

Article

Using a Cellular Automata-Markov Model to Reconstruct Spatial Land-Use Patterns in Zhenlai County, Northeast China

Yuanyuan Yang ^{1,2,3}, Shuwen Zhang ^{2,*}, Jiuchun Yang ², Xiaoshi Xing ³ and Dongyan Wang ¹

¹ College of Earth Science, Jilin University, 2199 Jianshe Street, Changchun 130061, China; E-Mails: sophiayangyuanyuan@hotmail.com (Y.Y.); wang_dy@jlu.edu.cn (D.W.)

² Northeast Institute of Geography and Agroecology, Chinese Academy Sciences, 4888 Shengbei Street, Changchun 130102, China; E-Mail: yangjiuchun0830@163.com

³ Center for International Earth Science Information Network (CIESIN), Earth Institute, Columbia University, P.O. Box 1000 (61 Route 9W), Palisades, NY 10964, USA; E-Mail: xxiaoshi@ciesin.columbia.edu

* Author to whom correspondence should be addressed; E-Mail: zhangshuwen@neigae.ac.cn; Tel.: +86-431-8554-2246.

Academic Editor: Xiangzheng Deng

Received: 16 January 2015 / Accepted: 24 April 2015 / Published: 5 May 2015

Abstract: Decadal to centennial land use and land cover change has been consistently singled out as a key element and an important driver of global environmental change, playing an essential role in balancing energy use. Understanding long-term human-environment interactions requires historical reconstruction of past land use and land cover changes. Most of the existing historical reconstructions have insufficient spatial and thematic detail and do not consider various land change types. In this context, this paper explored the possibility of using a cellular automata-Markov model in 90 m × 90 m spatial resolution to reconstruct historical land use in the 1930s in Zhenlai County, China. Then the three-map comparison methodology was employed to assess the predictive accuracy of the transition modeling. The model could produce backward projections by analyzing land use changes in recent decades, assuming that the present land use pattern is dynamically dependent on the historical one. The reconstruction results indicated that in the 1930s most of the study area was occupied by grasslands, followed by wetlands and arable land, while other land categories occupied relatively small areas. Analysis of the three-map comparison illustrated that the major differences among the three maps have less to do with the simulation model and more to do with the inconsistencies among the land categories during the study period.

Different information provided by topographic maps and remote sensing images must be recognized.

Keywords: cellular automata-Markov model; three-map comparison; historical reconstruction; spatial pattern; land use and land cover change; Northeast China

1. Introduction

Decadal to centennial land use and land cover change (LUCC) has been consistently singled out as a key element and an important driver of global environmental change [1–4]. LUCC could significantly affect key aspects of Earth system functioning [5], playing an essential role in balancing energy use. Land use activities, e.g., afforestation, deforestation, excessive reclamation, grazing, excavation and abandonment as well as urbanization, all influence runoff, evapotranspiration, the distribution of the precipitation among the soil waters and the land surface energy budget [6,7]. Ecologists increasingly recognize that a full understanding of ecosystems should be based on the analysis of ecosystem functioning over long time scales [8] as historical land use legacies have a strong and sometimes over-riding influence on the dynamics of present-day ecosystems, exhibiting a time-lagged response. Thus, understanding long-term human-environment interactions is essential to understanding changes in terrestrial ecosystems and this requires historical reconstruction of past land cover changes [9,10].

Historical reconstruction of land use/land cover (LULC) aims to reproduce information concerning past land use, not only the quantity of land use/cover in a historical period, but also its spatial distribution. Recently, improved remote sensing technology has made feasible the continuous observation of land cover. However, remotely-sensed data have only existed for the last four decades at most, following the advent of the first land satellite, LandSat-1, launched in 1972. Prior to that, other data sources must be relied on, which may cover the global scale but often inconsistently. In this context, increasing numbers of researchers have made efforts to reconstruct historical LULC based on prime data sources and research approaches, and substantial progress in gathering historical land change data has been made both at the global and regional scales [9,11–25]. This has been discussed in detail in a review of historical LULC reconstruction methods [26]. However, most existing historical LULC reconstructions do not sufficiently meet the requirements of climate assessments due to insufficient spatial and thematic detail and the lack of consideration of various land change types [27]. Current research mainly focuses on revising the historical land-cover quantity. There are few historical land-cover datasets with high spatial accuracy resulting in ineffective applications of the study results in simulations. The building of historical spatial land use and land cover datasets with high resolution is a timely research direction [26]. In addition, most current studies do not thematically represent 100% of the land area, and ignore the consideration of competing land categories and land conversion types. For example, current research mainly focuses on cultivated land, wetlands and forestland and it does not provide information on other land categories like settlements, water and other land.

In reconstructing historical LULC, it is highly worthwhile to use a reconstruction model. Land-use and land-cover change (LUCC) modeling as a scientific field is rapidly advancing, as land use change is among the most important human influences on the environment. The cellular automata (CA)-Markov

model is applicable to spatial land use simulations and land cover reconstructions [28], which could be capable of converting the quantitative results of the Markov chain into spatially explicit outcomes by means of a CA function [29]. This approach is also capable of simulating several land categories simultaneously. However, transition matrixes describing various land-covers from $t - 1$ to t might not be totally consistent with those in the period between t and $t + 1$, due to different changes among various land categories over time. Thus, the transition matrixes were modified by analyzing land use and its change in recent decades which was also used to define the rules for the cells of CA, assuming that current spatial pattern of land use is dynamically dependent on a historical pattern.

Evaluation of model performance often requires comparison of model simulations with observed outcomes. The proportion of observations classified correctly is perhaps the most commonly used measurement to compare two different expressions of a set of categories, for example, to compare land-cover categories expressed in a map and to reference data collected for the map's accuracy assessment [30]. It has become customary in the remote-sensing literature to report the kappa index of agreement along with the correct proportion, especially for purposes of accuracy assessment, since kappa compares two maps that show a set of categories. However, kappa indices are useless, misleading and flawed for the practical applications in the land use modeling because the kappa indices attempt to compare accuracy to a baseline of randomness which is not a reasonable alternative for map construction [30]. Based on the above analysis, Pontius, Jr. *et al.* [30–34] proposed the three-map comparison methodology to reveal the accuracy of the land change model *versus* a null model that predicts complete persistence. The two measures of quantity disagreement and allocation disagreement illuminate a much more enlightened path.

In the past century, western Jilin Province, having experienced a large-scale population migration and land reclamation process, has been one of the regions in China that reveals dramatic LUCC. This change will have a profound impact on the regional ecological environment and the balance of energy, with important regional research value. Here, considering the richness of regional LUCC data, we take Zhenlai County, located in northwestern Jilin Province, as an example. In this paper, our main objectives are: (1) to reconstruct historical land use in Zhenlai County in the early period of reclamation (the 1930s) using a $90\text{ m} \times 90\text{ m}$ CA-Markov model to produce a backward projection according to the land use/cover change rule from 1954 to 2005; (2) to validate reconstruction results using three-map comparison by classifying pixels as one of four types: null successes, hits, misses and false alarms.

2. Materials and Methods

2.1. Study Area

Zhenlai County (N45°28'-N46°18', E122°47'-E124°04', Figure 1), as a typical farming-pastoral ecotone, is located in northwestern Jilin Province, Northeast China, where it occupies the northernmost part of the province, bordering Heilongjiang to the east and Inner Mongolia to the west. It is under the administration of Baicheng City. Historically the region was the nomadic land of the Mongol princes, and inhabitants were not allowed to reclaim it until the enactment of the “lifting a ban on reclaiming” policy during the late Qing Dynasty (1902). The county was established in 1910 as Zhendong County,

and in 1947 Laibei County was merged into this region, which was renamed Zhenlai County. It has experienced a relatively complete and dramatic LUCC during the past century.

The county has a variety of geomorphologic types and features a terrain that is high in the northwest and low in the southeast. Its northwest is adjacent to the Greater Hinggan Mountain, its central area is mostly rolling hilly land, and its east and south surround the Nenjiang River and the Tao'er River, respectively, forming a fertile flood plain on the banks of both rivers. The major soil types are chernozem, alluvium soil, alkali soil and meadow soil [35].

Climatically, the region is subject to a temperate continental monsoon climate with distinct seasons, as it is located in a mid-latitude inland area. The mean annual rainfall is 402.4 mm, unevenly distributed over time, while the mean annual evaporation is 1755.9 mm, about four times as much as the mean annual rainfall. Thus, the low amount of precipitation and the high amount of evaporation mainly result in a drought-prone climate in the study area, especially in spring. The mean annual temperature is around 4.9 °C.

