J. Geogr. Sci. (2009) 19: 36-48
DOI: 10.1007/s11442-009-0036-x

Gridding cropland data reconstruction over the agricultural region of China in 1820

LIN Shanshan, ZHENG Jingyun, *HE Fanneng

Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China

Abstract: Recent studies have demonstrated the importance of LUCC change with climate and ecosystem simulation, but the result could only be determined precisely if a high-resolution underlying land cover map is used. While the efforts based satellites have provided a good baseline for present land cover, what the next advancement in the research about LUCC change required is the development of reconstruction of historical LUCC change, especially spatially-explicit historical dataset. Being different from other similar studies, this study is based on the analysis of historical land use patterns in the traditional cultivated region of China. Taking no account of the less important factors, altitude, slope and population patterns are selected as the major drivers of reclamation in ancient China, and used to design the HCGM (Historical Cropland Gridding Model, at a 60 km×60 km resolution), which is an empirical model for allocating the historical cropland inventory data spatially to grid cells in each political unit. Then we use this model to reconstruct cropland distribution of the study area in 1820, and verify the result by prefectural cropland data of 1820, which is from the historical documents. The statistical analyzing result shows that the model can simulate the patterns of the cropland distribution in the historical period in the traditional cultivated region efficiently.

Keywords: approach; gridding data; Chinese historical cropland records

1 Introduction

Land-use activities—whether converting natural landscapes for human use or changing management practices on human-dominated lands—have transformed a large proportion of the planet's land surface (Foley *et al*., 2005). Large scale land-cover change has an important impact on atmospheric composition, climate change, regional water circle, carbon circle, environmental quality, species diversity and the frangibility and productivity of terrestrial ecosystem etc. (IGBP, 2005). But the impact of land system on earth system is complicated and nonlinear, such as the simulation of environmental effect, especially the climate effect, of land-cover change remains very rough. So many critical problems are still unsolved, and

---

Received: 2008-04-14 **Accepted:** 2008-08-25
Foundation: National Natural Science Foundation of China, No.40471007; Innovation Knowledge Project of CAS, No.KZCX2-YW-315

Author: Lin Shanshan (1982–), Master, specialized in historical environmental change. E-mail: summer.lin33@gmail.com
***Corresponding author:** He Fanneng, E-mail: hefn@igsnrr.ac.cn

how to obtain land-cover data with high resolution and high precision is the basic solution to these problems. Though the remote sensing technology has made it easy to get the land-cover data, it can only provide information within decades, but the historical land-cover data are still very deficient.

China has a long cultivation history with numerous historical records available for historical LUCC research. Recently, efforts have been made to quantitatively reconstruct Chinese historical land-use, especially since the Qing Dynasty (Ge *et al.*, 2002; Ge *et al.*, 2004; He *et al.*, 2002; He *et al.*, 2008). However, in order to understand the influence of LUCC change on the terrestrial biosphere, we need geographically explicit data in order to account for the heterogeneity in climate, soil, vegetation cover, and croplands within political units. Hence we need to develop a method to "pixelize" the socioeconomic data (Ramankutty *et al.*, 1999). Ramankutty *et al.* (1999) and Goldewijk *et al.* (2001) have designed two such methods respectively, and reconstructed two datasets of global land-cover change over the last three centuries (at a 5 min resolution). But in view of the regional scale, the database they used and the gridding method are all too rough. And this study is an exploratory attempt to conduct a more suitable gridding approach to reconstruct cropland distribution in ancient China.

In this study, we first select and quantify the major drivers of the historical cropland distribution, and conduct an empirical model to pixelize the historical cropland data. In a case study, this model is used to reconstruct the pixelized cropland dataset based on provincial cropland records in 1820. We then statistically test the results with the prefectural records in 1820 to check the efficiency of the model. The gridding model conducted in this study can well introduce the numerous cropland data into the climate model or the other integrated global change models.

2 Study area

This study is conducted in the traditional cultivated region in China. Since we take the 25th year of Jiaqing Period (1820) as a study case, the administrative boundary in 1820 is used to demarcate the study area, approximately including 17 provinces but not the region to north of the Great Wall, Taiwan prefecture then (Figure 1).

3 Methods

3.1 General description of HCGM

Cropland data were mainly recorded in political units in ancient China, from statistic to grid maps, HCGM (historical cropland gridding model) is essentially a method to allocate the inventory data into grid cells in each political unit. HCGM is an empirical model, including a series of algorithms, and operates in geographic information system (GIS). It aims to be used in the climate models, so in order to correspond to the RCM (regional climate model) (Wei *et al.*, 2000), we set the model resolution at 60 km×60 km.

Figure 1 Location of the study area

3.2 Selecting the major reclamation drivers in ancient China

Generally, the cropland distribution is co-influenced by natural (topography, heat, water, soil, vegetation etc.) and social (population, economy, agricultural policy, war etc.) factors. However, the major drivers may be quite different in different regions or different space scales in the same region.

