Front. Earth Sci. 2016, 10(2): 245–252
DOI 10.1007/s11707-015-0522-7

RESEARCH ARTICLE

Simulating land use change by integrating landscape metrics into ANN-CA in a new way

Xin YANG (✉)[1], Yu ZHAO[2], Rui CHEN[2], Xinqi ZHENG[3]

1 School of Computer Engineering, Qingdao Technological University, Qingdao 266520 ,China
2 Institute of Policy and Management, Chinese Academy of Sciences, Beijing 100190, China
3 School of Information Engineering, China University of Geosciences, Beijing 100083, China

Abstract Landscape metrics are measurements of land-use patterns and land-use change, but even so, have rarely been integrated into land-use change simulation models. This paper proposes a new artificial neural network-cellular automaton by integrating landscape metrics into the model. In this model, each cell acquires unique landscape metric values. The landscape metric values of each cell are actually the landscape metric values of land use type in its neighborhood, which takes the cell as center. The calculation of landscape metrics ensures that those of each cell can represent cellular spatial environmental characteristics. The model is used to simulate land use change in the Changping district of Beijing, China. Comparisons of the simulated land use map with the actual map show that the proposed model is effective for land use change simulation. The validation is further carried out by comparing the simulated land use map with that simulated by an artificial neural network-cellular automaton model, which has not been integrated with landscape metrics. Results indicate that the proposed model is more appropriate for simulating both quantity and spatial distribution of land use change in the study area.

Keywords land use change, landscape metrics, cellular automata, artificial neural network

1 Introduction

Landscape metrics (LM) are measurements for characterizing land-use patterns, including their composition, distribution, and fragmentation. LM have often been used for depicting the results of land use change (Seto and Fragkias, 2005; Serra et al., 2008; Liu et al., 2010; Feng et al., 2011). Researchers have integrated LM into land use change models to improve simulation accuracy. Liu et al. (2014) calculated a landscape expansion index (LEI) based on LM to identify urban growth type, and then generated different transition rules for various growth types. Li et al. (2013) calibrated an urban expansion simulation model according to LM.

In previous research on Land use Change (LUC), LM were incorporated into a Markov-Cellular Automata (CA) model (LM- Markov-CA model) (Yang et al., 2014). Relationships between non-transition probabilities (probability of the current, persistent land use type) and LM are discovered in the model and are then used to generate non-transition probability maps, which are used as suitability maps of Markov-CA. Two points from this work should be particularly noted: 1) in a sub-region, LM-based non-transition probabilities are the same for all cells with the same land use type; 2) in the model, the final probability of LUC is determined by the smaller values between LM-based probability and other conditions-based probability.

In this paper, a LUC model that incorporates LM into ANN (artificial neural network)-CA is proposed. The ANN-CA model has been widely used in dynamic spatial modeling (Mahajan and Venkatachalam, 2009; Pijanowski et al., 2014). It has an advantage due to its ability to detect relationships in data by emulating the brain's ability to sort patterns through the interconnected systems of many neurons (Arekhi and Jafarzadeh, 2014; Tayyebi and Pijanowski, 2014). In addition, ANN is very effective at handling incorrect and inferior data, and capturing non-linear, complex features in modeling processes. Therefore, it has been generally accepted as being capable of achieving results with superior modeling accuracy (Li and Yeh, 2002).

In the proposed model, LM and other conditions (such as distance-based variables) are used together to generate transition probabilities of land use. These land-use patterns could possibly affect human behavior toward land use in

Received August 25, 2014; accepted March 5, 2015

E-mail: qdlgyx@126.com

the future (Mitsuda and Ito, 2011), thus impacting future land use change. Taking Nanchang as an example, the results of the Zeng et al. (2004) study show that land-use patterns can affect urban sprawl. In contrast to the LM-Markov-CA model, the final probability of LUC is determined by probabilities based on LM and other conditions. Another difference between the models is that in the proposed model, each cell has unique LM values.

