Applied Geography 52 (2014) 67–77

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

Reconstruction of historical arable land use patterns using constrained cellular automata: A case study of Jiangsu, China

Ying Long [a,b], Xiaobin Jin [a,c,*], Xuhong Yang [c], Yinkang Zhou [a,c]

[a] Natural Resources Research Center of Nanjing University, Nanjing 210023, China
[b] Beijing Institute of City Planning, Beijing 100045, China
[c] School of Geographic and Oceanographic Sciences, Nanjing University, Nanjing 210023, China

Keywords:
Constrained cellular automata
Historical arable land
Jiangsu
LUCC
Spatial allocation

A B S T R A C T

The reconstruction of arable land patterns over historical periods is one of critical research issues in the study of land use and land cover change (LUCC). Taking into account the continuous distribution of arable land and spatial constraints, this paper proposes a constrained cellular automata model to reconstruct historical arable land patterns. The paper describes model establishment, parameter calibration, and results validation in detail. The model was applied to Jiangsu Province, China, and was compared with a conventional spatial allocation method. The results showed that the methodology developed in this study can more objectively reflect the evolution of the pattern of arable land over historical periods, in terms of similarity with contemporary pattern, than the spatial allocation methods and can provide an effective basis for the historical study of arable land.

Introduction

Land use and land cover change (LUCC) forms one of the core elements of global environmental change (IGBP, 2001). Large-scale and long-term LUCC have profound effects on atmospheric composition, climate change, nutrient cycling, ecosystems, and more (Foley, Ruth, & Gregory, 2005). The effect of human activities on the earth has increased, especially in the past 300 years, and the resulting changes in the global environment have been profound. The impact of changes in land cover over historical periods should be included during the building of models that accurately simulate global environmental change (Turner et al., 1995). The International Geosphere–Biosphere Program (IGBP) and the International Human Dimensions Program on Global Environmental Change (IHDP) co-sponsored LUCC research in the mid-1990s. That research emphasized that various means could be adopted to reconstruct a detailed history in the change of land use (Ge, Dai, He, Pan, & Wang, 2008; Lambin et al., 2001; Lambin & Veldkamp, 2005), thus creating an upsurge in research in the area of land cover changes over historical periods. Reconstruction of land cover history, especially reconstruction of high-precision spatial data of arable land cover, has been receiving much attention by scholars (Zhu, Cui, & Miu, 2012).

Reconstruction of arable land use, especially reconstruction of data on arable land during historical periods is divided into two types of reconstruction: quantity reconstruction and spatial pattern reconstruction. Quantity reconstruction is used to obtain quantitative statistics related to arable land dynamics (Zhu et al., 2012). Quantity reconstruction mainly reflects the overall trends and regional differences during historical periods; but is also an important foundation for pattern reconstruction. Spatial pattern reconstruction refers to restoration of the spatial distribution of the total amount of arable land. It simulates and creates the spatial distribution of arable land, based on certain spatial allocation principles and combined with the land quantity data (Zhu et al., 2012). Pattern reconstruction provides an important basis for in-depth analysis of LUCC and the effects of ecology, climate, and environment. Quantity reconstruction is relatively mature scientifically speaking; it mainly consists of collections of, corrections to, and amendments to multi-source historical data, and when amended, trends such as food production, population, and cultivation were taken as affecting indicators to set up a correction system of the amount of arable land. However, the reconstruction of patterns still requires in-depth study.

* Corresponding author. School of Geographic and Oceanographic Sciences, Nanjing University, 163 Xianlin Road, Nanjing 210023, PR China. Tel.: +86 13512541166 (mobile).
E-mail address: jinxb@nju.edu.cn (X. Jin).

http://dx.doi.org/10.1016/j.apgeog.2014.05.001

This paper focuses on patterns of reconstruction and provides quantity reconstruction results. Among existing studies related to reconstructing arable land patterns, the gridded global land-cover datasets SAGE and HYDE established by Ramankutty and Foley (1999) and Goldewijk (2001), Goldewijk, Beusen, and Van (2011) respectively, are the most cutting edge products. SAGE reconstructed the global distribution of arable land in a $0.5^\circ \times 0.5^\circ$ spatial resolution for the past 300 years based on the modern pattern of global land use. The latest version of HYDE simulated historical changes of global arable land and pasture at a higher spatial resolution ($5' \times 5'$) over the past 12,000 years using a more detailed algorithm. These two studies have served as excellent references for subsequent studies. Many scholars used or improved these methods for more in-depth data reconstruction, including worldwide research conducted by Pongratz, Reick, and Raddatz (2008) who reconstructed the global distribution of arable land and pasture of 800–1700 AD using 1700 as the base year. In addition, Kaplan, Krumhardt, and Zimmermann (2009) conducted a region-wide study and reconstructed the forest cover in Europe since 1000 BC. Among studies that reconstructed patterns of arable land in China, Liu and Tian (2010) established a representative study using the entire cultivation intensity data set for Chinese traditional agricultural regions in 1820 in China. Other studies are limited to some local regions. For example, Lin, Zheng, and He (2008) reconstructed the arable land pattern in traditional agricultural areas on mainland China for six historical time periods. Ye, Fang, and Ren (2009) reconstructed the number of arable land types in each county of Northeast China over the past 300 years. Li, He, and Zhang (2011) reconstructed datasets of arable land in the years 1671 and 1827 in Yunnan Province. Li, He, and Chen (2012) reconstructed the pattern of land distribution in the southwest region of China.

Overall, current reconstruction results on historical arable land all deeply explore available information on arable land, and in these studies, the abundance of historical data on China was used and given full attention (Lin et al., 2008). In most of existing studies, the patterns reconstruction followed the quantity reconstruction; first, environmental, socio-economic background, and historical data were used to reconstruct the overall amount of land use during historical periods, and then modern patterns of land use and the impact of factors were used to establish appropriate algorithms for spatial allocation (Zhu et al., 2012). The basis of these methods is a top-down spatial allocation based on land suitability. Three different types of spatial allocation exist. First, the total dependent-method performs simple allocation, taking the modern land-use as the entire impact factor. Second, the partially dependent-method performs allocation by taking the pattern of modern arable land as a boundary condition, while considering population, terrain and other spatial factors that are used for the evaluation of land suitability. Third, the dynamic dependent-method adds a dynamic weight on the basis of the partially dependent method. In this case, the weight of the modern agriculture pattern gradually decreases while going back in time and the weight of spatial factors gradually increased while going back in time to reflect the effects of time on spatial factors affecting the historical pattern (Zhu et al., 2012).

In the literature, many scholars have summarized and screened out spatial factors that affect land use changes. These factors are generally classified into three groups, socio-economic drivers (e.g. increasing population and booming economy), biophysical drivers (e.g., soil condition and slope) and proximate forces (e.g. distance to roads and human settlements) (Turner et al., 1995). For instance, Verburg, De Koning, Kok, Veldkamp, and Bouma (1999) included biophysical and socio-economic driving forces in their CLUE modeling framework. Here we focus on literature related to China since we are reconstructing patterns of Chinese arable land. Li et al. (2011) argued that the process of development in arable land was influenced by terrain, heat, water, soil, vegetation and other natural factors as well as demographic, economic development, agricultural policy, war and other social factors. Liu, Liu, and Xia (1995) took elevation, erosion, content of organic matter, pH, soil texture, topsoil thickness, soil depth, and drainage condition as the main factors in an evaluation of arable land suitability. Overall, factors affecting the suitability of arable land can be divided into two categories: natural factors and human factors. The natural factors include terrain slope, elevation, heat, moisture, intensity of soil erosion, climate, potential productivity, temperature conditions (accumulated), effective soil thickness, exposed bedrock (bare rock per total area), soil texture, hydrological and drainage conditions, soil salinization (ratio of salinization area), irrigation (distance to water source), vegetation index, soil pH, soil organic content and topsoil thickness. Human factors include demographic information, economic development, agricultural policies, wars and famines.