3.2.1 Natural drivers

Firstly, our study area is located in the monsoon region of eastern China. Though the heat and water condition is critical to agriculture, their influence is mainly on the agrotype, cropping system and agricultural disasters, not on the cropland distribution. Secondly, after hundreds/thousands of years of deforestation and cultivation, there is barely any natural vegetation left in our study area as the traditional cultivated region. The fertilizing is common, even a great deal of low-productive soil has been cultivated. Thus we can pick out the topography as the most important impact factor at first, which includes altitude, slope, relief and slope aspect. Considering our study resolution (60 km×60 km), the influence of the relief and the slope aspect is negligible. So the altitude and slope can be further screened out as the major natural driver of the cropland distribution.

3.2.2 Human drivers

Though the flat place is suitable for cultivation, the historical land development is not a shot-time job, and the social factors, such as population situation, economy, agricultural policy, war etc., played significant roles in such progresses. The relation between population situation and the cropland area and pattern was the tightest in these factors: in a certain region, the growth of population usually indicated the increase of cropland area; at the scale of the whole study area, the population migration brought the extension of cultivated region. The cropland area depended mainly on the quantity of labor force in ancient China. Though the productivity level could improve the efficiency and the output, since ironware began to be agricultural tools around 2500 years ago, the technology had been barely improved. The increasing need for grain could only rely on the extension of cropland and the increase of labor force. So in the certain period, the change of population could reflect the process of cultivation very well. Policy and war could also influence the quantity and distribution of cropland besides population, but their impact firstly on the laborers, and then caused the cropland abandoned or extended. So this kind of impact had actually reflected on the change of population.

Considering the population data, there are numerous records of population and cropland area, which are the most important data source for current researcher to study the ancient Chinese society. However, the research of historical cropland was much less systematic and elaborate compared with population history. Therefore, population distribution as the major driver of the distribution of historical cropland is representative and feasible.

3.3 Model inputs

HCGM requires the following input variables: historical cropland inventory data and population density, the inventory unit needs not be determined, but the population resolution should be higher than the cropland data. Besides them, grid terrestrial proportion is also required.

3.4 Model process

HCGM has two procedures (Figure 2): TOPOGRAPHY, which simulates the impact of natural drivers on the cropland distribution; ALLOCATION, which is a series of algorithms to combine the impact of natural and human drivers, and allocates the inventory cropland data into the grid cells.

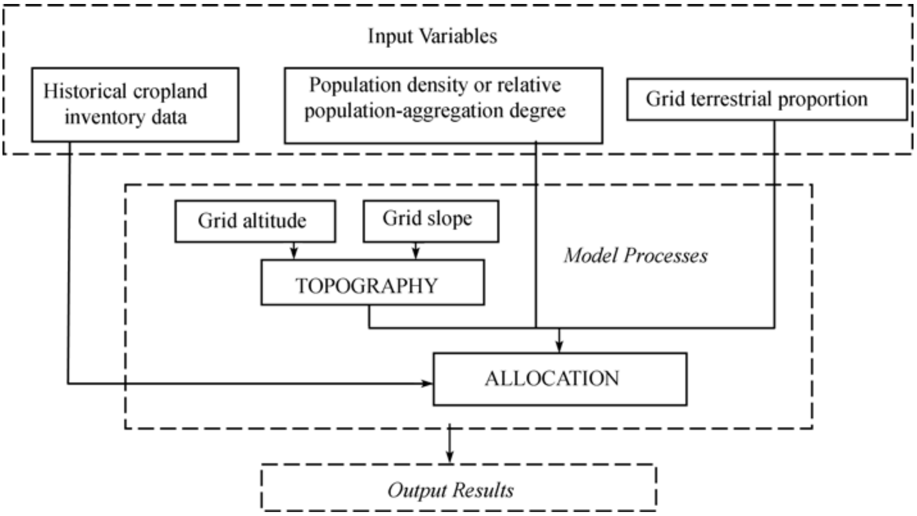

Figure 2 Structure of HCGM

3.4.1 TOPOGRAPHY

In order to quantify the impact of altitude and slope on cropland distribution, we use contemporaneous data to analyze the relation between altitude/slope and the cropland distribution in the procedure TOPOGRAPHY.

Firstly, in view of altitude, it is mostly below 2000 m in the study area. Most of China's cropland is concentrated in places to the east of the Taihang Mountain, Wushan Mountain and Xuefeng Mountain, where the altitude is mostly below 500 m. There are different reclamation patterns at different altitudes, including the proportion of cropland, the farming system and the multi-cropping condition etc., for the reason of the vertical difference of heat and water condition. But within different ranges of altitude, the significance of such variations is quite different. The altitude can only impact cropland distribution when it reaches certain height at which the growth of crop begins to be limited by the water and heat condition.

