SCIENCE CHINA
Earth Sciences

•RESEARCH PAPER•

https://doi.org/10.1007/s11430-020-9866-2

Gridded reconstruction of the population in the traditional cultivated region of China from 1776 to 1953

Xuezhen ZHANG[1,5], Fahao WANG[1,2], Weidong LU[3], Shicheng LI[4] & Jingyun ZHENG[1,5*]

[1] *Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China;*
[2] *Faculty of Geographic Sciences, Beijing Normal University, Beijing 100875, China;*
[3] *Center for Historical Geographical Studies, Fudan University, Shanghai 200433, China;*
[4] *Department of Land Resource Management, School of Public Administration, China University of Geosciences, Wuhan 430074, China;*
[5] *College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China*

Received May 24, 2021; revised August 29, 2021; accepted November 5, 2021; published online December 17, 2021

Abstract Using modern census and environmental factor data, this study first identified the environmental factors that significantly affect the population distribution through Geodetector analysis and then constructed a population spatial distribution model based on the random forest regression algorithm. Finally, with this model and historical population data that were examined and corrected by historians, gridded population distributions with a spatial resolution of 10 km by 10 km in the traditional cultivated region of China (TCRC, hereafter) were reconstructed for six time slices from 1776 to 1953. Using the reconstruction dataset, the spatiotemporal characteristics of the population distribution were depicted. The results showed that (1) the environmental factors that significantly affected the population density differences among counties in the TCRC mainly consisted of elevation, slope, relief amplitude, distances to the nearest prefectural and provincial capitals, distance to the nearest river and the climatology moisture index. (2) Using the census data of 1934 counties in the TCRC in 2000 and the above-mentioned environmental factor data, a random forest regression algorithm-based population spatial distribution model was constructed. Its determination coefficient (R^2) is 0.81. In 88.4% of the counties (districts), the relative errors of the model predictions were less than 50%. (3) From 1776 to 1953, the total population in the study area showed an uptrend. Prior to 1851, the population increased mainly in the Yangtze River Delta. During this period, the number of grid cells in which the population densities were greater than 500 persons per km^2 increased from 292 to 683. From 1851 to 1953, the population increased extensively across the study area. In the North China Plain and the Pearl River Delta, the number of grid cells in which the population densities were greater than 500 persons per km^2 increased from 36 to 88 and from 4 to 35, respectively. The spatial clustering pattern of the population distribution varied temporally. The potential reasons included the shifts in economic development hot spots, traditional beliefs, wars, famine, and immigration policies. (4) Between our reconstructions and the HYDE dataset, there are large differences in the data sources, selected environmental factors and modeling methods. As a consequence, in comparison to our reconstructions, there were fewer populations in the eastern area and more populations in the western area from 1776 to 1851 and more populations in urban areas and fewer populations in rural areas after 1851 in the HYDE dataset.

Keywords Population, Traditional cultivated region of China (TCRC), Historical periods, Gridded reconstruction

Citation: Zhang X, Wang F, Lu W, Li S, Zheng J. 2021. Gridded reconstruction of the population in the traditional cultivated region of China from 1776 to 1953. Science China Earth Sciences, 64, https://doi.org/10.1007/s11430-020-9866-2

* Corresponding author (email: zhengjy@igsnrr.ac.cn)

1. Introduction

Human activities have profoundly modified natural landscapes and atmospheric compositions in the last 300 years and are important driving forces of global environmental change (IPCC, 2007). It is estimated that 42–68% of the global land area has been modified by human activities over the last 300 years (Hurtt et al., 2006). Greenhouse gas (GHG) emissions induced by anthropogenic land use/cover changes as well as agricultural activities account for approximately 20–25% of total emissions globally (Searchinger et al., 2018). The strength of human activities is closely related to population density. The spatial variability of population density determines the spatial pattern of anthropogenic disturbance strength on the environment. The spatial distribution of the population has been used as important primary data in studies of historical land use/cover changes (He et al., 2018; Fang et al., 2020; Kaplan et al., 2011) and historical water resource development and utilization (Qin et al., 2019). It is thereby important to reconstruct the spatial distributions of historical populations in the fields of historical geography, environmental evolution, and global changes.

