Chin. Geogra. Sci. 2015 Vol. 25 No. 5 pp. 629–643
doi: 10.1007/s11769-015-0753-2

Springer Science Press
www.springerlink.com/content/1002-0063

Reconstructing Spatial Distribution of Historical Cropland in China′s Traditional Cultivated Region: Methods and Case Study

YANG Xuhong[1], GUO Beibei[1], JIN Xiaobin[1], LONG Ying[1, 2], ZHOU Yinkang[1]

(1. *College of Geographic and Oceanographic Sciences*, *Nanjing University*, *Nanjing* 210023, *China*; 2. *Beijing Institute of City Planning*, *Beijing* 100045, *China*)

Abstract: As an important part of land use/cover change (LUCC), historical LUCC in long time series attracts much more attention from scholars. Currently, based on the view of combining the overall control of cropland area and ′top-down′ decision-making behaviors, here are two global historical land-use datasets, generally referred as the Sustainability and the Global Environment datasets (SAGE datasets) and History Database of the Global Environment datasets (HYDE datasets). However, at the regional level, these global datasets have coarse resolutions and inevitable errors. Considering various factors that influenced cropland distribution, including cropland connectivity and the limitation of natural and human factors, this study developed a reconstruction model of historical cropland based on constrained Cellular Automaton (CA) of ′bottom-up′. Then, an available labor force index is used as a proxy for the amount of cropland to inspect and calibrate these spatial patterns. Applied the reconstruction model to Shandong Province, we reconstructed its spatial distribution of cropland during 8 periods. The reconstructed results show that: 1) it is properly suitable for constrained CA to simulate and reconstruct the spatial distribution of cropland in traditional cultivated region of China; 2) compared with ′SAGE datasets′ and ′HYDE datasets′, this study have formed higher-resolution Boolean spatial distribution datasets of historical cropland with a more definitive concept of spatial pattern in terms of fractional format.

Keywords: traditional cultivated region; historical cropland; reconstruction; constrained Cellular Automaton (CA); Shandong Province

Citation: Yang Xuhong, Guo Beibei, Jin Xiaobin, Long Ying, Zhou Yinkang, 2015. Reconstructing spatial distribution of historical cropland in China′s traditional cultivated region: methods and case study. *Chinese Geographical Science*, 25(5): 629–643. doi: 10.1007/s11769-015-0753-2

1 Introduction

A large amount of research has proven that the land use/cover change (LUCC) resulting from human activities can bring significant effects on regional and even global climatic/ecological changes (Goldewijk and Navin, 2004; Ye and Fu, 2004; Foley *et al*., 2005). With the development and improvement of technologies and productivity, such effect has become much more conspicuous since the industrial revolution in the western world (Cao *et al*., 2013). In history, particularly in the past 300 years, LUCC has been an important parameter used to conduct long-time-scale climate simulation study, diagnose climate forming mechanisms, identify the sensitivity of the climatic system to the nature and human forces, and predict future climate change and so on. As the global change study continuingly intensified, LUCC draws extensive attention from the international academic world (Feddema *et al*., 2005; Voldoire *et al*., 2007; Shi *et al*., 2007). In the 1990s, a LUCC research project jointly initiated by the International Geosphere-Biosphere Programme (IGBP) and the Interna-

Received date: 2014-01-17; accepted date: 2014-05-12
Foundation item: Under the auspices of National Basic Research Program of China (No. 2011CB952001), National Natural Science Foundation of China (No. 41340016, 412013860)
Corresponding author: JIN Xiaobin. E-mail: jinxb@nju.edu.cn

tional Human Dimensions Programme (IHDP) on Global Environmental Change stressed that the history of land use changes in the past must be reconstructed by all necessary means, which triggered a wave of research into the land cover changes in history. Reconstruction of the data on land cover in history, particularly the high-precision spatial data, has drawn extensive attention from scholars (Ramankutty and Foley, 1999; Goldewijk, 2001; Zhu *et al*., 2012; Tian *et al*., 2014).

Representative achievements of the researches intended for the reconstruction of the spatial pattern of land use include the Sustainability and the Global Environment (SAGE dataset) and the Historical Database of the Global Environment (HYDE dataset) established by Ramankutty and Foley (1999), Goldewijk (2001) and Goldewijk *et al*. (2011), respectively. SAGE dataset, based on the modern pattern of global land use, reconstructed the distribution of global cropland in the period from AD 1700 to AD 1992 with a spatial resolution of 0.5° × 0.5°. HYDE dataset has so far been issued in four versions, where the latest version (HYDE 3.1) takes into account the factors such as population, terrain slope, distance to rivers, urban areas distribution pattern, forest land and potential vegetation, and employs a more complex algorithm to simulate the history of changes experienced by the global cropland and grassland in a continuous period since contemporary, with a higher spatial resolution of 5′ × 5′ and over a greater time span (past 12 000 years). Some scholars have argued that these two datasets seem to be very rough when used in regional simulation researches, in terms of either cropland amount estimation or spatial distribution pattern, and that the data acquired can only be applied at the global scale and can not be used as the basis for regional researches (Li *et al*., 2010). Nonetheless, the above-mentioned studies have provided certain reference data and usable approaches for the related studies, and many scholars can use or modify such approaches to make more intensive data reconstruction researches. For example, Pongratz *et al*. (2008), on the basis of the above-mentioned datasets, used the land use pattern in AD 1700 as the reference to reconstruct the distribution pattern of the global cropland and grassland in the period from AD 800 to AD 1700 with a spatial resolution of 0.5° × 0.5°; Ray and Pijanowski (2010) used artificial neural network, GIS technologies and the step-by-step land use conversion method to reconstruct the conversion between construction land, cropland and forest land along Muskegon River in Michigan. As for Chinese researches, Liu and Tian (2010) used the modern land use pattern, the statistic data of higher resolution and more credible precision to reconstruct the spatial pattern of cropland, forest land and construction land of the entire country in the continuous period from AD 1700 to AD 2005 with a spatial resolution of 10 km × 10 km. Due to the restrictions of data, most of the researches are focused on certain regions. For example, Lin *et al*. (2008) built the agricultural population gravity and land topography gravity model, where population and topographic slope are the factors that affect the cropland distribution pattern in history, and gridded the spatial distribution pattern of cropland in six historical period sections of the traditional agricultural area of China. Li *et al*. (2011) used the MODIS land cover remote sensing data to introduce the historical cropland space gridding method and reconstruct the dataset of the cropland in Yunnan in 1671 and 1827. He *et al*. (2012a) employed the gridding method invented by Lin *et al*. (2008) to rectify the amount on population and land, and then reconstruct the spatial distribution pattern of cropland during the mid-Northern Song Dynasty (AD1004–1085).

From the perspective of methodology, the above-mentioned studies mostly employ a ′top-to-bottom′ method that proceeds from quantity reconstruction to spatial pattern reconstruction. Cropland amount in a historical period is obtained from historical archives, the factors that affect spatial distribution of cropland are screened and quantified, land arability is determined, and then the historical amounts of cropland are spatially matched to areas based on levels of arability. However, the principle of continuous distribution of cropland suggests that a piece of land surrounded by cropland is more likely to be cultivated, which is consistent with the concept of Cellular Automata (CA) modeling. Based on the modeling concept of the constrained CA (Wu, 1998; Li and Ye, 1999; Yeh and Li, 2001), we construct a model that treats the natural environment and cultural environment in the region as constraint conditions. This article also takes into account the effects of factors such as cell states and neighborhood to simulate the historical spatial changes in cropland.