In the process of reconstructing patterns of arable land across the landscape, in addition to considering the suitability of arable land (in the form of spatial factors), one also needs to consider the principle of continuous distribution of arable land, that is, to consider the places that are surrounded by arable land are more likely to be arable land. This is consistent with the cellular automata (CA) modeling concept. CA is a bottom-up research tool based on complex adaptive systems (CAS), and can be used to simulate spatiotemporal dynamic processes occurring in a complex system. The basic concept of CA is that the status of any particular cell at the next time period is affected by its own status and the status of its neighboring cells. Constrained CA can simulate more realistic spatial dynamic processes than other methods by considering other constraints in addition to a simple CA and its neighborhood. Constrained CA models have been extensively applied in modeling urban expansion thanks to their capability to simulate dynamic spatial processes from a bottom-up perspective in the real world. For instance, Li and Yeh (2000) modeled sustainable urban development by integrating constrained CA and GIS for the city of Dongguan, China. Ward, Murray, and Phinn (2003) integrated spatial optimization and CA for modeling urban expansion in southeast Queensland, Australia. White, Straatman, and Engelen (2004) developed a CA based on an integrated dynamic spatial simulation model for the entire Netherlands. Lagarias (2012) developed a CA model for simulating urban sprawl in Thessaloniki, Greece. Long, Shen, and Mao (2012) used a CA model to simulate urban expansion in Beijing, China, and identified policies that were necessary for implementing planned urban patterns. Moghadam and Helbich (2013) applied a Markov chains-CA integrated model for simulating urbanization process of Mumbai, India. The spatial factors to be considered in the spatial allocation process of the reconstruction of patterns of arable land can be included in constrained CA in the form of (spatial) constraints. Additionally, a constrained CA model has a strong spatial computing power; it can effectively simulate complex systems in a bottom-up approach, and has been shown to have advantages in studies on urban expansion, land use change, and other fields (Long et al., 2012). Therefore, the constrained CA method has the ability to consider land contiguous development principles and multiple spatial factors in the reconstruction of patterns of historical arable land.

Based on previous research of constrained CA and historical reconstruction of the patterns of arable land, this study reconstructs arable land patterns using constrained CA, and applies the approach in Jiangsu Province, China. Existing constrained CA studies have focused on forecasting the process of future urban expansion. In this study, on the contrary, we focus on using constrained CA for backcasting historical arable land. To the best of our knowledge, no constrained CA-based reconstruction research has been conducted on the historical land use so this study provides a new perspective for it. The remaining four parts of this paper include four parts. Study area and data section describes the study area and the data used for the study. Methods section focuses on modeling, parameter identification, and validation of the method in reconstruction of arable land pattern based on constrained CA. Results and discussion section shows simulation results of applying the model to Jiangsu Province, and results of the model validation. Conclusions section discusses the potential contribution of this study and future studies, as well as makes concluding remarks.

Study area and data

Study area

Thus study encompasses the entire scale of present-day Jiangsu Province in the center of the eastern coastal areas of China (Fig. 1). This lies in the downstream region of the Yangtze River, west of and adjacent to the Yellow Sea, northwest of Zhejiang and Shanghai, east of Anhui, and south of Shandong. Jiangsu Province covers a total area of 103,000 km^2, and is densely covered by a network of canals, rivers, streams and other waterways. By 2011, the area had 47,200 km^2 of arable land or 45.8% of the entire study area.

Data

Spatial distribution of arable land in 1980 and 1960

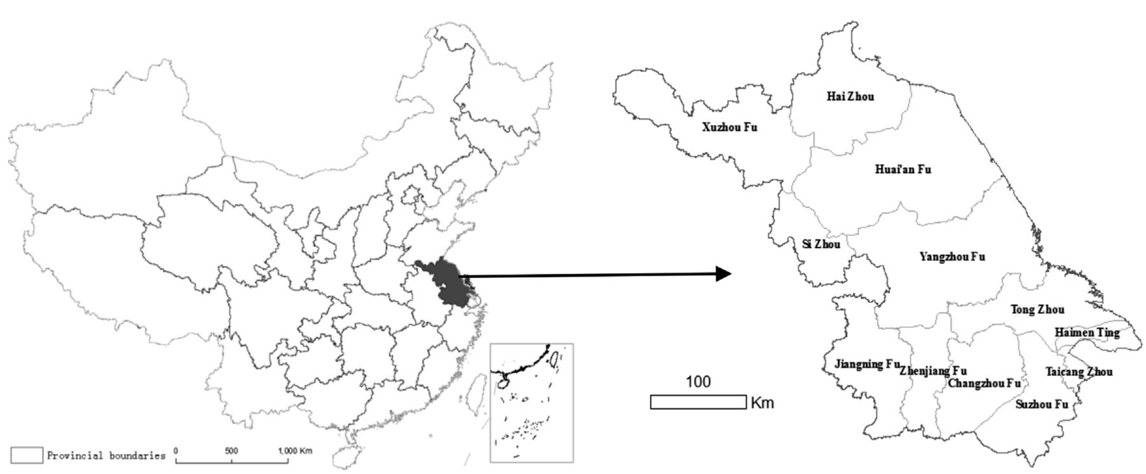
In this paper, 1980 was used as the base year for the reconstruction of land use patterns and is based on land use data for Jiangsu Province in 1980. These data were interpreted from remote sensing images which was downloaded from http://www.geodata.cn; we extracted the spatial distribution of arable land of Jiangsu Province in 1980 (Fig. 2(a)). Total arable land in 1980 was 72,338 km^2.

In addition to arable land in 1980, we interpreted arable land in Dec 1960 from CORONA remote sensing images from the USA that one can download from http://earthexplorer.usgs.gov/. This was used for calibrating the arable land transition rule during 1980–1960 for reconstructing the historical pattern of arable land use. Because of poor data availability, arable land in 1960 was limited to a part of Jiangsu Province with an area of 10,278 km^2 (Fig. 2(b)), and a total arable land of 5664 km^2. The area of arable land in the same region was 8289 km^2 in 1980 with an expansion from 1960 to 1980 of 3319 km^2.

The total amount of arable land in typical years

Total arable land data for certain years was selected to determine the total arable land at various reconstruction points in time, as the exogenous variables of constrained CA. The arable land data from the Qing Dynasty came mainly from three sources: (a) official records and local history books such as “Qing Yi Tong Zhi,” and “Rebuilt Jiaqing Tong Zhi; ” (b) widely cited research results, such as “Chinese dynasties account, farmland, land tax statistics” (Liang, 2008); (c) research results related to regional history and geography, such as “Jiangsu and Anhui provinces land use and its driving force mechanism” (Zhao, 2005). Recovery and reconstruction of arable land data from the Qing Dynasty includes two aspects: the correction system of registered rebellion data and the validation system of revised arable land data. The correction system of registered rebellion data analyses the impact factor leading to untrue registers of land by estimating the extent of the influence of a variety of factors in different periods. We established a factor correction table and turned registered rebellion data into actual acres of arable land, and initially obtained corrected data of arable land. The validation system of revised arable land data, based on preliminary revised data, further examines and revises it in terms of the population base and the reclamation trends (Cao, Jin, & Zhou, 2013). Taking into account the availability of data and policies related to arable land, the years selected in the Qing Dynasty were 1661, 1820, and 1887.