Demographics are usually measured by the total population of each administrative region; hence, it is difficult to depict the spatial heterogeneity of population density within the region. It is also difficult to match natural environmental data, which are usually expressed based on grid cells; therefore, it is unfavorable for deeply understanding the relationships between the population and natural resources as well as the environment. To address this issue, gridded modeling of population spatial distribution is widely used as a downscaling method (Bai et al., 2013; Sorichetta et al., 2015; Leyk et al., 2019). Although there are immensely diverse methods with which the population spatial distribution is modeled, their mathematical foundations are summarized as spatial statistics and modeling, referring to constructing a statistical theory-based mathematical model to present the relations between population density (dependent variable) and environmental factors (independent variables) in the spatial dimension. The differences among the existing methods are mainly exhibited in two aspects: (1) The mathematical model and (2) environmental factors. In terms of the mathematical model frame, there are linear models (Yang et al., 2013; Tan et al., 2018) and nonlinear models (Linard et al., 2017; Li et al., 2018). In terms of the independent variable, there are models that use only geographical coordinates as the independent variables (Wang et al., 2010) as well as models in which multiple types of environmental factors are applied as the independent variables (Klein Goldewijk et al., 2010; Fang and Jawitz, 2018). In terms of the number of independent variables, some models use only one environmental factor as an independent variable (Deville et al., 2014), while other models incorporate multiple environmental factors as independent variables (Tatem, 2017; Han et al., 2019; Yang et al., 2019). It is well known that the population distribution is affected by multiple environmental factors, including not only natural factors but also human factors, and that the effects of these factors are generally observed in complex, nonlinear ways. Therefore, multivariate nonlinear models are widely applied when modeling population spatial distribution.

In comparison to studies focusing on gridded spatial distribution modeling of modern populations, few studies of historical populations, particularly for populations in the traditional cultivated region of China (TCRC) over the last 300 years, have been conducted. There are two potential reasons. On the one hand, the availability of historical demographic data is limiting and are not compared directly to modern data. The historical populations were usually measured by households (*Ding* and *Hu*, in Chinese Pinyi) rather than persons. Moreover, administrative boundaries have undergone tremendous changes, and many issues have arisen regarding the authenticity of historical population data during the last 300 years. Therefore, to obtain population data that can be compared with modern census data, many meticulous examinations and corrections are needed. On the other hand, it is limited by the models. As mentioned above, multivariate nonlinear models are widely applied in gridded spatial distribution modeling of populations. Such models usually require multiple independent variables, most of which (such as the vegetation index, impervious surface area, night light index, etc.) are unavailable in historical periods. Recently, Xue et al. (2019) reconstructed the spatial distributions of the populations in Suzhou in 1776 and 1820 with a suitability model. However, this model used a large amount of explicit data for urban areas in historical periods, and this type of data is almost unavailable over the entire geographic domain of the TCRC. Additionally, in this model, there are a large number of empirical parameters. These parameters are suitable for a particular region rather than for large domains; therefore, there would be large uncertainties when applying the parameters used by Xue et al. (2019) to other regions. Wang et al. (2020) conducted an exploratory study on the girded spatial distribution modeling of the population in Gansu Province with a random forest regression model, which is a nonlinear model, and independent variables including city, terrain, climate, and river factors. His approach not only obtained small prediction errors but also used independent variables that are easily available in historical periods. It is practical for gridded population spatial distribution modeling for historical periods. With this approach, Wang (2020) carried out gridded reconstruction of the population for the 17 provinces in central and eastern China in the Qing Dynasty.

