



Predicting land use changes in northern China using logistic regression, cellular automata, and a Markov model

Min Wang¹ · Longyan Cai² · Hui Xu³ · Shengli Zhao³

Received: 4 July 2018 / Accepted: 18 November 2019 / Published online: 13 December 2019
© Saudi Society for Geosciences 2019

Abstract

Land use changes are complex processes affected by both natural and human-induced driving factors. This research is focused on simulating land use changes in southern Shenyang in northern China using an integration of logistic regression, cellular automata, and a Markov model and the use of fine resolution land use data to assess potential environmental impacts and provide a scientific basis for environmental management. A set of environmental and socio-economic driving factors was used to generate transition potential maps for land use change simulations in 2010 and 2020 using logistic regression. An average relative operating characteristic value of 0.824 was obtained, indicating the validity of the logistic regression model. The logistic–cellular automata (CA)–Markov model was calibrated by comparing the actual and simulated land use maps of 2010. A match of 83.7% was achieved between the predicted and actual maps of 2010, which represented a satisfactory calibration. This indicated that the integration of the logistic regression, CA, and Markov model has a high potential for simulating land use change in northern China. The calibrated hybrid model was implemented to obtain a land use map for 2020. The results showed a new wave of suburban development in the southwestern, west, and northwestern parts of the study area during 2010–2020. In addition, urban expansion has been accelerating, which is very likely to exacerbate the extensive environmental pollution currently existing in this area. Moreover, rapid urban expansion has resulted in significant decreases in forest areas and agricultural lands.

Keywords Land use change · Logistic regression · Markov model · Cellular automata · Environmental influence

Introduction

Land use change is a complex process that is affected by human activities and natural environmental changes (Arsanjani et al. 2011; Etemadi et al. 2018; Liu et al. 2017b; Memarian et al. 2012; Meyer and Turner 1994; Wang et al. 2019a; Watson 2000; Yang et al. 2014). After China's reform and opening policies were implemented in 1978, this country has experienced rapid land use change. Rapid industrialization

and urbanization have resulted in significant environmental impacts, such as water and air pollution (Lin and Zhu 2018; Peng et al. 2016; Shao et al. 2006; Wang et al. 2019b). This is especially true in many cities, such as Shenyang city in northern China, which is a key internal trading center and has a reputation as an industry leader in China (Geng et al. 2013). Therefore, a study of land use change in this area is urgently needed to facilitate environmental management.

Modeling is an important technique for studying land use dynamics (Lambin 1997; Mustafa et al. 2018; Omrani et al. 2017; Zheng et al. 2015). Rapid advances in geospatial models have made it increasingly possible to simulate land use change. Different geospatial models, such as CLUE, GEOMOD, cellular automata (CA), Markov chain model, and SLEUTH, have been used to assess land use change (Arsanjani et al. 2013; Dickinson and Henderson-Sellers 1988; Dietzel and Clarke 2006; Hagenauer and Helbich 2018; Veldkamp and Fresco 1996; Verburg et al. 2002). CA and Markov chain models are the most common approaches used for simulating land use dynamics. It is difficult to simulate the spatial pattern of land use change using Markov chain

Responsible Editor: Nilanchal Patel

✉ Longyan Cai
lonycal@126.com

¹ College of the Environment and Ecology, Xiamen University, Xiamen 361102, China

² Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China

³ China Mobile Group Liaoning Company Limited, Shenyang 110179, China

models (Ye and Bai 2008). However, CA models with powerful spatial computing can be used to predict spatial variations in land use (Sang et al. 2011). Therefore, it is expected that the integration of CA and Markov chain models may have potential for projecting land use change.

Worldwide use of CA–Markov models for predicting land use change has significantly increased in recent years (Al-sharif and Pradhan 2014; Arsanjani et al. 2011; Chen et al. 2013; Fu et al. 2018; Kamusoko et al. 2009; Luo et al. 2015; Mishra and Rai 2016; Mitsova et al. 2011; Naboureh et al. 2017; Vaz et al. 2012; Wang et al. 2019c; Yang et al. 2014). With the rapid development of urbanization, we believe that the use of a CA–Markov model for predicting land use change will be expected to increase in the future. According to the methods used to generate probability surfaces of the driving factors, studies on land use change modeling can be divided into two categories, namely, multicriteria evaluation (MCE) and logistic regression methods. MCE is a common approach to generate probability surfaces of driving factors (Fu et al. 2018; Kamusoko et al. 2009; Ku 2016; Vaz et al. 2012; Zhou et al. 2012). For example, Kamusoko et al. (2009) simulated future land cover changes (up to 2030) based on an MCE–CA–Markov model to assess rural sustainability in Zimbabwe. The results indicated that if the current land cover trends continued without holistic sustainable development measures, severe land degradation would occur. Zhou et al. (2012) successfully assessed regional land salinization resulting from biophysical and human-induced influences using an MCE–CA–Markov model in the Yinchuan Plain in northwest China in 2009. Logistic regression methods are frequently utilized approaches to generate probability surfaces of driving variables (Arsanjani et al. 2013; Hamdy et al. 2016; Islam et al. 2018; Liu et al. 2015; Siddiqui et al. 2017; Sun et al. 2018; Wang et al. 2019c). For example, Arsanjani et al. (2013) analyzed suburban expansion in the metropolitan area of Tehran in Iran using a logistic regression–CA–Markov model. Siddiqui et al. (2017) simulated urban growth dynamics of an Indian metropolitan area using a logistic regression–CA–Markov model. Compared with the MCE method, logistic regression is an easier and more efficient method for generating suitability maps (Arsanjani et al. 2013; Fu et al. 2018; Islam et al. 2018; Liu et al. 2015; Siddiqui et al. 2017; Sun et al. 2018; Wang et al. 2019c); therefore, this method has become increasingly popular for land use change simulations worldwide.

In this research, we used a model that combines logistic regression, CA, and Markov chain analysis to evaluate the changes in eight land use types in southern Shenyang in northern China under various driving forces including environmental and socio-economic factors. This area is experiencing rapid urbanization and environmental problems are increasing, such

as haze events. The goal of this study was to evaluate the potential of using a logistic regression–CA–Markov model for estimating land use changes in a cool temperate region and simulate future land use changes in order to provide reference data for environmental management.

Materials and methods

Study area

The study area is located in central Liaoning province (122°41′~123°80′E, 41°20′~42°29′N) in northern China (Fig. 1) and covers an area of approximately 8.42×10^5 ha. There are four distinct seasons, and it is mild in the spring and autumn with a hot summer (25~31 °C) and cold winter (–25~–30 °C). The average annual precipitation is about 646 mm. There are 12 districts and the population is nearly 6.9 million. Cultivated land is the predominant land use and accounted for 69.1% of the study area in 2010. This area has experienced rapid urbanization and land use change. The built-up land has increased from 12.6% (2000) to 17.4% (2010). The economy of this area is based on heavy industry and the area is one of China's largest industrial centers. It is home to an extensive industrial system including electronics, textiles, chemicals, metallurgy, and food industries (Geng et al. 2013).

Data and data processing

The main data sources were digital land use maps from 2000, 2005, and 2010. The digital map data for the study area were classified into eight classes: residential land, urban traffic, mining lease, industrial land, cultivated land, forest land, undisturbed (desert and bare land), and water. The focus areas of this study are not only the assessment of urban expansion but also the detection of changes in industrial land, urban traffic, and residential land. In addition, two categories of driving forces are expected to explain land use change, namely, (i) environmental factors including elevation, slope, aspect, and precipitation and (ii) socio-economic factors including population density, distance to roads, water, tourist attractions, town, as well as gross domestic product (GDP). Elevation, slope, and aspect were derived from a 30-m digital elevation model (DEM) obtained from the U.S. NASA website (<http://reverb.echo.nasa.gov/reverb/>). The other data used in this study, such as the road, water, and town layers were all obtained from the statistical yearbook and the local land department.