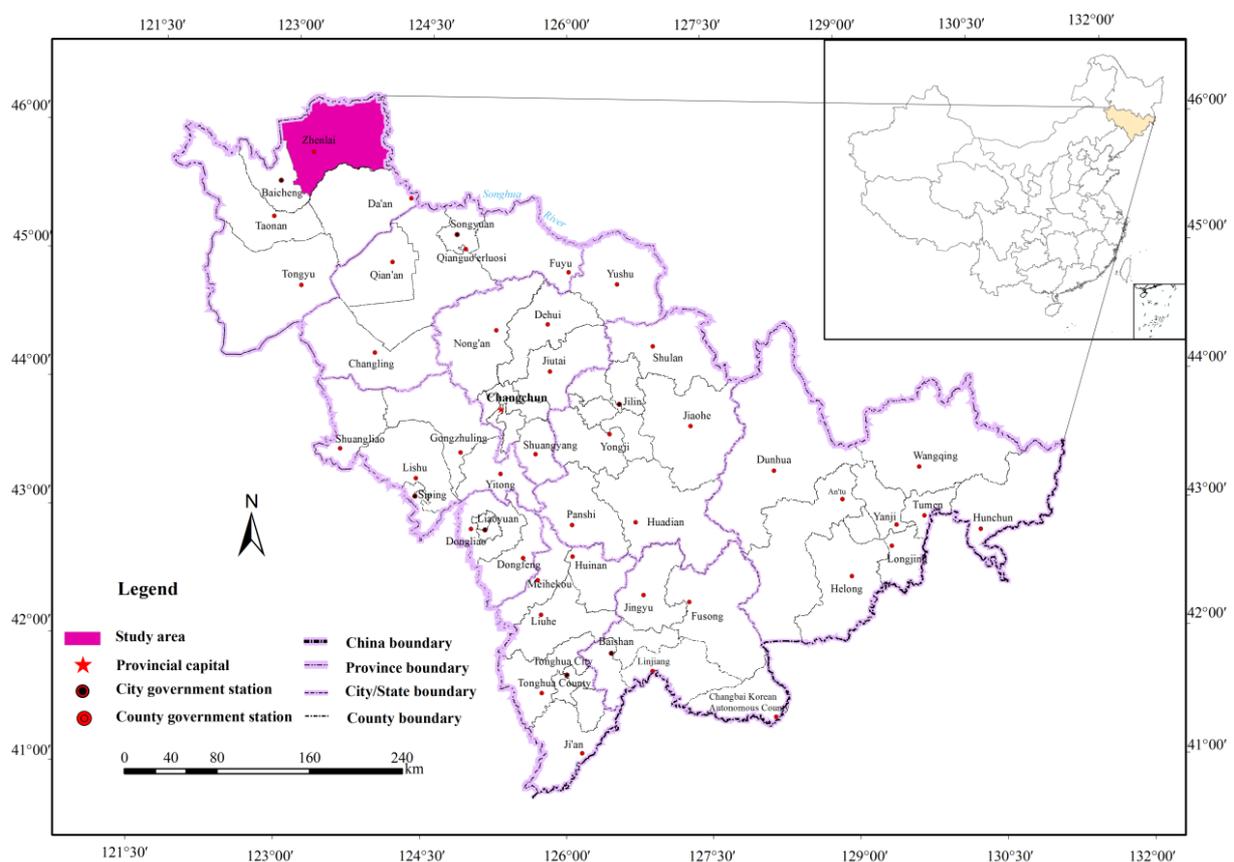


Figure 1. The study area, Zhenlai County.

2.2. Data

One Landsat Multispectral Scanner (MSS) and two Landsat Thematic Mapper (TM) images were selected pertaining to the years 1976, 2000, and 2005. Then the land use data were interpreted from the above remote sensing images which were downloaded from the United States Geological Survey (USGS) website (<http://glovis.usgs.gov/>). Meanwhile, our research team reconstructed the spatiotemporal distribution of land use and land cover in 1954 by making use of topographic maps and physical

environmental background maps, including those of terrain, climate, geology, soil, vegetation, hydrology, and socioeconomic statistical data [36,37]. Soil data was digitized from the “*Local Record of Zhenlai County*” [35]. Slope raster was generated from 90 m raster DEM data which is from the Shuttle Radar Topography Mission (SRTM) data. The data for model validation was digitized from topographic maps dating back to the 1930s, drawn to scale at 1:100,000.

2.3. Classification System

There are 14 land categories in the topographic maps in the 1930s according to the land use and land cover legends (Figure 2). The features of various land-covers in the topographic maps are described as below: (1) blank areas without any symbols represents rainfed land and it is always stocked with scattered grassland and rural settlements, *etc.*; (2) forestland in the topographic maps contains deciduous forests, orchards, coniferous forests, low pinewood and forests of unknown species, and there is no distribution of the last four land-cover types in the study area; most of the deciduous forests are located around settlements; (3) grassland is often mixed with other land covers and its boundaries are not easy to determine; besides, most of grassland in these maps is often judged as wildland, resulting in limited grassland symbols in the maps; (4) river have certain clear boundaries and it is easy to digitize them from other land use types; (5) lakes and ponds are described in the maps with the same symbols, and with certain boundaries; (6) settlements contain urban and rural construction, which are often surrounded by arable land; (7) wetlands have a clear distribution and location with certain boundaries and are sometimes mixed with wildland; (8) wildland mentioned here contains not only unused land (sand, saline-alkali land, bare land), but also meadow and prairie areas.

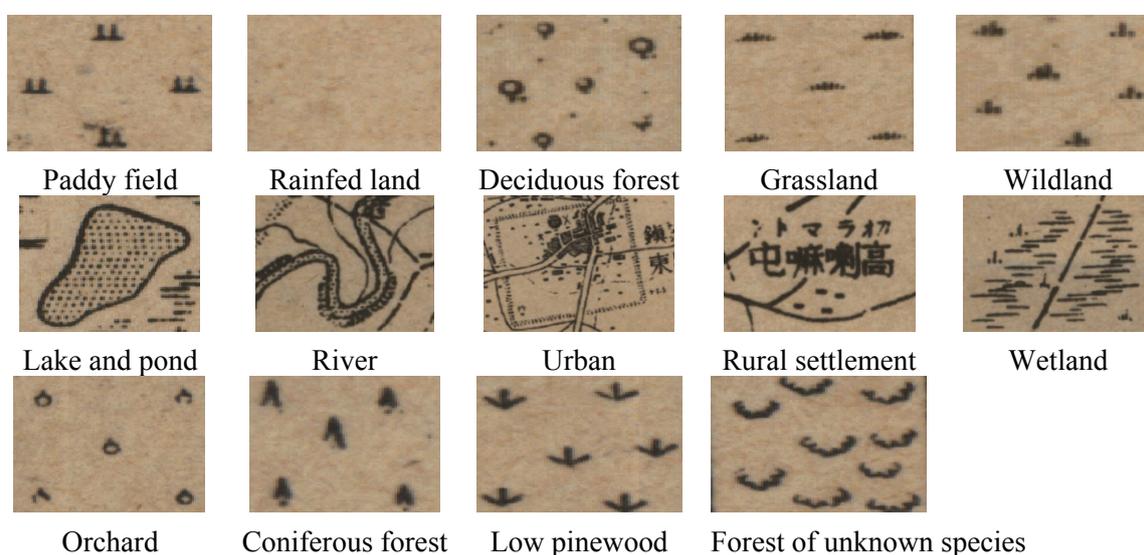


Figure 2. Legends of topographic maps at scale 1:100,000 in the 1930s.

As for the remote sensing images, in this study we took the land use classification system used by the Chinese Academy of Sciences that includes 20 subcategories (Table 1) [38]. Due to the relative maturity of the remote sensing image interpretation technique, it will not be explained in detail here.

The land-cover types on topographic maps and remote sensing images are often different, thus we must first produce a map series with unified contents. To be able to make comparisons over time the

maps therefore had to be thematically generalized. Taking into account both the local characteristics and the predominant land use classification system used in China [38], the available land classes were aggregated into seven suitable land categories for this study: arable land, forest land, grassland, water, settlement, wetland and other unused land (Table 1).

Table 1. Detailed classification system for land-use types.

Code	Name of the land categories	Types in remote sensing images	Types in topographic maps
1	arable land	paddy field rainfed land	paddy field rainfed land
2	forest land	closed forest land shrubbery sparse wood land other forest land	deciduous forest, coniferous forest, low pinewood – – orchards, forest of unknown species
3	grassland	high coverage grassland moderate coverage grassland low coverage grassland	grassland wildland
4	water	river lake swag beachland	river lake, swag – –
5	settlement	urban rural settlement other construction	urban rural settlement –
6	wetland	wetland	wetland
7	other unused land	sand saline-alkali land bare land	–

2.4. Methods

2.4.1. CA-Markov Model

The CA-Markov model integrates both the CA and Markov models. Spatiotemporal Markov Chain models can be used to model changes over time among land use categories through the use of transition probabilities while spatial dynamics are controlled by local rules through a cellular automata (CA) mechanism considering either neighborhood configuration or transition probabilities, thus coupling of transition matrices and CA models on change in land-covers or landscape over space and time can provide a better understanding of the historical landscape's imprint on today's plant diversity. This approach is able to convert the quantity results of the Markov Chain by means of a CA function to spatially explicit outcomes [39]. The CA-Markov land use change model can be used for spatial land use simulations and land-cover reconstructions and this approach is capable of simulating several land categories simultaneously.

