., 2010a; Lin et al., 2011; Dai and Zhang, 2013; Jiang et al., 2015). In addition, several studies have introduced components that depict self-organization in logistic regression (Duan et al., 2004; Verburg et al., 2004a; Pan et al., 2011; Wang et al., 2014) to increase the simulation accuracy of CLUE-S. However, few studies have simultaneously addressed the above two effects in CLUE-S modeling to simulate land-use changes.

This study mainly aims to emphasize the importance of considering the spatial autocorrelation and self-organization effects in the context of CLUE-S modeling for simulating land-use changes. First, a regression method (NE-autologistic regression) was developed to improve CLUE-S by simultaneously incorporating a spatial autocorrelation variable (Wu et al., 2010b; Jiang et al., 2015) and components that depict self-organization (Duan et al., 2004; Verburg et al., 2004a) into logistic regression. Then, the Zengcheng District of Guangzhou, China was selected as the study area. The land-use data of 2001, 2005, and 2009, as well as a set of factors that drive land-use changes were used to demonstrate that NE-autologistic regression was superior to logistic regression, autologistic regression, and logistic regression by considering only self-organization (NE-logistic regression) in predicting the probabilities of land-use changes with spatial autocorrelation and self-organization, and thereby to validate the improved CLUE-S model. Finally, land-use patterns in 2020 were simulated under three different scenarios using the improved model, and land-use changes from 2009 to 2020 were analyzed under different scenarios to provide a valuable reference for relevant policy-makers.

2 Study area and data

2.1 Study area

The Zengcheng District (between 23°05′–23°37′N and 113°32′–114°00′E), which has an area of 1614.83 km^2, is located in the east of Guangzhou, China and governs six towns and three sub-districts (Fig. 1). The terrain of Zengcheng mainly consists of low hills and alluvial plains, and the terrain is low in the south and high in the north. Zengcheng has experienced rapid urbanization and economic growth since 2001 as a result of the urban development strategy of Guangzhou, known as "eastern advance, western union, southern expansion, northern optimization, and central region adjustment". This process has led to significant land-use changes in this area. In particular, arable land and woodland have been reduced, whereas land under construction has increased. The Overall Plan for Land Use in Zengcheng (2010–2020) suggests that the protection of arable land and other agricultural lands is severe, the contradiction between supply and demand for construction land is prominent, and the coordination of land uses is facing a considerable challenge in Zengcheng. Therefore, the simulation of land-use changes in this area is significant for the rational planning and management of regional land resources.

2.2 Data sources and processing

The data sets used in this study include land-use data, topographic data, and socioeconomic data. The 2005 and 2009 land-use change survey data of the study area were

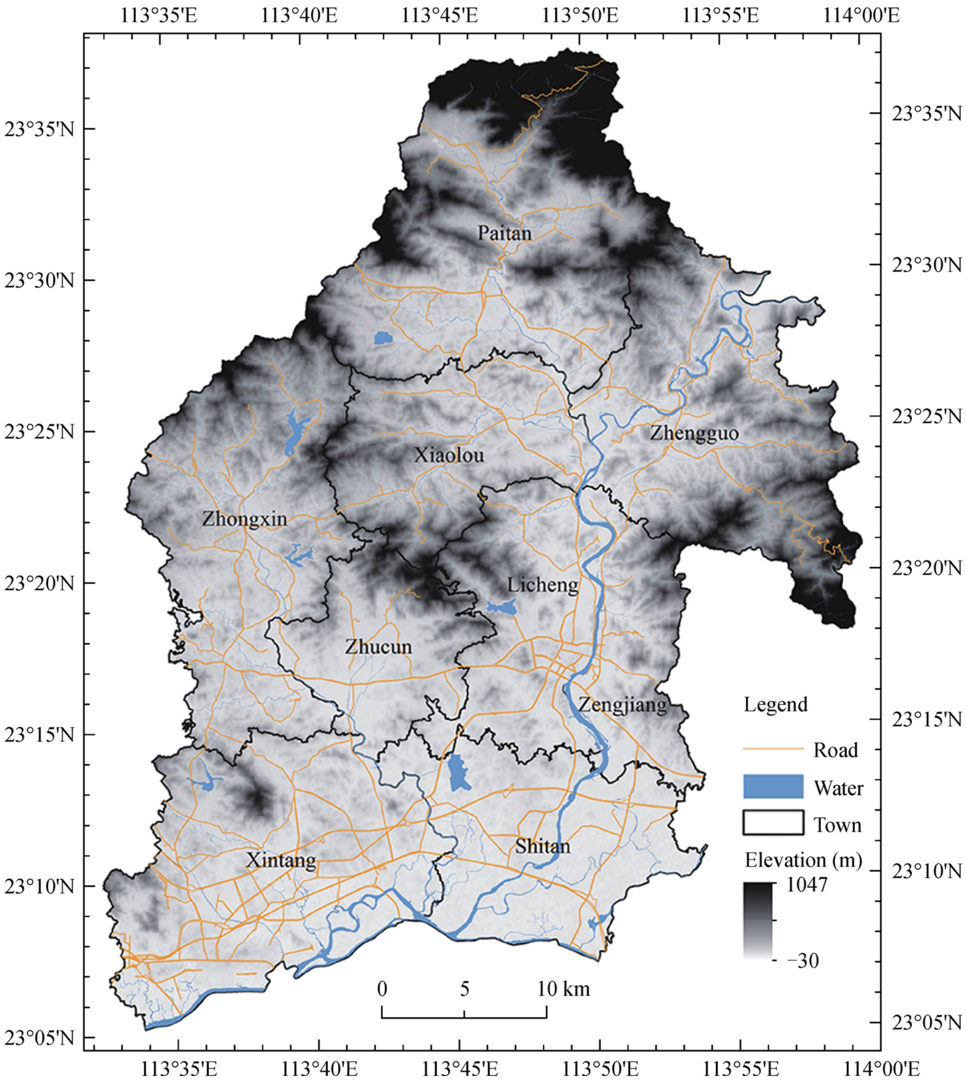

Fig. 1 Location of the study area.

obtained from the Guangzhou Municipal Land Resources and Housing Administrative Bureau. To fulfill the requirement for CLUE-S modeling, namely, that the area proportion of each considered land-use type in the region must be over 1%, the original land-use data were reclassified and grouped into six types: arable land, woodland, residential/industrial land, traffic land, water area, and unused land. In addition, the 2001 land-use data derived from the Landsat ETM+ image were integrated into the aforementioned six types. The 1:50,000 digital elevation model (DEM) with a resolution of 30 m for this area in 2009 was derived from the International Scientific Data Service Platform (http://www.datamirror.csdb.cn). The relevant socioeconomic data were collected from the Zengcheng Statistical Yearbooks in 2001, 2005, and 2009.