Modeling approach

In this study, we integrated a logistic regression model, CA, and Markov chain analysis to predict the expected land cover

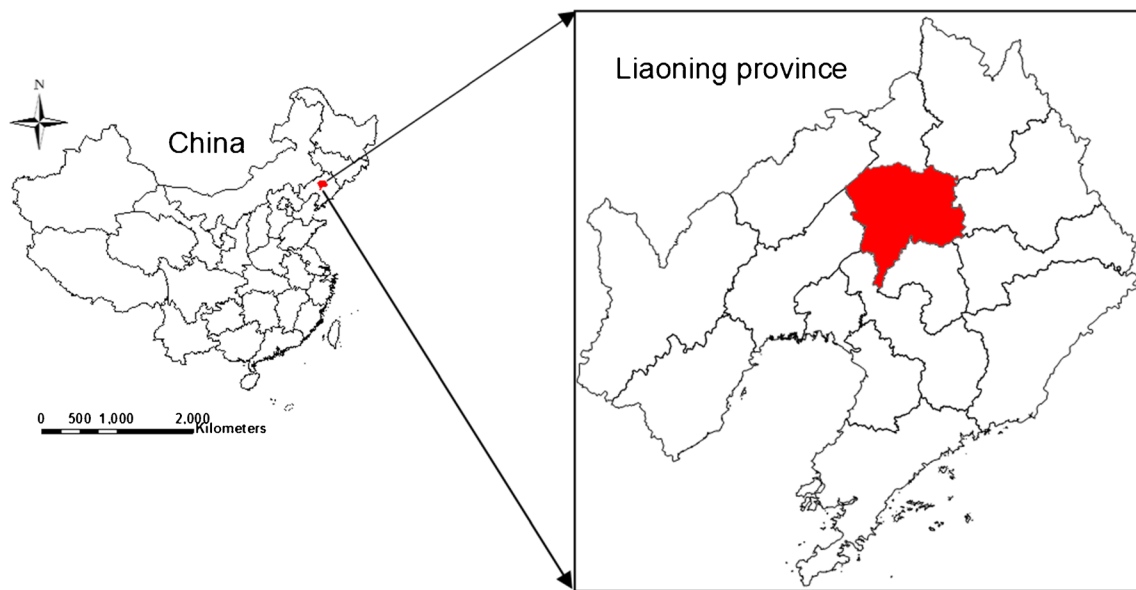


Fig. 1 The geographic location of the study area (red area)

(2010 and 2020); IDRISI software was used. The CA–Markov model requires a land cover dataset to represent the initial states, a Markov transition matrix, a group of land use suitability images, a number of iterations, and a contiguity filter. Specifically, two land use maps at different time points were used to calculate the probabilities of transition in land use using a Markov chain model; (1) a matrix of transition probabilities between the 2000 and 2005 maps was used to predict the land use in 2010 using the year 2005 as a starting point. (2) Similarly, the 2000 and 2010 land use maps were used to calculate the probabilities of transition to predict the land use in 2020. We used the year 2010 as a starting point. The maps of driving variables were created using logistic regression. Subsequently, the resulting probability surfaces of the dependent variables (different land types) were used to estimate the degree of change based on the CA–Markov model. A standard 5×5 contiguity filter was used as the neighborhood definition in the simulations; 10 CA iterations were used to predict the spatial pattern in the study area. During each iteration, the pixels with the highest transition probability and highest suitability for a particular land type were changed to a new land type whereas the pixels with lower probabilities and lower suitability remained unchanged; 30-m resolution spatial data were used as the input to the model.

The overall prediction accuracies (OPA) and kappa index (Rosenfield and Fitzpatrick 1986) were used to assess the model performance. The OPA is defined as follows:

$$OPA(\%) = \frac{\text{Simulated land use} \cap \text{Actual land use}}{\text{Actual land use}} \quad (1)$$

The OPA provides a measure of the similarity between the simulated land use and the actual land use and ranges from 0

to 100%. The closer the OPA is to 100%, the more accurate the model is.

Logistic regression

Logistic regression is used when the dependent variable is a binary variable (0 and 1) and the independent variables are continuous and categorical variables (Long 1997; MacCullagh and Nelder 1989). The dependent variable in a logistic regression model represents the probability that a particular theme will be in one of the categories (Arsanjani et al. 2013). The basic assumption is that the probability of the dependent variable takes the value of 1 (positive response) and follows the logistic curve and its value can be estimated with the following formula:

$$P(y = 1|X) = \frac{\exp(\sum BX)}{1 + \exp(\sum BX)} \quad (2)$$

where P is the probability of the dependent variable; X represents the independent variables, $X = (x_0, x_1, x_2, \dots, x_k)$, $x_0 = 1$; B represents the estimated parameters, $B = (b_0, b_1, b_2, \dots, b_k)$.

In order to linearize the model and remove the 0/1 boundaries of the original dependent variable (probability), the following transformation is commonly applied:

$$P' = \ln(P/(1-P)) \quad (3)$$

This transformation is referred to as the logit or logistic transformation. Note that after the transformation

P' can theoretically assume any value between minus and plus infinity (Lewicki and Hill 2006). By performing the logit transformation on both sides of the above logit regression model, we obtain the standard linear regression model:

$$\ln(P/(1-P)) = b_0 + b_1*x_1 + b_2*x_2 + \dots + b_k*x_k + error_term \quad (4)$$

The main driving forces determining land use change were investigated in this study and logistic regression was used to create probability surfaces to determine the most probable sites that were developed. The dependent variable is the area of change (e.g., change from non-residential land to residential land); it has a binary form where a value of 1 indicates a change to residential land a value of and zero indicates areas of no change within a time period (Fig. 2). The independent variables have been described in “Data and data processing”. Figure 3 shows the spatial representation of the independent variables. The relative operating characteristic (ROC) method was used to validate the performance of the approach (Alsharif and Pradhan 2013; Pontius and Schneider 2001). ROC = 1 indicates a perfect fit and ROC = 0.5 indicates a random fit. In this study, stratified random sampling was chosen to eliminate spatial autocorrelation.

CA–Markov model

The Markov chain model is a random process model that describes how likely it is that one state (t_1) changes

into another state (t_2) (Houet and Hubert-Moy 2006). Based on the Bayes’ theorem of conditional probability, the land use change is calculated using the following formula (Sang et al. 2011):

$$S(t+1) = P_{ij} \times S(t) \quad (5)$$

where $S(t)$ and $S(t+1)$ are the system statuses at the time of t and $t+1$. P_{ij} is the transition probability matrix in a state and is calculated as follows:

$$P_{ij} = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1n} \\ P_{21} & P_{22} & \dots & P_{2n} \\ \dots & \dots & \dots & \dots \\ P_{n1} & P_{n2} & \dots & P_{nn} \end{bmatrix} \quad (6)$$

$$0 \leq P_{ij} < 1 \text{ and } \sum_{j=1}^n P_{ij} = 1, (i, j = 1, 2, 3 \dots n)$$

However, in a Markov chain model, the spatial distribution of the land cover categories are unknown (Ye and Bai 2008). To address this problem, the CA–Markov model was developed to add a spatial dimension to the model using CA. CA has been widely used to simulate urban sprawl (Arsanjani et al. 2013; Dietzel and Clarke 2004; Guan et al. 2011; Torrens 2006; White and Engelen 1993). A CA model is an agent or object that has the ability to change its state based on the application of a rule that relates the new state to its previous state and those of its neighbors (Eastman 2009; Surabuddin Mondal et al. 2013). The CA model combined with powerful spatial computing can be used to predict the spatial variation. It has been demonstrated that the integration of a Markov chain model and CA was effective for predicting land use changes (Behera et al. 2012).