2.4.2. Three-Map Comparison

The validation of LUCC model requires a three-map comparison where the maps are reference time 1, reference time 2, and simulation time 2. Comparison between the reference map of time 1 and the reference map of time 2 characterizes the observed changes in the maps, which reflect the land use dynamics. Comparison between the reference map of time 1 and the simulated map of time 2 characterizes the model's predicted change, which reflects the behavior of the model. Validation comprehends the agreement (or error) resulting from the comparison between the simulation time 2 map and the reference (or real) time 2 map. In this case, the accuracy of model is assessed by determining the four components of correctness and error, namely null successes (correct due to observed persistence predicted as persistence), hits (correct due to observed change predicted as a change), misses (errors due to observed changes predicted as persistence) and false alarms (errors due to observed persistence predicted as a change), and the summary statistics for error due to quantity (EQ , Equation (4)) and error due to allocation (EA , Equation (5)) [31,34]. The type of error due to quantity reflects the error due to the model's less than perfect prediction of the quantity of net change. This error is not influenced by spatial allocation. The type of error due to allocation is associated with the imperfect ability of the model to allocate change pixels across the landscape. This error derives from the model's spatial allocation and is associated with the independent variables that the model uses. The error might be sensitive to modification of the spatial allocation algorithm. In addition, figure of merit (FOM , Equation (6)) and three ratio indices (Equations (7)–(9)) that quantify the amount of hits, misses and false alarms relative to the observed change were also determined. The figure of merit (FOM) is calculated dividing the hits between the addition of hits, misses and false alarms and, in the case of the models that simulate several categories, removing the partial hits from numerator. This measure allows to assess the cell-to-cell coincidence between simulated and real maps in a more realistic way than more common metrics such as kappa indexes or overall accuracy which are usually calculated using the entire surface area [40]. FOM ranges from 0%, meaning no overlap between observed and predicted change, to 100%, meaning perfect overlap between observed and predicted change [32,41]:

$$OC \equiv \text{Observed Change} = M + H \quad (1)$$

$$PC \equiv \text{Predicted Change} = H + F \quad (2)$$

$$T \equiv \text{Total Error} = M + F \quad (3)$$

$$EQ = |PC - OC| = |(H + F) - (M + H)| = |F - M| \quad (4)$$

$$EA = T - Q = (M + F) - |F - M| = 2 \times \text{Minimum}(F, M) \quad (5)$$

$$FOM = \frac{H}{H + M + F} \times 100 \quad (6)$$

$$HOC = \frac{H}{H + M} \quad (7)$$

$$MOC = \frac{M}{H + M} \quad (8)$$

$$FOC = \frac{F}{H + M} \quad (9)$$

where H , M and F are the hits, misses and false alarms; HOC , MOC and FOC are, respectively, the ratio of hits, misses and false alarms to the observed change, which is the summation of the hits and misses.

Based on the above methods, the flowchart of the technical route of this land use reconstruction is shown in Figure 3.

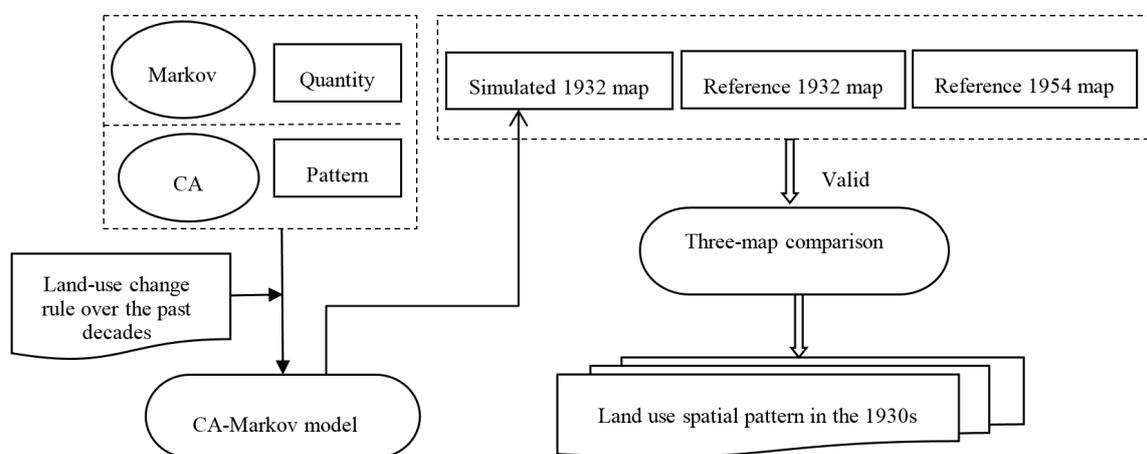


Figure 3. Flowchart of the technical route of land use reconstruction.

3. Results

GIS modeling was undertaken using the program module CA-MARKOV of the software IDRISI Selva (17.02). According to the analysis of land use change over the past 60 years using published land cover data based on topographic and environmental background maps and also remotely sensed images including Landsat MSS and TM pertaining to the years 1954, 1976, 2000, and 2005 [42], we get the following basic assumptions of the model: (1) current spatial pattern of land use is dynamically dependent on a historical pattern, and unchanged land cover during 1954–2005 had actually also existed in the 1930s; (2) reclamation of arable land from 1932 to 1954 was derived from grassland; (3) factors for land suitability do not change over time due to data availability. In all the digital raster maps, each grid cell represents an area $90\text{ m} \times 90\text{ m}$ which is considered to best represent the land-cover in the study area. This precision could avoid losing the data with high resolution in this study although it could result in data redundancy. The parameters of CA-Markov land use change model were as follows:

(1) The input and initial map was set to represent the seven land use types from 1954. Here, some researchers [43] assumed that the hydrological network remained static through the modeled period. In this study, the water body, with clear certain boundaries in the topographic maps in the 1930s, had higher spatial resolution than that from the abovementioned assumption if we consider the water dynamics resulted from climate change, so the input map in this research was masked by the water layer derived from the topographic maps at the scale of 100,000.

(2) Transition matrixes were revised between 1954 and 1976 according to the rules governing LUCC in the study area over the past decades (from 1954 to 2005), to meet the demand between 1932 and 1954. Based on our former study about the LUCC rate and the available data in the study area, monitoring of land use and land cover changes has been done at four time nodes: 1954, 1976, 2000, and 2005 [42]. The spatiotemporal changes during 1954–2005 in Zhenlai County [42] showed that people would prefer

not to utilize the nearby forests and wetlands until all the grasslands that were suitable for farming were completely reclaimed. Hence, we add the grid numbers for wetland converted into arable land from 1954 to 1976 to the grassland.

(3) We used two types of criteria, factors and constraints, where a factor signifies a continuous degree of fuzzy membership (in the range of 0–255), and constraints to limit the alternatives altogether (*i.e.*, fuzzy membership is either 0 or 1) [44]. The factors for suitability maps—soil class, slope, distances from settlements, rivers and roads—were selected for this research (Figure 4A1–A5). Slope raster was generated from 90 m raster DEM data. In addition, we digitized settlements, rivers and roads layers from the 1930s topographic maps, calculated their Euclidean distances and then dispersed the distance value to the range of 0–1, where 0 represents the nearest distance while 1 represents the farthest distance.

Also, in accordance with the principle of spatial autocorrelation [45], it is more likely to develop the same land use type near a certain land cover, thus we also used the distance factor of each land cover based on the land use maps in 1954 (Figure 4B1–B6).

Here, two constraints were considered, water and unchanged land-cover during 1954–2005. Since the other six land cover types cannot typically grow on water bodies, water layers digitized from topographic maps in the 1930s (Figure 4C1) was subject to the same constraints as all the other land use/land cover types, except itself. The unchanged land cover in the study period was relatively stable and certain in spatial distribution over long-term landscape development, and we assume the unchanged land cover over the past 60 years had actually also existed in the 1930s, whereas the present land use pattern was dynamically dependent on the historical one. Thus, one land use type could not grow where other land covers had existed at that time (Figure 4D1–D6).

(4) The neighborhood filter type chooses the “Standard 5×5 contiguity filter”.

(5) The number of iterations was 22, depending on the length of the modeled section and the degree of change described by the transition matrix.

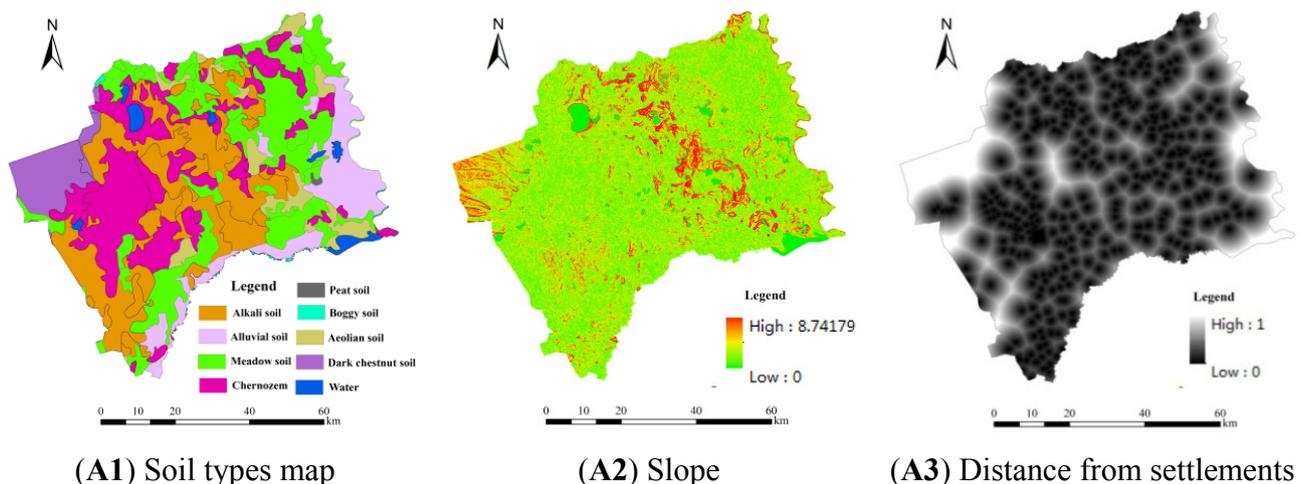


Figure 4. Cont.