The proposed model also differs from that of Li et al. (2013). In that their model incorporates LM into the calibration procedure of the LUC model. In the calibration, a special fitness function is designed based on LM to affect the weights of urban expression factors. The proposed model uses LM as factors of LUC, and differentiates transition probabilities of land use in various sub-regions by incorporating their different landscape patterns.

To demonstrate the feasibility and advantages of the proposed model, a case study of Changping, a district of Beijing, is presented.

2 Methodology

2.1 Landscape metrics

LM are quantitative descriptions of landscape patterns (Hu and Dong, 2013), using various indexes for these descriptions (McGarigal and Cushman, 2002). If all indexes are integrated into an LUC model, the simulation must be very complex. In this paper, we chose the percentage of landscape (PLAND) and patch density (PD) for land use type for incorporation into the ANN-CA model.

PLAND is the percentage of landscape occupied by the type, and is an index for quantifying landscape composition. It was chosen because the land use class change of a patch may be influenced by the abundance of patches with the same land use class in the surrounding landscape. A patch with high PLAND may be more difficult to change to other land use classes. This index is calculated as Eq. (1).

$$\mathrm{PLAND}_i = \frac{\sum_{j=1}^{n} a_{ij}}{A} \times 100, \tag{1}$$

where i represents patch type, j the number of patches, a_{ij} the area (m^2) of patch ij, A the total landscape area (m^2), and PLAND_i the PLAND of i.

PD is a fundamental index that represents landscape configuration. It expresses the number of patches per unit within an area and best describes landscape fragmentation. Holding total area constant, a landscape with greater density would be considered more fragmented than one with lower density. A land use type with a high PD in the landscape may be more easily changed to other types. This index is calculated as Eq. (2).

$$\mathrm{PD}_i = \frac{n}{\sum_{j=1}^{n} a_{ij}}, \tag{2}$$

where PD_i is the PD of i.

2.2 ANN-CA based LUC Model

In an ANN-CA model, the CA provides a spatiotemporal framework for LUC simulation; ANN is used to discover local transition rules of CA, which are the transition probabilities of all land use types (Li and Yeh, 2002; Lin et al., 2011).

In this work, a feed-forward neural network with a back-propagation algorithm was devised. The ANN was designed with one input, one hidden, and one output layer (Isik et al., 2013). Training and simulation of the ANN was cell-based, and each cell had a set of n attributes as inputs to the neural network. It was assumed that these attributes determined the land use change probabilities. The ANN had an equal number of hidden neurons and input neurons. The number of output neurons was dependent on the number of land use types. A neuron in the output layer corresponded to a land use type. The value of a neuron in the output layer represented the transition probability from an existing type to the corresponding land use type (Dai et al., 2005).

The network maintained the same structure in the training and simulation phases. Each neuron had signal collection and activation processes within that structure. In the input layer, each neuron accepted a value, and then generated an output value as input to all the neurons in the hidden layer. The neurons in the hidden layer subsequently processed the signals and generated output values for the neurons in the output layer.

2.3 ANN-CA model integrated with LM

When integrated into ANN-CA, LM are attributions of a cell as the neurons in the ANN input layer. The key point of the integration is the calculation of the cellular LM values. As mentioned in Section 2.1, the two LM, PLAND, and PD are the characteristics for each land use type, which means that they are the attributions of that type in a region. Therefore, the two LM, PLAND, and PD must be transformed into cell characteristics from land use type characteristics.

Essentially, the transformation is taking the neighborhood of a cell as a landscape and setting LM values of the cellular land use type in the landscape as the cellular LM values. Here, the neighborhood of a cell is denoted as its sub-region. The sub-region takes the cell as its center, ensuring the sub-region can best represent the characteristics of its spatial environment. Clearly, the sub-region size influences the LM value.

The calculation of PLAND for each cell is as follows: 1) count the area of the cellular land use type in its sub-region, and denote the result as Area_1; 2) count the area of the sub-region and indicate the result as Area_2; 3) calculate the value of PLAND by dividing Area_1 by Area_2 and multiplying by 100. The focal statistics tool in ArcGIS software can count the number of cells in the neighborhood for all map cells in one operation. It was used to realize the batch computing of PLAND.