Ten typical biophysical and socioeconomic factors that were assumed to affect land-use changes were selected according to knowledge of the study area, data availability, and related studies. These factors were distance to the nearest road (*DRoad*), distance to the nearest water (*DWater*), distance to the nearest rural settlement center (*DRSettlement*), distance to the nearest town center (*DTown*), population density (*PDens*), residential site density (*RSDens*), elevation (*elevation*), slope (*slope*), aspect (*aspect*), and curvature (*curvature*). The data for the *DRoad*, *DWater*, *DRSettlement*, and *DTown* variables were generated from the land-use data. The data for the *elevation*, *slope*, *aspect*, and *curvature* variables were extracted from the DEM. We set 20 resolutions ranging 50 m to 1000 m, with 50 m steps, for comparisons because the CLUE-S model was based on high-resolution (typically higher than 1 km) spatial data (Verburg et al., 2002; Hu et al., 2013). The original *PDens* and *RSDens* data collected based on administrative areas (towns/sub-districts) were interpolated into raster data of the 20 resolutions using the Inverse Distance Weighted method, which was proven useful in the spatialization of socioeconomic statistical data (Li et al., 2012; Dai and Zhang, 2013). All the other driving factors and spatial data of the six land-use types were also rasterized to the 20 resolutions and transformed into ASCII formats to satisfy the input requirement of CLUE-S. Experiment results showed that the best simulation resolution in the study area was 150 m. Therefore, the simulation spatial scale was set as 150 m × 150 m in this study.

The raster data of 150 m resolution for all land-use types and driving factors were analyzed using logistic, autologistic, NE-logistic, and NE-autologistic regression methods to calculate land-use change probabilities. The 2009 land-use pattern was simulated based, respectively, on the 2001 and 2005 land-use data by individually applying the results obtained using the four methods in CLUE-S modeling. The relative operating characteristic (ROC) method was used to assess the efficiency of the four regression models. The proportion of correctly simulated cells and the kappa index were used to evaluate the simulation accuracy by comparing the simulated maps with the actual land-use map in 2009.

3 Methods

3.1 CLUE-S model parameter settings

CLUE-S modeling requires setting the following parameters prior to land-use allocation.

1) Land-use demands. In CLUE-S modeling, the area demands for different land-use types require prediction using other methods. In this study, the annual land-use demands for 2002–2004 and 2006–2008 were calculated through linear interpolation based on the land-use data in 2001, 2005, and 2009. Future land-use demands under different scenarios were set according to varying conditions.

2) Spatial characteristics. The analysis of spatial characteristics is the core of CLUE-S, which aims to determine the relationships between land use and its drivers, and consequently calculate the probability of occurrence of each land-use type at each grid cell (i.e., location suitability). Logistic regression is often used in the CLUE-S model to address this issue. The general form of logistic regression is

$$\log\left(\frac{P_{i,k}}{1-P_{i,k}}\right)=\beta_{0,i}+\sum_{n=1}^{m}\beta_{n,i}X_{n,k}, \tag{1}$$

where $P_{i,k}$ is the probability of land-use type i occurring in cell k; $X_{n,k}$ is the value of the n^{th} driving factor in cell k; $\beta_{0,i}$ is the regression constant; and $\beta_{n,i}$ is the estimated coefficient of the n^{th} driving factor for land-use type i.

In addition to logistic regression, we also used autologistic, NE-logistic, and NE-autologistic regression methods to analyze the spatial characteristics of CLUE-S modeling to show the possible spatial autocorrelation and self-organization effects in land-use data and compared the precision of these methods in evaluating land-use suitability.

3) Land-use conversion rules and elasticity. The conversion rules for the CLUE-S model account for the possibility of conversion between different land-use types via a conversion matrix. The six land-use types selected for this study can be converted into one another based on actual conversions from 2001 to 2009. The conversion elasticity (ELAS) reflects the stability degree of a land-use type and is typically set based on expert knowledge of actual situations and reiterative model calibration ranging from 0 to 1 (Verburg et al., 2002; Hu et al., 2013). A high ELAS value for a land-use type indicates difficult conversion from this type to other types. In this study, the ELAS parameters for arable land, woodland, residential/industrial land, traffic land, water area, and unused land in 2001–2009 were set as 0.8, 0.8, 0.9, 0.9, 0.8, and 0.8, respectively, while those in 2009–2020 were tuned according to the defined scenarios.

4) Spatial policies and restrictions. Spatial policies and restrictions can be set in CLUE-S to specify restricted areas where land-use changes are prohibited. In this study, six areas, namely, Gaotang, Lian'an, Baidong, Baihualin, Yujia, and Zengtang Reservoirs, were defined as restricted areas given the actual situations in Zengcheng.

3.2 Autologistic regression

The autologistic regression model incorporates an autocorrelation variable into logistic regression to account for spatial autocorrelation (Wu et al., 2010b; Lin et al., 2011; Dai and Zhang, 2013; Jiang et al., 2015). This model can be expressed as

$$\log\left(\frac{P_{i,k}}{1-P_{i,k}}\right)=\beta_{0,i}+\sum_{n=1}^{m}\beta_{n,i}X_{n,k}+\beta'\,Autocov_{i,k}, \tag{2}$$

where $Autocov_{i,k}$ is the spatial autocorrelation variable of land-use type i in grid k and is determined using Eq. (3).

$$Autocov_{i,k}=\left(\sum_{j=1}^{N_k}y_{i,j}w_{j,k}\right)/\sum_{j=1}^{N_k}w_{j,k}, \tag{3}$$

where N_k is the number of cells adjacent to cell k; $y_{i,j}$ denotes the presence or absence of land-use type i in cell j (presence: 1; absence: 0); and $w_{j,k}$ is the spatial weight coefficient measured via the inverse of the Euclidean distance between cells j and k (Wu et al., 2010b; Jiang et al., 2015).