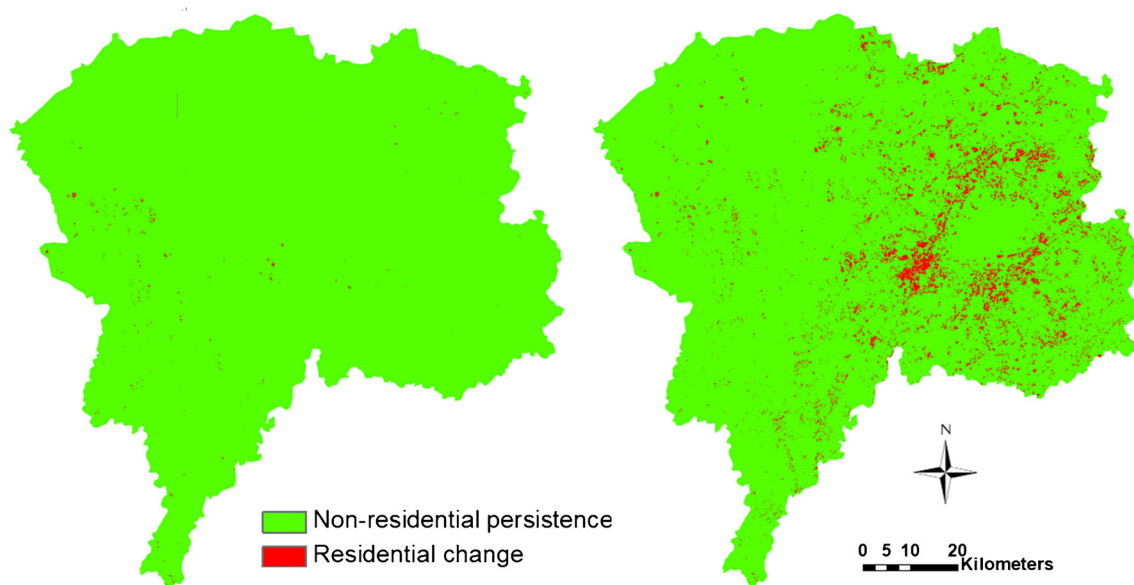
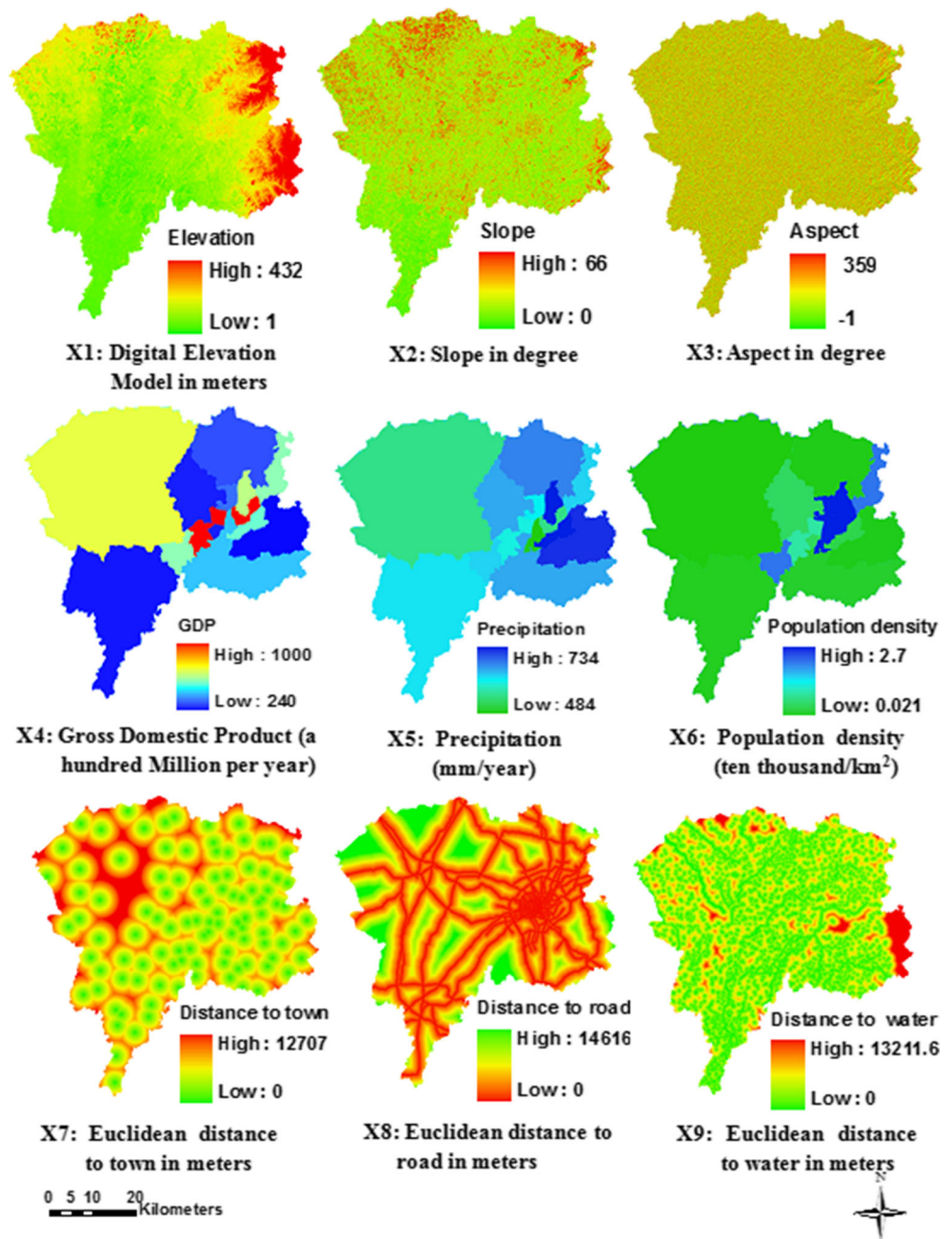


Fig. 2 Dependent variable Y; change to residential land during 2000–2005 (left) and 2000–2010 (right) (no change: green cells; change to residential land: red cells)

Fig. 3 Spatial representation of the independent variables



Results

Description of land use change during 2000–2010

The land use change from 2000 to 2010 is determined to quantify the extent and location of change. The changes are shown in Table 1 and Fig. 4.

As shown in Table 1, the main land use types in the area are cultivated land and residential land, which accounted for 85.5% of the total area in 2010. Residential land (including urban residential, and rural settlements) increased 37.4% from 2000 to 2010, whereas cultivated land showed

a decreasing trend (− 6.7%), reflecting the rapid quick decline in cultivated land resources. In the period 2000–2010, approximately 41,576.9 ha of cultivated land changed into other land types. Mining lease, industrial land, urban traffic, and undisturbed land also increased in this period whereas forest land and water exhibited fewer changes. The most apparent trend is the expansion of built-up land including residential land (37.4%), urban traffic (29.3%), and industrial land (10.1%); most changes were located in the northeastern, eastern, and southeastern parts of the study area. Additionally, changes were observed in the main core of the metropolis (Fig. 4).