The calculation of PD for each cell is as follows: 1) count the number of patches (a patch consists of eight-connected cells with the same land use type) with the cellular land use type in its sub-region, and denote the result as Num_1; 2) count the area of cellular land use type in its sub-region, and indicate the result as Area_2; and 3) calculate the value of PD by dividing Num_1 by Area_2. A tool was programmed with Matlab software for the batch computing of PD.

Given the above description, structure of the ANN-CA model integrated with LM (denoted as ANN-CA-LM model) is described in Fig. 1. In this model, social factors, economic factors, physical factors, and LM can be set as the neurons in the ANN input layer. Probabilities of land use types were neurons in the ANN output layer.

During the simulation, the land use change of a cell was determined by comparing transition probabilities generated by neurons in the output layer. Land use changed from an existing type to that associated with the highest transition probability. If the land use types had the same highest transition probability, randomness was generated to make the decision. The simulation was carried out by running CA and an ANN for a sufficient number of steps.

3 Case study

To demonstrate the feasibility of the ANN-CA model integrated with LM, (denoted as an ANN-CA-LM model), a case study was conducted to predict LUC in the Changping district, Beijing, China. This district was selected as a study area due to its significant spatial differences in landscape patterns and the data availability. The population of Changping is rapidly increasing due its proximity to the outer borders of central Beijng. In addition, the land use of the area has changed dramatically due to the influence of human environmental systems, while also benefitting from a growing economy.

3.1 Data sources

The data used to provide actual urban areas in this case study included three land use maps generated from classifications of Landsat TM5 images with a spatial resolution of 30 m, acquired during the summers of 1988, 1998, and 2008. Each map was treated as a cellular space, with each pixel representing a cell. Thus, each cell represented an area of 30 m × 30 m, or 900 m^2. The maps contained five land use/land cover types: Forest Land (FL), Cultivated Land (CuL), Construction Land (CoL), Water (WL) and Other Unused Land (OUL). Two maps from 1988 and 1998 (shown in Fig. 2) were used to capture the transition rules and historical development trends. The 2008 map was used to test the simulation.

In addition to the two LM, a total of ten spatial variables were used as attributes of the cells. Table 1 lists the details of these variables, including a series of distance-based variables, neighborhood conditions, and physical attributes, which were derived from remote sensing and GIS data. Studies have shown that these variables are closely related to probabilities of land use changes.

Traffic data and urban and town areas (in GIS format) for 1988 and 1998 were prepared to derive the distance-based variables. The Euclidean Distance function of ArcGIS was used to obtain the Euclidean distance raster map for the distance-based variables. A standard 7×7 contiguity filter was used to derive the neighborhood conditions. It was

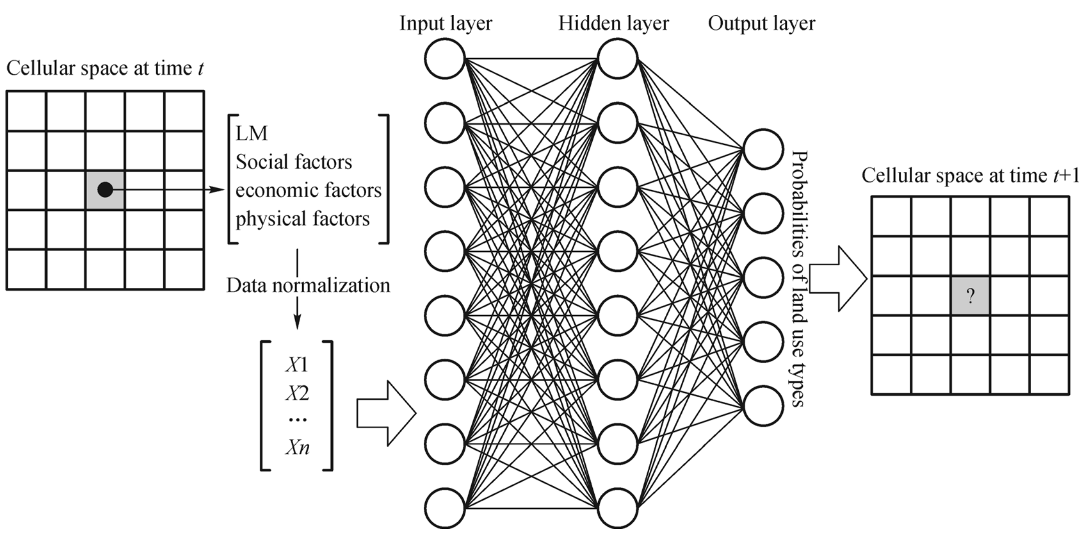

Fig. 1 Processing architecture of ANN-CA model integrated with LM.