3.3 NE-logistic and NE-autologistic regression models

Land-use change processes usually include spontaneous process, self-organizing process, and the competition process between land-use types (Chen et al., 2002; Duan et al., 2004). CLUE-S can simulate the spontaneous and competition processes (Verburg et al., 2002) ; however, it cannot address the self-organizing process (Duan et al., 2004; Wang et al., 2014). For land-use systems in urban and urban fringe areas, such as Zengcheng District, the three processes should be considered simultaneously (Duan et al., 2004). Duan et al. (2004) and Verburg et al. (2004a) indicated that neighborhood enrichment factors could represent the self-organization of land-use change. The neighborhood enrichment factor ($F_{k,i,d}$) of land-use type i in the neighborhood d of cell k is defined as (Duan et al., 2004; Verburg et al., 2004a)

$$F_{k,i,d}=\frac{n_{i,d,k}/n_{d,k}}{N_i/N}, \tag{4}$$

where d is the radius of the neighborhood from the central cell k ($d = 1$ denotes a 3 × 3 neighborhood, $d = 2$ denotes a 5 × 5 neighborhood, …); $n_{i,d,k}$ is the number of cells of land-use type i in the neighborhood d of cell k; $n_{d,k}$ is the total number of cells in the neighborhood area; N_i is the number of cells for land-use type i in the entire study area; and N is the number of all cells in the study area.

$F_{k,i,d}$ reflects the proportion of land-use type i occurring in the neighborhood d of cell k relative to the proportion of this land-use type in the entire study area; that is, the relative enrichment of land-use type i in the neighborhood d of cell k, and its details can be found in Verburg et al. (2004a).

The introduction of neighborhood enrichment factors into Eq. (1) generates a regression model (denoted as NE-logistic regression) that considers self-organization, as shown in Eq. (5). We incorporated the enrichment factors into Eq. (2) to establish a new regression model (denoted as NE-autologistic regression), as indicated in Eq. (6).

$$\log\left(\frac{P_{i,k}}{1-P_{i,k}}\right)=\beta_{0,i}+\sum_{n=1}^{m}\beta_{n,i}X_{n,k}+\sum_{i=1}^{N_0}\beta''_{i,d}F_{k,i,d}, \tag{5}$$

and

$$\log\left(\frac{P_{i,k}}{1-P_{i,k}}\right)=\beta_{0,i}+\sum_{n=1}^{m}\beta_{n,i}X_{n,k}+\beta'\,Autocov_{i,k}+\sum_{i=1}^{N_0}\beta''_{i,d}F_{k,i,d}, \tag{6}$$

where $\beta''_{i,d}$ is the coefficient to be estimated for $F_{k,i,d}$; and N_0 is the number of land-use types, which is equal to six in this study (code 1: arable land; code 2: woodland; code 3: traffic land; code 4: residential/industrial land; code 5: water area; and code 6: unused land).

3.4 Evaluation of regression and simulation results

The ROC method is typically used to evaluate the goodness of fit of a logistic regression model (Pontius and Schneider, 2001). The ROC value is between 0.5 and 1. The higher the ROC value, the better the fit of a regression model. An ROC value above 0.7 generally implies good explanatory power and reliable prediction accuracy of a regression model (Hu et al., 2013; Liu et al., 2014a), while an ROC value above 0.9 indicates excellent model fit (Lin et al., 2011; Jiang et al., 2015).

The kappa index is frequently used to assess the simulation accuracy of land-use change models (Lin et al., 2011; Hu et al., 2013; Jiang et al., 2015). This index measures the degree of agreement between simulated and actual land-use maps (Lin et al., 2011; Zheng et al., 2015), which is generally calculated using

$$\text{Kappa}=(P_0-P_c)/(1-P_c), \tag{7}$$

where P_0 is the simulated proportion of agreement, and P_c is the expected proportion of agreement resulting from chance and is equal to $1/n$ (n is the number of land-use types). The kappa value varies between 0 and 1. When the kappa value exceeds 0.7, the simulation results can be accepted; a higher kappa value represents a more accurate simulation (Liu et al., 2014a).

4 Results

4.1 Comparison of logistic, autologistic, NE-logistic, and NE-autologistic regression results

The value of the spatial autocorrelation variable *Autocov* of each land-use type was calculated using Eq. (3). The NE-autologistic regression trials at different distances (d) of neighborhood showed that at a distance of two cells, the resulting ROC values for all land-use types were all the highest (i.e., best model fit). Hence, d = 2 (5 × 5 neighborhood) was taken in Eq. (4) to calculate the enrichment factors. Then, the logistic, autologistic, NE-logistic, and NE-autologistic regression models were used to estimate the coefficients between land uses and their drivers (Tables 1, 2, 3, and 4; where "–" denoted that "variables were insignificant at the 0.05 level and were excluded") and to predict land-use change probabilities.

Tables 1, 2, 3, and 4 show the following results. 1) All the ROC values of the four regression models for all the land-use types were above 0.7, which indicated that the spatial patterns of all the land-use types could be well explained by the identified driving factors. 2) The ROC values of logistic regression for all the land-use types were lower than those of the other three regression models, thereby suggesting that autologistic regression, NE-logistic

Table 1 Logistic regression results for different land-use types

Variable	Arable land	Woodland	Traffic land	Residential/industrial land	Water area	Unused land
DRoad/km	0.001	0.002	–0.013	–0.006	0.002	0.001
DWater/km	0.001	0.002	–	–	–0.001	–
DRSettlement/km	–0.004	0.001	0.002	–0.006	0.006	0.003
DTown/km	0.001	–	–	–0.003	0.001	–
PDens/$(\text{people}\cdot\text{km}^{-2})$	–0.020	0.007	0.008	0.020	0.015	–0.007
RSDens/%	–	–7.419	–	–4.318	–	12.377
Elevation/m	–0.177	0.068	–	–0.064	–0.050	–0.064
Slope/(°)	–0.241	0.175	–0.072	–0.064	–0.232	–0.135
Aspect/(°)	–0.024	0.037	–	0.005	–0.021	–0.008
Curvature	–3.741	0.460	–0.911	–	–2.823	–2.020
Constant	0.815	–1.106	–3.031	–1.165	–2.676	–2.548
ROC	0.818	0.848	0.757	0.834	0.824	0.782