Secondly, referring to surface slope, the difference of slope has large different statues of the water and soil loss of the cropland, so it can impact the pattern of agricultural land-use and the situation of reclamation.

Referring to the classification of altitude and slope in "*Agricultural Land Use in China*" (Table 1) (Sun *et al*., 2003), we work out the average altitude $D(i)$ and average slope $S(i)$ (at a 60 km×60 km resolution) based on 1 km DEM database[1] and 1 km slope database[2] and the reclamation ratio (at a 60 km×60 km resolution) based on 1 km land-use database[3].

Table 1 The classification of altitude and slope of the traditional cultivated region in China

Classification of altitude (m)	Defined altitude in calculation (m)	Classification of slope (°)	Defined slope in calculation (°)
≤100	100	≤2	2
100–250	250	2–6	6
250–500	500	6–15	15
500–750	750	15–25	25
750–1000	1000	>25	45
1000–1500	1500		
1500–2000	2000		
2000–3000	3000		
>3000	4000		

The result shows that the cropland in the area above 3500 m mainly situated in the west Sichuan and north Yunnan occupies less than 0.1% of the whole cropland in the study area, and the reclamation ratio is below 5%, thus the altitude of 3500 m can be regarded as the upper limit of the altitude range of cropland distribution.

The correlation analysis (Table 2) shows that there is significant negative correlation between altitude/slope and reclamation ratio (the correlation coefficient is –0.488/–0.694) in the area below 3500 m. By further analysis we can find that there is a high coherence be-

---

[1] Data source: the global 1km DEM, from www.geodata.cn

[2] The 1 km slope data is calculated according to 1:250,000 DEM database.

[3] The 1 km land-use data is calculated according to national 1 km land-use database in the 1980s from www.geodata.net, which is derived from the 1:100,000 land-use database in the 1980s

Table 2 The correlations between altitude $D(i)$, slope $S(i)$ or reclamation ratio $C(i)$ of the traditional cultivated region in China

	$D(i)$–$C(i)$	$S(i)$–$C(i)$	$D(i)$–$S(i)$
Correlations coefficient	–0.488	–0.694	0.736
Sig. (2-tailed)	0.000	0.000	0.000
Partial correlations coefficient while $S(i)/D(i)$ is controlled	0.046	–0.567	
Sig. (2-tailed)	0.180	0.000	

tween the altitude and the slope (the correlation coefficient is 0.736): North China Plain, the middle and lower reaches of the Yangtze River and Sichuan Basin are the lowest and the flattest, while the southern hilly area and the northern mountainous area are relatively higher and steeper, and the highest part is situated in the Yunnan–Guizhou Plateau which is also the steepest. Thus partial correlation analysis is required to reveal the relationship between altitude/slope and reclamation ratio more precisely. The result shows that while altitude $D(i)$ is under control, slope $S(i)$ and reclamation ratio $C(i)$ are still significant in negative correlation (the partial correlation coefficient is –0.567), but while slope $S(i)$ is under control, the partial correlation coefficient between altitude $D(i)$ and reclamation ratio $C(i)$ is only 0.046 (the 2-tailed sig. is 0.180), so the impact of altitude on cropland distribution can be considered insignificant. Thus slope $S(i)$ can be selected to present the impact of natural driver on cropland distribution in the area below 3500 m.

The study area as a traditional cultivated region has been reclaimed maximized under the huge pressure of the population, so the contemporaneous distribution of cropland can be considered as the potential cultivated region before the agricultural exploitation. We can quantify the impact of the slope on cropland distribution very well by fitting the relationship between contemporaneous grid reclamation ratio and slope.

Firstly, the reclamation ratio $C(i)$ is standardized to facilitate the comparison:

$$\alpha(i)=C(i)/Max(C(i)) \tag{1}$$

In formula (1), $\alpha(i)$ represents the cultivability, and $Max(C(i))$ represents the maximum of all the reclamation ratio. The larger $\alpha(i)$ the higher is cultivability the grid, whose range is [0, 1]. We find that $\alpha(i)$ changes with $S(i)$ as obvious inverse regression by curve fitting (Figure 3), and the fitting formula is :

$$\hat{\alpha}(i)=0.047+1.504/S(i) \tag{2}$$

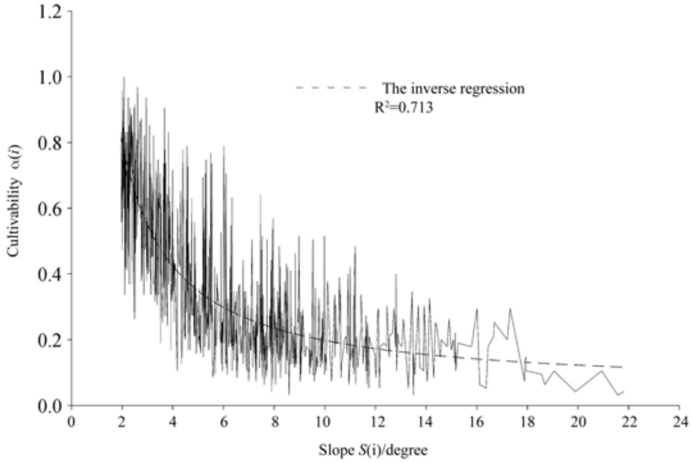

Figure 3 The inverse regression of cultivability ($\alpha(i)$) and slope ($S(i)$) of the traditional cultivated region in China