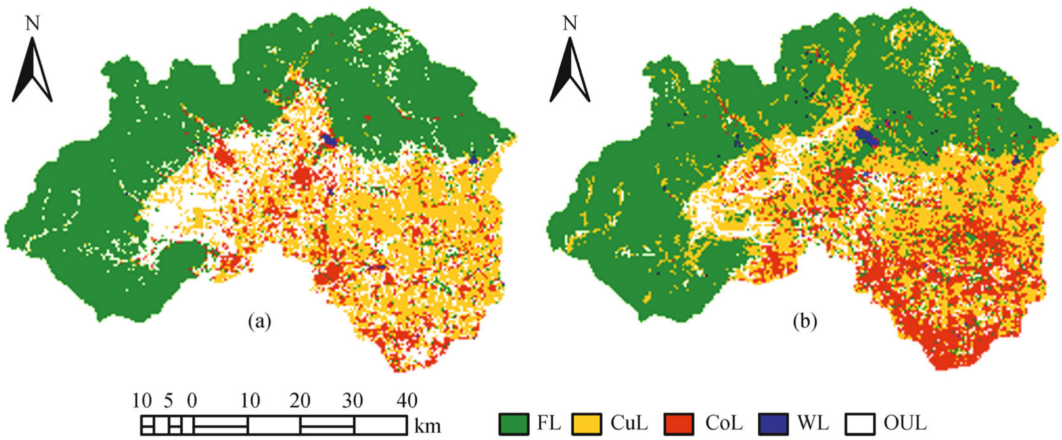

Fig. 2 Land use maps. (a) 1988 land use map; (b) 1998 land use map. FL = forest land, CuL = cultivated land, CoL = construction land, WL = water, OUL = other unused land.

Table 1 Spatial variables used for the conditional items of local transition rules

	Spatial variable	Short name
Physical attribute	Cellular initial state	IS
Distance-based variables	Distance from the cell to the major urban areas	Dist(MU)
	Distance from the cell to the closest town areas	Dist(CT)
	Distance from the cell to the closest road	Dist(CR)
	Distance from the cell to the closet railway	Dist(CW)
Neighborhood conditions	Amount of cells used as forest	Num(FL)
	Amount of cells used as cultivated land	Num(CuL)
	Amount of cells used as construction land	Num(CoL)
	Amount of cells whose states are water	Num(WL)
	Amount of cells used as other unused land	Num(OUL)

determined that other types of contiguity filters should not be used because earlier research indicated that a combination of small cell and neighborhood sizes generated improper expressions of land use transitions (Pan et al., 2010).

3.2 Simulation process

Training the ANN by using an ANN-CA model is essential before the simulation. The 1988 and 1998 data were both used to train the ANN and to obtain the transition rule for CA. It was unwise to use the entire data-set for training because it was so large. A total of 50,000 random samples were proportionally selected from different land use classes. The ten spatial variables, listed in Table 1, were easy to obtain, yet the LM calculation presented a challenge.

The calculation method of cell LM was introduced in Section 2.3. It was difficult to determine the size of the specific sub-region for cells for this calculation. The influence landscape patterns have on land use change may differ based on size variance. Even though this work did not focus on sub-region size for calculating LM, it must be thoroughly analyzed in future research. The size of the sub-region was set to 301 (width) $\times$ 201 (height), representing an area of $301 \times 201 \times 900\,m^2$, or $54\,km^2$.

The PLAND and PD of each cell in the 1988 land use map were obtained as shown in Figs. 3(a) and 4(a) and in the 1998 land use map as shown in Figs. 3(b) and 4(b)), which were used for simulating the 2008 land use map.