There was always high population density and intense human activity in the TCRC. This study took the TCRC as

the study area, which consists of 18 provinces in the Qing Dynasty. Based on the abovementioned studies, this study attempted to quantitatively analyze the relations between population spatial distribution and environmental factors and then construct a population spatial distribution model based on the random forest regression algorithm. Finally, using the historical demographics data examined and corrected by historians, we carried out the gridded reconstruction of populations with grid sizes of 10 km by 10 km for the TCRC from 1776 to 1953, and using the reconstruction data, depicted the spatiotemporal characteristics of the population distribution. It is expected to provide primary data for research on population spatial distribution patterns and human-land relations as well as their evolution in historical periods.

2. Material and method

2.1 Study area

The study area covers the domain in south of the Inner Mongolia Plateau and in east of the Qinghai-Tibet Plateau (excluding Taiwan Prefecture, as shown in Figure 1). It consists of 18 provinces in the Qing dynasty. As the main region of Chinese cultivated civilization, this region has always been the most densely populated area in China since ancient times. The main body of the study area is in a temperature and subtropical climate regime. It is dominated by a humid and semihumid monsoon climate, with rain and heat occurring over the same period, and the annual total precipitation is in the range of 400–2000 mm. The terrain is mainly characterized by high land in the west and low land in the east. In the east, there are mainly plains and hills, such as the North China Plain, Middle-Lower Yangtze Plain and low mountains in the southeast. In the west, there are mainly plateaus and basins, such as the Loess Plateau, Yunnan-Guizhou Plateau, and Sichuan Basin. Most of the study area is below 2000 m above sea level.

2.2 Data sources

This study used three categories of data consisting of demographic data, administrative division data and environmental factor data. Among them, environmental factor data included terrain, climate and river data.

(1) Demographic data: The county-level demographic data for 2000 were derived from the Standardized Database of Chinese Population Distribution (Yang, 2016). The original data were derived from the Fifth National Population Census in China, and the spatial units of the measurements were municipal districts and counties (autonomous prefectures). The historical prefecture-level population data included six time slices, i.e., 1776, 1820, 1851, 1880, 1910, and 1953, which cover three typical periods, namely, the period when

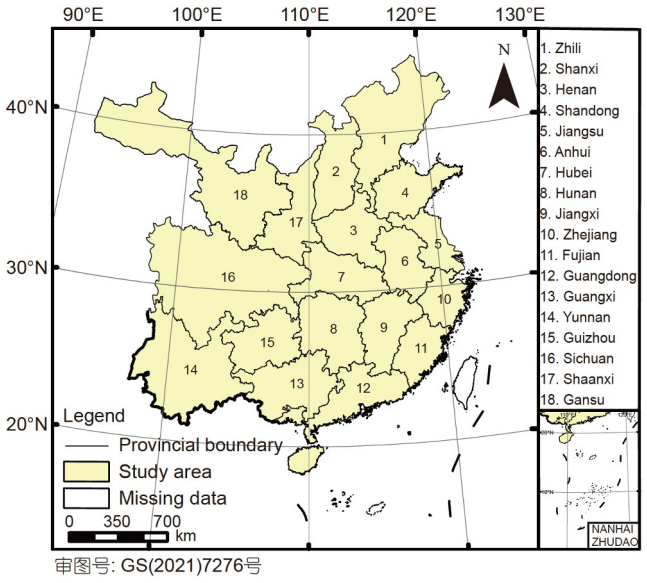


Figure 1 The domain of the study area in 1820 (yellow shaded area) (provided by the China Historical Geographic Information System).

the population grew largely in the Qing Dynasty, the period of the Republic of China and the early days of the People's Republic of China. These data were provided by the Center for Historical Geographical Studies at Fudan University (Cao, 2001). In comparison to the original demographic data measured by households (*Ding* and *Hu* in Chinese Pinyin) for the early Qing Dynasty (i.e., prior to 1776), these data were produced through examination and correction by historians and, hence, took the person as the measurement unit, which is consistent with modern demographic data (Lu, 2014). Additionally, after being examined and corrected, the population data at each time slice were uniformly recorrected to match the prefecture-level administrative regions in 1820 from the *Gazetteer of the Great Qing Unification*.