Table 1 Area of land use changes during 2000–2010

	ML	FL	IL	CL	WT	UT	RL	UD
2000 (ha)	201.7	85,974.0	675.1	622,931.1	25,374.5	5318.6	100,651.6	935.0
2010 (ha)	244.8	86,781.0	743.1	581,354.2	26,617.3	6879.6	138,333.4	1108.3
Changed area (ha)	43.1	806.9	68.0	-41,576.9	1242.8	1561.0	37,681.8	173.3
Change ratio (%)	21.4	0.9	10.1	-6.7	4.9	29.3	37.4	18.5

ML mining lease, *FL* forest land, *IL* industrial land, *CL* cultivated land, *WT* water, *UT* urban traffic, *RL* residential land, *UD* undisturbed (desert and bare land)

Quantification of land change and transition potential maps

The Markov chain model was used to analyze the land cover images in two time periods and output a transition probability matrix, a transition area matrix, and a set of conditional probability images. The probability matrix shows the probability of change of the land cover categories into other land cover categories (Table 2). For example, during 2000–2010, there was a 22.76% chance that cultivated land would transition into residential land and only a 1.98% chance that cultivated land would turn into forest land. During the same period, there was a 15.65% chance that forest land would transition into cultivated land. The transition area matrix lists the number of pixels that are expected to change from each land cover type to each other land cover type over the specified number of time units (Table 3). Figure 5 shows the Markov transition areas and suitability maps of the major four land cover classes in an area of central Liaoning province in northern China. An average ROC value of 0.824 was obtained, which verifies the validity of the logistic regression model to predict the most

probable areas of development. These results were used for the subsequent land use change predictions.

Land use change prediction results

We compared the simulated land use map of 2010 derived from the CA–Markov model with the actual land use map of 2010. A match of 83.7% was achieved between the simulated and actual maps of 2010, which represents a satisfactory calibration. The Kappa index was 0.86, indicating a very good agreement between the simulated and observed land cover. Thus, the approach used to predict the land cover in the future (i.e., 2020). Figure 6 b illustrates the spatial pattern of land cover in 2020.

Table 4 shows the areas of land use change occurring in the time period 2010–2020. The forest land and cultivated land exhibited significant reductions, for example, the cultivated land decreased by 18.0%. Much of the cultivated land transitioned into other land types (Table 3). Approximately 104,673.7 ha of cultivated land changed into other land types in 2010–2020. There is an increasing trend in the built-up area

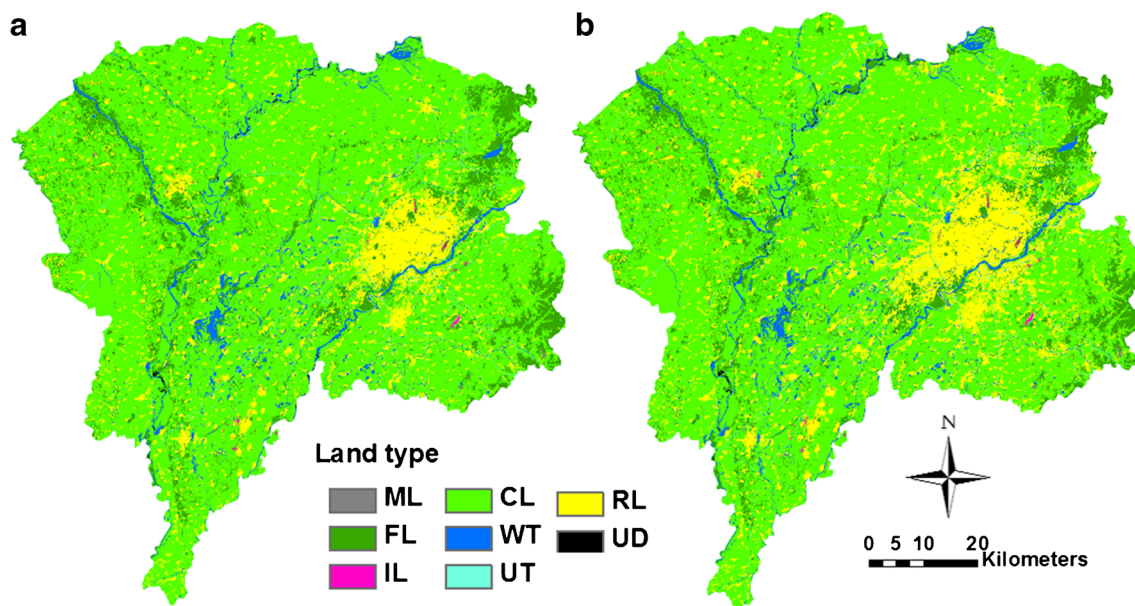


Fig. 4 Land use map of **a** 2000 and **b** 2010. Note: *ML*—mining lease, *FL*—forest land, *IL*—industrial land, *CL*—cultivated land, *WT*—water, *UT*—urban traffic, *RL*—residential land, and *UD*—undisturbed (desert and bare land)

Table 2 Markov transition probability matrix

		ML	FL	IL	CL	WT	UT	RL	UD
Probability value of 2010 based on transition matrix of 2000–2005	ML	0.8500	0.0214	0.0214	0.0214	0.0214	0.0214	0.0214	0.0214
	FL	0.0035	0.8395	0.0000	0.1480	0.0037	0.0001	0.0052	0.0000
	IL	0.0000	0.1299	0.8484	0.0217	0.0000	0.0000	0.0000	0.0000
	CL	0.0000	0.0246	0.0017	0.8470	0.0128	0.0529	0.0609	0.0000
	WT	0.0000	0.0190	0.0000	0.1294	0.8433	0.0000	0.0083	0.0000
	UT	0.0000	0.0000	0.0000	0.1501	0.0000	0.8499	0.0000	0.0000
	RL	0.0000	0.0085	0.0000	0.1217	0.0170	0.0028	0.8500	0.0000
	UD	0.0000	0.0000	0.0000	0.1652	0.0000	0.0000	0.0016	0.8331
Probability value of 2020 based on transition matrix of 2000–2010	ML	0.8500	0.0214	0.0214	0.0214	0.0214	0.0214	0.0214	0.0214
	FL	0.0014	0.8201	0.0002	0.1570	0.0108	0.0013	0.0090	0.0003
	IL	0.0000	0.0139	0.8037	0.0835	0.0000	0.0000	0.0989	0.0000
	CL	0.0001	0.0168	0.0005	0.7889	0.0073	0.0072	0.1783	0.0010
	WT	0.0000	0.0507	0.0003	0.1015	0.8197	0.0008	0.0138	0.0133
	UT	0.0000	0.0313	0.0000	0.0251	0.0000	0.8497	0.0940	0.0000
	RL	0.0000	0.0033	0.0009	0.0103	0.1351	0.0033	0.8459	0.0011
	UD	0.0000	0.0079	0.0000	0.0536	0.0013	0.0000	0.1842	0.7530

ML mining lease, FL forest land, IL industrial land, CL cultivated land, WT water, UT urban traffic, RL residential land, UD undisturbed (desert and bare land)

including residential land (63.6%), urban traffic (54.4%), and industrial land (29.4%). The urban expansion is more extensive during 2010–2020 than during 2000–2010. There is a new wave of suburban development in the southwestern, western, and northwestern parts of the study area. Significant changes were observed in the main core of the metropolis (Fig. 6).