There were a total of 12 variables for each cell, and consequently, there were 12 neurons in the input layer. The number of neurons within the hidden layer was also set to 12. Five land use types resulted in five neurons in the output layer. No explicit transition rules were required for an ANN-CA model. The only task was to train the ANN using 50,000 samples. Then, the trained ANN was used to generate transition probabilities from 1998 to 2008, which were provided for CA to simulate the land use map in 2008.

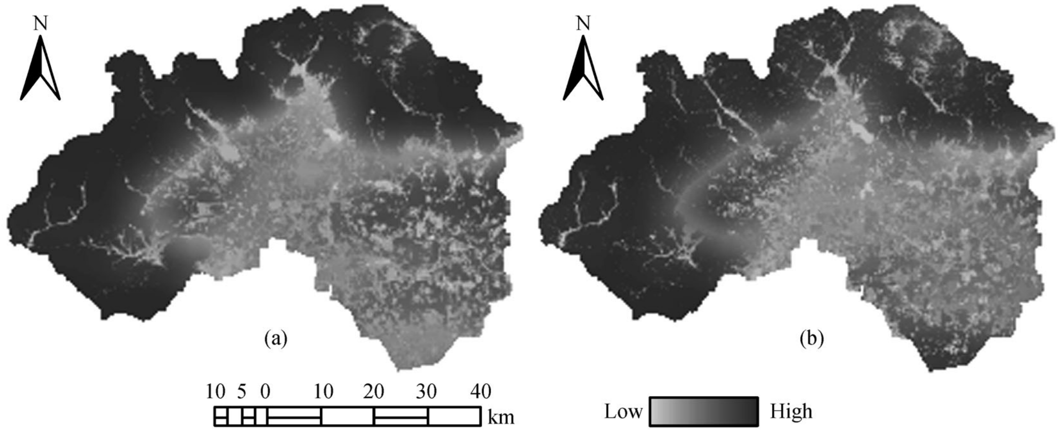

Fig. 3 PLAND value maps (a) 1988 PLAND value map, (b) 1998 PLAND value map. PLAND = the percentage of landscape.

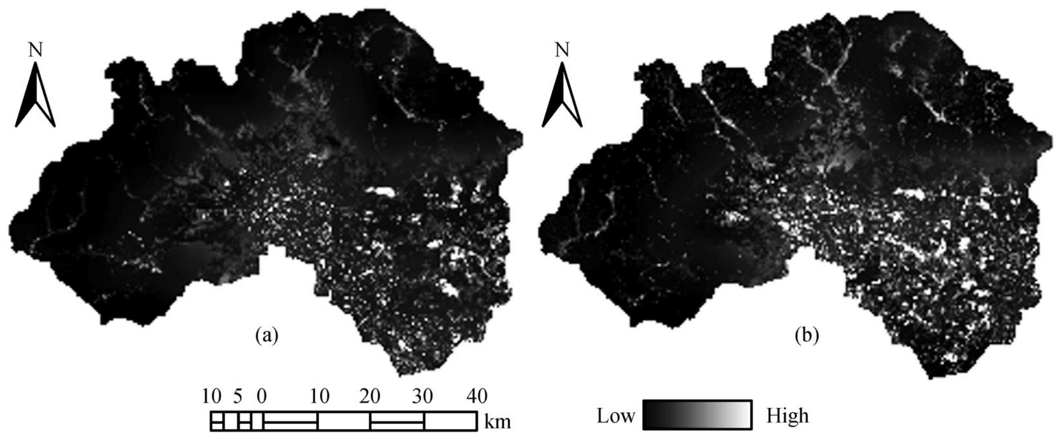

Fig. 4 PD value maps. (a) 1988 PD value map; (b) 1998 PD value map. PD = patch density.

3.3 Results and discussion

The simulated land use map in 2008 was generated, as shown in Fig. 5(a). The simulated result was analyzed by comparing it with the actual map in 2008, as shown in Fig. 5(b).