(2) Administrative division data: For the modern periods, the county-level administrative division data were derived from the 1:1 million national basic geographic information database (http://www.webmap.cn/). For the historical periods, the administrative division data are at the prefecture level and refer to that in 1820. The original data were from the *Gazetteer of the Great Qing Unification*. The capital locations of each administrative region were recorrected by following the administration evolution information obtained from the *evolution table of China's modern political divisions* (Zhang, 1987). Both corrected historical administrative division data and capital location data are derived from the China Historical Geographic Information System (http://www.people.fas.harvard.edu/~chgis/).

(3) Terrain data: The terrain factors include the elevation, slope, and relief amplitude. The original data were the ASTER GDEM data provided by NASA (https://www.nasa.gov/), with a spatial resolution of 30 m by 30 m. After being

projected and transformed in ArcGIS software, the elevation and slope were extracted. The relief amplitude was calculated as the difference between the highest and lowest elevations within a window of 5 km by 5 km.

(4) Climate and river data: The climate data refer to the climatic moisture index and are provided by the Resources and Environmental Science and Data Center of the Chinese Academy of Sciences (http://www.resdc.cn/). The climatic moisture index reflects the ratio of climatology meanly surface water income (i.e., precipitation) to water expenditure (i.e., evaporation and runoff) over several decades (Xu and Zhang, 2017). The river data refer to the spatial distribution data for river systems in modern periods and are provided by the National Earth System Science Data Center (http://www.geodata.cn/). Historical river data refer to those for 1820, which have been examined and corrected by historians and are provided by the China Historical Geographic Information System.

2.3 Approach

First, the influence degree of environmental factors on population spatial distribution was quantified, and those significantly affecting the population spatial distribution were selected using county-scale data in 2000 with the Geodetector method. Then, a random forest regression algorithm-based model predicting population density with environmental factors as independent variables was constructed, and its accuracy was verified. Next, using this model, the spatial distribution of the historical population was reconstructed with contemporaneous environmental factors and demographic data that were examined and corrected. Finally, through spatial autocorrelation analysis, the characteristics and evolution of the spatial distributions of the population in the TCRC from 1776 to 1953 were depicted, and the similarities and differences between our reconstructions and the HYDE population dataset were revealed by comparing the datasets with each other.

2.3.1 The selection of environmental factors

The factors that affect the spatial distribution of a population are diverse. In previous studies, environmental factors were usually selected through single-factor evaluations and correlation analyses. These methods ignore the complex interactions among factors, and it is hence difficult to select dominant and independent environmental factors. In this study, a Geodetector analysis was applied to select environmental factors. Geodetector analysis is a statistical method that is based on the theory of spatial heterogeneity to reveal the potential causes of geographical phenomena. The degree of influence of each independent variable on the dependent variable is measured by the similarity between the spatial distributions of the independent variable and the dependent variable. The degree of influence is defined as the q value and is in the range of 0 to 1. The closer the value is to 1, the stronger the explanatory ability of the independent variable on the dependent variable is (Wang and Xu, 2017). This method can be used not only for numerical data but also for qualitative data and can be used to detect the interactions among environmental factors. It has been widely applied in factor selection and factor interaction research (Luo et al., 2016; Liu and Li, 2017).

In this study, environmental factors were selected following three criteria: (1) The factor has significant effects on the population spatial distribution; (2) the factor was approximately unvariable over the past 300 years; and (3) the factor can be measured quantitatively. Finally, two types of factors were included in the Geodetector analysis. One type refers to natural factors such as terrain, climate and river factors, including the surface elevation above sea level, slope, relief amplitude, climatic moisture index and the distance to the nearest level-1 to level-3 rivers. The other type refers to city-related factors represented by distance to the capital cities, such as the distance to the nearest prefectural capital and the distance to the nearest provincial capital.

2.3.2 Training of the RFRM-based population distribution model and its verification

The random forest regression model (RFRM) is a machine learning algorithm and is usually realized through classification and regression trees (CART) (Breiman, 2001). The RFRM can express the nonlinear relationships between the population and environmental factors well. It has been widely used in gridded distribution modeling of populations for modern periods (Stevens et al., 2015; Ye et al., 2019; Wang et al., 2019). Herein, the county-level census data and environmental data of 1934 counties within the TCRC in 2000 were used as a sample set to train the RFRM model. Population density was the dependent variable, and the environmental factors selected by the Geodetector analysis were the independent variables.