Table 2 Autologistic regression results for different land-use types

Variable	Arable land	Woodland	Traffic land	Residential/industrial land	Water area	Unused land
DRoad/km	0.001	0.002	–0.009	–	0.001	0.001
DWater/km	–	0.001	–	–0.001	–0.003	–
DRSettlement/km	0.003	0.004	0.003	0.003	0.004	0.004
DTown/km	–	–	–	–0.001	–	–
PDens/$(\text{people}\cdot\text{km}^{-2})$	–0.032	–	0.010	0.015	0.004	–
RSDens/%	–	– 3.608	–	–3.302	–	4.419
Elevation/m	–0.086	0.047	–	–0.026	–0.023	–0.037
Slope/(°)	–0.121	0.134	–0.030	–	–0.055	–0.048
Aspect/(°)	–0.017	0.039	–	0.006	–0.013	–0.007
Curvature	–4.483	0.482	–0.911	–	–2.553	–2.197
Autocov	5.822	4.732	3.608	6.657	6.050	4.826
Constant	–5.221	–5.719	–5.601	–7.030	–6.805	–5.800
ROC	0.914	0.883	0.849	0.944	0.933	0.899

regression, and NE-autologistic regression were more efficient than logistic regression in predicting land-use probability distribution. From the ROC values, autologistic regression was superior to NE-logistic regression but slightly inferior to NE-autologistic regression, thereby proving that the regression analysis that considered spatial autocorrelation was more explanatory than the analysis that considered only self-organization for regional land-use patterns. Moreover, the regression analysis that involved both effects provided the highest level of explanation. Accordingly, NE-autologistic regression exhibited the best goodness of fit and the highest prediction accuracy for land-use suitability. 3) Each land-use type was correlated with *Autocov* and several enrichment factors, which showed that both spatial autocorrelation and self-organization effects existed in regional land-use data and played a role in explaining regional land-use patterns. Therefore, the two effects should be considered simultaneously when predicting the suitability of regional land uses.

4.2 Model validation

We simulated the land-use map of 2009 based respectively on the land-use data of 2001 and 2005 by applying the four regression results in CLUE-S modeling separately. We then verified simulation accuracy by comparison between the simulated and actual maps for 2009. The results showed that the overall proportions of correctly simulated cells using logistic-CLUE-S, autologistic-CLUE-S, NE-logistic-CLUE-S, and NE-autologistic-CLUE-S models were 75.46%, 84.32%, 78.26%, and 86.53%, respectively, for the 2001–2009 simulations, and 78.52%, 88.43%, 81.95%, and 89.04%, respectively, for the 2005–2009 simulations. These results proved that the spatial allocation

Table 3 NE-logistic regression results for different land-use types

Variable	Arable land	Woodland	Traffic land	Residential/industrial land	Water area	Unused land
DRoad/km	–	0.001	–0.010	0.001	–	0.001
DWater/km	0.001	–	–	–	–0.001	–
DRSettlement/km	–	–0.002	0.002	–0.005	–	–
DTown/km	–	–	–	–	–	–
PDens/(people·km^{-2})	–0.003	–	0.006	–	–0.011	–
RSDens/%	–3.081	–	–	–2.933	–	1.779
Elevation/m	–0.059	0.023	–	–0.047	–0.014	–0.026
Slope/(°)	–0.114	0.055	–0.054	–	–0.115	–0.069
Aspect/(°)	–0.019	0.037	–	–	–0.016	–0.010
Curvature	–5.310	0.324	–0.784	–	–3.205	–2.723
NE_1	2.621	–0.630	–0.137	–0.337	–0.842	–0.292
NE_2	–1.228	–1.669	–0.311	–1.602	–0.653	–0.603
NE_3	–	–0.056	–1.197	–0.181	–0.112	–
NE_4	–0.178	–0.430	–	1.921	–0.094	–0.193
NE_5	–	–0.343	–0.084	–0.161	–	–
NE_6	–0.142	–0.188	–0.137	–0.202	–0.184	–
Constant	–3.065	–3.392	–4.250	–2.686	–3.056	–4.018
ROC	0.874	0.857	0.820	0.941	0.905	0.786

Note: NE_1, NE_2, NE_3, NE_4, NE_5, and NE_6 represent the neighborhood enrichment factors of arable land, woodland, traffic land, residential/industrial land, water area, and unused land, respectively.

Table 4 NE-autologistic regression results for different land-use types

Variable	Arable land	Woodland	Traffic land	Residential/industrial land	Water area	Unused land
DRoad/km	–	0.001	–0.009	–	–	–
DWater/km	0.001	–	–	–	–0.002	–
DRSettlement/km	–	–0.002	0.002	–0.005	–	–
DTown/km	–	–	–	–	–	–
PDens/(people·km^{-2})	–0.004	–	0.007	–	–0.009	–
RSDens/%	–2.767	–	–	–2.530	–	3.421
Elevation/m	–0.062	0.024	–	–0.039	–0.017	–0.022
Slope/(°)	–0.105	0.055	–0.053	–	–0.175	–0.066
Aspect/(°)	–0.017	0.037	–	–	–0.014	–0.008
Curvature	–4.704	0.283	–	–	–2.412	–2.299
Autocov	4.986	8.617	4.979	5.563	5.925	4.873
NE_1	0.493	–0.631	–0.128	–0.372	–0.866	–0.302
NE_2	–1.334	–2.961	–0.335	–1.784	–0.743	–0.727
NE_3	–	–0.053	–0.633	–0.196	–0.109	–
NE_4	–0.176	–0.435	–	0.173	–0.114	–0.231
NE_5	0.042	–0.344	–0.080	–0.173	–	–
NE_6	–0.134	–0.191	–0.136	–0.223	–0.173	–
Constant	–3.427	–3.404	–4.439	–2.802	–3.289	–4.088
ROC	0.919	0.898	0.854	0.951	0.939	0.899

Note: NE_1, NE_2, NE_3, NE_4, NE_5, and NE_6 represent the neighborhood enrichment factors of arable land, woodland, traffic land, residential/industrial land, water area, and unused land, respectively.

accuracy of NE-autologistic-CLUE-S was the highest for both periods. In addition, the kappa index was used to further assess simulation accuracy. The kappa values of the logistic-CLUE-S, autologistic-CLUE-S, NE-logistic-CLUE-S, and NE-autologistic-CLUE-S models were 0.71, 0.81, 0.74, and 0.84, respectively, for the 2001–2009 simulations, and 0.74, 0.86, 0.78, and 0.87, respectively, for the 2005–2009 simulations. These results indicated that the NE-autologistic-CLUE-S model demonstrated the best performance. As shown in Fig. 2, a strong agreement is observed between the two simulated maps of the NE-autologistic-CLUE-S model and the actual map in 2009. Therefore, the application of NE-autologistic regression to CLUE-S modeling can increase simulation accuracy, and the CLUE-S model improved via this regression is more suitable for simulating land-use changes in Zengcheng.