Visual comparisons with the 2008 actual land use map indicated that land use classes in the simulated map were similar to corresponding types in the actual map. However, three deviations were observed in the simulated map. The first was in the north of the district where in the actual map, there were some small CuL patches that were not simulated by the proposed model. A second was the greater area of OUL shown in the simulated map. The third was observed in the central area of the district where there was a lesser amount of CoL.

The accuracy of the simulation results were then analyzed quantitatively, using the CROSSTAB model in IDRISI in the process. Simulated area, actual area, error rate, and Kappa Index of Agreement (KIA) are listed in Table 2. Error rate was used to describe the quantitative accuracy, and KIA was used to describe the spatial accuracy. The results showed that the accuracy of the quantity and spatial distributions were both satisfactory. Error rates for FL, CuL, and CoL were particularly low at – 0.0204 and 0.0733, respectively. KIA for FL, CuL, CoL, and WL were at 0.805, 0.599, 0.596, and 0.389 respectively, which were all acceptable. These results indicated that the ANN-CA model integrated with LM could predict future land use patterns objectively and accurately.

3.4 Model validation and comparison

In this section, the simulation result discussed in Section 3.3 is compared with that simulated by the ANN-CA model without integrated LM to validate the advantage of the proposed model. Spatial variables other than the LM, as used in the local transition rules of the ANN-CA model, were the same as those used in the ANN-CA-LM model.

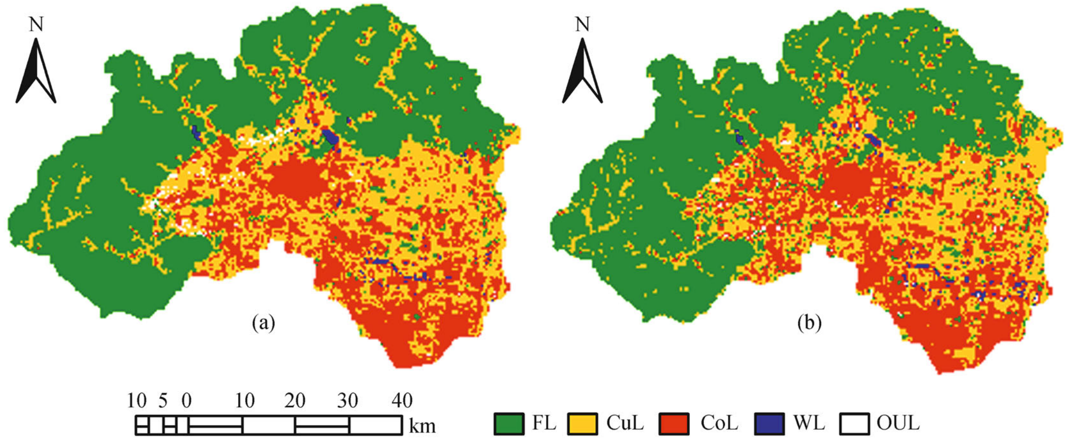

Fig. 5 Land use maps. (a) Simulated land use in 2008; (b) actual land use in 2008. FL = forest land, CuL = cultivated land, CoL = construction land, WL = water, OUL = other unused land.

Table 2 Comparison of accuracy for simulated area and actual area

	Simulated area/km^2	Actual area/km^2	Error rate	KIA
FL	653.16	666.75	−0.0204	0.805
CuL	379.97	354.01	0.0733	0.599
CoL	285.68	316.11	−0.0963	0.596
WL	8.58	10.73	−0.2004	0.389
OUL	24.6	4.39	4.6036	0.181

The land use map simulated by the ANN-CA model in 2008 is shown in Fig. 6.

Simulated land use areas, error rates, and KIA of the ANN-CA model are listed by category in Table 3. From the table we can see that, overall, error rates and KIAs of the land use map simulated by the ANN-CA were both smaller than those simulated by ANN-CA-LM. In terms of quantitative accuracy, error rates for CuL, CoL, WL, and OUL in the map simulated by ANN-CA-LM were all smaller than those simulated by ANN-CA. In terms of spatial accuracy, KIAs for all land use types of the ANN-CA model were smaller than those of ANN-CA-LM.