Discussion

In this study, logistic regression, a Markov chain model, and a CA model were integrated to simulate urban expansion. Logistic regression is one of the most frequently used approaches to determine the suitability of a particular land use

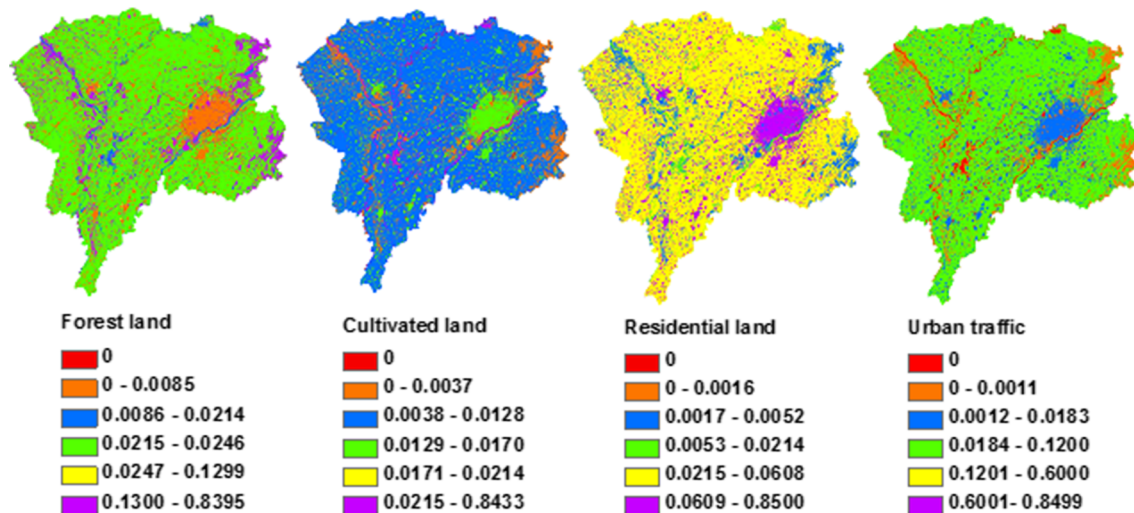
type in land use change modeling (Verhagen 2007). Our results indicate that logistic–CA–Markov model shows good predictive performance for land use changes in a cool temperate region, such as our study site located in northern China. The results are similar to those of other previous studies using the logistic regression–CA–Markov models to predict land use changes (Arsanjani et al. 2013; Memarian et al. 2012; Siddiqui et al. 2017; Sun et al. 2018; Wang et al. 2019c). Moreover, we should point out that different thematic resolutions (2-, 4-, 5-, 6-, 8- and 10-class land use maps) were used to simulate land use changes in the aforementioned studies. For example, Arsanjani et al. (2013) integrated logistic regression, Markov chain, and CA models to simulate urban expansion (5-classes). An agreement of 89.0% was observed between the simulated and actual maps. In our study, the logistic

Table 3 Transition area matrix based on the Markov model for 2020 (ha)

		ML	FL	IL	CL	WT	UT	RL	UD
2020	ML	208.1	5.2	5.2	5.2	5.2	5.2	5.2	5.2
	FL	118.8	71,166.4	16.7	13,627.7	938.2	108.6	780.3	24.1
	IL	0.0	10.4	597.2	62.1	0.0	0.0	73.5	0.0
	CL	55.0	9762.4	267.0	458,613.3	4225.4	4186.4	103,682.6	562.0
	WT	0.0	1349.1	7.2	2701.0	21,818.6	20.1	367.8	353.5
	UT	0.0	215.5	0.0	172.4	0.0	5845.3	646.5	0.0
	RL	0.0	462.2	123.6	1422.7	18,694.1	454.3	117,021.2	155.4
	UD	0.0	8.8	0.0	59.4	1.4	0.0	204.2	834.6

ML mining lease, FL forest land, IL industrial land, CL cultivated land, WT water, UT urban traffic, RL residential land, UD undisturbed (desert and bare land)

Transition probability areas for 2010 land use change prediction using Markov model



Probabilities surface resulting from logistic regression

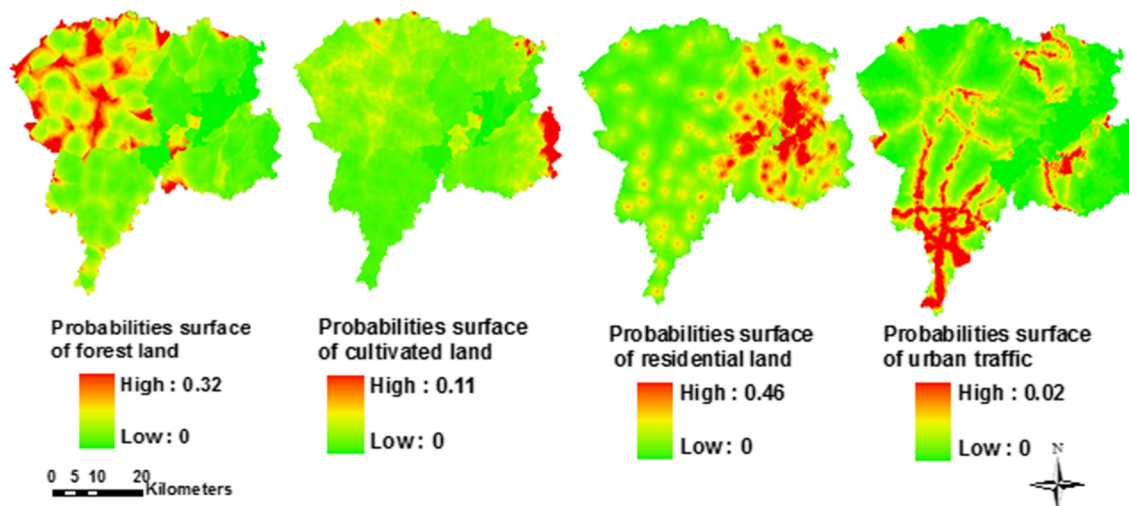


Fig. 5 An illustrative example of the Markov transition probability areas and suitability maps of major four land covers for 2010

regression–CA–Markov model was used to predict land use change for 8-classes. The overall accuracy was 83.7%. Memarian et al. (2012) simulated land-use/cover changes (10-classes) using a logistic regression–CA–Markov model and a match of 89.9% was achieved between the predicted and actual maps. These results demonstrate that a higher thematic resolution does not result in lower OPA. This further emphasizes the potential of the logistic regression–CA–Markov model for land use change prediction.

The model outputs indicate that our study area is facing unprecedented challenges due to rapid urbanization. Based on the latest city plans of the Shenyang government, the

urbanization rate will reach 90% in 2030. This situation will further exacerbate the extensive environmental pollution currently existing in this area (Fu et al. 2011; Liu et al. 2017a; Wang et al. 2013b; Yue and Du 2010). For example, the built-up land (consisting of industrial land, urban traffic areas and residential land) accounted for 12.67% of the total area in 2000 (Table 1) and increased significantly and reached 28.25% in 2020 (Table 4). As urbanization progresses, the area of impervious surfaces, such as pavement, rooftops, and compacted soil increases (Choe et al. 2002; Ferreira et al. 2013); as a result, runoff volumes from terrestrial catchments increase (Arnold Jr. and Gibbons 1996; Coker et al. 2018;