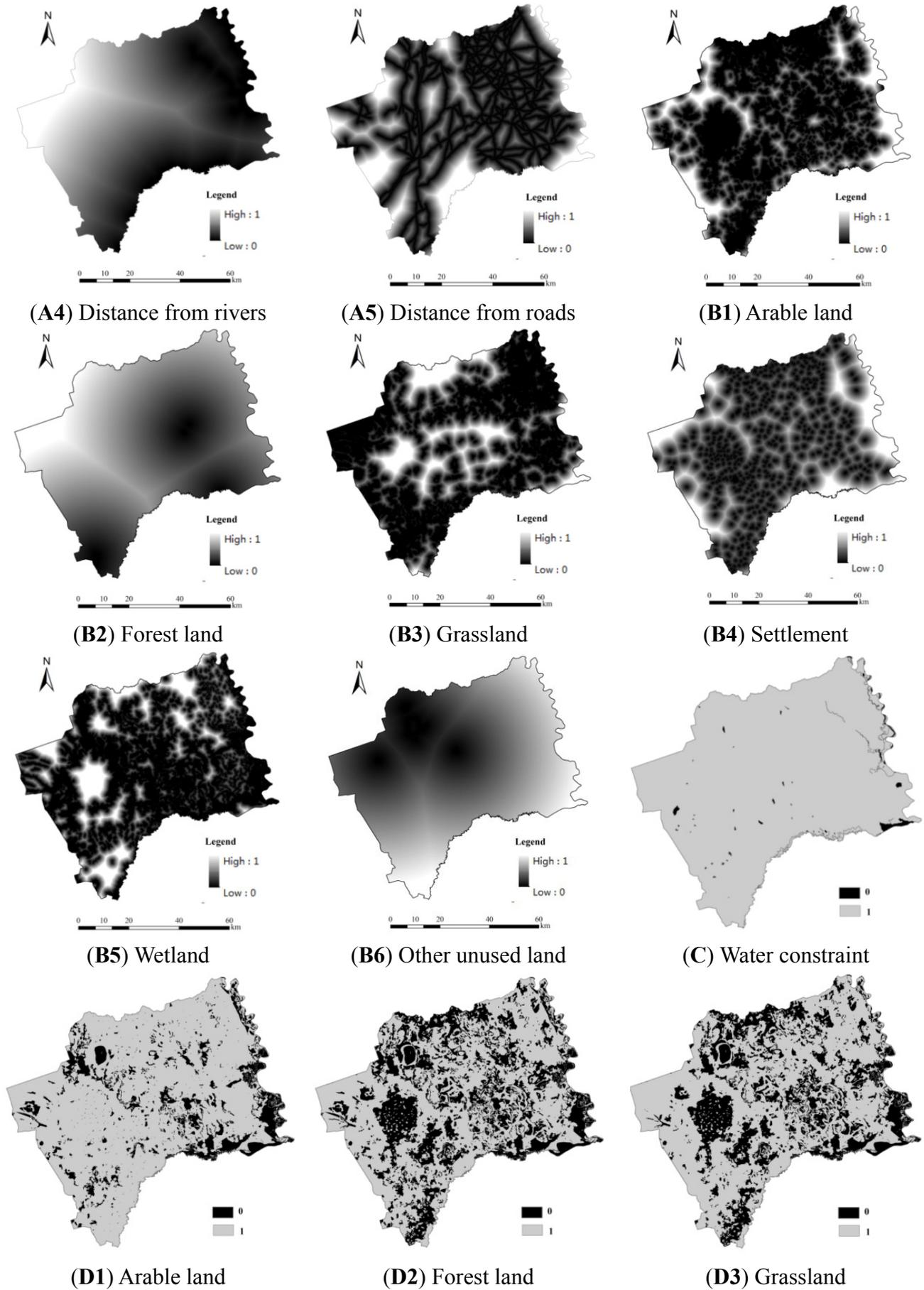


Figure 4. Cont.

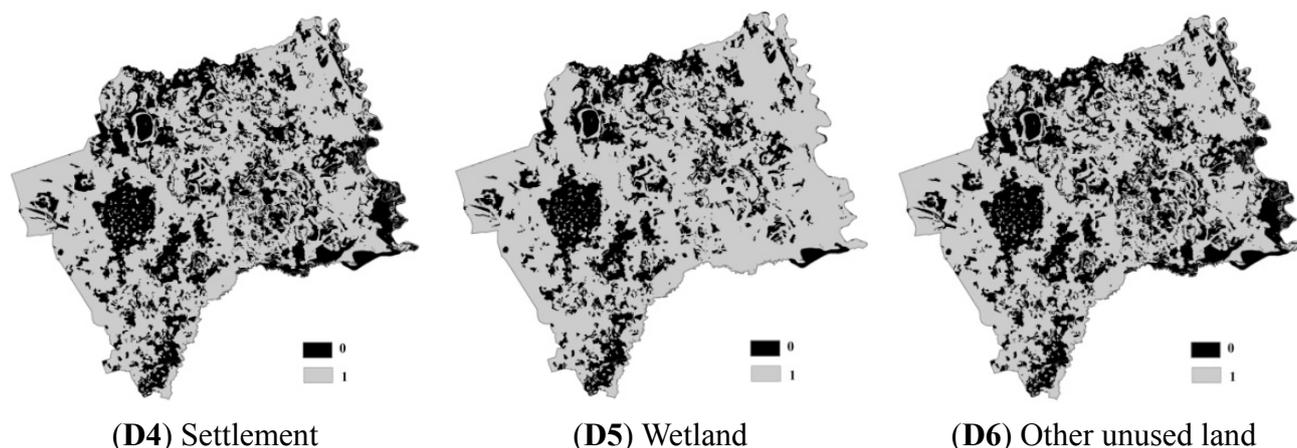


Figure 4. Factors for suitability maps (A1)–(A5); spatial auto correlation distance factors (B1)–(B6); and constraint images for (C) water and (D1)–(D6) unchanged land-covers.

We used the MCE process involving criteria of varying importance in accordance with decision makers, with information about the relative importance of the criteria. This is usually obtained by assigning a weight to each. Here, the weights assigned to different factors (Table 2) were obtained by Saaty’s Analytical Hierarchy Process (AHP). The larger the weight, the more important is the criterion in the overall utility.

Table 2. Factors and their weights used for construction of suitability maps.

Factors		Arable land	Forest	Grassland	Settlement	Wetland	Other unused land
Driving factors	Soil	0.249	0.139	0.051	0.042	0.268	0.103
	Slope	0.036	0.025	0.071	–	0.246	–
	Distance from river	0.096	–	0.099	–	–	–
	Distance from roads	0.092	0.072	–	0.124	–	–
	Distance from settlement	0.226	0.427	–	0.514	–	–
Spatial autocorrelation factors	Arable land	0.301	–	–	–	–	–
	Forest land	–	0.337	–	–	–	–
	Grassland	–	–	0.426	–	0.164	0.124
	Settlement	–	–	–	0.320	–	–
	Wetland	–	–	0.189	–	0.247	0.205
	Other unused land	–	–	0.164	–	0.075	0.568

Figure 5 illustrates the suitability maps for various land categories and Figure 6 shows the result of simulated land-use in the study area in the 1930s. Grassland dominates the largest patches of land cover (210,543.23 ha, 39.61% cover) in the land-use map simulated by the spatial CA-Markov model. This makes sense, as a large number of immigrants came to this area to reclaim wildland or grassland, with anthropogenic activities intensifying after the enactment of “lifting a ban on reclaiming” policy in Northeast China in the late Qing Dynasty. The second largest area is wetland (154,914.97 ha, 29.14% cover). The proportion of arable land accounted as part of the total area is 27.69% (147,177.74 ha). Most of the arable land is located in flat areas—central and western areas and the north in Zhenlai County, while it is distributed in a scattered fashion through other areas. The area of water bodies is

15,798.28 ha, consisting of 2.97% of land cover in the study area. Due to the small areas, it is not so easy to identify settlements (1,672.47 ha, 0.31% cover), other unused land (873.35 ha, 0.16% cover), and forest (626.12 ha, 0.12% cover) in the simulated map.