The traditional cultivated region refers to the district mostly located in the second- and third-level terrain

ladders in the terrain pattern of China which is to the east of the Hu Huan-yong population line (from Aihui County, Mohe County today, in Heilongjiang Province to Tengchong County in Yunnan Province) and to the south of Yanshan Mountain of Hebei Province. In terms of administration organization system, this region refers to the 17 inland provinces that existed as administrative divisions during the 25th year of the reign of Emperor Jiaqing in the Qing Dynasty (AD 1820). This region was the primary socio-economic area of China during the period and had a relatively high concentration of cropland. Over 300 years, despite the dynastic transitions and warfare during which farming had to be discontinued, the peripheral boundaries of cropland did not change significantly. Spatial changes of the cropland were due to temporary discontinued farming and restorative reclamation. This region is the key area in the research on reconstruction of historical spatial patterns of cropland in China (Lin *et al*., 2008; He *et al*., 2012b). This article is to use Shandong Province which is located within the traditional agricultural area as the study area. We use the modern cropland spatial pattern, the historical amount of cropland, land reclamation suitability, and historical population within a constrained CA model to reconstruct the 1 km × 1 km spatial pattern of cropland.

2 Methodology

Reconstruction of historical cropland dataset consists of quantity and spatial pattern reconstruction (Li *et al*., 2011; Zhu *et al*., 2012). The former refers to the statistical information on historical amount of cropland obtained by reading historical archives, and the latter refers to the process where the historical spatial distribution pattern of the cropland is restored on the basis of certain principles, assumptions, or approaches for spatial distribution (Ye *et al*., 2009; Cao *et al*., 2013).

As for expansion of cropland in China, all Chinese studies have indicated that the amount of cropland in China during the period from the Early Qing Dynasty to 1980 increased in an oscillating pattern and reached its peak in 1978 to 1980 (Zhou, 2001; Ge *et al*., 2003; Zhang *et al*., 2003; Feng *et al*., 2005). Afterwards, the amount of cropland decreased considerably, thus the cropland boundary in 1980 can be viewed as the peripheral boundary of its historical distribution range.

Reclamation of cropland is highly dependent on natural conditions and usually proceeds first from land parcels that are easy to reclaim and then to land blocks that are difficult to reclaim. The land units with preferable natural conditions for farming are reclaimed earlier than the ones with unfavorable conditions (He *et al*., 2003). Only when the population pressure increases or a major disaster happens can the land units with high terrain, steep slopes, and/or low fertility be treated as reclaimable land (Li *et al*., 2011). Furthermore, the land units surrounding a cropland are preferable for reclamation given considerations of return on investment and farming convenience.

The reconstruction of the historical pattern of cropland in China′s traditional agricultural region was done using the land pattern of 1980 as the maximum potential distribution range for the cropland in the past. The model projects past cropland via backward simulation starting with 1980. Based on the land reclamation suitability in a region, distribution of modern cropland, and total cropland and land boundaries in certain period, the cropland transition rules were established and a constrained CA modeling method was used to reconstruct the spatial pattern of cropland. A flow chart of the process is shown in Fig. 1.

2.1 Reconstruction model of historical spatial pattern of cropland

2.1.1 Assumptions and conception of model

The constrained CA model used in this model has some basic assumptions that are similar to models used in the literature (Ramankutty and Foley, 1999; Bai *et al*., 2007; Pongratz *et al*., 2008; Lin *et al*., 2008; Li *et al*., 2011; He *et al*., 2012a). They are as follows: 1) there are similarities between the historical cropland spatial pattern and the contemporary spatial pattern; 2) the most unsuitable farming cells were turned into non-cropland first (from contemporary to history); 3) cells surrounded by high ratio of non-cropland were turned into non-cropland first; 4) the range of historical cropland does not exceed the scope of contemporary cropland, in that it is impossible for an cropland in history not be cultivated in modern times; 5) factors influencing land reclamation suitability do not change over time (due to data availability).

Identification of the factors that affect spatial distribution patterns of cropland is the basis for building the

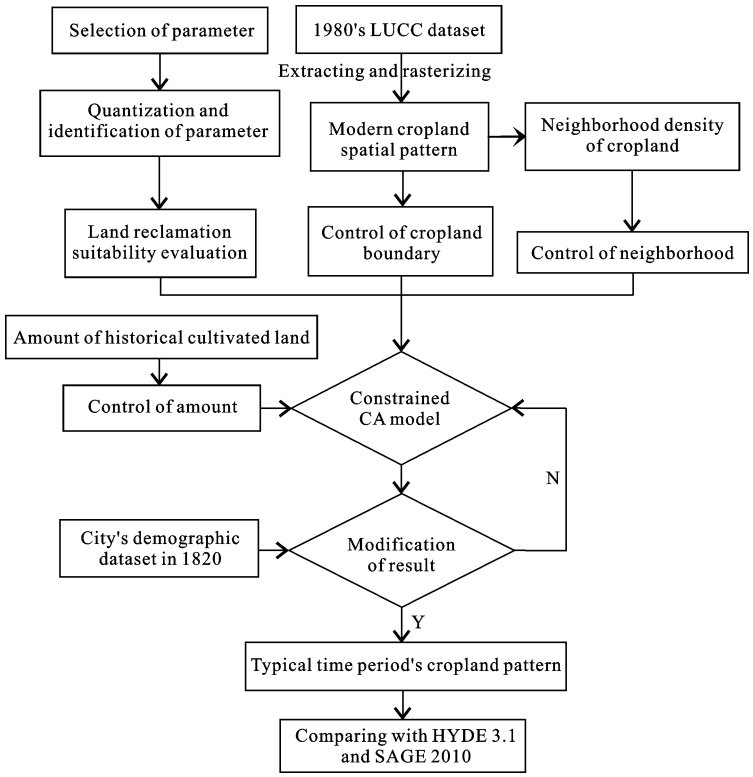

Fig. 1 Flow chart of reconstructing historical cropland spatial pattern based on constrained CA

model. We make use of the results from prior efforts to achieve this aim (Liu *et al*., 1995; Jiao and Liu, 2004; Lin *et al*., 2008; Li *et al*., 2011; Li *et al*., 2012; He *et al*., 2012a; Xie *et al*., 2013). This research concludes two types of factors that affect the spatial distribution of cropland: natural factors (terrain elevation, topographic slope, river system, soil quality, *etc*.) and human factors (population, economic level, agricultural policies, wars, *etc*.). In addition, the pattern of cropland can also be affected by accidental factors or random events. In order to make the simulation result more realistic, a random factor needs to be introduced into the constrained CA model to address uncertainty in the evolution of cropland.

Based on the above-mentioned information and the constituents of the CA model (Wu, 1998; Li and Ye, 1999; Yeh and Li, 2001), the constrained CA model used for reconstruction of the spatial pattern of cropland in history can be abstracted as this equation:

$$State_i^{t-1} = f\begin{pmatrix} Natural_i^t\,(Slope_i, Elevation_i \ldots, River_i, Erosion_i, Quality_i), \\ Human_i(Population_i, Policy_i \ldots, War_i), \\ Neighbor_i^t, State_i^t, Rand_i^t \end{pmatrix} \quad (1)$$

where the $State^{t-1}{}_i$ is the cell states in the grid unit i at the time t–1, which is synthetically determined by natural factors (including $Slope_i$, $Elevation_i$, $River_i$, $Erosion_i$, and $Quality_i$ *etc.*, which represent respectively the variables value of slope, elevation, river, soil erosion and soil quality in the grid unit i at the time t, the same below), Human factors (including $Population_i$, $Policy_i$, and War_i *etc*., which represent respectively the variables value of population, policy, and war in the grid unit i at the time t), surrounding neighborhood ($Neighbor_i^t$, which is the neighborhood-based variable value of grid unit i at the time t), cell status ($State_i^t$, which is the cell states in the grid unit i at the time t), and random disturbance factors ($Rand_i^t$, which represent the random variable value of grid unit i at the time t). Here, the backward deduction is based on the current period.