In formula (2), in order to be discriminated from the real $\alpha(i)$, $\widehat{\alpha}(i)$ is used to represent the fitted value inverse slope $S(i)$, which means that the slower the slope the more suitable to be reclaimed. Referring to the historical agricultural development, such area has been often prior to be reclaimed. As the value of $\widehat{\alpha}(i)$ is a reflection of the gravitation of the reclamation and influenced by only topography factor (slope), it can be called as topography-gravitation for reclamation.

3.4.2 ALLOCATION

We use some simple algorithms to allocate the inventory cropland data in procedure ALLOCATION (Figure 4). Firstly, we assume that the spatial resolution of population dataset is at grade *h* (such as prefecture) while cropland dataset at grade *k* (such as province). The pattern of population density can well indicate the variation of cultivation in different areas when the spatial resolution of population dataset is higher than cropland dataset. But the increase in population density will stop bringing obvious increase of cropland after the cultivation reaching a certain degree, so there is some limitation to reflect the situation of cultivation by the population density. For example, the population density in the North China Plain is lower than the middle and lower reaches of the Yangtze River, but not the reclamation ratio. In order to eliminate the influence of such limitation, the comparison

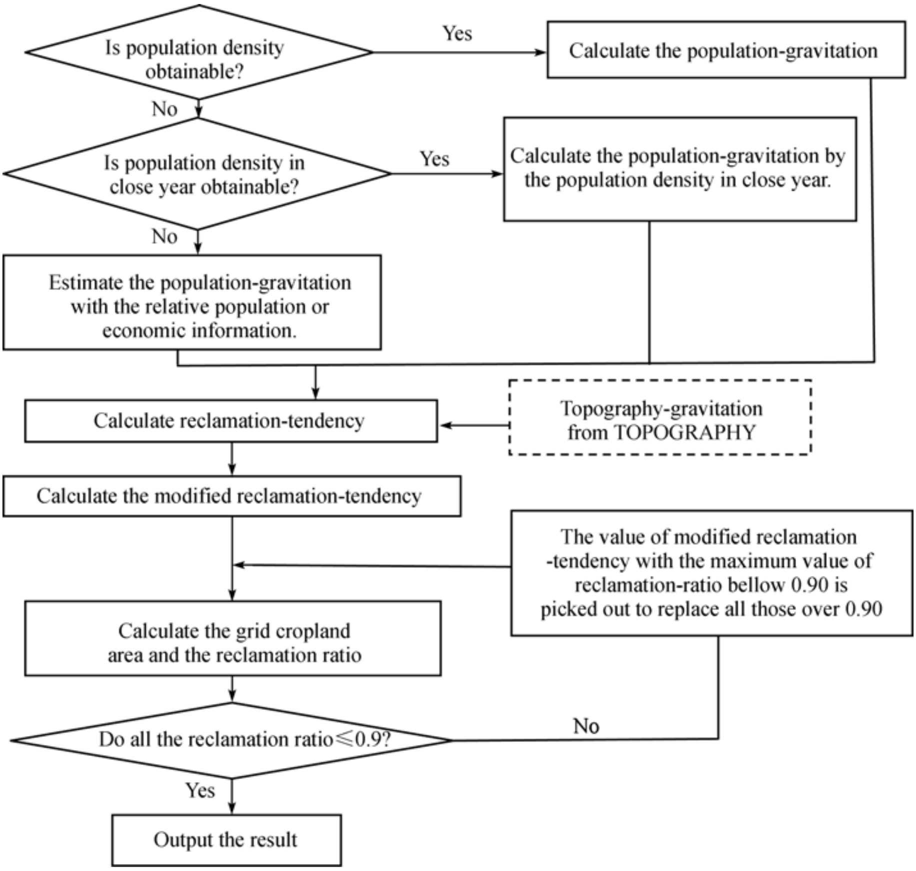

Figure 4 Structure of procedure ALLOCATION

needs to be downscale from the whole study area to the administrative region k, where the population density and reclamation ratio of the administrative region h are much more positively correlated. The population density of region h_m in the year t_v is assumed as $P(h_m,t_v)$, and standardized as:

$$\beta(h_m,k_n,t_v)=P(h_m,k_n,t_v)/Max(P(h_m,k_n,t_v)) \tag{3}$$

In formula (3), $\beta(h_m,k_n,t_v)$ represents the relative population-aggregation degree of region h_m, belonging to the region k_n in the year t_v, and the range of the value is [0, 1], $Max(P(h_m,k_n,t_v)$ represents the maximum value of $P(h_m,t_v)$ within the region k_n. So the value of β can show the pattern of population distribution, thus indicating the intension of cultivation positive-relatively only within the same region k_n. As the value of $\beta(i)$ is a reflection of the gravitation of the population, it can be called as population-gravitation.