Little variation in arable land data is available during the Republic of China. Researchers have been forced to use limited sources of data source, e.g. “China’s agricultural profile estimates” (Zhang, 1933) and “China’s land-use statistics” (Pu, 1933). Nevertheless, researchers mostly used the “quoted convergence” correction method citing related literature as the main data based on considering the link to arable land data. The arable land data during the Republic of China in this paper mainly came from 1933 data available in “China’s statistical analysis of land issues” (Statistics Bureau of the National Government, 1936).



(a) (b)

Fig. 1. Location of Jiangsu in the modern China (a) and prefectural boundaries during the Qing Dynasty (b) of Jiangsu Province, China.

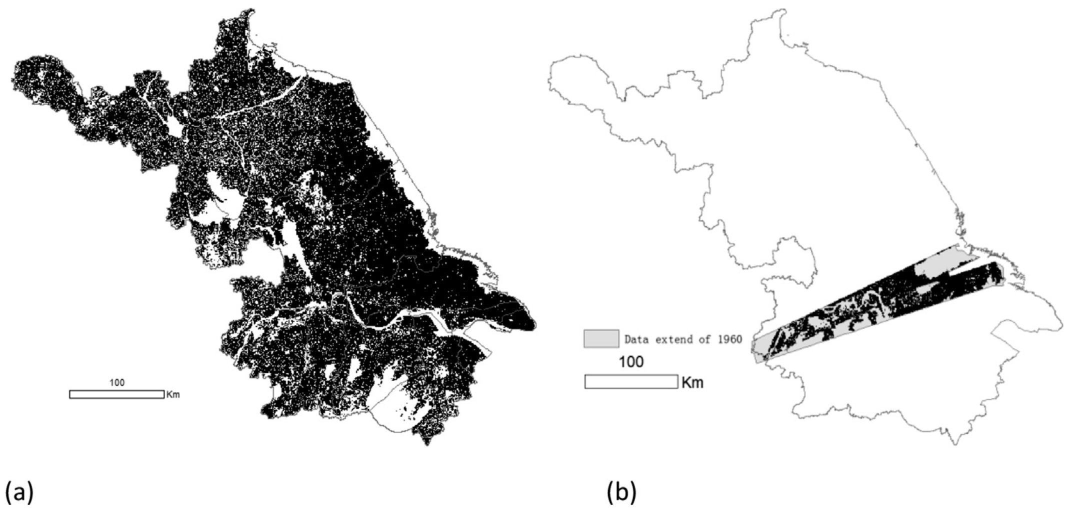


Fig. 2. Arable land in 1980 (a) and 1960 (b). Note: black represents arable land.

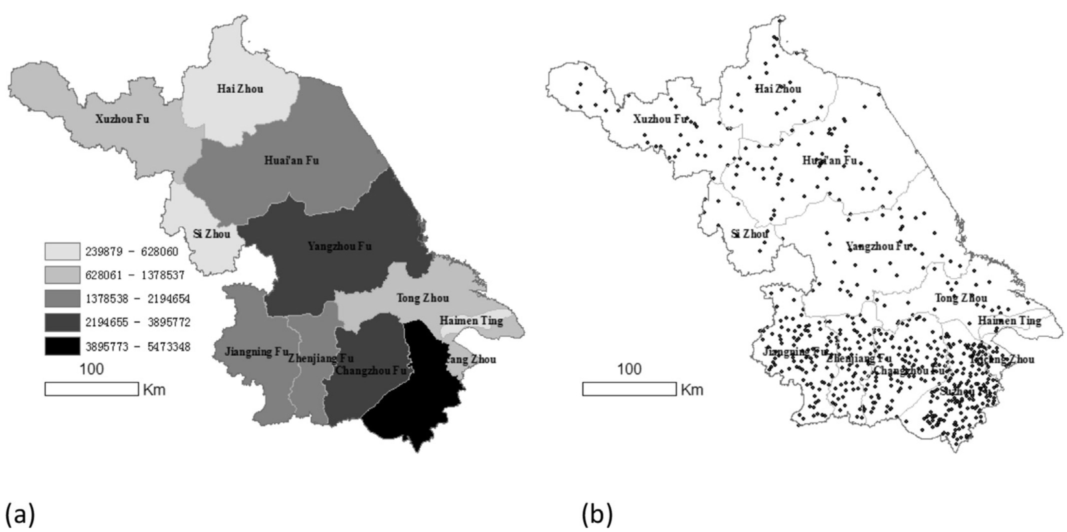


Fig. 3. Prefectural population (a) and human settlements (b) distribution of Jiangsu Province in 1820.

Based on the procedures discuss above, the total arable land in typical years have been reconstructed as follows: 43,175 km^2 in 1661, 54,479 km^2 in 1820, 59,544 km^2 in 1887 and 56,864 km^2 in 1933 according to the study by Cao et al. (2013).

Population and human settlements

Population and human settlement data are used for validating reconstructed arable land patterns. We selected the population and spatial distribution of human settlements of Jiangsu Province in 1820 for validation, which was taken from the Harvard University CHGIS database (CHGIS, 2007). The prefectural boundaries in 1820 do not completely overlap with the present-day Jiangsu Province boundary. For minor differences, those boundaries along the Jiangsu borders were expanded or shrunk to the modern borders, yet the population of each boundary will remain unchanged. For those prefectures with a large area outside the modern borders, however, their population in 1820 was adjusted using the area-weighting method (Fig. 3).

Spatial factors (constraints in CA)

We selected the following spatial factors affecting the reconstruction of patterns of arable land while taking into account the natural climate, historical agricultural conditions, research scales and data availability in the study area. The flat plains of the Jiangsu Province study area allowed us to not consider slope as a factor. To facilitate comparison between factors, all factors needed to be standardized to either 0 to 1 or 0 and 1 (Fig. 4).

(1) Intensity of soil erosion: minor water erosion was set to 1, mild and moderate water erosion and erosion caused by human engineering were set to 0.
(2) Soil pH: pH from 6.5 to 7.5 were set to 1; and any pH values outside this range were set to 0.
(3) Content of soil organic matter: high soil organic matter content is favorable for farming. Organic matter content within the range of 0–13.99 was rescaled to 0–1.
(4) Distance to the nearest human settlement: based on the distribution of human settlements in 1820, the distance to the nearest human settlement of all cells within the study area was calculated and normalized to 0–1.
(5) Distance to the nearest water body: the distance to the nearest water body in 1820 of all cells within our study area was calculated and normalized to 0–1.

Methods

Model assumptions and conceptual model

A Historical Arable-land Reconstruction Model (HARM) was constructed based on a constrained CA. The basic assumptions of model are: (1) similarities exist between the historical and the contemporary arable land patterns; (2) the most unsuitable farming cells would be turned into non-arable land first (from contemporary to historical times); (3) cells surrounded by a high ratio of non-arable land would be turned into non-arable land first; (4) the range of historical arable land does not exceed the scope of contemporary arable land, in that it is impossible for land that was historically arable not be arable land in modern times (this works for the historical period before 1980 in Jiangsu, when major urbanization had not occurred in China)[1]; (5) land suitability related factors do not change over time as a result of data availability, a situation that other studies also faced.

The basic elements of a HARM based on constrained CA are as follows:

(1) Lattices: the entire Jiangsu Province;
(2) Cells: cell size was 1 km × 1 km, with a total of 103,395 cells in the study area;
(3) Cell states: $V = 1$ indicates arable lands; $V = 0$ indicates non-arable lands;
(4) Transition rules will be specifically addressed in the next section, using a multi-criteria evaluation (MCE) method;
(5) Neighborhoods: Moore neighborhood, 3 cells × 3 cells, a total of eight neighboring cells;
(6) Discrete time: one iteration in simulation equals one year in the real world;
(7) Constraints: the five aforementioned spatial factors: soil organic matter (SOM) content, intensity of soil erosion (EROSION), soil pH (PH), minimum distance to the nearest human settlement (DSETTLEMENT) and minimum distance to the nearest water body (DRIVER).