Interestingly, the improvement in the overall simulation accuracy of the NE-autologistic-CLUE-S model with both spatial autocorrelation and self-organization effects is minor compared with that of autologistic-CLUE-S considering autocorrelation only. The main reason is as follows: although self-organization existed in both woodland and unused land, its influencing degrees on the location suitability of woodland and unused land were low (their ROC values were only increased by 1.06% and 0.51%, respectively, using NE-logistic regression instead of logistic regression). In addition, woodland accounted for over 60% of the total area in Zengcheng during 2001–2009. This reduced the increment rate of the overall accuracy of simulating all land-use types using the NE-autologistic-CLUE-S model to a certain extent. However, the ROC values for arable land, traffic land, residential/industrial land, and water area were increased by 6.85%, 8.32%, 12.83%, and 9.83%, respectively, when using NE-logistic regression instead of logistic regression, which indicates that the self-organization effect significantly contributed to the suitability of all the four land-use types, particularly residential/industrial land. Consequently, self-organization should be considered in addition to spatial autocorrelation when simulating the four land-use types. We did not exclude insignificant self-organization that existed in woodland and unused land in simulating these two land-uses to verify the proposed model as comprehensively as possible.

4.3 Simulation of future land-use changes under different scenarios

4.3.1 Establishment of future scenarios

Three scenarios, namely, the natural growth (NG), ecological protection (EP), and economic development (ED) scenarios, were established in this study to simulate land-use changes in Zengcheng from 2009 to 2020.

1) The NG scenario follows the historical land-use change trend and conversion settings during 2001–2009. In this scenario, the annual area demand for each land-use type from 2009 to 2020 was predicted using the Markov model according to the transition probability matrix of 2001–2009.

2) The EP scenario assumes that land-use changes will obey related ecological protection policies, e.g., arable land, woodland, and water area should be effectively protected, and construction land size should be reasonably controlled. In this scenario, the annual area demand of each land-use type during 2009–2020 was determined by adjusting the results of the NG scenario based on the eco-environmental protection planning policies indicated in the Urban Master Plan of Zengcheng (2008–2020) and the Overall Plan for Land Use in Zengcheng (2010–2020). The ELAS parameters for arable land, woodland, residen-

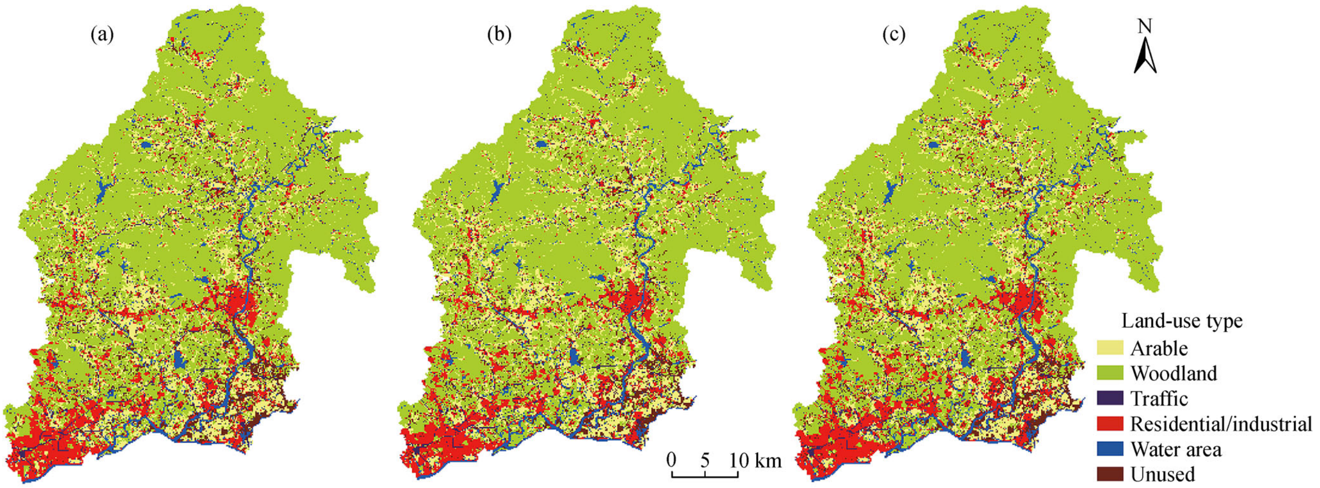

Fig. 2 The actual 2009 land-use map (a), and simulated 2009 maps by NE-autologistic-CLUE-S based on the land-use data of (b) 2001 and (c) 2005.

tial/industrial land, traffic land, water area, and unused land were set as 0.9, 0.9, 0.9, 0.9, 0.9, and 0.7, respectively.

3) The ED scenario aims to ensure sufficient supply of construction land, including residential/industrial land and traffic land, in this study. On the basis of the 13th Five-Year Plan for Economic and Social Development in Zengcheng (2016–2020) and the 2020 Guangzhou Overall Development Strategy, the area proportions of residential/industrial land and traffic land in Zengcheng in 2020 will reach 15% and 5%, respectively, while those of arable land, woodland, water area, and unused land will be 11.24%, 60%, 4.37%, and 4.39%, respectively. The land-use area demands for other years were obtained via linear interpolation. The ELAS parameters for arable land, woodland, residential/industrial land, traffic land, water area, and unused land were set as 0.7, 0.7, 0.9, 0.9, 0.8, and 0.7, respectively.

Table 5 presents the land-use area demands under the three different scenarios in 2020.