Though there are countless records of population in the Chinese historical documents, the population data had been often statistical-defined differently and intermingled due to the deficiency of population census data or governmental statistical report or others just like in modern times. So it is difficult to obtain the integrated and precise historical population density (or quantity) data in many years in the historical period. But the requirement of our gridding model on the information of population is not the absolute population density but the relative population-aggregation degree, population-gravitation β within a certain region according to the formula (3), so we can estimate β in those years without proper population data by the following approach.

(1) The year population data in close period is obtainable: t_2 is assumed as the studied year without population data, while t_1 the close year with proper population data. Generally, the population-distribution pattern is stable in a certain administrative region within a short period except for war, epidemic or great natural disaster. Thus the population-gravita- tion in the year t_2 can be calculated with the population data in year the t_1 directly:

$$\beta(h_m,k_n,t_2)\approx\beta(h_m,k_n,t_1)=P(h_m,k_n,t_1)/Max(P(h_m,k_n,t_v)) \tag{4}$$

(2) The year population data in close period is not obtainable: population-gravitation $\beta(h_m,k_n,t_v)$ can be estimated directly based on the situation of population distribution or economic situation in such case.

Reclamation-tendency is used to represent the integrative impact of natural driver (topography-gravitation) and the social driver (population-gravitation):

$$\delta(i,h_m,k_n,t_v)=\hat{\alpha}(i)\bullet\beta(h_m,k_n,t_v) \tag{5}$$

In formula (5), $\delta(i,h_m,k_n,t_v)$ represents reclamation-tendency, while $\hat{\alpha}(i)$ and $\beta(h_m,k_n,t_v)$ represent topography-gravitation and population-gravitation respectively. According to formulas (2), (3) and (5), $\delta(i,h_m,k_n,t_v)$ is negatively related to the slope and positively to the relative population-aggregation degree, which means that the flatter the topography and denser the population, the greater the reclamation-tendency. Because $\hat{\alpha}(i)$ is constant with time, the value of δ is controlled only by $\beta(h_m,k_n,t_v)$.

We must notice that the water area is not taken into account in topography-gravitation derived from formula (5). Though most rivers and lakes occupy negligible part of the land surface at a 60 km×60 km resolution in our study area, the impact of some large lakes and oceans (the grids on the coastline) cannot be ignored. Besides, the grids on the national boundary are also not up to 3600 km^2. Thus formula (5) needs to be modified: the real terrestrial proportion of the grid is assumed as $\varepsilon(i,\ t_v)$, and the modified reclamation-tendency as $\delta'(i,h_m,k_n,t_v)$:

$$\delta'(i,h_m,k_n,t_v)=\delta(i,h_m,k_n,t_v)\bullet\varepsilon(i,t_v)=\hat{\alpha}(i)\bullet\beta(h_m,k_n,t_v)\bullet\varepsilon(i,t_v) \tag{6}$$

Because the cropland area $A(k_n,t_v)$ of the region k_n is known, in order to work out the grid cropland area $X(k_n,t_v)$, the key point is to derive the proportion, $\theta(i,h_m,k_n,t_v)$, of every grid in the region k_n it belonged. $\theta(i,h_m,k_n,t_v)$ is driven by the modified reclamation-tendency $\delta'(i,h_m,k_n,t_v)$, the larger-valued $\delta'(i,h_m,k_n,t_v)$ the greater proportion within the same region k_n, so $\theta(i,h_m,k_n,t_v)$ can be obtained by the method of calculating the proportion coefficient directly:

$$\theta(i,h_m,k_n,t_v)=\delta'(i,h_m,k_n,t_v)/\sum_i \delta'(i,h_m,k_n,t_v) \tag{7}$$

According to $\theta(i,h_m,k_n,t_v)$, cropland area $X(h_m,k_n,t_v)$ and reclamation ratio $D(i,t_v)$ can be easily worked out:

$$X(i,k_n,t_v)=\theta(i,k_n,t_v)\bullet A(k_n,t_v) \tag{8}$$

$$D(k_n,t_v)=X(i,k_n,t_v)\bullet area(i) \tag{9}$$

In formula (9), $area(i)$ represents the area of grid, which is 3600 km^2.