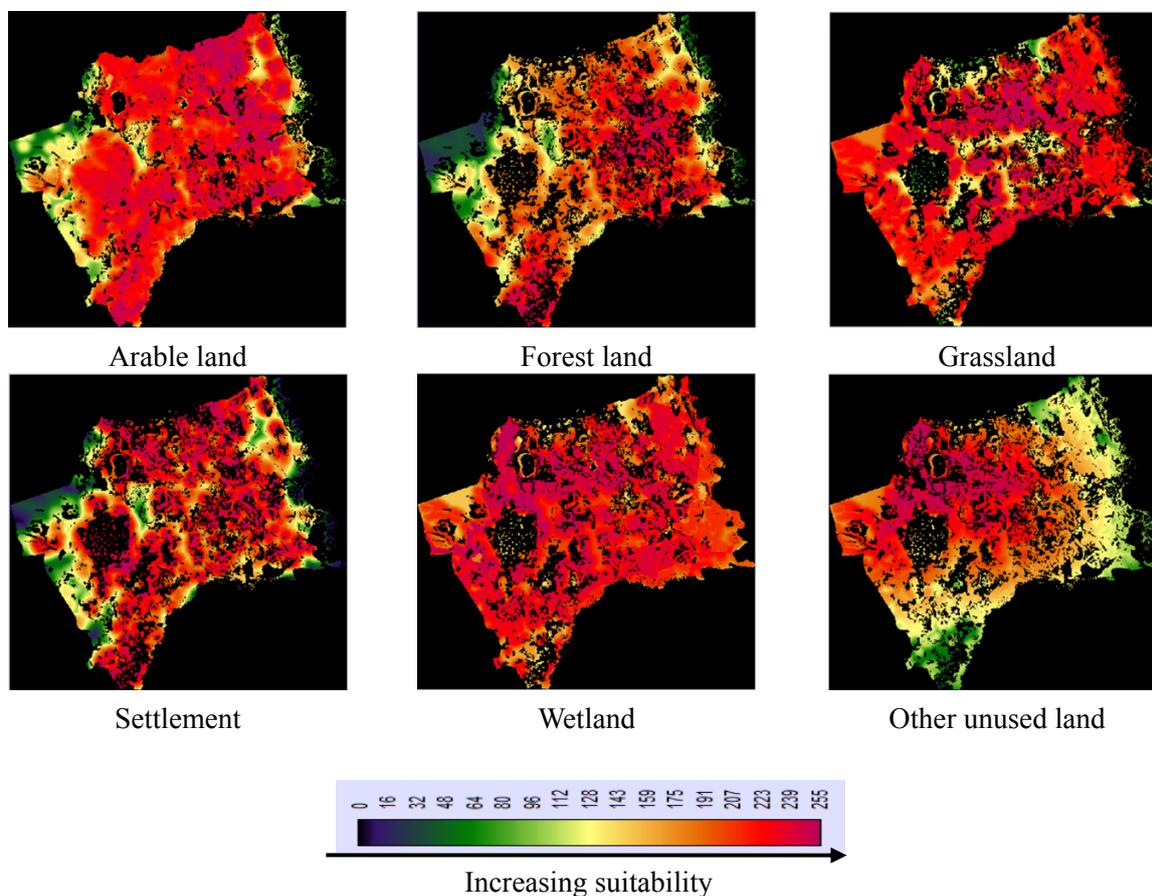


Figure 5. Suitability maps for various land categories.

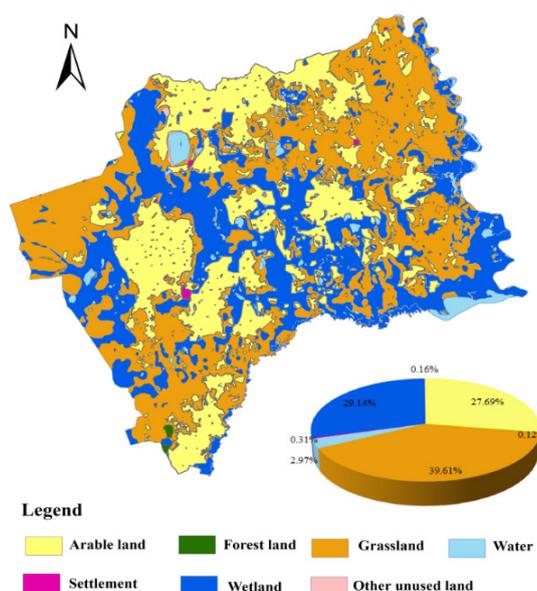


Figure 6. Land-use maps in the 1930s simulated by the spatial CA-Markov model.

4. Discussion

The main purpose of applying CA-Markov model in this study is to produce backward projection and then to reconstruct historical land use and land cover in the 1930s in Zhenlai County. A common approach in testing the reliability of historical reconstructions is to compare them with information from independent sources [46]. Detailed historical maps are very valuable materials for land cover reconstruction because historical land cover can be digitized directly from them. In our study we compared our model results contemporary with the historical topographic map based on the fact that 1:100,000 topographic maps of that time could meet the need of cartography and analyze land use on the regional scale [47]. To validate the model, we used the three-map comparison methodology to identify all possible types of prediction successes and errors based on three maps: the observed 1954 land-use map, the observed 1932 land-use map digitized from topographic maps, and the predicted 1932 land-use map. As there are six land categories in the digitized topographic maps, wetland and other unused land were aggregated into unused land both in the reference 1954 map and simulated 1932 map. Thus, to assess prediction accuracy, six land categories were available: arable land, forest land, grassland, water bodies, settlement, and unused land.

Figure 7A shows the percentage of area covered by each land category at reference 1954, reference 1932 and simulated 1932, respectively. The pie-charts demonstrate that the largest categories in each land use map are grassland, arable land and unused land. The areas of arable land, forest and settlement simulated in the CA-Markov model are similar to those in the digitized topographic maps, especially arable land. These pie-charts give useful information about the quantity of each land category, but they do not offer any details concerning individual transitions between categories. Therefore we overlay the reference 1954 map with the reference 1932 map and the reference 1954 map with the simulated 1932 map to analyze the reference change and the simulated change between 1954 and 1932, respectively. Two matrices were produced presented in Table 3. Each matrix has a total column at the right that gives the stock of each category at the initial time (1954), and a total row at the bottom that gives the stock of each category at the final year (1932).

Figure 7B,C and Table 3 illustrate the gross losses and gross gains, mainly distributed in the middle of the study area and the areas near rivers and lakes. The total change area for the reference change is 306,161.00 ha while it is only 100,101.93 ha for the simulated change. It is especially important to note that arable land and settlements increased from 1932 to 1954 both in the reference change and the simulated change while grassland had a significant loss, indicating that arable land expanded at the expense of grassland due to the population growth. The proportion of water bodies changed slightly and fluctuated from 1.27% in reference 1932 and 2.97% in simulated 1932 to 4.89% in 1954, reflecting a slight increase. Water bodies were mainly influenced by natural factors, especially affected by climate change in the absence of irrigation facilities such as reservoirs. Some research showed that 1936–1959 was a rainy period in the past century in Northeast China, resulting in the increase of water bodies. In addition, unused land surged from 1932 to 1954. A large amount of grassland was converted into other unused land from 1932 to 1954 in both change maps, especially in the reference change, indicating environmental degradation in Zhenlai County during the study period. The spatial reconstruction of the location of settlements over long periods is difficult due to the negligible size compared to other land cover classes and its variation in site selection preferences over time [43]. This research focused on

natural factors and excluded other drivers (e.g., cultural and socioeconomic forces), consequently, the results of reconstruction of settlement areas were unsatisfactory as there are many difference about its gross gains and gross losses between the reference change and the simulated change.