Status transition rules

The specific transformation rules for the HARM (Equation (1)) use *Land Amount* as the total number of arable land cells to-be-reduced during the year T_2 to the year T_1 ($T_2 > T_1$), and where $stepNum^t$ is the number of cells reduced in iteration t reflecting the arable land dynamics as the macro constraint, ij is the cell's coordinate, s_{ij}^t is arable land suitability for cell ij, w is the variable coefficient, p_g^t is the initial transition potential, $p_{g\,\max}^t$ is the maximum value of p_g^t across the entire lattices, α is the dispersion parameter ranging from 1 to 10 used to adjust arable land conversion speed, p^t is the final transition probability, *inStepID* is the sub-loop ID, V_{ij} is the cell status, $p_{\min}^t$ is the minimum of final transition probability in different sub-loops within each iteration, and its value is continuously updated in the sub-loop. The rules could guarantee there would be *stepNum* cells translated from arable lands to non-arable lands in each CA iteration, according to the arable land suitability calculated by spatial constraints and the neighborhood effect.

[1] In the over 300 years from the earlier period of Qing Dynasty to 1980, the total area of arable land in China increased by 160% and reached its peak in 1978–1980. The total area of arable land significantly decreased since 1980. For this, we regarded 1980 as our starting time point for construction. This has been indicated in the modified manuscript.

$$
\begin{aligned}
&1.\ LandAmount = \sum_t stepNum^t \\
&2.\ s_{ij}^t = x_0 + x_1{*}\mathrm{SOM}_{ij} + x_2{*}\mathrm{EROSION}_{ij} + x_3{*}\mathrm{PH}_{ij} \\
&\quad + x_4{*}\mathrm{DSETTLEMENT}_{ij} + x_5{*}\mathrm{DRIVER}_{ij} + x_N^{*}{*}neighbor_{ij}^t \\
&3.\ p_g^t = \frac{1}{1+e^{-s_{ij}^t}} \\
&4.\ p^t = \exp\left[\alpha{*}\left(\frac{p_g^t}{p_{g\,\max}^t} - 1\right)\right] \\
&5.\ \text{for } inStepID = 1 \text{ to } stepNum \\
&\text{if } p_{ij}^t = p_{\min}^t \text{ then } V_{ij}^{t+1} = 0 \\
&p_{ij}^t = 1 \\
&p_{\min}^t \text{ update} \\
&\text{next } inStepID
\end{aligned}
\tag{1}
$$

The above status transition rules are different from those proposed by Wu (2002): $p_c^t = p_g{*}con(s_{ij}^t = suitable){*}\Omega_{ij}^t$. In his status transition rules, p_g is the observed global potential which does not change across simulation iterations, *con* () converts the state of suitable land into a binary variable, and Ω_{ij} is a neighborhood evaluation function. The weight of neighborhood effect in Wu (2002) cannot be calibrated using historical observed datasets. According to Long, Mao, and Dang (2009), the simulation results are sensitive to the weighting parameter x_N of the neighborhood effect. In our paper, the weight of the neighborhood effect was calibrated (see Model calibration section).

Model calibration

The parameters needed to be calibrated in the HARM included $stepNum^t$, x_k ($k = 0$–5), and x_N, in which various approaches can be adopted. $stepNum^t$ is assumed to be constant through the entire simulation period, and can be calculated as follow:

$$
stepNum = \left|\frac{C_{T_2} - C_{T_1}}{T_2 - T_1}\right| \tag{2}
$$

where C_{T_2} and C_{T_1} are the total cell number of arable land in the year T_2 and T_1, respectively.

The weights x_k for spatial constraints can be retrieved by a logistic regression. For the dependent variable, among all none arable land cells in T_1, the expanded cells from the year T_1 to T_2(reduced from the year T_2 to T_1 as well) are set as 1, and other cells are set as 0. Independent variables correspond to the spatial factors, while considering that the dependent variable is 0 or 1, which is not normally distributed. Therefore, we used logistic regression to identify the coefficients (weights) of five spatial factors.

Keeping the identified x_k static, x_n can be calibrated using the *MonoLoop* method (for details see Long et al., 2009), with x_n continually sampled from 0 to $x_{n,\max}$ with an interval as $x_{n,\max}/M$. $x_{n,\max}$, and can be set based on the user's experience; M is set at 200 in this paper. The sampled x_n and the already calibrated weights x_k are used as the input variables for the constrained CA model. The simulated arable land patterns will be compared on a cell-by-cell basis with the observed arable land patterns to obtain the *Kappa* index. In the present paper, the *Kappa* index is calculated to analyze the degree of similarity (goodness-of-fit) between the observed and reconstructed arable land patterns for each cell. The x_n with the