The first step in training the RFRM is to randomly select a certain number of samples from the sample set to construct a training subset. This study used a bootstrap method to construct the training subset, the size of which was exactly the same as the sample set. The second step is to construct a decision tree that follows a binary tree structure and grows recursively from the root node to the leaf node without pruning. The key point of this step is to determine the dominant factor controlling the bifurcation structure. In this study, two environmental factors were randomly selected from all the environmental factors, one of which was further selected as the dominant factor controlling the bifurcation at each node with the CART method following the principle of least variance (Rodriguez-Galiano et al., 2014; Song et al., 2016).

By repeating the above process, 300 decision trees were constructed. Therefore, given a set of environmental data, 300 predictions could be obtained. Finally, by following the idea of unbiased estimations based on large samples, the mean value of the 300 prediction results was calculated and used as the final prediction.

The leave-one-out method was used to verify the accuracy and stability of the RFRM (Vehtari et al., 2017). For the data of 1934 sampled counties, 1933 counties were extracted to train the model; then, the model was used to predict the population density of the left-out county. This method was carried out 1934 times, and the population density of each county was predicted. Then, the accuracy and stability of the model were verified by comparing predictions to measurements. The determination coefficients (R^2), relative error and root mean square error were used to quantify the accuracy of the model, and the reduction of error (*RE*) and the coefficient of efficiency (*CE*) were used to quantify the stability of the model. The *RE* and *CE* values range from negative infinity to 1; the closer to 1, the more stable the model is (Gou et al., 2015).

2.3.3 *Gridded reconstruction of population distribution in the TCRC from 1776 to 1953*

With the abovementioned RFRM, into which historical environmental factor data were incorporated, the historical population density can be predicted for each grid at sizes of 10 km by 10 km. It is notable that because the RFRM was trained using census data for the year 2000, its parameters actually represent the quantitative relationships between the population and environmental factors in 2000. Since the total population has increased prominently during the last 300 years, the predictions are largely different from the historical demographic data. However, the relative weights of the predictions among the grid cells are useful because they represent the habitability, which is determined jointly by the selected environmental factors. Therefore, the RFRM-predicted population density was readjusted using the ratio of the historical demographic data to the predicted total population for each prefectural area, as follows:

$$P'_{ij} = P_{ij} \times \frac{S_i}{\sum_{j=1}^{n} D_{ij} P_{ij}}, \tag{1}$$

where P'_{ij} and P_{ij} represent the readjusted population density and RFRM-predicted population density for grid cell j within the domain of prefecture i; S_i represents the historical total population of prefecture i; and D_{ij} represents the land area of grid cell j within the domain of prefecture i.

In the process of gridded allocation, the area weight method was utilized to recalculate the populations of grid cells that were shared by more than one administrative region:

$$P''_j = \frac{\sum_{i=1}^{n} P'_{ij} D_{ij}}{D_j}, \tag{2}$$

where j denotes grid cells that were shared by more than one ($i \geq 2$) administrative region; P''_j denotes the readjusted population density for grid cell j; D_{ij} denotes the actual land area of grid cell j belonging to prefecture i; D_j denotes the total land area of grid cell j; and P'_{ij} denotes the population density derived from eq. (1) for grid cell j belonging to prefecture i.

Moreover, for any grid cell that was partly occupied by water bodies, i.e., rivers and lakes, the relative population weights of large lakes and first- and second-level rivers were set to 0 according to the actual historical distribution of rivers and lakes, since there would be no people living on a water body. Based on the actual land area, the population in each grid was recorrected at the prefectural scale following the method of Lin et al. (2009).