Further investigation was carried out through a combination of visual and quantitative comparisons. The landscape pattern of FL in southern Changping was clearly different from that in the north. The ANN-CA-LM simulated area of FL was smaller than the actual area, and that simulated by ANN-CA was larger than the actual area. A comparison of Figs. 5(a) and 6, shows that the number of FL fragments in the south of the map simulated by ANN-CA-LM was smaller than that simulated by ANN-CA. These differences in the simulated results prove that integration of LM did change the transition probabilities according to the landscape. Another obvious difference between the simulated maps was the spatial distribution of OUL. In the east of the map simulated by

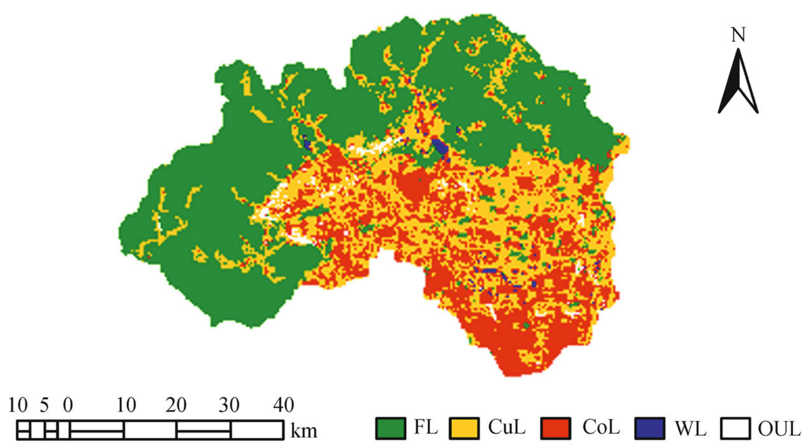

Fig. 6 Simulated land use map in 2008 using ANN-CA model without LM. FL = forest land, CuL = cultivated land, CoL = construction land, WL = water, OUL = other unused land.

Table 3 Simulation result of ANN-CA model without LM

	Simulated area/km^2	Error rate	KIA
FL	674.41	0.0114	0.714
CuL	383.59	0.0836	0.536
CoL	255.65	−0.1912	0.499
WL	8.55	−0.2032	0.398
OUL	29.79	5.7859	0.041

ANN-CA-LM, there was almost no OUL, but with ANN-CA, some remained. As seen in Fig. 2(b), greater OUL fragmentation was observed in eastern Changping 1998, thus demonstrating that integrating the LM caused the transition probabilities to decrease.

Comparison and analysis showed that, with the same spatial variable restrictions, the model proposed in this paper enhanced the ANN-CA model by integrating LM. The model had better simulation performance than that of ANN-CA. The comparison proved that the integration method of LM into the LUC model improved simulation accuracy for both quantity and spatial distributions.

4 Conclusions

Land use change simulation can better assess development impacts, preparation of land use plans, and search for optimal land use patterns. The simulation enables rural and urban planners to provide the public with the facilities and services necessary to sustain development. Simulation accuracy is an important issue. Land-use pattern is the embodiment of the spatial heterogeneity of land use, and can affect future land use change. The integration of LM (the measurement of landscape pattern) is effective for improving the land use change simulation model.

This paper proposed a land use change simulation model by integrating LM into ANN-CA. In the model, the LM were integrated with other spatial variables to differentiate transition probabilities of land use. In the model, LM were transformed from attributions for each land use type to attributions at cell level by setting cellular LM equal to the LM of the cellular land use class in a specific sub-region, which took the cell as its center. The proposed model considered variations of landscape pattern within the study area to differentiate the probabilities of land use change and ensured that the LM of each cell best represented the characteristics of its spatial environment.

The proposed model was successfully used in the simulation of land use change in the Changping district of Beijing, China. The simulation result was compared with that from the ANN-CA model. This comparison indicated that the proposed model was superior to ANN-CA in both spatial and quantitative accuracy. It appeared that the integration of LM into the land use change simulation model improved simulation accuracy and that the integration method was effective.