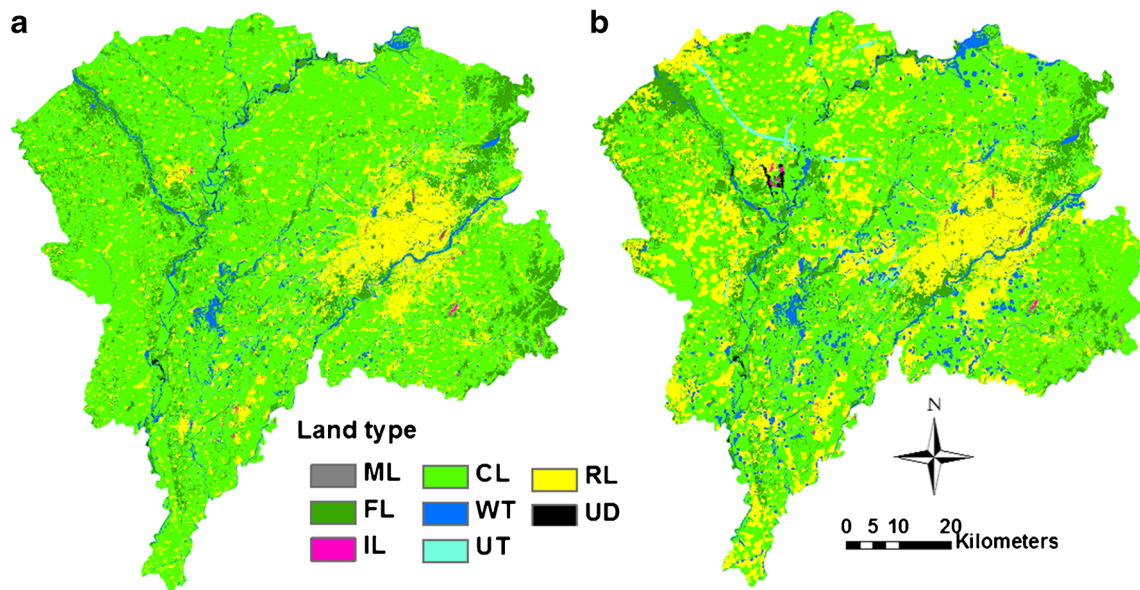


Fig. 6 Land use map of **a** 2010 and **b** simulated land use map of 2020. Note: ML—mining lease, FL—forest land, IL—industrial land, CL—cultivated land, WT—water, UT—urban traffic areas, RL—residential land, and UD—undisturbed (desert and bare land)

Desta et al. 2019; Shuster et al. 2005) and pollutants (heavy metals, nutrients, and organic compounds) increase in the water after precipitation events (Arnold Jr. and Gibbons 1996; Carey et al. 2013; Coker et al. 2018; Islam et al. 2015; Wu et al. 2016). In addition, there is also a significant increase in industrial land areas (Tables 1 and 4). An increasing number of factories discharge sewage into waterbodies, resulting in large challenges for water quality management in the future. Moreover, as shown in Tables 1 and 4, the large increase in urban traffic areas and residential land will cause extensive and dangerous air pollution (Liu et al. 2017a; Xue et al. 2016), such as smog/haze because many pollutants (e.g., PM_{2.5}, PM₁₀, SO₂, and NO₂) that affect human health are discharged from industry, car exhaust fumes (Moldovan et al. 1999), and coal-fired stoves in winter (Buhre et al. 2005; Xiao et al. 2015). For example, Liu et al. (2017a) found that SO₂ emissions in core urban areas were significantly higher than those in the surrounding urban areas during the heating season in Shenyang.

Our results also suggest that forest land and cultivated land have been shrinking owing to the urban expansion

and the current trend will continue (Tables 3 and 4). Therefore, it is urgent to strengthen the protection of forest land and cultivated land, prevent the indiscriminate use of these areas, and promote sustainable use. Protection of these lands will facilitate socially sustainable development in this region, where short-term economic benefits should be balanced with long-term environmental and economic sustainability.

Although the logistic regression–CA–Markov is an effective technique for investigating land use change, there are several factors that significantly affect the prediction uncertainties in land use modeling. First, previous studies have shown that the main factors affecting model prediction accuracy are the driving forces (Arsanjani et al. 2013; Memarian et al. 2012; Park et al. 2011). Thus, the choice of an optimal set of driving forces can improve model prediction accuracy. In our study, several environmental and socio-economic factors were considered. However, a limited number of driving factors may have resulted in errors in the estimation of land use changes. Hence, additional socio-economic factors should be considered in

Table 4 Area of land use change during 2010–2020

	ML	FL	IL	CL	WT	UT	RL	UD
2010 (ha)	244.8	86,781.0	743.1	581,354.2	26,617.3	6879.6	138,333.4	1108.3
2020 (ha)	391.4	82,980.3	961.8	476,680.5	42,192.1	10,620.9	226,305.5	1929.2
Changed area (ha)	146.6	−3800.7	218.7	−104,673.7	15,574.8	3741.3	87,972.1	820.9
Change ratio (%)	59.9	−4.4	29.4	−18.0	58.5	54.4	63.6	74.1

ML mining lease, FL forest land, IL industrial land, CL cultivated land, WT water, UT urban traffic, RL residential land, UD undisturbed (desert and bare land)

land change modeling, such as the distance to hospitals and the distance to markets to improve model prediction accuracy. Second, land development policy has been inconsistent in recent years for the revitalization of the old industrial base in Northeast China (Wang et al. 2013a), which may also have caused uncertainty in the logistic regression–CA–Markov model predictions. For example, if a plot of land has been allocated for the purpose of developing a garden, the developer may instead choose to turn the garden into a factory. Finally, logistic regression techniques suffer from certain limitations, such as spatial autocorrelation of the independent variables (Arsanjani et al. 2013; Fotheringham et al. 2000; Hu and Lo 2007; Smith 1994), which may lead to errors in the suitability images (e.g., suitability images for predicting the land use map of 2020) and subsequently cause uncertainty in land use modeling. In summary, land use modeling is a complex process that is affected by natural factors, human-induced driving factors, and model limitations and may result in prediction errors of land use changes; therefore, model outputs of interest should be used with caution.

Conclusion

We used land use maps from different time periods (2000, 2005, and 2010) and integrated the logistic regression, CA, and a Markov model to successfully simulate land use changes in Shenyang city in northern China. The results indicate that the hybrid model has good potential to simulate land use changes. However, there are many significant uncertainty factors that affect the prediction accuracy of land use change, such as socio-economic driving forces, model limitations, and land development policy and result in simulation errors of the land use dynamics. Consequently, in order to improve the prediction accuracy, the uncertainty factors should be thoroughly considered and addressed when predicting land use dynamics.

The model outputs showed that our study area is facing unprecedented challenges due to rapid urbanization, which is likely to exacerbate the significant environmental pollution currently existing in this area. This poses a great challenge for environmental management in the future, and adaptation measures (e.g., low impact development and use of clean energy) should be implemented. In addition, forest land and cultivated land have been shrinking due to urban expansion. Therefore, it is urgent to prevent the indiscriminate use of forest land and cultivated land and promote the sustainable use of land. The results of this study are expected to provide input for land use management and environmental protection in this area.