Reclamation ratio should range from 0 to 1 theoretically, but it will not reach 100% practically. According to the cropland distribution in the 1980s, the maximum value of reclamation ratio $D(i)$ is 0.95, and more than 99.5% is under 0.90. Because cropland distribution in modern times is similar to one under saturated cultivation, the value of reclamation ratio in ancient China can be considered no larger than 0.90.

Since the increase of population density would stop bringing obvious increase of cropland after the cultivation reaching a certain degree, the reclamation ratio $D(i)$ may be overvalued in the extremely densely-populated area according to the above-mentioned method, so the result should be further modified by the following method:

(1) 0.90 is set as the upper limitation of reclamation ratio $D(i)$.

(2) The value of reclamation ratio $D(i)$ is controlled by the modified reclamation-tendency $\delta'(i,h_m,k_n,t_v)$ according to the formulas (6)–(9), so it can be modified by inducing the value of the latter in the grids over 0.90. The specific approach is followed as: the maximum value of the reclamation-ratio $D(i)$ above 0.90 is picked out to replace all the value over 0.90, then $X(i,k_n,t_v)$ and $D(k_n,t_v)$ are recalculated, and the process is looped until all the value of $D(k_n,t_v)$ is under 0.90.

4 Case study and model test

We apply HCGM in the 25th year of Jiaqing period in the Qing Dynasty (1820), in which the records is relatively integrated, and then compare the simulated result with the documental prefectural cropland data to statistically test the model efficiency.

4.1 Reconstructing cropland distribution in 1820

The data of cropland and population are derived from Ge's (2004) cropland dataset of the last three centuries and Cao's (Table 3) (2001) prefectural population dataset from 1776 to 1953, and the documental prefectural cropland data is derived from *Jiaqing Chongxiu Yitong Zhi* (the revised edition of *Comprehensive Geography in Jiaqing Period*), which is

Table 3 Cropland area in the traditional cultivated region of China in 1820 (Ge *et al*., 2004)

Province	Cropland area in Qing Dynasty (hm^2)	Province	Cropland area in Qing Dynasty (hm^2)	Province	Cropland area in Qing Dynasty (hm^2)
Zhili	5966908	Jiangsu	5380440	Fujian	1354320
Shanxi	3632080	Zhejiang	2003120	Jiangxi	2702640
Shaanxi	4536560	Hunan	3331010	Sichuan	5500820
Henan	9714830	Hubei	3825910	Guizhou	1885960
Shandong	6652990	Guangxi	3267910	Yunnan	2469200
Anhui	4669210	Guangdong	2569300		

Annotation: Because of the restriction of the study area, Taiwan prefecture is not include in Fujian province, Chengde prefecture and the three prefectures of Koubei are not included in Zhili province. So the cropland data of these areas should be deducted from the relevant provincial data in Ge's dataset. The cropland area of the northern part of Zhili and Shanxi provinces is derived from *Land use/cover change in Rehe-Chahar-Suiyuan area in last 300 years*[1], and the cropland area of Fujian is not re-calculated due to the absence of exact information of Taiwan prefecture.

secondarily derived from *Statistics of Registered Permanent Residence*, *Plowlands and Land Taxes in Past Dynasties in China* (Liang, 1980). The resolution of the prefectural population data is higher than the provincial cropland data, which satisfy the condition of the model application. Besides, both of the population and cropland data are based on *Jiaqing Chongxiu Yitong Zhi*, which avoids too large errors due to the difference of statistical definition.

The grid cropland area and reclamation ratio in 1820 are calculated by the Chinese historical cropland data gridding model (Figure 5), some relevant problems are solved by the following methods.

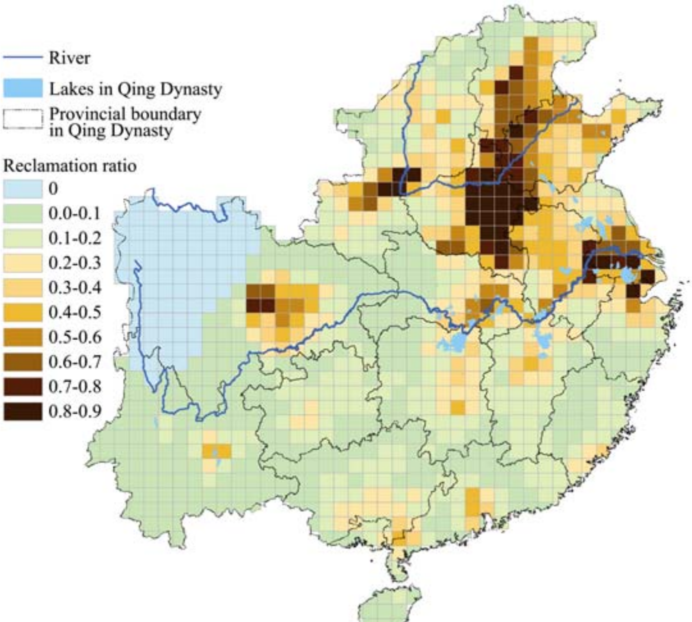

Figure 5 Reclamation ratio of the traditional cultivated region of China in 1820 (at a resolution of 60 km×60 km)

---

[1] Tian Yanyu. Beijing: Institute of Geographic Sciences and Natural Resources Research, CAS. 2004. (master degree thesis)

(1) In this case, h_m represents the prefectural region while k_n represents the provincial region.