4.3.2 Simulation of land use in 2020

After validating the improved CLUE-S model, the validated NE-autologistic-CLUE-S model was used to simulate the future land-use patterns in Zengcheng in 2020 under the three scenarios (Fig. 3).

The comparison of Fig. 3 with Fig. 2(a) indicates that land-use changes from 2009 to 2020 under the three scenarios could be concluded as follows.

1) Under the NG scenario, traffic land and residential/industrial land would continue to increase, and the increased area would mainly come from arable land and woodland, whereas the spatial extents and sizes of arable land and woodland would decrease. Water area and unused land would decrease slightly and would be mainly converted into traffic land. The simulation results indicated that the expansion of traffic land and residential/industrial land would mainly occur in the southwestern part of Xintang Town, the central part of Shitan Town, and the central and eastern parts of Licheng Sub-district. Moreover, the decrease of arable land and woodland in the areas surrounding the towns would be significant.

2) Under the EP scenario, arable land would decrease, but the decrease rate would be considerably less than that under the other two scenarios, and the decreased area would be mainly converted into woodland and partly into traffic land. Unused land would decrease significantly and would be mainly converted into woodland, traffic land, and residential/industrial land, and partly into water area. By contrast, woodland and water area would exhibit a slight increase. The simulation results also indicated that traffic land and residential/industrial land would increase mainly around their existing locations and would have considerably less increase rates under this scenario compared with the other two scenarios. Overall, arable land in the central and northern areas of Zengcheng, as well as woodland and

Table 5 Land-use area (km^2) demands under three scenarios in 2020

Scenario	Arable land	Woodland	Traffic land	Residential/industrial land	Water area	Unused land
NG	196.84	997.22	75.29	192.07	66.88	86.53
EP	221.78	1,013.90	69.56	179.28	71.21	59.10
ED	181.48	968.90	80.74	242.22	70.67	70.82

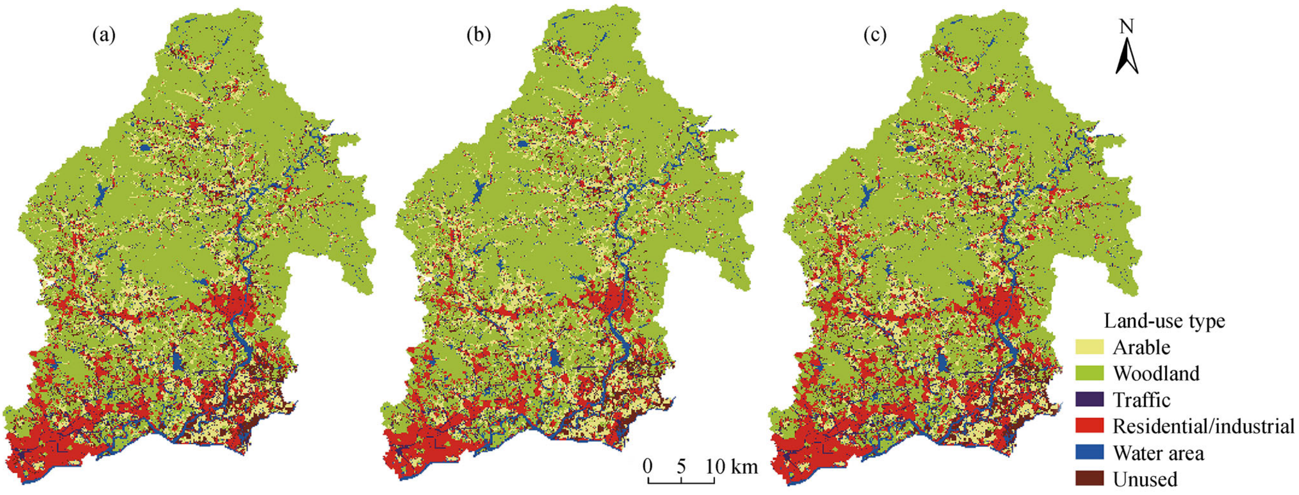

Fig. 3 Simulated land-use maps in Zengcheng in 2020 under three scenarios: (a) natural growth scenario; (b) ecological protection scenario; (c) economic development scenario.

water area in the entire region, would be effectively protected; the sizes of traffic land and residential/industrial land would be reasonably controlled under this scenario.

3) Under the ED scenario, traffic land and residential/industrial land would expand dramatically at the expense of considerable areas of arable land and woodland around towns and rural settlements, thereby resulting in a significant loss of these land-use types. In particular, the residential/industrial land in each town/sub-district would have different degrees of expansion, and such expansion would be especially evident in the southwestern part of Xintang Town and the central junction between Licheng and Zengjiang Sub-districts. However, unused land would decrease substantially, particularly in Shitan Town, and the decreased area would be mainly converted into residential/industrial land. Water area would exhibit an insignificant decrease.

5 Conclusions and discussion

This research developed a NE-autologistic regression method to improve CLUE-S by simultaneously incorporating a spatial autocorrelation variable and neighborhood enrichment factors that depict self-organization into logistic regression. The Zengcheng District of Guangzhou, China was selected as the case study area, and the historical land-use data of 2001, 2005, and 2009 were used to validate the improved CLUE-S model. Then, three future land-use scenarios in 2020, the NG scenario, EP scenario, and ED scenario, were simulated using the improved model. The main conclusions are as follows.

1) In this study, the proposed NE-autologistic regression performed better than logistic regression, NE-logistic regression, and autologistic regression in predicting land-use change probabilities. This result proved that NE-autologistic regression was more feasible to use compared with logistic, autologistic, or NE-logistic regression in modeling land-use changes with both spatial autocorrelation and self-organization effects. It could simultaneously show the two effects in land-use data.

2) The application of NE-autologistic regression to CLUE-S modeling improved simulation accuracy. The comparison of the simulated maps based respectively on land-use data in 2001 and 2005 with the actual land-use map in 2009 validated the improved CLUE-S model. The spatial allocation accuracy and kappa values of NE-autologistic-CLUE-S were higher than the values of logistic-CLUE-S, autologistic-CLUE-S, and NE-logistic-CLUE-S for the two time simulations, thereby demonstrating that the NE-autologistic-CLUE-S model provided the best simulation and NE-autologistic regression could increase the simulation accuracy of CLUE-S. Therefore, the CLUE-S model improved via NE-autologistic regression was more reliable and suitable for simulating regional land-use changes.