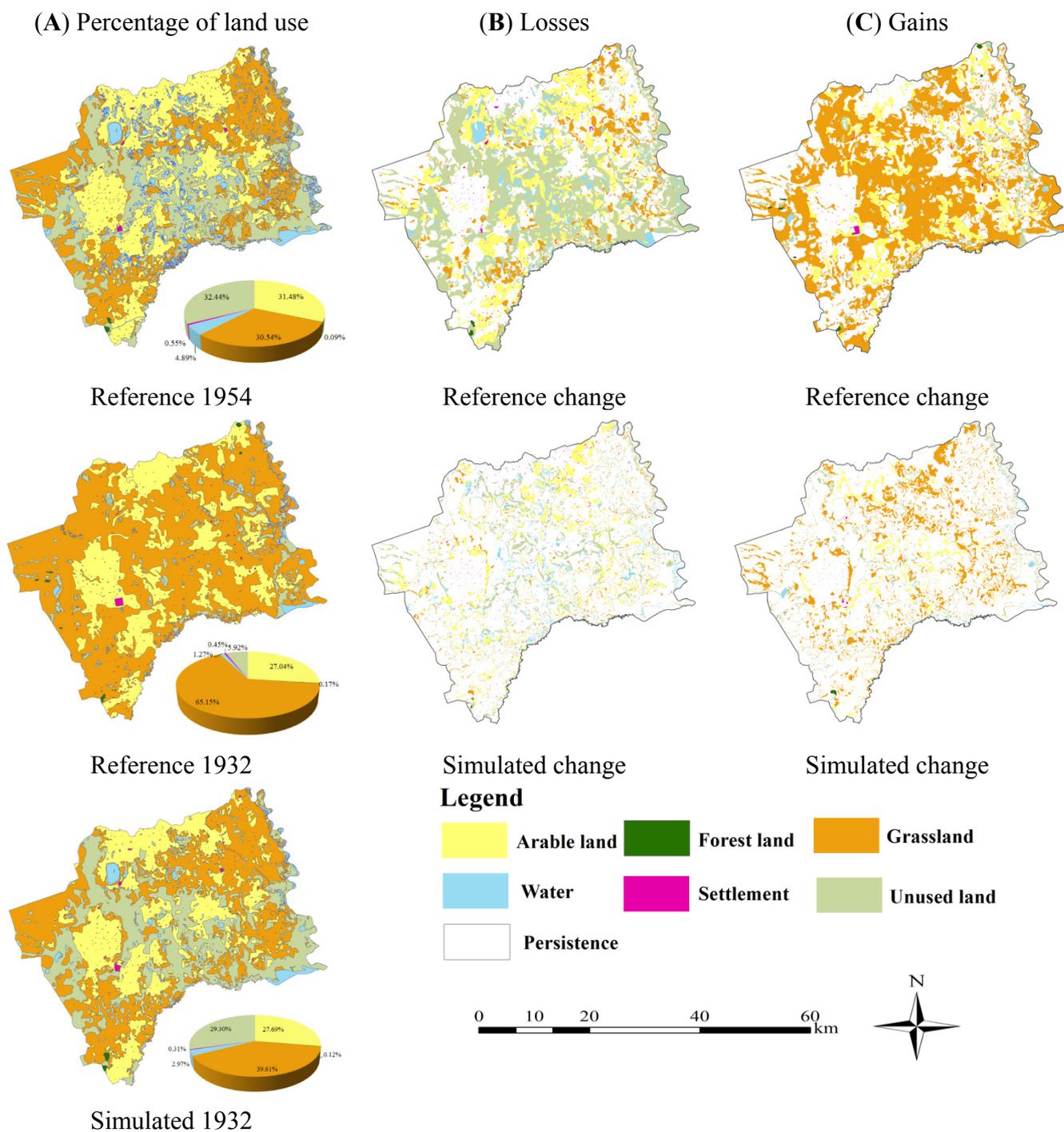


Figure 7. Percentage of land categories at three time points and changes during two time intervals in Zhenlai County.

Table 3. Area counts (ha) of persistence on the main diagonal (underlined) and change off the main diagonal from 1954 to 1932: reference change (in italics) and simulated change (in bold).

		Final year (1932)					Initial	Gross loss	
		Arable land	Forest	Grassland	Water	Settlement	Unused land	total	
	Arable land	<i>85202.00</i>	122.19	76904.66	461.23	1026.98	3638.65	167355.71	82153.71
		129195.45	0.56	34966.03	522.39	509.08	1735.09	166928.60	37733.15
	Forest	<i>143.27</i>	<i>76.74</i>	252.40	0.00	16.24	0.00	488.65	411.91
		50.24	369.53	35.54	0.00	2.51	27.82	485.63	116.11
Initial year (1954)	Grassland	<i>24937.32</i>	<i>534.41</i>	<i>123704.28</i>	<i>1234.24</i>	<i>334.95</i>	<i>11626.66</i>	<i>162371.86</i>	<i>38667.58</i>
		2017.96	238.00	152382.29	1325.70	82.71	7307.26	163353.92	10971.63
	Water	<i>3788.96</i>	<i>0.00</i>	<i>16556.56</i>	<i>2797.54</i>	<i>64.72</i>	<i>2792.23</i>	<i>26000.01</i>	<i>23202.47</i>
		2125.27	0.00	4155.68	11048.21	13.39	8459.47	25802.01	14753.80
	Settlement	<i>1422.64</i>	<i>0.00</i>	<i>1192.72</i>	<i>2.93</i>	<i>275.85</i>	<i>50.44</i>	<i>2944.58</i>	<i>2668.73</i>
		1219.25	3.89	670.57	1.70	969.07	78.82	2943.29	1974.22
	Unused land	<i>28250.58</i>	<i>148.62</i>	<i>127728.90</i>	<i>2263.89</i>	<i>664.62</i>	<i>13388.77</i>	<i>172445.37</i>	<i>159056.60</i>
		12478.79	3.04	19363.19	2612.29	95.72	137539.71	172092.73	34553.02
Initial total		<i>143744.76</i>	<i>881.97</i>	<i>346339.52</i>	<i>6759.83</i>	<i>2383.36</i>	<i>31496.74</i>	<i>531606.17</i>	<i>306161.00</i>
		147086.96	615.01	211573.28	15510.29	1672.48	155148.16	531606.17	100101.93
Gross Gain		<i>58542.77</i>	<i>805.23</i>	<i>222635.24</i>	<i>3962.29</i>	<i>2107.51</i>	<i>18107.97</i>	<i>306161.00</i>	–
		17891.51	245.48	59191.00	4462.08	703.41	17608.45	100101.93	–

Figure 8 and Table 4 show the comparison of the observed change with the predicted change and distinguishes four types of correctness and error. This map was created using the predicted land cover map of 1932 which was overlaid with the reference 1954 to reflect persistence *versus* change. The simulated 1932 land use map had null successes, hits, misses, and false alarms relative to the whole study area of 38.05%, 14.52%, 43.07% and 4.36%, respectively. Observed change (*OC*) occurs on 57.59% of the land use, whereas the predicted change (*PC*) on 18.88% of the land use. Relative to the whole land area, it had a total *EQ* (error due to quantity) of 38.71%, *EA* (error due to allocation) of 8.72% and a total error of 47.43%. This illustrates that there is a minor allocation disagreement and a major quantity disagreement. If a goal were to decrease error due to allocation, then it would be necessary to find additional spatially-explicit with the spatial distribution of the land change processes [31]. If a goal were to decrease error due to quantity, then it would be necessary to use multiple methods to improve the data accuracy in reconstructing quantity of land use and land cover. The total error is smaller than the observed change, so this model is more accurate than its null model, *i.e.*, a model of no change. It had *HOC*, *MOC* and *FOC* ratios of 0.252, 0.748 and 0.076, respectively. Besides, the figure of merit (*FOM*) was 23.44%, higher than in some of the case studies [32,33].

The analysis above illustrated that the major differences among the three maps have less to do with the simulation model and more to do with the inconsistencies among the land categories between 1954 and 1932. The applications that have a large figure of merit are the applications that use the correct or nearly correct net quantities for the categories in the prediction map. The reason causing a large number of hits in this model is that there are different definitions between the reference map 1932 and the simulated map 1932.

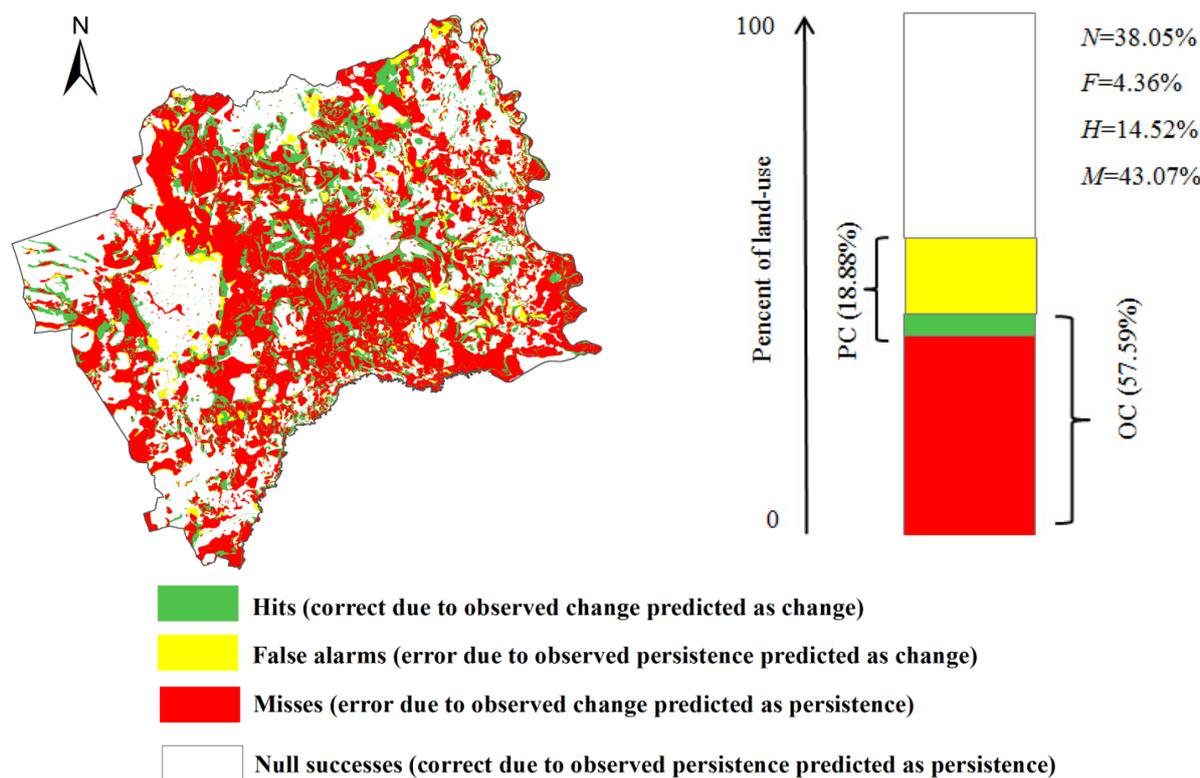


Figure 8. Prediction correctness and error based on 1954 (reference), 1932 (reference) and 1932 (simulated) land-use maps.