Natural factors are region-specific and relatively stable features, while some human factors are highly variable and difficult to quantify and spatialize. The research scale and the available data were also con-

strained. Therefore, this article uses slope, elevation, river, soil erosion, and soil quality as the dominant factors that determine reclamation suitability of land and use population as the correction factor for spatial distribution of historical cropland. Equation (1) is simplifies to be Equation (2).

$$State_i^{t-1} = f(Q^t{}_i, Neighbor_i^t, State_i^t, Rand_i^t) \quad (2)$$

where the $State_i^{t-1}$ of grid unit i at the time t–1 is synthetically determined by $Q^t{}_i$, neighborhood-based variable ($Neighbor_i^t$), cell status ($State_i^t$) and random disturbance factor ($Rand_i^t$). $Q^t{}_i$ represents the contribution probability of land reclamation suitability of grid unit i at the time t, which will be described in detail below.

According to properties of CA model, the basic elements of this research′s constrained CA model are as follows:

(1) Lattices: the entire study area;

(2) Cells: cell size was 1 km × 1 km, spatial cells are rasterized by maximum area method in GIS;

(3) Cell states: *cell* = 1 means cropland, *cell* = 0 means non-cropland;

(4) Transition rules: it will be specifically addressed in the next section, a multi-criteria evaluation (MCE) form;

(5) Neighborhoods: Moore neighborhood, 3 cells × 3 cells, a total of eight neighboring cells;

(6) Terminate the iteration: the amount cell of cropland equals to the amount of cropland in typical year (exogenous variable)

(7) Constraints: the five aforementioned spatial factors: slope, elevation, river, soil erosion and soil quality.

2.1.2 Decision criterion

Spatial criterion of reclamation suitability variable: in agricultural society, people always prefer to reclaim flat and fertile land. Only when population pressure and people′s demands for agricultural products keep increasing, will people gradually reclaim high, steep and less fertile land? Therefore, in reconstructing spatial pattern of historical cropland, it is possible to reduce cropland by excluding cells which are less suitable for reclamation.

Land reclamation suitability of grid unit i at the time t (S_i^t) is calculated by multi-factor synthesis method (Wu, 1998), and parameter values of a and γ_1–γ_5 are obtained by Logistic regression analysis, where $Q^t{}_i$ is the contribution probability of land reclamation suitability ($S^t{}_i$), and a, γ_1–γ_5 are the regression parameters respectively for constant, *Slope*, *Elevation*, *River*, *Erosion* and *Quality*.

$$S_i^t = a + \gamma_1 \times Slope + \gamma_2 \times Elevation + \gamma_3 \times River + \gamma_4 \times Erosion + \gamma_5 \times Quality \quad (3)$$

$$Q_i^t = \frac{1}{1 + e^{-s_i^t}} \quad (4)$$

Spatial criterion of the neighborhood-based variable: in terms of return on investment, continuous intensive cultivation, and cultivation convenience, it is more likely that land blocks around cropland will be reclaimed. Thus, isolated cropland with low reclamation suitability within the modern pattern is more likely to have been reclaimed later than less isolated land with higher reclamation suitability. We believe that such cells should be removed first in the process of reconstructing cropland. Following the methods of relevant studies (Long *et al*., 2009; 2010), the neighborhood-based variable ($Neighbor_i^t$) adopts the ratio of the number of cells with the property of cropland in the Moore neighborhood to the total number of neighboring cells 8 at time-step t. See the following equation:

$$Neighbor_i^t = \frac{\sum_{i=1}^{8} cell_i^t}{3 \times 3 - 1} \quad (5)$$

where $cel\,l_i^t$ of grid unit i at the time t is the cell states in cropland dataset.

Compound decision criterion for reconstructing spatial pattern of historical cropland: spatial evolution of historical cropland is the result of a compound decision of land reclamation suitability, cell neighborhood, and a random disturbance factor. From the spatial criteria of the above factors and reference to previous studies (Liu *et al*., 2006; Zhang *et al*., 2008), we obtained the compound decision criterion for spatial evolution of historical cropland:

$$P_i^t = \exp\left[\omega\left(\frac{R_i^t}{R_{g\max}^t} - 1\right)\right] \quad (6)$$

$$State_i^{t-1} = \begin{cases} 0 & if \;\; P_i^t < Threshold^t \\ 1 & Else \end{cases} \quad (7)$$

where P_i^t is the final probability for the unit cell i to be reduced to non-cropland at t. The smaller the value is, the higher the probability that unit cell i may be reduced

to non-cropland. R^t_{gmax} is the maximum value in potential reclamation cell set composed of unit cells which are involved in the operation at t; ω is discrete parameter, with a value range of 1–10. $State_i^{t-1}$ represents the state at t–1 of unit cell i involved in the operation at t, i.e., whether it is to be maintained as cropland. $Threshold^t$ is the threshold of state transition. As the reduction rule of cropland is nonlinear and fluctuant, this value is gradually increased at every time of loop iteration (Equation (8)). R_i^t represents the potential reclamation probability of cell i under the combined action of contribution probability of land reclamation suitability and cell neighborhood (Equation (9)).

$$Threshold^t = Threshold^{t+1} + \theta \tag{8}$$

$$R_i^t = (\alpha \times Q_i^t + \beta \times Neighbor_i^t) \times \left[1 + (Rand_i^t - 0.5)/\varepsilon\right] \tag{9}$$

where $Threshold^{t+1}$ represent the threshold of state transition at the time $t + 1$; θ is the additive constant of the threshold; Q^t_i and $Neighbor^t_i$ are described as above; α and β are the weight parameters of Q^t_i and $Neighbor^t_i$, respectively ($\alpha + \beta = 1$); β and α are identified by the Monoloop method (Long *et al*., 2009). Based on equal difference principle, adjust β from 0 to β_{max} (maximum weight coefficient, set based on experience), and the weight parameters obtained under maximum *Kappa* index are α and β. The parameter $Rand_i^t$ is the random disturbance variable from 0 to 1, representing influence of humanity factors, such as agricultural policy and war, on spatial diffusion of cropland. ε is a constant which indicates disturbance degree.

In conducting loop iterations in Python with the above model, the amount of cropland in the historical spatial pattern is generated by backward time–steps and decreases progressively. It is possible to identify the pattern of historical cropland in a corresponding year if the amount of reconstruction is equal to the amount of cropland in aimed year (exogenous variable). This represents the potential distribution of historical cropland under the premise of certain social economy, level of productive forces, and amount of cropland.

2.2 Model calibration and validation

After generating a potential spatial distribution pattern of cropland in typical year with the above model, it is necessary to inspect and correct the reconstructed results to make the simulated result consistent with reality.

Before 1980, agricultural production in China was dominated by small–scale cultivation by farmers with limited natural resources. This entailed family–based cultivation dependent on the local labor force. Thus, the amount of historically cropland that is reconstructed in various districts and cities should comply with the local labor supply level. We establish a verification function based on the ability of the labor force to cultivate cropland, as shown in the equation:

$$\mu_{n,j} = \frac{P_{n,j} \times m \times lc}{A_{n,j}} \tag{10}$$

where $P_{n,j}$ refers to the total population of city j in year n; m refers to the proportion of population engaged in agricultural labor; lc refers to the cropland area available for labor force per unit; $A_{n,j}$ refers to the amount of cropland in city j in the year n generated by the model; $\mu_{n,j}$ represents the index of available cropland for the labor force in city j in the year n. When $\mu_{n,j}>1$, it indicates that the local labor force level can support the amount of cropland to be reconstructed. Otherwise, it is necessary to correct the discrete parameter ε of the above model.