2.3.4 *Analysis of the agglomeration patterns of historical population distribution*

Spatial autocorrelation analysis is usually applied to explore spatial agglomeration patterns and to reveal the spatial nonrandom distribution characteristics of geographic phenomena (Legendre, 1993), which are measured by the Moran Index, i.e., Moran's I for short. Moran's I is a statistical metric to quantify the spatial adjacency or the similarity of adjacent unit attributes, with a range from −1 to 1. The closer to 1, the more similar; the closer to −1, the more opposite. A value of zero indicates that no correlation relationship exists between the adjacent units (Liu et al., 2015; Wang and Yang, 2019). According to the spatial scale, for which Moran's I is suitable, there is global Moran's I and local Moran's I:

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij} (X_i - \overline{X})(X_j - \overline{X})}{\sum_{i=1}^{n} (X_i - \overline{x})^2 \sum_{i=1}^{n} \sum_{j=1}^{n} (W_{ij})}, \tag{3}$$

where X_i and X_j denote the population densities of grid i and grid j, respectively, $\overline{X}$ denotes the mean population density of the whole study area, n denotes the number of grids in the whole region, and W_{ij} denotes the elements of the spatial weight matrix (W_{ij}=1 represents spatial adjacency while 0 represents nonadjacency).

$$I_i = Z_i \sum_{j=1}^{n} W_{ij} Z_j, \tag{4}$$

where Z_i and Z_j are the standardized measurements of grid i and grid j, respectively. Under a given confidence interval, if both I_i and Z_i are positive, grid i is considered to be in the high-high clustering quadrant (H-H); if I_i is positive and Z_i is negative, grid i is in the low-low clustering quadrant (L-L); if I_i is positive and Z_j is negative, grid i is in the high-low clustering quadrant (H-L); If I_i and Z_j are negative, grid i is in the low-high clustering quadrant (L-H).

3. Result

3.1 The selection of environment factors

The Geodetector analysis shows that all four categories of factors, including terrain, city, climate, and river factors, significantly impact the spatial distribution of the population. Among them, the impacts of terrain, city, and climate factors are significant at the level of 0.01, and the impacts of river factors are significant at the level of 0.1. As shown in Table 1, the degree of impact of terrain is the most prominent and that of city is secondarily prominent; the degree of impacts of climate and river factors are the third. In terms of q statistics, the q values of terrain factors are in the range of 0.37–0.55, which is higher than that ranging from 0.21 to 0.37 for cities. The q values of the climatic moisture index and river factor are as low as 0.02, which is one order of magnitude lower than those of terrain and cities. This result suggests that within the domain of the TCRC, the spatial variabilities of population density are highly consistent with those of terrain but weakly consistent with those of the effects of cities. The explanatory ability of terrain to the spatial variability of population density is stronger than that of city factors.

Among the q values of terrain factors, the q value of elevation reaching up to 0.55 was the highest, and q values of slope and relief amplitude, both of which are approximately 0.37, are comparable with each other. This result indicates that the explanatory ability of elevation to the spatial variability of population density was stronger than those of slope and relief amplitude. On behalf of the city factors, the q value of the distance to the nearest prefectural capital was 0.37, which was slightly higher than the q value of 0.21 of the distance to the nearest provincial capital city. This result indicates that the explanatory ability of the radiation effects of prefectural capital cities was stronger than that of provincial capital cities. The climatic moisture index and the distance to the nearest river also significantly impacted the spatial variation in population density. There are usually dense populations in wetter climate regimes and closer to rivers, and vice versa. However, the degree of impact of these factors on the population distribution was very weak.

Table 1 The q values of the single-factor detector in Geodetector[a)]

Environmental factors	q value	p value
Elevation	0.55***	0.00
Slope	0.37***	0.00
Relief amplitude	0.37***	0.00
Distance to the nearest prefectural city	0.37***	0.00
Distance to the nearest provincial capital	0.21***	0.00
Moisture index	0.02***	0.01
Distance to the nearest river	0.02*	0.09

a) * and *** denote significance at the 0.1 and 0.01 confidence levels, respectively.