(2) The provincial cropland data and prefectural population data in 1820 are loaded on the administrative map of 1820, which then overlaid with the grid map at a 60 km×60 km. The administrative borderline is processed according to the following principles because of the unmatched boundary of the administrative region and the grid: ① the grid shared by several provinces is ascribed to the province with the largest part of the grid. ② the population-gravitation $\beta(h_m,k_n,t_v)$ of a grid that shared by several prefectures within a certain province is calculated by area-weighted method.

(3) The real terrestrial proportion $\varepsilon(i, t_v)$ of the grid on the boundary of the study area, including the national boundary and the coastline, is calculated directly by GIS software. Because the change of rivers from the middle Qing Dynasty to modern times is negligible, the area of river surface can be approximately calculated by the contemporaneous river-distribution map. The surface area of some lakes that has significantly changed is calculated by overlaying the lake-distribution map with the grid map.

4.2 Model test

The simulated reclamation ratio cannot be compared with the documental data of *Jiaqing Chongxiu Yitong Zhi* since the cropland data has been modified in Ge's dataset. But the modification aimed to join the data of the Qing Dynasty, the Republic of China and modern times, thus it only occurred at the provincial scale. So though the provincial cropland data differs from the Ge's dataset to the documental data in *Jiaqing Chongxiu Yitong Zhi*, the proportion of a certain subject prefecture should be close between the simulated and recorded result. We can evaluate the efficiency of the model by testing this hypothesis. Though the cropland data in the Chinese historical document has been widely questioned, for example, the opinion like "the cropland data of the Ming and Qing dynasties was actually the unit of tax" has been commonly accepted (He, 1988; Zhou, 2001), the distortion was similar within a certain province, so the documental data can reflect the relative distribution of cropland factually and is feasible to test the simulation result.

$\hat{F}(h_m)$ and $F(h_m)$ are assumed to represent the simulated and documental value of the prefectural cropland area proportion to the provincial cropland area, and we can estimate whether there is significant difference between them by paired sampled T test .

After normal transforms, the difference value $E(h_m)$ between $\hat{F}(h_m)$ and $F(h_m)$, also the absolute error subjected to normal distribution. The null hypothesis of the paired T test of $\hat{F}(h_m)$ and $F(h_m)$ is that there is no significant difference between the population distribu-

Table 4 Paired sample T test between $F'(h_m)$ and $\hat{F}(h_m)$ after transformation to normality

Paired differences					t	df	Sig. (2-tailed)
Mean (%)	Std. deviation	Std. error mean	95% confidence interval of the difference				
			lower	upper			
–0.043	0.673	–0.043	–0.013	0.041	–1.008	246	0.314

tions of $\widehat{F}(h_m)$ and $F(h_m)$. According to the testing result (Table 4), the t statistics is −1.008, and the two-tailed significance is 0.314, larger than the significant level 0.05, thus the null hypothesis cannot be refused, which means that the difference between $\widehat{F}(h_m)$ and $F(h_m)$ is not statistically significant.

The above-mentioned paired T test is actually the test on whether the absolute error is 0, in order to analyze the error resources better, the analysis on relative error is also required due to the large variation of $F(h_m)$ (the variation coefficient is 0.87):

$$E^*(h_m)=(\widehat{F}(h_m)-F(h_m))/F(h_m) \tag{10}$$

According to Figure 6, the prefectures with larger relative error are mainly low-cultivated, which are mostly situated in Shaanxi, Shanxi, Sichuan, Yunnan and Guangxi provinces. The impact factors in these regions are more complicated. For example, the west Shaanxi was much less suitable for reclamation because of drought, and Guangxi and Yunnan were mainly inhabited by minority, which can not only influence the pattern of land use, but also the distortion extent of the cropland data in historical documents. Taking the Guangnan prefecture in Yunnan province, which has a large relative error of 61, as an example, the cropland area recorded in *Jiaqing Chongxiu Yitong Zhi* was only 6142 mu, occupying 0.06% of the whole cropland area of Yunnan province, but no population data record because of the minority inhabitant, which can also indicate that the documental cropland data should be much lower than the actual one. According to the region that $F(h_m)$ is bigger than 2%, the range and the mean of $E^*(h_m)$ are reduced to [–0.83, 3.22] and 0.059 separately. Such errors are acceptable to estimate the distribution of historical cropland.