3 Case Study

3.1 Study area

Shandong Province is located along the eastern coast of China near the lower reaches of the Huanghe (Yellow) River and the northern central region of the Beijing–Hangzhou Grand Canal. During the Ming and Qing dynasties, Shandong encompassed an area that included most parts of modern Shandong Province (except the counties of Dongming, Ningjin and Qingyun); the counties of Fanxian and Taiqian in Henan Province; as well as Guantao County, Qiuxian County, the southwest of Gucheng County, and the southeast of Haixing County in Hebei Province. Considering the need for connecting historical data with modern data and research results, Shandong Province under modern administrative division is selected as the study area. With a land area of 1.57×10^5 km^2, plain regions (including basin and tableland) cover 64% of the area and 35% is covered by mountain land and hills. This region is characterized by a temperate monsoon climate. With adequate sunshine and moderate rainfall, it has traditionally been the major

grain production base in China. This paper proposes a reconstruction period from 1661 to 1933. The starting point of backward modeling was the status of land use in the study area in 1980 as obtained by remote sensing interpretation (http: //www.geodata.cn.). The cropland area in 1980 was determined to be 1.10×10^5 km^2, and it occupied 69.66% of the total study area. The cropland was well distributed, except in the hilly areas in the center and northeast where the amount of cropland is low. Since cultivation areas on isolated islands are difficult to identify and have low probability of existence, they are not initially considered (Fig. 2).

3.2 Data resources and processing

Given the considerations of the typical scale of cropland elements, feasibility of model calculations, the scale of relevant studies (i.e., climate simulation and carbon cycle), as well as convenience in superposing with existing data, the spatial resolution of this research is set as 1 km × 1 km.

(1) Amount of historical cropland: cropland from 1661 to 1933 was based on the research of Cao *et al*. (2013) combined with the 1933 data from ′China′s Statistical Analysis of Land Issues in 1936′.

(2) Demographic data: demographic data of Shandong Province in 1820 are adopted from the Chinese Population Geographic Information System (CPGIS) which was obtained from the Harvard University CPGIS database. It is amended to current divisions by the administrative area ratio method.

(3) Topographic data: the DEM was provided by Data Sharing Infrastructure of Earth System Science (http:// www.geodata.cn.).

(4) River network and lake data: basic geographic data of the river network and lakes of China were obtained from the National Geomatics Center of China (1 ∶ 4 000 000). They are divided into grades 1 to 5 according to river lever (http: //www.geodata.cn.).

(5) Soil erosion data: the results of the second national soil erosion research by remote sensing are adopted (http: //www.geodata.cn.). 1 km × 1 km spatial grids are generated by the maximum area method.

(6) Physicochemical data of soil: the data of the China Soil Scientific Database of the Institute of Soil Science, Chinese Academy of Sciences are adopted. The analysis indices include pH, organic matter, total N,

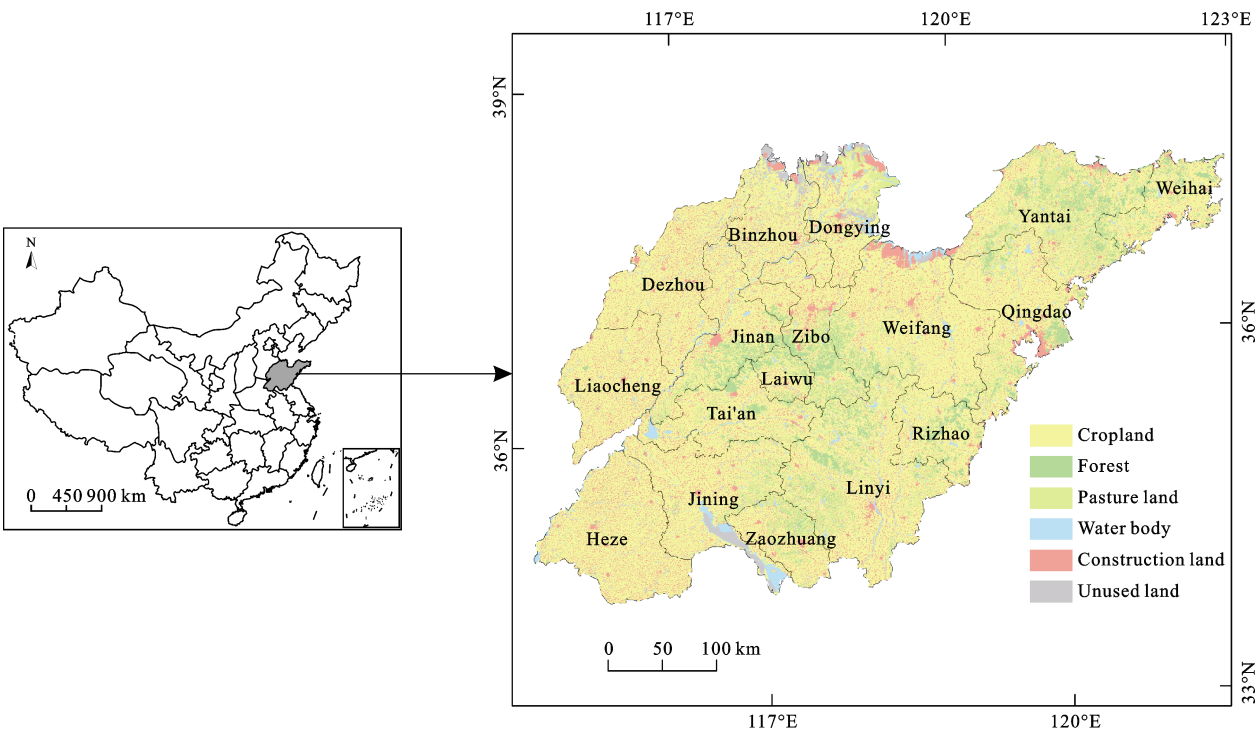

Fig. 2 Land use cover of Shandong Province in 1980

total P, and total K (surface soil 0–20 cm; spatial resolution 2 km × 2 km).

3.3 Quantification and identification of parameters

3.3.1 Parameters quantification

Normalization processing is required for all factors to facilitate computer regression analysis and eliminate dimensional disturbance. Land reclamation suitability is set as follows: 1 is the most suitable and 0 is the least suitable.

(1) Slope standardization

Topographic slope plays an important role in deciding the land use type. Generally, the higher the slope is, the lower the arability will be. Thus, the slope is normalized using the following equation.

$$Slope = \frac{\max(s_i) - s_i}{\max(s_i)} \tag{11}$$

where *Slope* is the normalization value of slope; s_i is the mean elevation of grid *i*; and max (s_i) is the maximum value of slope within the research area.

(2) Elevation standardization

Heat and moisture vary with the elevation, which causes a vertical differentiation of land reclamation. When the elevation reaches a certain altitude, the hydrothermal conditions become the dominant factor restricting the growth of crops (Lin *et al*., 2008; Li *et al*., 2011; Li *et al*., 2012). Thus, the distribution of cropland presents obvious vertical zonality. The elevation is normalized using the following equation.

$$Elevation = \frac{\max(e_i) - e_i}{\max(e_i)} \tag{12}$$

where *Elevation* is the normalization value of elevation; e_i is the mean elevation of grid *i*; and max(e_i) is the maximum value of elevation within the research area

(3) River standardization

Water sources are among the key factors affecting agricultural cultivation, and their spatial distribution largely determines the spatial pattern of cropland. Land units closer to a river channel or lake have more convenient access to water which can be used for agricultural purposes. In the period dominated by small-scale farmers with low productivity, such land is more likely to be reclaimed and maintained as cropland due to lack of large-scale irrigation and water conservation measures (Liu *et al*., 2011). The river accessibility variable was subjected to weighted normalization processing using a theory of spatial influence based on exponential decay with distance (Zhang *et al*., 2008).

$$River = a_1 e^{-b_1 d_1} + a_2 e^{-b_2 d_2} + a_3 e^{-b_3 d_3} + a_4 e^{-b_4 d_4} + a_5 e^{-b_5 d_5} + a_6 e^{-b_6 d_6} \tag{13}$$

where *River* represents the river accessibility and is defined as the spatial distance exponential weight of a river or lake in relation to grid *i*; parameters a_1, a_2, a_3, a_4, a_5, and a_6 are respectively the influence weights of rivers of level 1 to level 5 and adjacent lakes, and the influence weights were determined via expert opinion in combination with the AHP method; b_1, b_2, b_3, b_4, b_5, and b_6 are respectively the spatial influence attenuation coefficients of rivers of all levels and lakes (Zhang *et al*., 2008); and d_1, d_2, d_3, d_4, d_5, and d_6 are respectively the Euclidean distances between grid *i* to rivers of all levels and lakes.