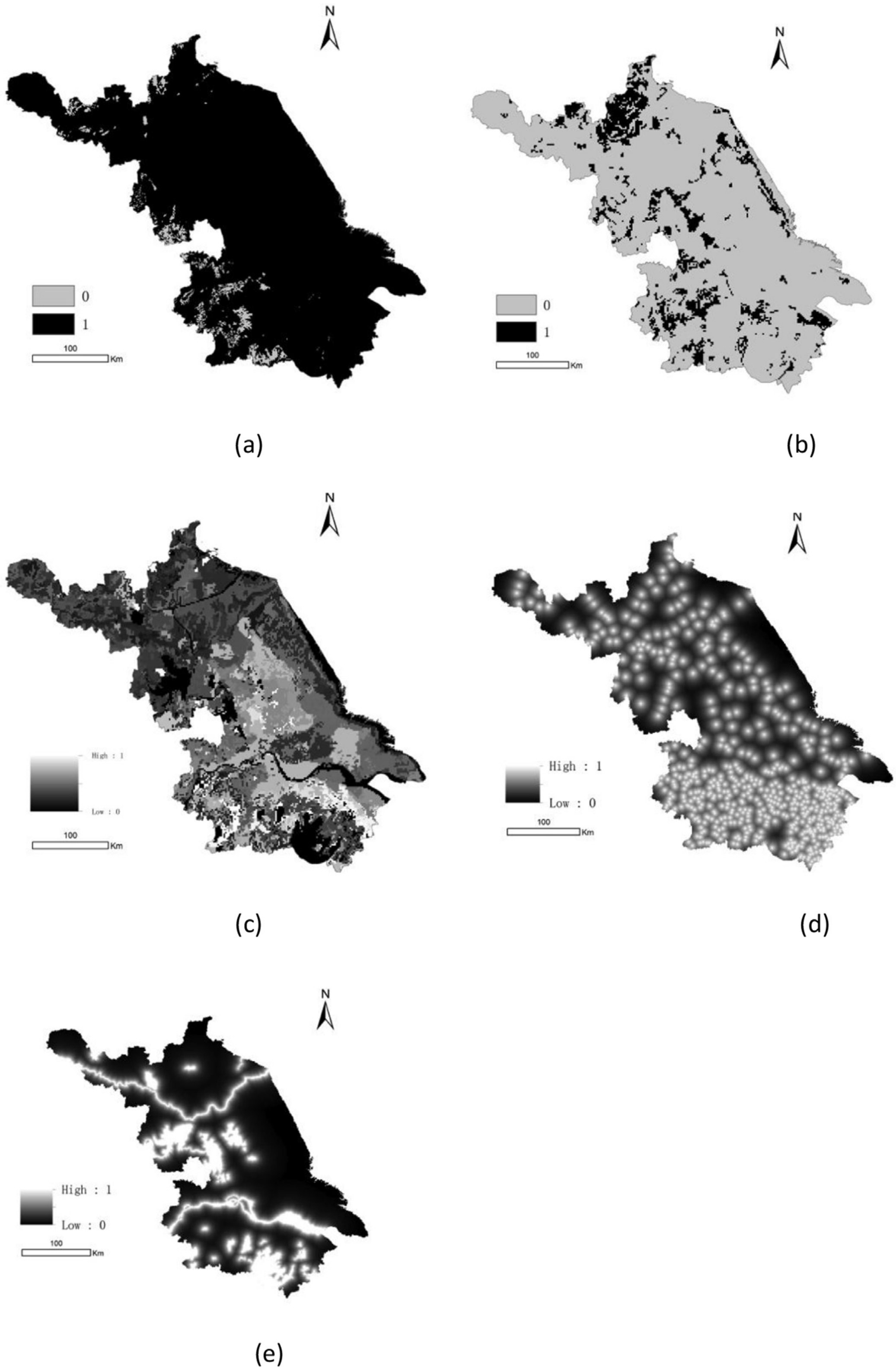


Fig. 4. Spatial factors: soil erosion (a), soil pH (b), soil organic matter (c), distance to settlements (d) and distance to water body (e). Note that Fig. a, b, and c are based on Shi, Yu, and Gao (2007), and Fig. d and e are based on CHGIS (2007).

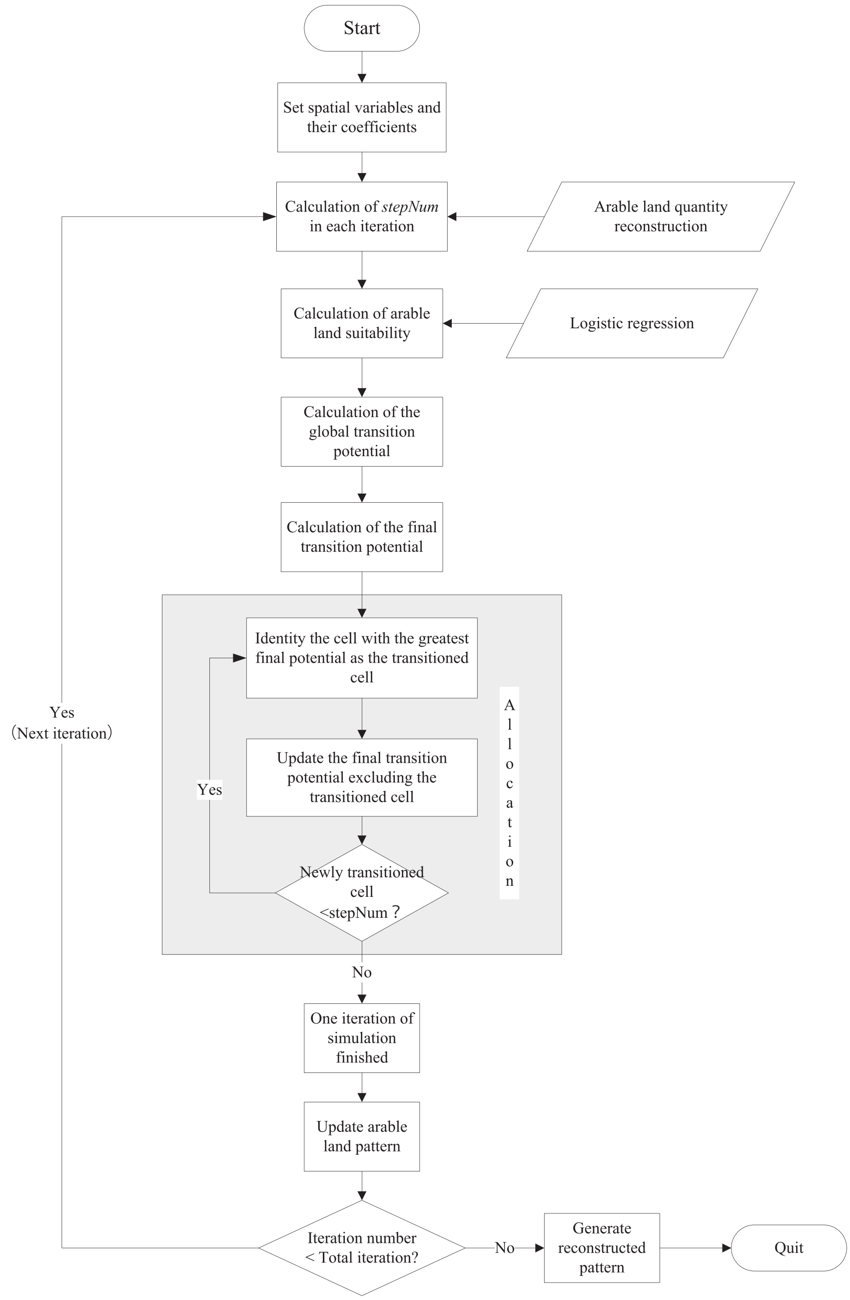


Fig. 5. Flow chart of the HARM.

maximum *Kappa* index will be specified as the final weight for the neighborhood effect.

Finally, $stepNum^t$, x_N^* with the best matching degree and x_0–x_5 obtained by logistic regression would be put together into status transition rules of the HARM to reconstruct patterns of historical arable land use.

Reconstruction of patterns of historical arable land use

The simulation procedures using the HARM were used to reconstruct the historical arable land patterns based on the established status transition rules (Fig. 5). First, we set the exogenous variables like total area of farmland in various historical points in time, spatial constraints and their corresponding coefficients for the model. Then, based on the macro socio-economic conditions, *stepNum* parameters were calculated for different periods of time, and also calculated the initial, global, and final probability of cell transition in the CA environment. Next, cells to be converted were iteratively selected in the allocation process in each iteration. In the last step, based on the target time of the simulation, we determined the number of iterations, completed the entire simulation process after several allocations, and obtained the reconstructed the patterns of arable land over a historical period.

In some cases, the total area of arable land for each sub-region within the entire study area is obtainable. In this condition, the HARM can be extended to the partition HARM. For example, each subregion works as a separate sub-model. When the sub-model runs into the sub-region's total arable land, it stops running. To ensure that the model runs to the end, the total arable land in each sub-region is consistent. The idea of partition simulation has also been reflected in the traditional spatial allocation method; this idea is expected to further enhance the pattern reconstruction precision by the HARM.

Model validation

Model validation is a key procedure that is used to test the applicability of the HARM. For constrained CA models used for simulating urban expansion, the urban layout interpreted by remote sensing images generally could be used for model validation. The condition for validating the HARM, however, is different in that the spatial distribution of arable land in historical time phases is generally not obtainable because of the long period of time in which remote sensing did not even exist. Direct data on the layout of arable land was rare in the history of China, but settlement distribution and population of each state capital can typically be obtained. Several methods can be used to verify the simulation results using these data, qualitatively and quantitatively:

(1) Qualitative validation: compare more small, high, low and other descriptive text in the ancient history books with the reconstruction results in the corresponding region at the sub-region level, while looking for trends. If the trends were basically the same, then the reconstruction results were better;
(2) Semi-qualitative validation: if a population dataset is rich in a sub-study area, then directly compare the number of different levels of settlements in the corresponding sub-study area with the reconstructed total arable land. If they are basically the same, then the reconstruction results were better;
(3) Quantitative validation (our recommended approach and used in validating HARM in this paper): if specific information is available on the number of households, perform correlation analysis between the number of households or population in the sub-region with total arable land. The higher the correlation coefficient is, the better the reconstruction. In addition, the patterns of known and reconstructed arable land could be compared in terms of average size, compactness, and scaling characteristics. If the underlying characteristics of model output follow the known arable land patterns, the model could be evaluated as being able of reconstructing the distribution of arable land in historical phases.