Table 4. Validation results of the three-map comparison.

Index	Value (%)	Index	Value (%)	Index	Value (%)	Index	Value
<i>H</i>	14.52	<i>OC</i>	57.59	<i>EQ</i>	38.71	<i>HOC</i>	0.252
<i>M</i>	43.07	<i>PC</i>	18.88	<i>EA</i>	8.72	<i>MOC</i>	0.748
<i>F</i>	4.36	<i>T</i>	47.43	–	–	<i>FOC</i>	0.076
<i>N</i>	38.05	<i>FOM</i>	23.44	–	–	–	–

H–Hits; *M*–Misses; *F*–False alarms; *N*–Null success; *OC*–Observed change; *PC*–Predicted change; *T*–Total change; *FOM*–Figure of merit; *HOC*–Ratio of hits to the observed change; *MOC*–Ratio of misses to the observed change; *FOC*–Ratio of false alarms to the observed change; *EQ*–Error due to quantity; *EA*–Error due to allocation.

It could be that any precise measurement of simulation accuracy is nearly impossible due to the inconsistencies among time points concerning the definitions of land categories in the maps. Most of the topographic maps used in this study were produced from 1932 to 1935 by Japan as a response to military requests while some of the work was performed by the Chinese military during the Manchukuo era. Thus, topographic maps in this study are limited primarily by the fact that they were created for a specific military purpose and by the methods used in elaborating them at the time when they were made, and they focus on certain land cover, such as settlements, water, *etc.*, while some land-use types are not described at all, such as other unused land. Each map shows its own land cover classes mainly based on its purpose and criteria. The features of grassland and unused land drawn in the 1930s’ topographic maps are distinctly different from those derived from remote sensing images, because of their different

intended uses. Grassland in the topographic maps is often mixed with other land covers and its boundary is not easy to be determined. Besides, most of grassland in these maps is often judged as wildland, resulting in limited grassland's symbol in maps. In addition, rainfed lands, the blank areas without any symbol described in maps, are often difficult to digitize from grassland and wildland, so it is very difficult to extract and digitize the spatially explicit grassland data. When applying modern mapping standards, unused land is generally underrepresented on historic maps. Thus, historical maps have a variety of limitations that must be considered to accurately interpret apparent land cover change. We must recognize that the topographic maps and remote sensing images can provide different information. In addition, with the small topographic relief in the study area, climate change, especially change in precipitation, has large impacts on unused land, especially on wetland area. Figure 6, Figure 7A and Figure 8 show that wetland area from the simulated 1932 map is larger than that in the reference 1932 map digitized from the topographic maps, thus a corresponding lesser area of grassland appears in the simulated map. If we consider the large impacts of precipitation on unused land, it will be consistent with the fact that 1900–1935 was a drier period in the past century in Northeast China [48,49]. Thus, the simulated model might be more accurate if we input climate data into the model. Furthermore, it is better for us to combine multi-source data due to the heterogeneous data sources. Blending different data can extend information about environmental change across a broad range of temporal and spatial scales. By combining different data sources, a more complete picture of land use and land cover change can often be gained.

5. Conclusions

The main objective of this work was to reconstruct land use/cover in the 1930s in Northeast China using a CA-Markov model in $90\text{ m} \times 90\text{ m}$ spatial resolution based on the assumption that current spatial patterns of land use are dynamically dependent on a historical pattern and factors for land suitability do not change over time. Then the three-map comparison methodology was used to validate the reconstruction map. We have chosen Zhenlai County in Northeast China as the target for study. The main conclusions are summarized as follows:

- (1) The CA-Markov land cover change model can be simultaneously applicable to spatial reconstructions of various land cover types. The results of historical reconstruction showed that grassland occupied the largest percentage of the study area, followed by wetland and arable land. Other land categories, however, occupied relatively small areas.
- (2) The total change area for the reference change between 1954 and 1932 is 306,161.00 ha while it is only 100,101.93 ha for the simulated change. Gross losses and gross gains were mainly distributed in the middle of the study area and the areas near rivers and lakes. Arable land expanded at the expense of grassland due to the fast population growth during this period. The proportional area of water bodies increased slightly due to the increase of the precipitation. A large amount of grassland was converted into other unused land from 1932 to 1954 in both change maps, especially in the reference change, showing environmental degradation in the study area.
- (3) The figure of merit of the model was 23.44%. The relative error due to allocation was 8.72% while the error due to quantity was 38.71% because of the inconsistencies among time points

concerning the definitions of categories in the maps. The major differences among the three maps have less to do with the simulation model and more to do with the inconsistencies among the land categories during the study period, especially for the grassland and unused land. The grassland in the topographic maps is often mixed with other land covers and its boundaries are not easy to determine. Besides, most of grassland in these maps is often judged as wildland, resulting in difficulty to extract and digitize the spatial explicit grassland data. It is important to choose a reference map with high accuracy in model validation using the three-map comparison methodology, however, it is very difficult for researchers to collect and obtain a suitable reference map in validation of reconstruction model due to the limitation of available historical data.

- (4) Historical topographic maps have a variety of limitations that must be considered to accurately interpret apparent land cover change. Each map shows its own land cover classes mainly based on its purpose and criteria. Different information provided by topographic maps and remote sensing images must be recognized, because of their intended uses. Blending different data can extend information about environmental change across a broad range of temporal and spatial scales. And then by combining multi-source data and information, a more complete picture of land use and land cover change can be obtained.

Acknowledgments

This research was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDA05090310), the National Natural Science Foundation of China (No. 41271416), and the State Scholarship Fund of China (201306170095). The paper was also completed with support from the Center for International Earth Science Information Network (CIESIN), Columbia University, USA. We would like to thank Alex de Sherbinin of CIESIN for his constructive suggestions, and thank Robert S. Chen of CIESIN for his help in the preparation of this paper. Also thanks to Elisabeth Sydor of CIESIN for her editing assistance.

Author Contributions

Yuanyuan Yang conceived and designed the research; Shuwen Zhang and Jiuchun Yang collected and processed the data from the remote sensing images and topographic maps. Xiaoshi Xing and Dongyan Wang provided some useful suggestions and modified the draft. Yuanyuan Yang wrote the paper. All authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Lambin, E.F. Modelling and monitoring land-cover change processes in tropical regions. *Prog. Phys. Geogr.* **1997**, *21*, 375–393.
2. Turner, B.L.; Lambin, E.F.; Reenberg, A. The emergence of land change science for global environmental change and sustainability. *Proc. Nat. Acad. Sci. USA* **2007**, *104*, 20666–20671.

3. Foster, D.R.; Swanson, F.; Aber, J.; Burke, I.; Brokaw, N.; Tilman, D.; Knapp, A. The importance of land-use legacies to ecology and conservation. *Bioscience* **2003**, *53*, 77–88.
4. Gragson, T.L.; Bolstad, P.V. Land use legacies and the future of southern Appalachia. *Soc. Nat. Resour.* **2006**, *19*, 175–190.
5. Lambin, E.F.; Turner, B.L.; Geist, H.J.; Agbola, S.B.; Angelsen, A.; Bruce, J.W.; Coomes, O.T.; Dirzo, R.; Fischer, G.; Folke, C.; *et al.* The causes of land-use and land-cover change: Moving beyond the myths. *Glob. Environ. Change* **2001**, *11*, 261–269.
6. Deng, X.Z.; Zhao, C.H.; Lin, Y.Z.; Zhang, T.; Qu, Y.; Zhang, F.; Wang, Z.; Wu, F. Downscaling the impacts of large-scale LUCC on surface temperature along with IPCC RCPs: A global perspective. *Energies* **2014**, *7*, 2720–2739.
7. Arora, V.K.; Montenegro, A. Small temperature benefits provided by realistic afforestation efforts. *Nat. Geosci.* **2011**, *4*, 514–518.
8. Oldfield, F.; Dearing, J.A.; Gaillard, M.J.; Bugmann, H. Ecosystem processes and human dimensions—The scope and future of HITE (Human Impacts on Terrestrial Ecosystems). *PAGES Newsl.* **2000**, *8*, 21–23.
9. Ramankutty, N.; Foley, J.A. Estimating historical changes in global land cover: Croplands from 1700 to 1992. *Glob. Biogeochem. Cycles* **1999**, *13*, 997–1027.
10. Petit, C.C.; Lambin, E.F. Long-term land-cover changes in the Belgian Ardennes (1775–1929): model-based reconstruction vs. historical maps. *Glob. Change Biol.* **2002**, *8*, 616–630.
11. Goldewijk, K.K.; Ramankutty, N. Land cover change over the last three centuries due to human activities: The availability of new global data sets. *GeoJournal* **2004**, *61*, 335–344.
12. Pongratz, J.; Reick, C.; Raddatz, T.; Claussen, M. A reconstruction of global agricultural areas and land cover for the last millennium. *Glob. Biogeochem. Cycles* **2008**, *22*, 1–16.
13. Hurtt, G.C.; Frohking, S.; Fearon, M.G.; Moore, B.; Shevliakova, E.; Malyshev, S.; Pacala, S.W.; Houghton, R.A. The underpinnings of land-use history: Three centuries of global gridded land-use transitions, wood-harvest activity, and resulting secondary lands. *Glob. Change Biol.* **2006**, *12*, 1208–1229.
14. Liu, M.; Tian, H. China's land cover and land use change from 1700 to 2005: Estimations from high-resolution satellite data and historical archives. *Glob. Biogeochem. Cycles* **2010**, *24*, 1–18.
15. Ye, Y.; Fang, X.Q. Land use change in Northeast China in the twentieth century: A note on sources, methods and patterns. *J. Hist. Geogr.* **2009**, *35*, 311–329.
16. Ye, Y.; Fang, X.Q. Spatial pattern of land cover changes across Northeast China over the past 300 years. *J. Hist. Geogr.* **2011**, *3*, 408–417.
17. Pontius, R.G., Jr.; Cornell, J.D.; Hall, C.A.S. Modeling the spatial pattern of land-use change with GEOMOD2: Application and validation for Costa Rica. *Agric. Ecosyst. Environ.* **2001**, *85*, 191–203.
18. Bolliger, J.; Schulte, L.A.; Burrows, S.N.; Sickley, T.A.; Mladenoff, D.J. Assessing ecological restoration potentials of Wisconsin (USA) using historical landscape reconstructions. *Restor. Ecol.* **2004**, *12*, 124–142.
19. Wulf, M.; Sommer, M.; Schmidt, R. Forest cover changes in the Prignitz region (NE Germany) between 1790 and 1960 in relation to soils and other driving forces. *Landsc. Ecol.* **2010**, *25*, 299–313.