(4) Erosion standardization

Soil erosion is the result of interaction and mutual restriction of geographical environment factors. Soil erosion is divided into water and wind erosion based on the *Standards for Classification and Gradation of Soil Erosion* (*SL 190–2007*) utilized in China. Different arability values are assigned to grids with different soil erosion types and grades by using Table 1 in combination with expert opinions.

(5) Selection and standardization of soil quality index

We select pH, organic matter, total N, total P, and total K to act as the soil physicochemical index which affects the spatial distribution of cropland.

Table 1 Land suitability value under different types and intensities of soil erosion

Type	Intensity level					
	Micro degree	Mid degree	Moderate degree	Strength degree	Severe degree	Extremely severe degree
Eroded by water	0.95	0.80	0.75	0.50	0.30	0.10
Eroded by wind	0.95	0.75	0.70	0.40	0.20	0.10

Note: Engineering erosion only presented in 1.45 km^2 in Shandong Province, thus, conclude it into its adjoining erosion to simplify the classification

First, we standardize the indices. We assign 1 to grid if its pH is 6.3–8.5 and 0 for other values. We then standardize the organic matter of soil, total N, total P, and total K using the following equation:

$$Quality_i = \frac{q_i}{\max(q_i)} \tag{14}$$

where $Quality_i$ is the standardized value of the corresponding index i; q_i is the initial value of corresponding index i; and max (q_i) is the maximum value of corresponding index i.

To determine the arability (*Quality*) of soil in each land unit, we apply an overlay analysis in ArcGIS 9.3 software to obtain a layer that represents a weighted combination of the indices as in the following equation:

$$Quality = \sum_{i=1}^{n} \rho_i \times Quality_i \tag{15}$$

Parameter ρ_i is the influence weight of corresponding index i and is calculated via AHP. ρ_1, ρ_2, ρ_3, ρ_4, and ρ_5 represent the index of the pH, organic matter of soil, total N, total P, and total K, respectively.

3.3.2 Parameter identification

Analyzing modern cropland spatial data is required to determine the relevant parameter values of state transition rules. In this paper, the cropland distribution range of the research area in 1980 is considered as the maximum of the historical cropland space distribution and is used for parameter identification:

(1) Parameter identification of basic data: basic data primarily refers to the river accessibility and soil quality variables. We calculate the spatial influence weights of the above two variables (Table 2) by combining expert opinions and AHP. The matrix consistency coefficients were 0.0403 and 0.0391, respectively, which are both less than 0.1 and meet the research requirements.

(2) We apply the Binary Logistic module of SPSS software for the regression of influence coefficients. The five land reclamation suitability factors (excluding the *Neighbor* variable) were the independent variables and the land use is selected to be the dependent variable. When the grid unit is cropland it is defined as 1, otherwise it is 0. The results are shown in Table 3. The result indicates that all factors pass inspection. The major factors affecting the spatial pattern in descending order of influence are: slope and elevation; soil erosion and soil quality; and water accessibility. Please note that the influence of these factors on the spatial pattern of cropland is positive mainly due to the quantification method of the arability factor. The prediction precision of logistic regression model can be measured by using it to predict actual data and calculating the proportion of correct classifications. The total accuracy of the modeling dataset is 81.40% and total accuracy of the inspection dataset is 79.60%, which indicate that the model prediction precision is high and the prediction capability is stable (Table 4).

Based on the above parameter settings, the spatial data patterns are shown in Fig. 3.

(3) Identification of α and β values: the starting year of backward modeling is set as 1980. We adjust α and β (the initial value is set as 0.25) values to generate the first cropland distribution pattern layer by applying the Monoloop method. We then identify the weight parameters, which are the values of α and β values (0.80, 0.20), where the maximum Kappa index exists.

(4) Result correction: we used the demographic dataset of cities from the CPGIS data and amount of cropland of each city simulated in 1820 to adjust values based on Equation (10). The parameter setting is done according to Cao *et al*. (2013), and the results are shown in Fig. 4. According to the analysis, the cropland indices

Table 2 Descriptive table of parameters

Parameter	Weight value	Parameter	Weight value
a_1	0.2996	ρ_1	0.0547
a_2	0.2267	ρ_2	0.2789
a_3	0.1799	ρ_3	0.1395
a_4	0.1039	ρ_4	0.3204
a_5	0.0841	ρ_5	0.2065
a_6	0.1059		
Consistency of judgment matrix	0.0403	Consistency of judgment matrix	0.0391