Results and discussion

Identification of model parameters

In this paper, the year 1980 was used as the base year for the HARM calibration. Table 1 shows the parameter calibration results that are established through the following steps.

(1) The macro parameter for the reconstruction for each typical year could be calibrated using the arable land in each typical year and that in the base year (1980) as well as for years between the base year and the typical year. Its values are 91 km^2 for simulation of the year 1661, 112 km^2 for 1820, 138 km^2 for 1887 and 329 km^2 for 1933.
(2) The weights of the five spatial constraints excluding x_n are calibrated using logistic regression in SPSS (see Table 1). All these variables are significant at the $P < 0.001$ level. Of all the factors, DSETTLEMENT, DRIVER and SOM are the three main factors for this model according to the weights regressed. The *Kappa* index between the observed arable land patterns in 1980 and the regressed patterns is 77.3%.
(3) x_n is identified as 11 using the *MonoLoop* method. The *Kappa* index between the observed and the simulated patterns using weights from both logistic regression and *MonoLoop* procedures is 81.3%. The significant increase of *Kappa* comparing with the one from logistic regression further indicates that the neighborhood effect plays an important role in arable land patterns. In addition, a *Kappa* value of greater than 80% also indicates our HARM can well replicate modern arable land patterns and provides a possible method of reconstructing historical arable land patterns.

Table 1
Calibrated parameters of HARM.

Name	Value
stepNum	91, 112, 138 and 329 for each typical year
x_0 (CONSTANT)	−1.370
x_1 (SOM)	2.734
x_2 (EROSION)	1.366
x_3 (PH)	−0.088
x_4 (DSETTLEMENT)	−5.186
x_5 (DRIVER)	−2.931
x_6 (Neighborhood effect)	11.000

Reconstructed arable land in typical years

Based on the model calibration results, we simulated arable land distribution at several historical stages using the HARM (not partition HARM here). Based on the total amount of arable land reconstruction (The total amount of arable land in typical years section), four typical years (1661, 1820, 1887 and 1933) were selected for the corresponding reconstruction of spatial layout of arable land (Fig. 6(a)–(d)).

Validation on the reconstructed results

Based on data availability, the reconstruction of arable land for 1820 was used for validation. We use CHGIS as the main data source for validation since the total population for each prefecture and the distribution of human settlements in 1820 are available in CHGIS (Population and human settlements section). We performed correlation analysis for "observed" population and amount of human settlements in each prefecture. The correlation coefficient is greater than 0.8, indicating the both datasets are significantly positively correlated. In this regard, we only used population of each prefecture for the further validation of the model. The validation results show that the correlation coefficient between the total arable land of each prefecture by the HARM and population of each prefecture in CHGIS was 0.12, denoting that these two were not correlated. Possible reasons may be as follows: (1) CHGIS datasets were created at a national scale, having limited accuracy at local levels such as Jiangsu Province in China; (2) The proportion of agricultural population (farmers) among all populations may vary across prefectures. This problem also applied for arable land quota per farmer. It would be difficult to translate total population into total arable land in each prefecture. When we inferred arable land using population in 1820 for prefectures, several prefectures were having total arable land significantly (2–3 times) greater than that in 1980, which was apparently not applicable. Given the above two reasons, it is difficult to use CHGIS to validate the results of the HARM reconstruction.

Therefore, the focus of validation in this article was on the comparison of the reconstruction results in 1820 by the HARM, and the partition HARM with the results of the conventional spatial

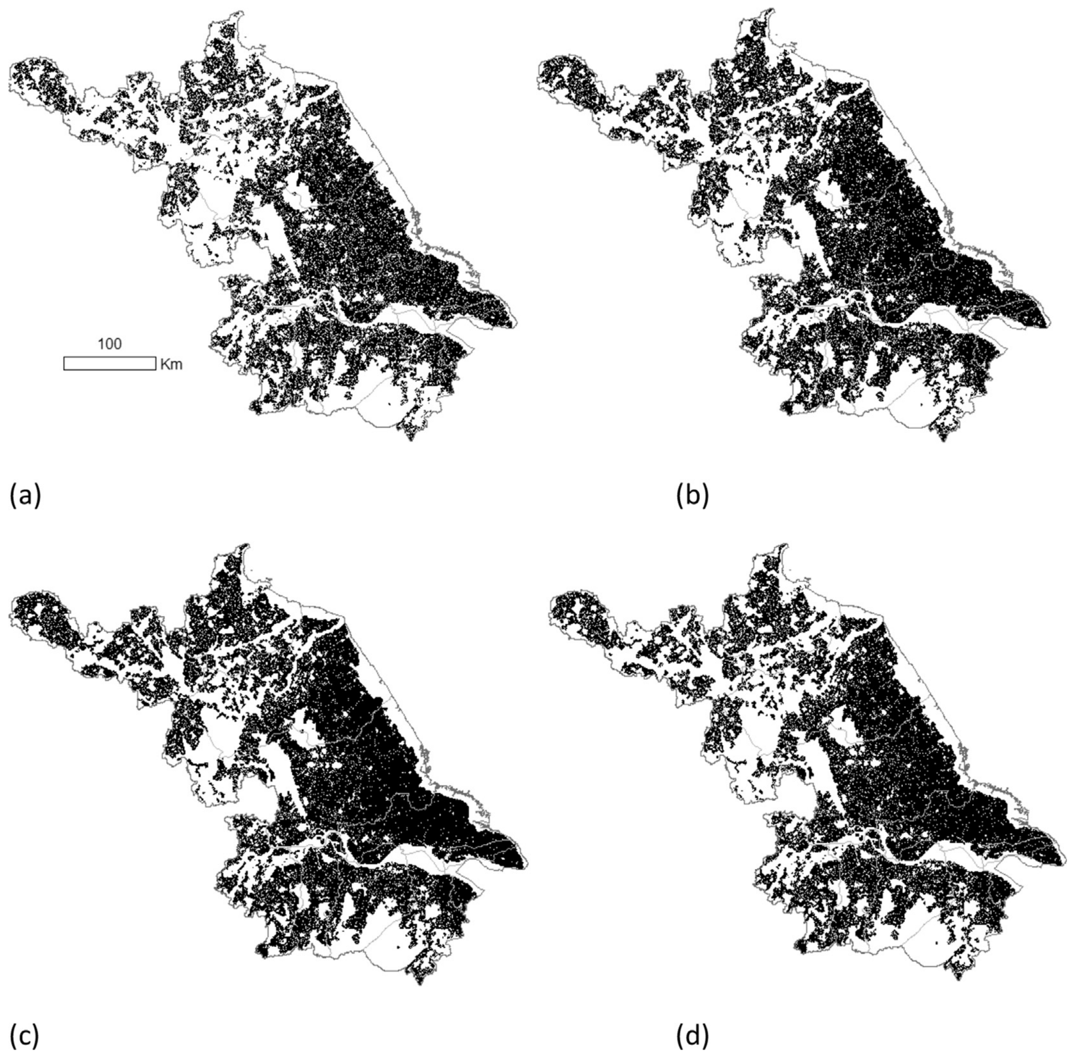


Fig. 6. Reconstructed arable lands in 1661 (a), 1820 (b), 1887 (c) and 1933 (d).

allocation model adopted by extensive studies as previously discussed in the introduction. Based on the calculated arable land suitability layer using calibrated coefficients for spatial factors in Identification of model parameters section, the reconstruction results obtained by the allocation model could be achieved by sorting suitability of cells and removing cells with lower suitability until reaching the total amount of arable land in 1820 (Fig. 7(a)). For using the partition HARM for reconstructing the arable land patterns in 1820 (Fig. 7(b)), the total amount of arable land in each prefecture could be inferred by assuming that the total amount of arable land is proportional to the population of each prefecture (available in CHGIS). Fig. 7(c) shows the results of reconstruction with the partition HARM.