20. Skaloš, J.; Weber, M.; Lipský, Z.; Trpáková, I.; Šantrůčková, M.; Uhlířová, L.; Kukla, P. Using old military survey maps and orthophotograph maps to analyse long-term land cover changes—Case study (Czech Republic). *Appl. Geogr.* **2011**, *31*, 426–438.
21. Bender, O.; Boehmerb, H.J.; Jens, D.; Schumacher, K.P. Using GIS to analyse long-term cultural landscape change in Southern Germany. *Landsc. Urban Plan.* **2005**, *70*, 111–125.
22. Boucher, Y.; Arseneault, D.; Sirois, L.; Blais, L. Logging pattern and landscape changes over the last century at the boreal and deciduous forest transition in Eastern Canada. *Landsc. Ecol.* **2009**, *24*, 171–184.
23. Hamre, L.N.; Domaas, S.T.; Austad, I.; Rydgren, K. Land-cover and structural changes in a western Norwegian cultural landscape since 1865, based on an old cadastral map and a field survey. *Landsc. Ecol.* **2007**, *22*, 1563–1574.
24. He, F.N.; Li, S.C.; Zhang, X.Z. The reconstruction of cropland area and its spatial distribution pattern in the mid-northern Song dynasty. *Acta Geogr. Sin.* **2011**, *66*, 1531–1539. (In Chinese)
25. Kumar, S.; Merwade, V.; Rao, P.S.C.; Pijanowski, B.C. Characterizing long-term land use/cover change in the United States from 1850 to 2000 using a nonlinear bi-analytical model. *AMBIO* **2013**, *42*, 285–297.
26. Yang, Y.Y.; Zhang, S.W.; Yang, J.C.; Chang, L.P.; Bu, K.; Xing, X.S. A review of historical reconstruction methods of land use/land cover. *J. Geogr. Sci.* **2014**, *24*, 746–766.
27. Fuchs, R.; Herold, M.; Verburg, P.H.; Clevers, J.G.P.W. A high-resolution and harmonized model approach for reconstructing and analyzing historic land changes in Europe. *Biogeosc. Discuss.* **2012**, *9*, 14823–14866.
28. Wang, S.Q.; Zheng, X.Q.; Zang, X.B. Accuracy assessments of land use change simulation based on Markov-cellular automata model. *Procedia Environ. Sci.* **2012**, *13*, 1238–1245.
29. Pontius, G.R., Jr.; Malanson, J. Comparison of the structure and accuracy of two land change models. *Int. J. Geogr. Inf. Sci.* **2005**, *19*, 243–265.
30. Pontius, R.G., Jr.; Millones, M. Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *Int. J. Remote Sens.* **2011**, *32*, 4407–4429.
31. Chen, H.; Pontius, R.G., Jr. Diagnostic tools to evaluate a spatial land change projection along a gradient of an explanatory variable. *Landsc. Ecol.* **2010**, *25*, 1319–1331.
32. Pontius, R.G., Jr.; Boersma, W.; Castella, J.C.; Clarke, K.; de Nijs, T.; Dietzel, C.; Duan, Z.Q.; Fotsing, E.; Goldstein, N.; Kok, K.; *et al.* Comparing the input, output, and validation maps for several models of land change. *Ann. Reg. Sci.* **2008**, *42*, 11–37.
33. García, A.M.; Santé, I.; Boullón, M.; Crecente, R. A comparative analysis of cellular automata models for simulation of small urban areas in Galicia, NW Spain. *Comput. Environ. Urban Syst.* **2012**, *36*, 291–301.
34. Chen, H.; Pontius, R.G., Jr. Sensitivity of a land change model to pixel resolution and precision of the independent variable. *Environ. Model. Assess.* **2011**, *16*, 37–52.
35. Zhang, Y.J.; Gao, Z.J.; Yu, H. *Local Record of Zhenlai County*; Jilin People's Press: Changchun, China, 1995. (In Chinese)
36. Bai, S.Y.; Zhang, S.W. The discussion of the method of land utilization spatial information reappearance of history period. *J. Arid Land Resour. Environ.* **2004**, *18*, 77–80. (In Chinese)

37. Bai, S.Y.; Zhang, S.W.; Zhang, Y.Z. Study on the method of diagnose the plowland spacial distribution in historical era. *Syst. Sci. Compr. Stud. Agric.* **2005**, *21*, 252–255. (In Chinese)
38. Liu, J.Y.; Liu, M.L.; Zhuang, D.F.; Zhang, Z.X.; Deng, X.Z. Study on spatial pattern of land-use change in China during 1995–2000. *Sci. China* **2003**, *46*, 374–384.
39. Pontius, R.G., Jr.; Shusas, E.; McEachern, M. Detecting important categorical land changes while accounting for persistence. *Agric. Ecosyst. Environ.* **2004**, *101*, 251–268.
40. Santé, I.; García, A.M.; Miranda, D.; Crecente, R. Cellular automata models for the simulation of real-world urban processes: A review and analysis. *Landsc. Urban Plan.* **2010**, *96*, 108–122.
41. Pontius, R.; Peethambaram, S.; Castella, J.-C. Comparison of three maps at multiple resolutions: A case study of land change simulation in Cho Don District, Vietnam. *Ann. Assoc. Am. Geogr.* **2011**, *101*, 45–62.
42. Yang, Y.Y.; Zhang, S.W.; Wang, D.Y.; Yang, J.C.; Xing, X.X. Spatiotemporal changes of farming-pastoral ecotone in Northern China, 1954–2005: A case study in Zhenlai County, Jilin Province. *Sustainability* **2014**, *7*, 1–22.
43. Poska, A.; Sepp, E.; Veski, S.; Koppel, K. Using quantitative pollen-based land-cover estimations and a spatial CA_Markov model to reconstruct the development of cultural landscape at Rõuge, South Estonia. *Veget. Hist. Archaeobot.* **2008**, *17*, 527–541.
44. Samat, N.; Hasni, R.; Elhadary, Y.A.E. Modelling land use changes at the peri-urban areas using geographic information systems and cellular automata model. *J. Sustain. Dev.* **2011**, *4*, 72–84.
45. Lin, N.F.; Tang, J.; Wang, X.G.; Bian, J.M.; Li, Z.Y.; Wang, C.Y. Constructions of a geo-information atlas-spatial analysis model toward a perspective of the spatial pattern changes in land use. *Quat. Sci.* **2009**, *29*, 711–723. (In Chinese)
46. Gimmi, U.; Bürgi, M.; Stuber, M. Reconstructing anthropogenic disturbance regimes in forest ecosystems: A case study from the Swiss Rhone valley. *Ecosystems* **2008**, *11*, 113–124.
47. Wang, Z.M.; Zhang, B.; Song, K.S.; Liu, D.W. Extracting land use information based on topographical map and knowledge rules. *Geo-inf. Sci.* **2008**, *10*, 67–73. (In Chinese)
48. Gao, P.; Mu, X.M.; Wang, F.; Wang, S.Y. Trend analysis of precipitation in Northeast China in recent 100 years. *J. China Hydrol.* **2010**, *30*, 80–84. (In Chinese)
49. Sun, F.H.; Yuan, J.; Lu, S. The change and test of climate in Northeast China over the last 100 years. *Clim. Environ. Res.* **2006**, *11*, 101–108. (In Chinese).