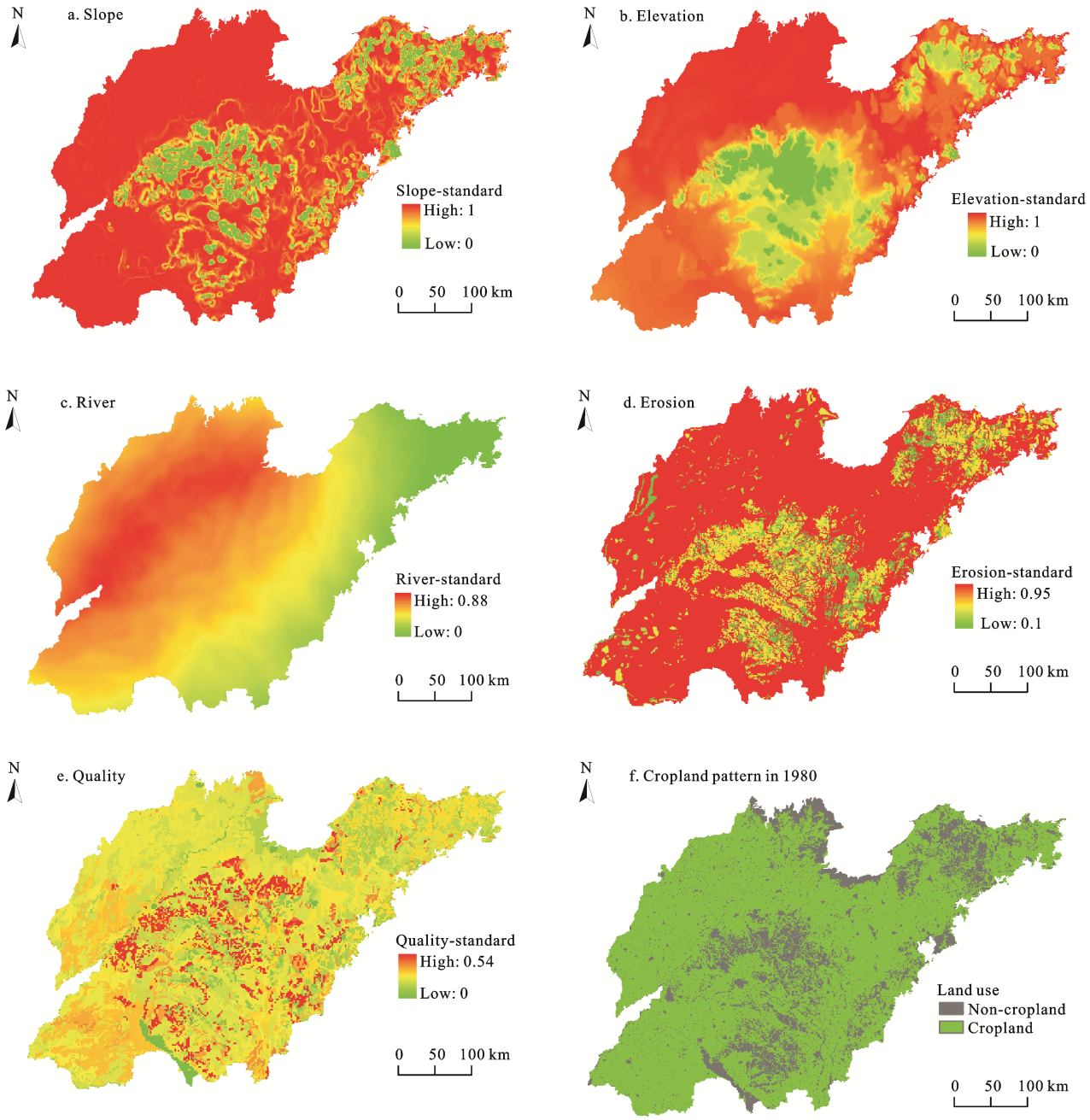

Fig. 3 Spatial pattern of influence factors to cropland reconstruction

Table 3 Result table of parameter identification

Parameters	B*	S.E.	Wald	Sig.
a	–11.587	0.133	7638	0
γ_1	4.719	0.150	985.38	0
γ_2	4.491	0.088	2575	0
γ_3	1.874	0.068	755.36	0
γ_4	2.882	0.045	4149	0
γ_5	2.084	0.085	596.83	0

Note: B* represents regression coefficient of variables

Table 4 Prediction accuracy by logistic

Observation cropland	Prediction Cropland		Accuracy (%)
	0	1	
0	7257	23863	23.3
1	4612	117160	96.2
Total accuracy (%)	61.14	83.08	81.4

Note: 1 means cropland, 0 means non-cropland

of labor force for each city (μ) within the study area are

primarily between 1 and 4. This indicates that a reasonable amount of cropland is distributed to each city. However, the μ values of Weihai and Laiwu cities are as high as 16. There are many mountainous regions and hills in Laiwu City with significant changes in topography. The cause of this outlier may be that the corrected demographic data are too large to conform to the historical reality of Laiwu City. For Weihai City, abnormal data may be due to mistaken correction of demographic data. Weihai City is close to the Huanghai Sea and more people fish rather than cultivate land. Besides, the rural settlement distribution pattern of Weihai City in 1820 in the Chinese Historical Geographic Information System (CHGIS) proves that the rural settlement of Weihai City was primarily on the coastline with little located inland.

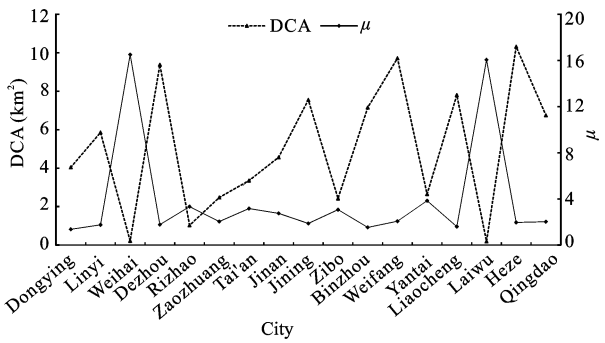

Fig. 4 Cropland index of labor force (μ) and distributed cropland area (DCA) in cities (1820)

3.4 Results and analyses

3.4.1 Results

Based on the land reclamation suitability and parameter values identified in the study area, the spatial distribution pattern of historical cropland over 300 years is created via backward modeling of historical cropland space. The spatial pattern distribution of cropland for that time period was then obtained according to the amount of historical cropland for each time period in the study area after correction with demographic data (Fig. 5).

According to the simulated spatial pattern of historically cropland, cultivation in Shandong spread to eastward and southward regions at first and then to the low hills and land along the sea. This situation is basically consistent with the historical change pattern of cropland in Shandong for the past 300 years. Due to the wars in the late Ming and early Qing dynasties (approximately AD 1628–1661), the cropland in Shandong was reduced sharply and was mainly located in the west and northwest along the bank of the Yellow River in the early Qing Dynasty. With a series of policies and measures encouraging reclamation issued by the emperors Kangxi, Yongzheng and Qianlong (Peng, 1990; Cheng, 2010), such as tax exemption for fragmented farm land and construction of new irrigation and water conservation projects, effective farm reclamation was conducted for the land abandoned in the late Ming Dynasty. As a result, the amount of cropland in Shandong was increased significantly and the spatial pattern showed extension towards the areas with flat terrain and higher fertility in the east and south (He *et al.*, 2005). After vigorous reclamation over the past several dynasties, practically all easily cropland during the Jiaqing and Daoguang periods was being utilized and population pressure on land resources increased. Shandong began to introduce the American crops such as sweet potato that could grow on less fertile land and the spatial pattern shows that the cropland extended towards Yantai City, Weihai City, and the central low mountain areas in Shandong as the land developed by outward extension was easy to dry (Cao *et al.*, 2013).

3.4.2 Comparison and analysis of results

The spatial distribution pattern of historical cropland is affected by numerous factors, the comparison data are inadequate, and the history can not be accurately verified. Thus, it is no longer suitable to compare a simulated result to the real pattern of historical cropland via the traditional point-to-point comparison method and overall comparison method for pattern spots. To further analyze the reliability of the simulated results, the spatial pattern is compared with international mainstream dataset. Selecting 1820 as the reference year, the results in this paper are compared to HYDE 3.1 (http://themasites.pbl.nl/en/themasite) and SAGE 2010 (http://www.sage.wisc.edu/iamdata). The spatial pattern in 1800 from the SAGE dataset was used since study periods did not coincide.

According to comparison between the results in this paper (Fig. 5e and Fig. 6a), due to the relatively low spatial resolution (0.5° × 0.5°), SAGE dataset is highly approximate with regard to the spatial pattern of cropland. The spatial pattern also presents evident land type mistakes, for example, Rizhao, Yantai and Weihai cities are replaced by the seas. Both spatial patterns have some similarities in where croplands are concentrated, but the spatial patterns varied in where croplands are