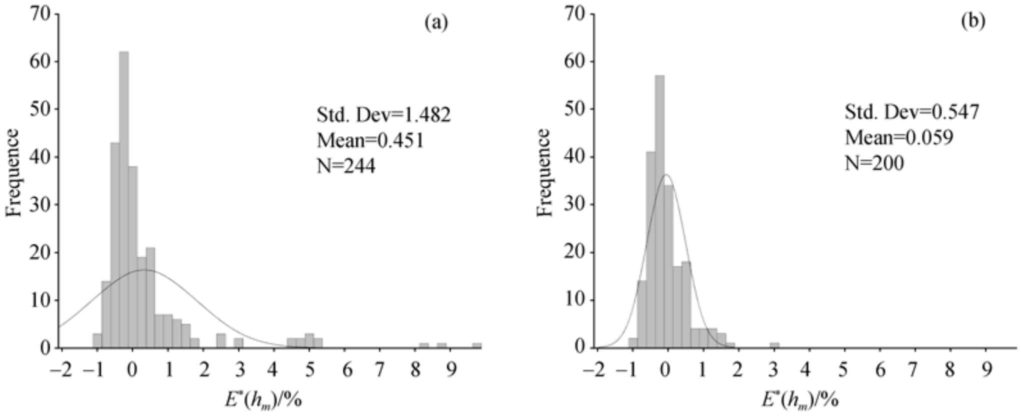

Figure 6 Histogram of relative error (a) all prefectures *(b) the prefectures of $F(h_m)$>2%

*The three values exceeding 10 are not included in order to facilitate the graphic display.

5 Conclusions and discussion

Historical cropland data gridding model aims to transfer the statistical historical cropland records to the spatial-explicit dataset, which is important to the simulation of historical land-cover change and climate change. The model designed especially for the Chinese traditional cultivated region in this paper is much more precise than the two existing datasets of global land-cover change.

Although the impact factors of the cropland distribution in the Chinese traditional cultivated region are multiplex, the effects vary in different spatio-temporal scales remarkably. Slope and population distribution are selected to represent the natural and social driver separately in this paper, which can not only simplify the model, but also can guarantee the efficiency of the simulation at a 60 km×60 km resolution.

The historical cropland data gridding model of the Chinese traditional cultivated region shows that the grid reclamation ratio is negatively related to the slope and positively related to the relative population-aggregation degree.

We choose 1820 as the study case to check the model for which it has much high-match data. The statistical test and error analysis shows that the model can change the statistical Chinese historical cropland data, which was recorded in political units, into the uniform and high-resolution gridding dataset.

References

Cao Shuji, 2001. Population History of China (Vol.5, Qing Dynasty Period). Shanghai: Fudan University Press. 690–720. (in Chinese)

Foley J A, Ruth D, Gregory P A *et al.*, 2005. Global consequences of land use. *Science*, 309: 570–574.

Ge Q S, Dai J H, He F N, 2004. Spatiotemporal dynamics of reclamation and cultivation and its driving factors in parts of China during the last three centuries. *Progress in Natural Science*, 14(7): 605–614.

Ge Quansheng, Zhao Mingcha, Zheng Jingyun, 2002. Land use change of China during the 20th century. *Acta Geographica Sinica*, 55(6): 198–706. (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.

He Bingdi, 1988. Textual Research and Estimation of Land Data in China. Beijing: China Social Science Press, 61–114. (in Chinese)

He Fanneng, Ge Quansheng, Dai Junhu, 2008. Forest change of China in recent 300 years. *Journal of Geographical Sciences*, 18(1): 59–72.

He Fanneng, Ge Quansheng, Zheng Jingyun, 2002. Reckoning the areas of urban land use and their comparison in the Qing Dynasty in China. *Acta Geographica Sinica*, 57(6): 709–716. (in Chinese)

IGBP Report No.53/IHDP Report No.19, 2005. GLP (2005) Science Plan and Implementation Strategy. Stockholm: IGBP Secretariat, 1–64.

Liang Fangzhong, 1980. Statistics of Registered Permanent Residence, Plowlands and Land Taxes in Past Dynasties in China. Shanghai: Shanghai People's Press. 401–413. (in Chinese)

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.

Sun Han, Shi Yulin, 2003. Agricultural Land Use in China. Nanjing: Phoenix Science Press, 3–84. (in Chinese)

Wei W C, Gong W, Wei H L, 2000. A regional model simulation of the 1991 severe precipitation event over the Yangtze–Huai river valley (Part I): Precipitation and circulation statistics. *Journal of Climate*, 13(1): 74–92.

Zhou Rong, 2001. A general inspection and re-appraisal on area under cultivation in the early period of the Qing. *The Journal of Chinese Social and Economic History*, (3): 39–49. (in Chinese)