From the comparison of results reconstructed by the three methods, clearly the arable land patterns obtained by our models, either HARM or the partition HARM, are more compact with fewer patches in contrast to the allocation model (Table 2), because the constrained CA considered the contiguous characteristics of arable land, while results from the allocation method are more decentralized. Compared to the HARM, the total amount of arable land in Huizhou by the partition HARM was small; this mainly occurred because Huizhou in the partition model is an independent model with reference to the results of the total allocation. Overall, the arable land reconstruction results by the partition HARM were much closer to the present-day arable land patterns in 1980, in terms of average compactness of all patches. In addition, the total arable land area in each prefecture was taken into account in the partition HARM. Considering these conditions, we recommend using the partition HARM to reconstruct historical arable land patterns.

In addition, we also used cluster analysis to measure the relationship between the patch size and frequency of various sizes of patches. On a logarithmic scale, this relationship should be linear. Hence, it can be used to validate cellular automata models (Vliet, White, & Dragicevic, 2009). The rank-size plots on the reconstruction results by the HARM and the partition HARM exhibit a significant rank-size pattern with large R-square values, indicating the applicability of our proposed two models (Fig. 8).

Conclusions

Concluding remarks

Based on related constrained CA research and existing research on the reconstruction of historical arable land, this study established a historical arable land reconstruction model (HARM) using constrained CA and applied the HARM to Jiangsu Province, China. We selected five constraints including soil pH, soil organic matter content, intensity of soil erosion, and distance to the nearest human

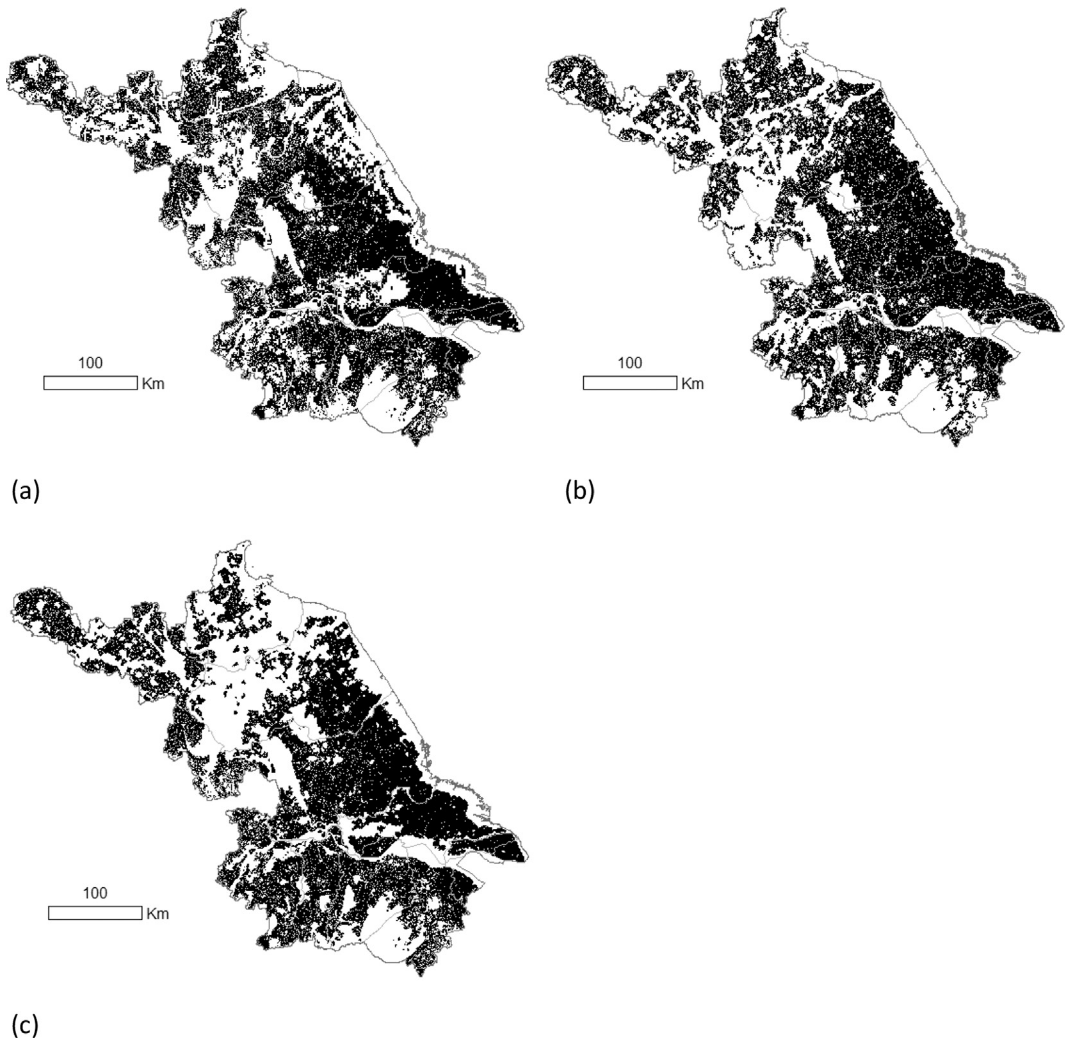


Fig. 7. Arable land reconstruction results for 1820 by: a: the allocation model; b: the HARM; c: the partition HARM.

settlements as well as distance to the nearest water body, and their relationships with the arable land distribution in 1980, as the transition rule of CA; these were quantitatively estimated using logistic regression. The reconstructed historical arable land patterns were compared with that created using the allocation method, with the HARM indicating a better result in terms of patch compactness. Compared to the conventional spatial allocation approach for reconstruction of spatial patterns of historical arable land, this study has the following characteristics:

(1) Borrowing ideas from urban expansion simulation, constrained CA have been initially applied for reconstructing historical arable land to analyze the contiguous development of arable land.
(2) Observed arable land expansion and several spatial factors were used to identify the objective transition rule of historical arable land while incorporating logistic regression, with the goal of avoiding the subjectivity in some existing studies.
(3) The constructed patterns can be dynamically visualized at ten year intervals.
(4) Compared to existing research, our patterns of reconstruction have high resolution (1 km grid) and each cell is associated with a unique land use (farmland or not), in contrast to the using the type of arable land proportion ongoing in each cell. Reconstruction results in other coarser scales could be aggregated based on the reconstructed 1-km patterns. Our high-resolution results, which considered local geography and historical situations in more detail, could be used to update data with low-resolution datasets such as globally available SAGE and HYDE.

Table 2
Comparison of arable land reconstruction results by various models.

Reconstruction result	Patch number	Average patch area in ha	Average compactness
Allocation	993	5477	0.51
HARM	451	12,007	0.57
Partition HARM	391	13,856	0.55
Arable land in 1980	1240	5826	0.54

Future work

While this explorative research has merit, several aspects deserve further research in the near future as follows.

(1) We are currently conducting research with the goal of delineating boundaries of the stable arable land across all of China. Different evolutionary rules are expected to exist for different regions that can be used by the HARM to achieve national scale reconstruction patterns of historical arable land.