References

- Al-sharif AAA, Pradhan B (2014) Monitoring and predicting land use change in Tripoli Metropolitan City using an integrated Markov chain and cellular automata models in GIS Arabian. *J Geosci* 7: 4291–4301. <https://doi.org/10.1007/s12517-013-1119-7>
- Alsharif AA, Pradhan B (2013) Urban sprawl analysis of Tripoli Metropolitan city (Libya) using remote sensing data and multivariate logistic regression model *Journal of the Indian Society of Remote Sensing*:1–15
- Arnold CL Jr, Gibbons CJ (1996) Impervious surface coverage: the emergence of a key environmental indicator. *J Am Plan Assoc* 62:243–258
- Arsanjani JJ, Helbich M, Kainz W, Darvishi Boloorani A (2013) Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion. *Int J Appl Earth Obs Geoinf* 21:265–275
- Arsanjani JJ, Kainz W, Mousivand AJ (2011) Tracking dynamic land-use change using spatially explicit Markov Chain based on cellular automata: the case of Tehran. *Int J Image Data Fusion* 2:329–345. <https://doi.org/10.1080/19479832.2011.605397>
- Behera MD, Borate SN, Panda SN, Behera PR, Roy PS (2012) Modelling and analyzing the watershed dynamics using cellular automata (CA)–Markov model—a geo-information based approach *Journal of earth system science* 121:1011–1024
- Buhre B, Elliott L, Sheng C, Gupta R, Wall T (2005) Oxy-fuel combustion technology for coal-fired power generation. *Prog Energy Combust Sci* 31:283–307
- Carey RO, Hochmuth GJ, Martinez CJ, Boyer TH, Dukes MD, Toor GS, Cisar JL (2013) Evaluating nutrient impacts in urban watersheds: challenges and research opportunities. *Environ Pollut* 173:138–149
- Chen X, Yu S-X, Zhang Y-P (2013) Evaluation of spatiotemporal dynamics of simulated land use/cover in china using a probabilistic cellular automata-Markov model. *Pedosphere* 23:243–255. [https://doi.org/10.1016/S1002-0160\(13\)60013-2](https://doi.org/10.1016/S1002-0160(13)60013-2)
- Choe J, Bang K, Lee J (2002) Characterization of surface runoff in urban areas. *Water Sci Technol* 45:249–254
- Coker ME, Bond NR, Chee YE, Walsh CJ (2018) Alternatives to biodiversity offsets for mitigating the effects of urbanization on stream ecosystems. *Conserv Biol* 32:789–797. <https://doi.org/10.1111/cobi.13057>
- Desta Y, Goitom H, Aregay G (2019) Investigation of runoff response to land use/land cover change on the case of Aynalem catchment. *Journal of African Earth Sciences* 153:130–143. <https://doi.org/10.1016/j.jafrearsci.2019.02.025>
- Dickinson RE, Henderson-Sellers A (1988) Modelling tropical deforestation: A study of GCM land-surface parametrizations Quarterly. *J R Meteorol Soc* 114:439–462
- Dietzel C, Clarke K (2006) The effect of disaggregating land use categories in cellular automata during model calibration and forecasting computers. *Environ Urban Syst* 30:78–101
- Dietzel C, Clarke KC (2004) Replication of spatio-temporal land use patterns at three levels of aggregation by an urban cellular automata. *Cellular Automata*. Springer, In, pp 523–532
- Eastman J (2009) Idrisi taiga Worcester. Clark University, MA
- Etemadi H, Smoak JM, Karami J (2018) Land use change assessment in coastal mangrove forests of Iran utilizing satellite imagery and CA–Markov algorithms to monitor and predict future change. *Environ Earth Sci* 77:208–213. <https://doi.org/10.1007/s12665-018-7392-8>
- Ferreira M, Lau S-L, Stenstrom M (2013) Size fractionation of metals present in highway runoff: beyond the six commonly reported species. *Water Environ Res* 85:793–805
- Fotheringham AS, Brunson C, Charlton M (2000) Quantitative geography: perspectives on spatial data analysis. Sage

- Fu B, Su J, Wu D, Li F, Xu C, Xie Y, Zhang Z (2011) Evaluation of temporal and spatial differences of water environment in Liaohe Basin. *Water Resour Prot* 27:5–8
- Fu X, Wang X, Yang YJ (2018) Deriving suitability factors for CA-Markov land use simulation model based on local historical data. *J Environ Manag* 206:10–19. <https://doi.org/10.1016/j.jenvman.2017.10.012>
- Geng Y, Sarkis J, Wang X, Zhao H, Zhong Y (2013) Regional application of ground source heat pump in China: a case of Shenyang. *Renew Sust Energ Rev* 18:95–102
- Guan D, Li H, Inohae T, Su W, Nagaie T, Hokao K (2011) Modeling urban land use change by the integration of cellular automaton and Markov model. *Ecol Model* 222:3761–3772
- Hagenauer J, Helbich M (2018) Local modelling of land consumption in Germany with RegioClust. *Int J Appl Earth Obs Geoinf* 65:46–56. <https://doi.org/10.1016/j.jag.2017.10.003>
- Hamdy O, Zhao S, Osman T, Salheen AM, Eid YY (2016) Applying a hybrid model of Markov chain and logistic regression to identify future urban sprawl in Abouelreesh. *Aswan: A Case Study Geosciences* 6. <https://doi.org/10.3390/geosciences6040043>
- Houet T, Hubert-Moy LL (2006) Modeling and projecting land-use and land-cover changes with Cellular Automaton in considering landscape trajectories. *EARSeL eProceedings* 5:63–76
- Hu Z, Lo CP (2007) Modeling urban growth in Atlanta using logistic regression computers. *Environ Urban Syst* 31:667–688. <https://doi.org/10.1016/j.compenvurbsys.2006.11.001>
- Islam K, Rahman MF, Jashimuddin M (2018) Modeling land use change using cellular automata and artificial neural network: the case of Chhunati wildlife sanctuary. *Bangladesh Ecological Indicators* 88: 439–453. <https://doi.org/10.1016/j.ecolind.2018.01.047>
- Islam MS, Ahmed MK, Raknuzzaman M, Habibullah-Al-Mamun M, Islam MK (2015) Heavy metal pollution in surface water and sediment: a preliminary assessment of an urban river in a developing country. *Ecol Indic* 48:282–291. doi:<https://doi.org/10.1016/j.ecolind.2014.08.016>
- Kamusoko C, Aniya M, Adi B, Manjoro M (2009) Rural sustainability under threat in Zimbabwe—simulation of future land use/cover changes in the Bindura district based on the Markov-cellular automata model. *Appl Geogr* 29:435–447
- Ku C-A (2016) Incorporating spatial regression model into cellular automata for simulating land use change. *Applied Geography* 69:1–9. <https://doi.org/10.1016/j.apgeog.2016.02.005>
- Lambin EF (1997) Modelling and monitoring land-cover change processes in tropical regions. *Prog Phys Geogr* 21:375–393
- Lewicki P, Hill T (2006) *Statistics: methods and applications*. Tulsa. Statsoft, OK
- Lin B, Zhu J (2018) Changes in urban air quality during urbanization in China. *J Clean Prod* 188:312–321. <https://doi.org/10.1016/j.jclepro.2018.03.293>
- Liu C, Yuan Z, Du Y, Qi X, Shi C, Wang N, Han X (2017a) Spatial distribution characteristics of air pollutants during heating season in Shenyang city. In: 2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCI), 8–11 Sept. 2017. pp 16–19. <https://doi.org/10.1109/CIAPP.2017.8167052>
- Liu X et al (2017b) A future land use simulation model (FLUS) for simulating multiple land use scenarios by coupling human and natural effects. *Landsc Urban Plan* 168:94–116. <https://doi.org/10.1016/j.landurbplan.2017.09.019>
- Liu Y, Dai L, Xiong H (2015) Simulation of urban expansion patterns by integrating auto-logistic regression, Markov chain and cellular automata models. *Journal of Environmental Planning and Management* 58:1113–1136. <https://doi.org/10.1080/09640568.2014.916612>
- Long JS (1997) Regression models for categorical and limited dependent variables vol 7. Sage
- Luo G, Amuti T, Zhu L, Mambetov BT, Maisupova B, Zhang C (2015) Dynamics of landscape patterns in an inland river delta of Central Asia based on a cellular automata-Markov model. *Reg Environ Chang* 15:277–289. <https://doi.org/10.1007/s10113-014-0638-4>
- MacCullagh P, Nelder JA (1989) *Generalized linear models* vol 37. CRC press,
- Memarian H, Balasundram SK, Talib JB, Sung CTB, Sood AM, Abbaspour K (2012) Validation of CA-Markov for simulation of land use and cover change in the Langat Basin. *Malaysia J Geogr Inf Syst* 4:542–554
- Meyer WB, Turner I (1994) *Changes in land use and land cover: a global perspective* vol 4. Cambridge University Press,
- Mishra VN, Rai PK (2016) A remote sensing aided multi-layer perceptron-Markov chain analysis for land use and land cover change prediction in Patna district (Bihar). *India Arab J Geosci* 9: 249. <https://doi.org/10.1007/s12517-015-2138-3>
- Mitsova D, Shuster W, Wang X (2011) A cellular automata model of land cover change to integrate urban growth with open space conservation. *Landsc Urban Plan* 99:141–153
- Moldovan M, Gomez MM, Palacios MA (1999) Determination of platinum, rhodium and palladium in car exhaust fumes. *J Anal At Spectrom* 14:1163–1169
- Mustafa A, Heppenstall A, Omrani H, Saadi I, Cools M, Teller J (2018) Modelling built-up expansion and densification with multinomial logistic regression, cellular automata and genetic algorithm computers. *Environ Urban Syst* 67:147–156. <https://doi.org/10.1016/j.compenvurbsys.2017.09.009>
- Naboureh A, Rezaei Moghaddam MH, Feizizadeh B, Blaschke T (2017) An integrated object-based image analysis and CA-Markov model approach for modeling land use/land cover trends in the Sarab plain. *Arab J Geosci* 10:259–216. <https://doi.org/10.1007/s12517-017-3012-2>
- Omrani H, Tayyebi A, Pijanowski B (2017) Integrating the multi-label land-use concept and cellular automata with the artificial neural network-based Land Transformation Model: an integrated ML-CA-LTM modeling framework. *GIScience & Remote Sensing* 54: 283–304. <https://doi.org/10.1080/15481603.2016.1265706>
- Park S, Jeon S, Kim S, Choi C (2011) Prediction and comparison of urban growth by land suitability index mapping using GIS and RS in South Korea. *Landsc Urban Plan* 99:104–114
- Peng C, Cai Y, Wang T, Xiao R, Chen W (2016) Regional probabilistic risk assessment of heavy metals in different environmental media and land uses: An urbanization-affected drinking water supply area. *Sci Rep* 6:37084. <https://doi.org/10.1038/srep37084> <https://www.nature.com/articles/srep37084#supplementary-information>
- Pontius RG, Schneider LC (2001) Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. *Agric Ecosyst Environ* 85:239–248
- Rosenfield GH, Fitzpatrick K (1986) A coefficient of agreement as a measure of thematic classification accuracy. *Photogramm Eng Remote Sens* 52:223–227
- Sang L, Zhang C, Yang J, Zhu D, Yun W (2011) Simulation of land use spatial pattern of towns and villages based on CA-Markov model. *Math Comput Model* 54:938–943
- Shao M, Tang X, Zhang Y, Li W (2006) City clusters in China: air and surface water pollution. *Front Ecol Environ* 4:353–361
- Shuster W, Bonta J, Thurston H, Warnemuende E, Smith D (2005) Impacts of impervious surface on watershed hydrology: a review. *Urban Water J* 2:263–275
- Siddiqui A, Siddiqui A, Maithani S, Jha AK, Kumar P, Srivastav SK (2017) Urban growth dynamics of an Indian metropolitan using CA Markov and logistic regression. *Egypt J Remote Sens Space Sci*. <https://doi.org/10.1016/j.ejrs.2017.11.006>
- Smith PA (1994) Autocorrelation in logistic regression modelling of species' distributions. *Glob Ecol Biogeogr Lett* 4:47–61. <https://doi.org/10.2307/2997753>