Although the proposed model yielded plausible simulation results for the study area, it is limited by the size of the neighborhood. In this paper, the problem of the size of the neighborhood was not addressed thoroughly. This must be addressed in future research.

Acknowledgements This study was supported by the China Postdoctoral Science Foundation (No. 2014M560120).

References

Arekhi S, Jafarzadeh A (2014). Forecasting areas vulnerable to forest conversion using artificial neural network and GIS (case study: northern Ilam forests, Ilam province, Iran). Arab J of Geosci, 7(3): 1073–1085

Dai E, Wu S H, Shi W Z, Cheung C K, Shaker A (2005). Modeling change-pattern-value dynamics on land use: an integrated GIS and artificial neural networks approach. Environ Manage, 36(4): 576–591

Feng Y X, Luo G P, Lu L, Zhou D C, Han Q F, Xu W Q, Yin C Y, Zhu L, Dai L, Li Y Z, Li C F (2011). Effects of land use change on landscape pattern of the Manas River watershed in Xinjiang, China. Environ Earth Sci, 64(8): 2067–2077

Hu R S, Dong S C (2013). Land use dynamics and landscape patterns in Shanghai, Jiangsu and Zhejiang. J Resour Ecol, 4(2): 141–148

Isik S, Kalin L, Schoonover J E, Srivastava P, Lockaby B G (2013). Modeling effects of changing land use/cover on daily streamflow: an artificial neural network and curve number based hybrid approach. Journal of Hydrology, 485(SI): 103–112

Li X, Lin J Y, Chen Y M, Liu X P, Ai B (2013). Calibrating cellular automata based on landscape metrics by using genetic algorithms. Int J Geogr Inf Sci, 27(3): 594–613

Li X, Yeh A G (2002). Neural-network-based cellular automata for simulating multiple land use changes using GIS. Int J Geogr Inf Sci, 16(4): 323–343

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

Liu X P, Li X, Chen Y M, Tan Z Z, Li S Y, Ai B (2010). A new landscape index for quantifying urban expansion using multi-temporal remotely sensed data. Landscape Ecol, 25(5): 671–682

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

Mahajan Y, Venkatachalam P (2009). Neural network based cellular automata model for dynamic spatial modeling in GIS. Computational Science and its Applications-ICCSA 2009 PT 1, 5592: 341–352

McGarigal K, Cushman S A (2002). Comparative evaluation of experimental approaches to the study of habitat fragmentation effects. Ecol Appl, 12(2): 335–345

Mitsuda Y, Ito S (2011). A review of spatial-explicit factors determining spatial distribution of land use/land-use change. Landsc Ecol Eng, 7

(1): 117–125

Pan Y, Roth A, Yu Z R, Doluschitz R (2010). The impact of variation in scale on the behavior of a cellular automata used for land use change modeling. Comput Environ Urban Syst, 34(5): 400–408

Pijanowski B C, Tayyebi A, Doucette J, Pekin B K, Braun D, Plourde J (2014). A big data urban growth simulation at a national scale: configuring the GIS and neural network based land transformation model to run in a high performance computing (HPC) environment. Environ Model Softw, 51: 250–268

Serra P, Pons X, Sauri D (2008). Land-cover and land-use change in a Mediterranean landscape: a spatial analysis of driving forces integrating biophysical and human factors. Appl Geogr, 28(3): 189–209

Seto K C, Fragkias M (2005). Quantifying spatiotemporal patterns of urban land-use change in four cities of China with time series landscape metrics. Landscape Ecol, 20(7): 871–888

Tayyebi A, Pijanowski B C (2014). Modeling multiple land use changes using ANN, CART and MARS: comparing tradeoffs in goodness of fit and explanatory power of data mining tools. Int J Appl Earth Obs Geoinf, 28: 102–116

Yang X, Zheng X, Chen R (2014). A land use change model: integrating landscape pattern indexes and Markov-CA. Ecol Modell, 283: 1–7

Zeng H, Jiang F, Li S J (2004). Impacts of urban landscape structure on urban sprawl: a case researches in Nanchang. Acta Ecol Sin, 24(9): 1931–1937