3) Under all three simulated scenarios, traffic land and residential/industrial land would increase, whereas arable land and unused land would decrease during the period 2009–2020, which exhibited similar change trends. The simulation results also presented apparent differences among the three scenarios. i) The simulated increase in both traffic land and residential/industrial land, as well as the loss of arable land, would be highest under the ED scenario and lowest under the EP scenario. The decrease rate of unused land would be highest under the EP scenario and lowest under the NG scenario. Moreover, woodland and water area would show a decline under the NG and ED scenarios but would increase slightly under the EP scenario. ii) From a spatial perspective, under the NG scenario, the expansion of traffic land and residential/industrial land would mainly occur in the southwestern part of Xintang Town, the central part of Shitan Town, and the central and eastern parts of Licheng Sub-district. By contrast, the decreased areas of arable land and woodland would be mainly located in the areas around towns. Under the EP scenario, arable land in the central and northern areas of Zengcheng, as well as woodland and water area in the entire region, would be effectively protected. Both traffic land and residential/industrial land would increase mainly around their existing locations, and their sizes would be reasonably controlled. Under the ED scenario, traffic land and residential/industrial land would expand dramatically by occupying considerable areas of arable land and woodland around towns and rural settlements, as well as part of unused land, particularly the expansion of residential/industrial land in the southwestern part of Xintang Town and the central junction between Licheng and Zengjiang Sub-districts would be the most significant.

The scenario simulation results could help policy-makers understand the future land-use change trend, find possible problems in land use, and accordingly develop rational decisions on land-use management. For example, the scenario simulations could identify hotspot areas that would be susceptible to land-use changes and indicate the trend of such changes, which would help policy-makers better formulate plans and manage these areas. From the simulated results, planners could reasonably allocate the demands for various land-use types and regulate their spatial layout in land-use planning revision.

This case study indicates the effectiveness of the improved CLUE-S model for land-use simulation. However, approximately 11% of the total cells were incorrectly simulated using this model during the validation period. The raster calculation analysis showed that these incorrectly simulated cells were more or less distributed in each town/sub-district but mainly concentrated in the central and southern areas of Zengcheng and mainly belonged to arable land, woodland, and residential/industrial land (this condition might be related to their high area proportions). In addition, some limitations were observed in this study. First, we only selected 10 typical driving factors to analyze the location suitability of land-use types. Some other factors, such as social diversities, personal preferences for housing locations, and policy changes, which might also affect land-use changes, were not considered because they could not be easily collected and transformed into quantitative spatial parameters for CLUE-S modeling. Furthermore, we merely used the inverse distance weighted interpolation to spatialize population density and residential density factors but did not use other approaches, such as kriging, spline, or polynomial methods. For future research, more factors and spatialization methods could be explored. Second, the relationships between land use and its drivers were estimated based on data of the base year. However, such relationships may vary with changes in certain driving factors (e.g., socioeconomic variables) at different stages; these changes also represent a flaw of the CLUE-S model. Hence, stable causalities between land use and appropriate drivers should be further investigated to enhance the reliability of simulations. Third, the setting of ELAS, which has an important effect on simulation results, was mainly based on the understanding of the actual land-use system, expert knowledge, and reiterative model calibration. Future studies should explore more effective methods to define the ELAS.

Acknowledgements This research was financially supported by the National Natural Science Foundation of China (Grant Nos. 41001078 and 41271060). The authors would like to thank the Editor and anonymous reviewers for their valuable comments on the manuscript.

References

Besag J (1972). Nearest-neighbour systems and the auto-logistic model for binary data. J R Stat Soc B, 34(1): 75–83

Castella J C, Verburg P H (2007). Combination of process-oriented and pattern-oriented models of land-use change in a mountain area of Vietnam. Ecol Modell, 202(3–4): 410–420

Chen J, Gong P, He C Y, Luo W, Tamura M, Shi P (2002). Assessment of the urban development plan of Beijing by using a CA-based urban growth model. Photogramm Eng Remote Sensing, 68(10): 1063–1071

Dai S P, Zhang B (2013). Land use change scenarios simulation in the middle reaches of the Heihe River Basin based on CLUE-S model: a case of Ganzhou district of Zhangye city. Journal of Natural Resources, 28(2): 336–348 (in Chinese)

Duan Z Q, Verburg P H, Zhang F R, Yu Z R (2004). Construction of a land-use change simulation model and its application in Haidian district, Beijing. Acta Geogr Sin, 59(6): 1037–1047 (in Chinese)

Hu Y C, Zheng Y M, Zheng X Q (2013). Simulation of land-use scenarios for Beijing using CLUE-S and Markov composite models. Chin Geogr Sci, 23(1): 92–100

Jiang W G, Chen Z, Lei X, Jia K, Wu Y F (2015). Simulating urban land use change by incorporating an autologistic regression model into a CLUE-S model. J Geogr Sci, 25(7): 836–850

Jiang Y, Liu J, Cui Q, An X H, Wu C X (2011). Land use/land cover change and driving force analysis in Xishuangbanna region in 1986–2008. Front Earth Sci, 5(3): 288–293

Li H X, Liu G H, Fu B J (2012). Estimation of regional evapotranspiration in Alpine area and its response to land use change: a case study in Three-River Headwaters region of Qinghai-Tibet plateau, China. Chin Geogr Sci, 22(4): 437–449

Lin Y P, Chu H J, Wu C F, Verburg P H (2011). Predictive ability of logistic regression, auto-logistic regression and neural network models in empirical land-use change modeling: a case study. Int J Geogr Inf Sci, 25(1): 65–87

Lin Y Z, Deng X Z, Li X, Ma E J (2014). Comparison of multinomial logistic regression and logistic regression: Which is more efficient in allocating land use? Front Earth Sci, 8(4): 512–523

Liu J Y, Deng X Z (2010). Progress of the research methodologies on the temporal and spatial process of LUCC. Chin Sci Bull, 55(14): 1354–1362

Liu M, Hu Y M, Chang Y, He X Y, Zhang W (2009). Land use and land cover change analysis and prediction in the upper reaches of the Minjiang River, China. Environ Manage, 43(5): 899–907