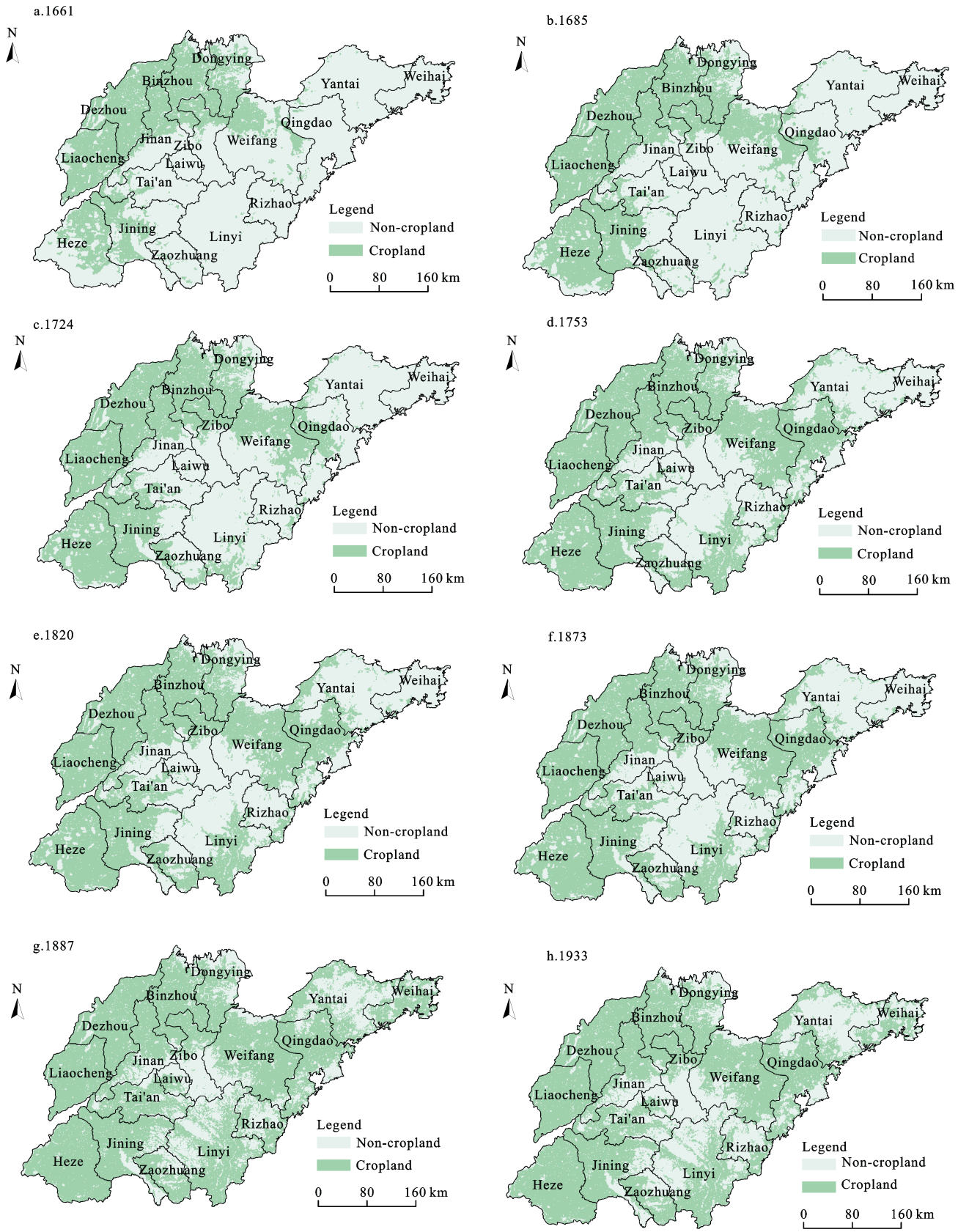

Fig. 5 Reconstruction results of cropland during 1661 to 1933

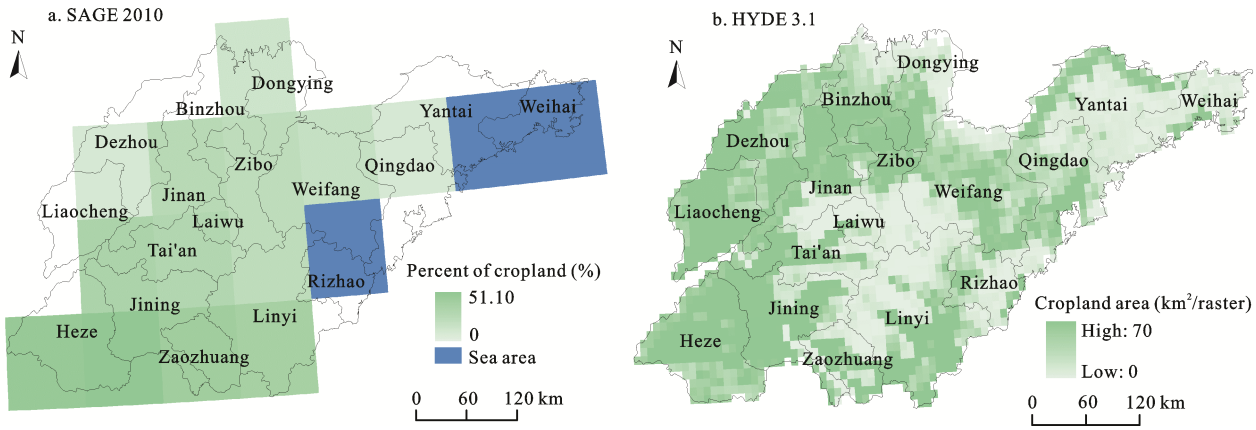

Fig. 6 Spatial pattern of cropland distributions between SAGE (2000) and HYDE 3.1 datasets in Shandong Province (1820)

scarce. For the hilly area in the central of Shandong, it was primarily grassland and forest land in the land use map of 1980, so the proportion of cropland in this area should not be significant in 1820.

Compared with the results in Fig. 5e and Fig. 6b, the spatial distribution patterns of cropland show a good similarity and higher consistency where croplands are concentrated and scarce. A major difference is that the latter has three low-value areas on the north bank of the Yellow River. According to Fig. 3, the region for the low-value areas has a higher value of the reclamation suitability factor and more is distributed evenly. In addition, according to the CHGIS dataset, the rural residential area on the northern bank of Yellow River was distributed evenly in 1820, so the opportunity to cultivate the lands should be equal. It is unreasonable that this region has three low-value areas of cultivation.

From the aforementioned comparison and analysis, the spatial pattern of historical cropland created by our model is more consistent with the objective reality. In addition, the Boolean spatial pattern generated has a higher spatial resolution and is more spatially explicit than the proportional dataset[0].

4 Conclusions

(1) We break from the traditional spatial distribution method of cropland, i.e., ′top-down′. The reconstruction model of the spatial pattern of historical cropland is established ′bottom-up′ with the introduction of the constrained CA model and assumptions regarding the continuity of cropland. The spatial pattern of historical cropland is backtracked under the restraint of modern pattern and the historical amount of cropland.

(2) By using the model established in this paper and taking Shandong as the study area, the simulated results show that cultivation in Shandong spread to the eastward and southward region at first and then to the low hills and regions along the sea. In addition, based on the comparison with the mainstream dataset, the spatial pattern of historical cropland created by the model is more consistent with the objective reality. This indicates that this model has certain applicability when backtracking the spatial pattern of historically cropland.

(3) Based on the modern pattern of cropland, reclamation suitability, neighborhood, and random disturbance factors, the spatial distribution pattern of historical cropland is reconstructed under the control of cropland amount for the typical time period. This spatial pattern is the most probable distribution pattern of historical cropland based on our rational analysis.

The model established in this paper does not consider the spatial differences of transition rules due to restrains in the availability of data. Transition rules could be made that correspond to different regions, and then provincial and even nationwide spatial patterns of historical cropland could be generated through synchronous or asynchronous partitioned operation after integration. The results of such a simulation are expected to be more reliable and this method needs to be investigated in future studies.

Acknowledgements

The authors thank Postdoctoral Zhang Honghui, Professor Cui Xuefeng and editors for helping me to finish this thesis perfectly.

References

Bai Shuying, Zhang Shuwen, Zhang Yangzhen, 2007. Digital rebuilding of LUCC spatial-temporal distribution of the last 100 years: taking Dorbod Mongolian Autonomous County in Daqing City as an example. *Acta Geographica Sinica*, 62(4): 427–436. (in Chinese)

Cao Xue, Jin Xiaobin, Zhou Yinkang, 2013. Research on cropland data recovery and reconstruction in the Qing Dynasty: method and case study. *Acta Geographica Sinica*, 68(2): 45–256. (in Chinese)

Cheng Fang, 2010. Commentary review on cultivation of land in Shandong of Qing Dynasty. *History Teaching*, (8): 32–36. (in Chinese)

Feddema J J, Oleson K W, Bonan G B *et al*., 2005. The importance of land-cover change in simulating future climates. *Science*, 310 (5754): 1674–1678. doi: 10.1126/science.1118160

Feng Zhiming, Liu Baoqin, Yang Yanzhao, 2005. A study of the changing trend of Chinese cultivated land amount and data reconstructing: 1949–2003. *Journal of Natural Resources*, 20(1): 35–43. (in Chinese)

Foley J A, DeFries R, Asner G P *et al*., 2005. Global consequence of land use. *Science*, 309(5734): 570–574. doi: 10.1126/ science.1111772

Ge Quansheng, Dai Junhu, He Fanneng *et al*., 2003. Cultivated land amount change and driving forces analysis of some provinces of china in past 300 years. *Advance in Natural Sciences*, 13(8): 825–832. (in Chinese)

Goldewijk K K, 2001. Estimating global land use change over the past 300 years: The HYDE database. *Global Biogeochemical Cycles*, 15(2): 417–433. doi: 10.1029/1999GB001232

Goldewijk K K, Navin R, 2004. Land-cover change over the last three centuries due to human activities: the availability of new global data sets. *Geojournal*, 61(4): 335–344. doi: 10.1007/ s10708–004–5050–z

Goldewijk K K, Beusen A, Van D G, 2011. The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12 000 years. *Global Ecology and Biogeography*, 20(1): 73–86.