- Sun X, Crittenden JC, Li F, Lu Z, Dou X (2018) Urban expansion simulation and the spatio-temporal changes of ecosystem services, a case study in Atlanta Metropolitan area, USA. *Sci Total Environ* 622–623:974–987. <https://doi.org/10.1016/j.scitotenv.2017.12.062>
- Surabuddin Mondal M, Sharma N, Kappas M, Garg PK (2013) Modeling of spatio-temporal dynamics of land use and land cover in a part of Brahmaputra River basin using Geoinformatic techniques *Geocarto Int* 28:632–656 doi:<https://doi.org/10.1080/10106049.2013.776641>
- Torrens PM (2006) Geosimulation and its application to urban growth modeling. *Complex artificial environments*. Springer, In, pp 119–136
- Vaz EdN, Nijkamp P, Painho M, Caetano M (2012) A multi-scenario forecast of urban change: a study on urban growth in the Algarve Landscape and Urban Planning 104:201–211
- Veldkamp A, Fresco L (1996) CLUE-CR: an integrated multi-scale model to simulate land use change scenarios in Costa Rica. *Ecol Model* 91:231–248
- Verburg PH, Soepboer W, Veldkamp A, Limpiada R, Espaldon V, Mastura SSA (2002) Modeling the spatial dynamics of regional land use: the CLUE-S model. *Environ Manag* 30:391–405. <https://doi.org/10.1007/s00267-002-2630-x>
- Verhagen P (2007) Case studies in archaeological predictive modelling: Proefschrift vol 14. Amsterdam University Press
- Wang C, Lei S, Elmore JA, Jia D, Mu S (2019a) Integrating temporal evolution with cellular automata for simulating land cover change. *Remote Sens*:11. <https://doi.org/10.3390/rs11030301>
- Wang E, Li Q, Hu H, Peng F, Zhang P, Li J (2019b) Spatial characteristics and influencing factors of river pollution in China. *Water Environ Res* 91:351–363. <https://doi.org/10.1002/wer.1044>
- Wang H, Stephenson SR, Qu S (2019c) Modeling spatially non-stationary land use/cover change in the lower Connecticut River Basin by combining geographically weighted logistic regression and the CA-Markov model. *Int J Geogr Inf Sci* 33:1313–1334. <https://doi.org/10.1080/13658816.2019.1591416>
- Wang Q, Li SP, Cui MX (2013a) Study on the status and development strategy of low-carbon economy in the northeast old industrial base. *Appl Mech Mater* 291:1455–1460
- Wang X, Cai M, Zhong B, Yao Y, Yin S, Wu D (2013b) Research on spatial characteristic of non-point source pollution in Liaohe River basin. *Huan jing ke xue* 34:3788–3796
- Watson RT (2000) Land use, land-use change, and forestry: a special report of the intergovernmental panel on climate change. Cambridge University Press
- White R, Engelen G (1993) Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns. *Environment and Planning A: Economy and Space* 25:1175–1199 doi:<https://doi.org/10.1068/a251175>
- Wu Q et al (2016) Contamination, toxicity and speciation of heavy metals in an industrialized urban river: implications for the dispersal of heavy metals. *Mar Pollut Bull* 104:153–161. <https://doi.org/10.1016/j.marpolbul.2016.01.043>
- Xiao Q, Ma Z, Li S, Liu Y (2015) The impact of winter heating on air pollution in China. *PLoS One* 10:e0117311. <https://doi.org/10.1371/journal.pone.0117311>
- Xue Y et al (2016) Trends of multiple air pollutants emissions from residential coal combustion in Beijing and its implication on improving air quality for control measures. *Atmos Environ* 142:303–312. <https://doi.org/10.1016/j.atmosenv.2016.08.004>
- Yang X, Zheng X-Q, Chen R (2014) A land use change model: Integrating landscape pattern indexes and Markov-CA. *Ecol Model* 283:1–7. <https://doi.org/10.1016/j.ecolmodel.2014.03.011>
- Ye B, Bai Z (2008) Simulating land use/cover changes of Nenjiang County based on CA-Markov model. *Computer And Computing Technologies In Agriculture, Volume I*. Springer, In, pp 321–329
- Yue Q, Du T (2010) Impact analysis of the industrial structure and distribution of Liaohe River Basin on the water quality. In: *Environmental Science and Information Application Technology (ESIAT), 2010 International Conference on*. IEEE, pp 263–266
- Zheng HW, Shen GQ, Wang H, Hong J (2015) Simulating land use change in urban renewal areas: a case study in Hong Kong. *Habitat Int* 46:23–34. <https://doi.org/10.1016/j.habitatint.2014.10.008>
- Zhou D, Lin Z, Liu L (2012) Regional land salinization assessment and simulation through cellular automaton-Markov modeling and spatial pattern analysis. *Sci Total Environ* 439:260–274. <https://doi.org/10.1016/j.scitotenv.2012.09.013>