Liu M, Hu Y M, Zhang W, Zhu J J, Chen H W, Xi F M (2011). Application of land-use change model in Guiding regional planning: a case study in Hun-Taizi river watershed, northeast China. Chin Geogr Sci, 21(5): 609–618

Liu M, Li C L, Hu Y M, Sun F Y, Xu Y Y, Chen T (2014a). Combining CLUE-S and SWAT models to forecast land use change and non-point source pollution impact at a watershed scale in Liaoning Province, China. Chin Geogr Sci, 24(5): 540–550

Liu X P, Li X, Liu L, He J Q, Ai B (2008a). A bottom-up approach to discover transition rules of cellular automata using ant intelligence. Int J Geogr Inf Sci, 22(11−12): 1247–1269

Liu X P, Li X, Shi X, Wu S K, Liu T (2008b). Simulating complex urban development using kernel-based non-linear cellular automata. Ecol Modell, 211(1−2): 169–181

Liu X P, Li X, Shi X, Zhang X H, Chen Y M (2010). Simulating land-use dynamics under planning policies by integrating artificial immune systems with cellular automata. Int J Geogr Inf Sci, 24(5): 783–802

Liu X P, Ma L, Li X, Ai B, Li S Y, He Z J (2014b). Simulating urban growth by integrating landscape expansion index (LEI) and cellular automata. Int J Geogr Inf Sci, 28(1): 148–163

Luo G P, Yin C Y, Chen X, Xu W Q, Lu L (2010). Combining system dynamic model and CLUE-S model to improve land use scenario analyses at regional scale: a case study of Sangong watershed in Xinjiang, China. Ecol Complex, 7(2): 198–207

Luo J, Zhan J Y, Lin Y Z, Zhao C H (2014). An equilibrium analysis of the land use structure in the Yunnan Province, China. Front Earth Sci, 8(3): 393–404

Overmars K P, de Koning G H J, Veldkamp A (2003). Spatial autocorrelation in multi-scale land use models. Ecol Modell, 164(2–3): 257–270

Overmars K P, Verburg P H, Veldkamp A (2007). Comparison of a deductive and an inductive approach to specify land suitability in a spatially explicit land use model. Land Use Policy, 24(3): 584–599

Pan Y, Liu Y H, Wang J, Yu Z R (2011). Non-point pollution control for landscape conservation analysis based on CLUE-S simulations in Miyun county. Acta Ecol Sin, 31(2): 529–537 (in Chinese)

Park J Y, Park M J, Joh H K, Shin H J, Kwon H J, Srinivasan R, Kim S J (2011). Assessment of MIROC3.2 hires climate and CLUE-S land use change impacts on watershed hydrology using SWAT. Trans ASABE, 54(5): 1713–1724

Pontius R G Jr, Schneider L C (2001). Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. Agric Ecosyst Environ, 85(1–3): 239–248

Stevens D, Dragicevic S (2007). A GIS-based irregular cellular automata model of land-use change. Environ Plann B Plann Des, 34(4): 708–724

Syartinilia S T, Tsuyuki S (2008). GIS-based modeling of Javan Hawk-Eagle distribution using logistic and autologistic regression models. Biol Conserv, 141(3): 756–769

Veldkamp A, Fresco L O (1996). CLUE: a conceptual model to study the conversion of land use and its effects. Ecol Modell, 85(2–3): 253–270

Verburg P H (2006). Simulating feedbacks in land use and land cover change models. Landsc Ecol, 21(8): 1171–1183

Verburg P H, de Nijs T C M, Ritsema van Eck J, Visser H, de Jong K (2004a). A method to analyse neighbourhood characteristics of land use patterns. Comput Environ Urban Syst, 28(6): 667–690

Verburg P H, Schot P P, Dijst M J, Veldkamp A (2004b). Land use change modelling: current practice and research priorities. GeoJournal, 61(4): 309–324

Verburg P H, Soepboer W, Veldkamp A, Limpiada R, Espaldon V, Mastura S S (2002). Modeling the spatial dynamics of regional land use: the CLUE-S model. Environ Manage, 30(3): 391–405

Wang Q, Meng J J, Mao X Y (2014). Scenario simulation and landscape pattern assessment of land use change based on neighborhood analysis and auto-logistic model: a case study of Lijiang River Basin. Geogr Res, 33(6): 1073–1084 (in Chinese)

Wear D N, Bolstad P (1998). Land-use changes in southern Appalachian landscapes: spatial analysis and forecast evaluation. Ecosystems (N Y), 1(6): 575–594

Wu F L (2002). Calibration of stochastic cellular automata: the application to rural-urban land conversions. Int J Geogr Inf Sci, 16(8): 795–818

Wu G P, Zeng Y N, Feng X Z, Xiao P F, Wang K (2010a). Dynamic simulation of land use change based on the improved CLUE-S model: a case study of Yongding county, Zhangjiajie. Geogr Res, 29 (3): 460–470 (in Chinese)

Wu G P, Zeng Y N, Xiao P F, Feng X Z, Hu X T (2010b). Using autologistic spatial models to simulate the distribution of land-use patterns in Zhangjiajie, Hunan Province. J Geogr Sci, 20(2): 310–320

Wu M, Ren X Y, Che Y, Yang K (2015). A coupled SD and CLUE-S model for exploring the impact of land use change on ecosystem service value: a case study in Baoshan district of Shanghai, China. Environ Manage, 56(2): 402–419

Yang J, Cai Y L, Yuan T (2014). Dynamic simulation on the spatio-temporal patterns of land use changes in Chongming County based on CLUE-S model. Chinese Agricultural Science Bulletin, 30(11): 258–264 (in Chinese)

Zhang P, Liu Y H, Pan Y, Yu Z R (2013). Land use pattern optimization based on CLUE-S and SWAT models for agricultural non-point source pollution control. Math Comput Model, 58(3–4): 588–595

Zheng H W, Shen G Q P, Wang H, Hong J K (2015). Simulating land use change in urban renewal areas: a case study in Hong Kong. Habitat Int, 46: 23–34

Zheng X Q, Zhao L, Xiang W N, Li N, Lv L N, Yang X (2012). A coupled model for simulating spatio-temporal dynamics of land-use change: a case study in Changqing, Jinan, China. Landsc Urban Plan, 106(1): 51–61