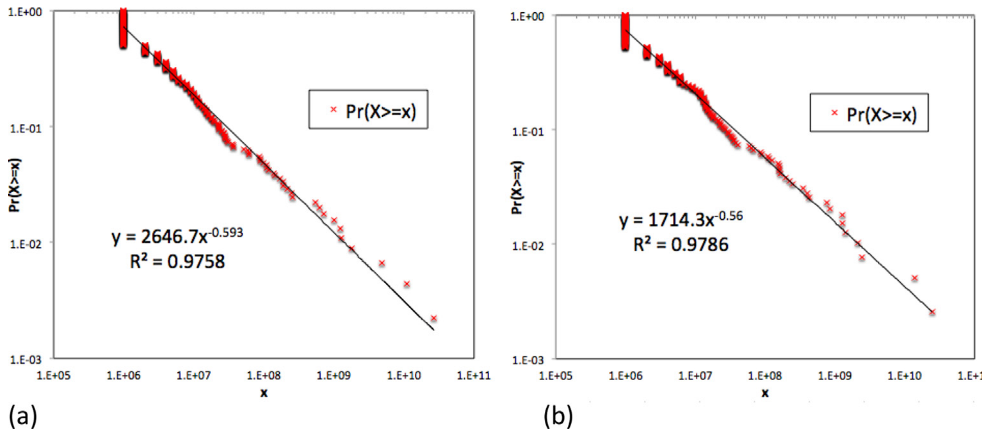


Fig. 8. Rank-size for the reconstruction results by the HARM (a) and the partition HARM (b). Note the unit of the x axis is m^2.

(2) Relevant policies/events regarding an increase or decrease of arable land at different stages in time could be introduced into the HARM, such as famine, large-scale reclamation, in order to obtain a more objective reconstruction pattern of arable land.
(3) In addition constrained CA could be used in historical reconstruction of arable land, and can also be used to predict or perform scenario analysis on the future patterns of arable land. The author anticipates carrying out such study, to establish past, present and future continuous arable land patterns.

Acknowledgments

We thank National Natural Science Foundation of China (No. 41340016) and National Basic Research Program of China (No. 2011CB952001) for their financial support of this research.

References

Cao, X., Jin, X. B., & Zhou, Y. K. (2013). Research on cropland data recovery and reconstruction in the Qing Dynasty: method and case study. *Acta Geographica Sinica, 68*(2), 245–256 (in Chinese).
CHGIS. (2007). Version 4. Cambridge: Harvard Yenching Institute.
Foley, J. A., Ruth, D., & Gregory, P. A. (2005). Global consequences of land use. *Science, 309*, 570–574.
Ge, Q. S., Dai, J. H., He, F. N., Pan, Y., & Wang, M. S. (2008). The study of change and carbon cycle of land use and land cover over the past 300 years. *Science in China (Series D: Earth Sciences), 38*(2), 197–210.
Goldewijk, K. K. (2001). Estimating global land use change over the past 300 years: the HYDE database. *Global Biogeochemical Cycles, 15*(2), 417–433.
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.
IGBP. (2001). Global change and earth system: a planet under pressure. *IGBP Science Series, 4*.
Kaplan, J. O., Krumhardt, K. M., & Zimmermann, N. (2009). The prehistoric and preindustrial deforestation of Europe. *Quaternary Science Reviews, 28*(27–28), 3016–3034.
Lagarias, A. (2012). Urban sprawl simulation linking macro-scale processes to micro-dynamics through cellular automata, an application in Thessaloniki, Greece. *Applied Geography, 34*, 146–160.
Lambin, E., Turner, B. L., Geist, H. J., Agbola, S. B., Angelsen, A., Bruce, J. W., et al. (2001). The causes of land-use and land-cover change: moving beyond the myths. *Global Environmental Change, 11*(4), 261–269.
Lambin, E., & Veldkamp, A. (2005). Key findings of LUCC on its research questions. *Global Change Newsletter, 63*, 12–14.
Li, K., He, F. N., & Zhang, X. Z. (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, S. C., He, F. N., & Chen, Y. S. (2012). Gridding reconstruction of cropland spatial patterns in Southwest China in the Qing Dynasty. *Progress in Geography, 31*(9), 1196–1203 (in Chinese).
Li, X., & Yeh, A. G. O. (2000). Modeling sustainable urban development by the integration of constrained cellular automata and GIS. *International Journal of Geographical Information Science, 14*(2), 131–152.
Liang, F. (2008). *Hukou, cropland and tax statistics for various dynasties of China*. Beijing: Zhonghua Book Company.
Lin, S. S., Zheng, J. Y., & He, F. N. (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*, GB3003. http://dx.doi.org/10.1029/2009GB003687.
Liu, Y. L., Liu, Y. F., & Xia, Z. F. (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, Y., Mao, Q., & Dang, A. (2009). Beijing urban development model: urban growth analysis and simulation. *Tsinghua Science and Technology, 14*(6), 787–794.
Long, Y., Shen, Z., & Mao, Q. (2012). Retrieving spatial policy parameters from alternative plans using constrained cellular automata and regionalized sensitivity analysis. *Environment and Planning B: Planning and Design, 39*(3), 586–604.
Moghadam, H. S., & Helbich, M. (2013). Spatiotemporal urbanization processes in the megacity of Mumbai, India: a Markov chains-cellular automata urban growth model. *Applied Geography, 40*, 140–149.
Pongratz, J., Reick, C., & Raddatz, T. (2008). A reconstruction of global agricultural areas and land cover for the last millennium. *Global Biogeochemical Cycles, 22*, GB3018. http://dx.doi.org/10.1029/2007GB003153.
Pu, K. (1933). *China land use statistics*. Beijing: The Commercial Press.
Ramankutty, N., & Foley, J. A. (1999). Estimating historical changes in global land cover: Croplands from 1700 to 1992. *Global Biogeochemcial Cycles, 13*(4), 997–1027.
Shi, X. Z., Yu, D. S., & Gao, P. (2007). Soil information system of China and its application. *Soil, 39*(3), 329–333 (in Chinese).
Statistics Bureau of the National Government. (1936). *Statistical analysis of chinese land issues*. Nanjing: Zhongzheng Publishing House.
Turner, B. L., Skole, D. L., Sanderson, S., Fischer, G., Fresco, L., & Leemans, R. (1995). *Land-use and land-cover change. Science/research plan*. Global change report.
Verburg, P. H., De Koning, G. H. J., Kok, K., Veldkamp, A., & Bouma, J. (1999). A spatial explicit allocation procedure for modelling the pattern of land use change based upon actual land use. *Ecological Modelling, 116*(1), 45–61.
Vliet, J. V., White, R., & Dragicevic, S. (2009). Modeling urban growth using a variable grid cellular automaton. *Computers, Environment and Urban Systems, 33*(1), 35–43.
Ward, D. P., Murray, A. T., & Phinn, S. R. (2003). Integrating spatial optimization and cellular automata for evaluating urban change. *The Annals of Regional Science, 37*(1), 131–148.
White, R., Straatman, B., & Engelen, G. (2004). Planning scenario visualization and assessment – a cellular automata based integrated spatial decision support system. In M. F. Goodchild, D. G. Janelle, & Z. G. Shrore (Eds.), *Spatially integrated social science* (pp. 420–442). Oxford University Press.
Wu, F. (2002). Calibration of stochastic cellular automata: the application to rural urban land conversions. *International Journal of Geographical Information Science, 16*, 795–818.
Ye, Y., Fang, X. Q., & Ren, Y. Y. (2009). The land use and land change cover of northern China over the past 300 years. *Science in China: Series D, 39*(3), 340–350.
Zhang, X. (1933). *China agriculture profile*. Nanjing: Jinling University (in Chinese).
Zhao, Y. (2005). *The driving force of land use change in the region of Su-Wan* (Doctoral dissertation). Shanghai: Fudan University (in Chinese).
Zhu, F., Cui, X. F., & Miu, L. J. (2012). China's spatially-explicit historical land-use data and its reconstruction methodology. *Progress in Geography, 31*(12), 1563–1573 (in Chinese).