He F N, Li S C, Zhang X Z, 2012a. Reconstruction of cropland area and spatial distribution in the mid-Northern Song Dynasty (AD1004–1085). *Journal of Geographical Sciences*, 22(2): 359–370. doi: 10.1007/s11442-012-0932-3

He Fanneng, Dai Junhu, Ge Quansheng, 2005. An analysis of reclamation trend in the early Qing Dynasty from the view point of Kangxi to Qianlong cultivation and reclamation policies. *Geographical Research*, 24(6): 878–888. (in Chinese)

He Fanneng, Li Shicheng, Zhang Xuezhen *et al*., 2012b. Comparisons of reconstructed cropland area from multiple datasets for the traditional cultivated region of China in the last 300 years. *Acta Geographica Sinica*, 67(9): 1190–1200. (in Chinese)

He Fanneng, Tian Yanyu, Ge Quansheng, 2003. Spatial-temporal characteristics of land reclamation in Guanzhong region in the Qing Dynasty. *Geographical Research*, 22(6): 687–697. (in Chinese)

Jiao Limin, Liu Yaolin, 2004. Application of fuzzy neural networks to land suitability evaluation. *Geomatics and Information Science of Wuhan University*, 29(6): 513–516. (in Chinese)

Li B B, Fang X Q, Ye Y *et al*., 2010. Accuracy assessment of global historical cropland datasets based on regional reconstructed historical data: a case study in Northeast China. *Science in China*: *Earth Science*, 40(8): 1048–1059. doi: 10. 1007/s11430–010–4053–5

Li Ke, He Fanneng, Zhang Xuezhen, 2011. An approach to reconstructing spatial distribution of historical cropland with grid-boxes by utilizing MODIS land cover dataset: a case study of Yunnan Province in the Qing Dynasty. *Geographical Research*, 30(12): 2281–2288. (in Chinese)

Li Shicheng, He Fanneng, Chen Yisong, 2012. Gridding reconstruction of cropland spatial patterns in Southwest China in the Qing Dynasty. *Progress in Geography*, 31(9): 1196–1203. (in Chinese)

Li Xia, Ye Jia′an, 1999. Constrained cellular automata for modeling sustainable urban forms. *Acta Geographica Sinica*, 54(4): 289–298. (in Chinese)

Lin Shanshan, Zheng Jingyun, He Fanneng, 2008. The approach for gridding data derived from historical cropland records of the traditional cultivated region in China. *Acta Geographica Sinica*, 63(1): 83–92. (in Chinese)

Liu M L, Tian H Q, 2010. China′s land cover and land use change from 1700 to 2005: estimations from high-resolution satellite data and historical archives. *Global Biogeochemical Cycles*, 24(3). doi: 10.1029/2009GB003687

Liu Minghao, Dai Zhizhong, Qiu Daochi *et al*., 2011. Influencing factors analysis and rational distribution on rural settlements in mountains region. *Economic Geography*, 31(3): 476–482. (in Chinese)

Liu X P, Li X, Yeh A G O, 2006. Multi-agent systems for simulating spatial decision behaviors and land use dynamics. *Science in China Series D*: *Earth Sciences*, 49(11): 1184–1194. doi: 10.1007/s11430-006-1184-9

Liu Yaolin, Liu Yaofang, Xia Zaofa, 1995. Land suitability evaluation based on fuzzy comprehensive judgment. *Journal of Wuhan Technical University of Surveying and Mapping*, 20(1): 71–75. (in Chinese)

Long Ying, Han Haoying, Mao Qizhi, 2009. Establishing urban growth boundaries using constrained CA. *Acta Geographica Sinica*, 64(8): 999–1008. (in Chinese)

Long Ying, Shen Zhenjiang, Mao Qizhi *et al*., 2010. Form scenario analysis using constrained cellular automata. *Acta Geographica Sinica*, 65(6): 643–655. (in Chinese)

Peng Yuxin, 1990. *Cultivation History In Qing Dynasty*. Beijing. China Agriculture Press, 1–285. (in Chinese)

Pongratz J, Reick C, Raddatz T *et al*., 2008. A reconstruction of global agricultural areas and land cover for the last millennium. *Global Biogeochem Cycles*, 22(3): 1–16. doi: 10.1029/ 2007 GB003153

Ramankutty N, Foley J A, 1999. Estimating historical changes in

global land cover: Croplands from 1700 to 1992. *Biogeochemical Cycles*, 13(4): 997–1027. doi: 10.1029/1999GB900046

Ray D K, Pijanowski B C, 2010. A backcast land use change model to generate past land use maps: application and validation at the Muskegon River watershed of Michigan, USA. *Journal of Land Use Science*, 5(1): 1–29. doi: 10.1080/17474230903150799

Shi Zhengguo, Yan Xiaodong, Yin Conghua *et al*., 2007. Effects of historical land cover changes on climate. *China Science Bulletin*, 52(12): 1436–1444. (in Chinese)

Tian H Q, Banger K, Bo T *et al*., 2014. History of land use in India during 1880–2010: large-scale land transformations reconstructed from satellite data and historical archives. *Global and Planetary Change*, 121: 78–88. doi: 10.1016/j.gloplacha.2014.07.005

Voldoire A, Eickhout B, Schaeffer M *et al*., 2007. Climate simulation of the twenty-first century with interactive land-use changes. *Climate Dynamics*, 29(2–3): 2–3. doi: 10.1007/s00382-007-0228-y

Wu F, 1998. SimLand: A prototype to simulate land conversion through the integrated GIS and CA with AHP-derived transition rules. *International Journal of Geographical Information Science*, 12(1): 63–82. doi: 10.1080/136588198242012

Xie Yaowen, Wang Xueqiang, Wang Guisheng *et al*., 2013. Cultivated land distribution simulation based on grid in middle reaches of Heihe River Basin in the historical periods. *Advances in Earth Science*, 28(1): 71–78. (in Chinese)

Ye Duzheng, Fu Congbin, 2004. Some advance in global change science study. *Bulletin of Chinese Academy of Sciences*, 19(5): 336–341. (in Chinese)

Ye Yu, Fang Xuiqi, Ren Yuyu *et al*., 2009. Northeast cultivated land cover changes over the past 300 years. *Science in China* (*Series D*), 39 (3): 340–350. (in Chinese)

Yeh A G O, Li X, 2001. A constrained CA model for the simulation and planning of sustainable urban forms by using GIS. *Environment and Planning B*: *Planning and Design*, 28(5): 733–753. doi: 10.1068/b2740

Zhang Guoping, Liu Jiyuan, Zhang Zengxiang, 2003. Spatial-temporal changes of cropland in china for the past 10 years based on remote sensing. *Acta Geographica Sinica*, 58(3): 323–332. (in Chinese)

Zhang Honghui, Zeng Yongnian, Jin Xiaobin *et al*., 2008. Urban land expansion model based on multi-agent system and application. *Acta Geographica Sinica*, 63(8): 869–881. (in Chinese)

Zhou Rong, 2001. A general inspection and re-appraisal on area under cultivation in the early period of Qing Dynasty. *Jianghan Tribune*, 9: 57–61. (in Chinese)

Zhu Feng, Cui Xuefeng, Miao Lijuan, 2012. China′s spatially-explicit historical land-use data and its reconstruction methodology. *Progress in Geography*, 31(12): 563–1573. (in Chinese)