This result suggests that moisture variations are within the range to which human production and livelihoods can adapt, and hence, its impact is weak. For the rivers, this may be explained by the fact that the study area was mostly located in humid and semihumid climate regimes, where river systems are highly developed; hence, the distances from most counties to rivers of the first three levels remained very short, and no differences could be detected.

On behalf of double-factor interactions (Table 2), most of the two-factor combinations may have stronger impacts than the individual factor. Among the combinations, the strongest impact resulted from the combination of the distance to the nearest prefectural city and elevation (q=0.69). At the secondary level, the two combinations were comparable in terms of their degrees of impact. They were the combination of the distance to the nearest provincial capital and elevation (q=0.64) as well as the combination of the slope and elevation (q=0.64). The impacts of the combinations of terrain and city factors were much stronger than those of any single terrain, climate or river factor. This finding further confirmed that the spatial distribution of the population was influenced by multiple factors, especially combinations of natural and human factors. Based on the results of the Geodetector analysis, terrain factors (i.e., elevation, slope, relief amplitude), city factors (i.e., distance to the nearest prefectural city, distance to the nearest provincial capital), climate factors (i.e., moisture index) and river factors (distance to the

Table 2 The q values of the interaction detection in Geodetector analysis[a)]

	Elevation	Slope	Ra	Dp	Dc	Mi	Dw
Elevation	0.55						
Slope	0.64	0.37					
Ra	0.64	0.38	0.37				
Dp	0.69	0.61	0.62	0.37			
Dc	0.64	0.47	0.48	0.46	0.21		
Mi	0.62	0.41	0.43	0.40	0.25	0.02	
Dw	0.58	0.39	0.40	0.41	0.28	0.09	0.02

a) Ra, relief amplitude; Dp, Distance to the nearest prefectural city; Dc, Distance to the nearest provincial capital; Mi, Moisture index; Dw, Distance to the nearest river

nearest level-1 to level-3 river) were selected to carry out gridded modeling of the population spatial distribution.

3.2 Verification of the RFRM-based population distribution model

The RFRM-based population distribution model with the abovementioned seven environmental factors as input could reproduce most of the spatial variability of the county-level population density in 2000. Figure 2 shows that the predicted population density is significantly positively correlated with the census data (r=0.9, p<0.001). This suggests that predictions could explain 81% of the total variance in the census data. Moreover, the coefficient of efficiency (*CE*) and reduction of error (*RE*) were 0.72 and 0.75, respectively, which suggested that the model was stable. However, notably, the predictions underestimated the values for very high-population-density areas and overestimated the values for very low-population-density areas. As a result, the spatial variance in the predicted population density is slightly smaller than that in the census data (Figure 2).

On behalf of the distribution shape of the predictions' relative errors, the relative errors approximately followed a normal distribution (Figure 3a). The counties where relative errors are less than 50% accounted for as much as 88.4% of the total counties, and the counties where relative errors are higher than 80% accounted for only 4.1% of the total. This result suggests that most predictions have small errors and a small number of predictions have large errors. This characteristic remains similar to the error distributions of output

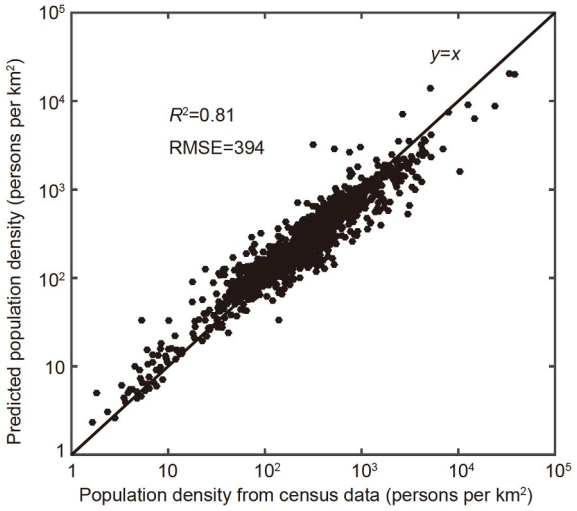
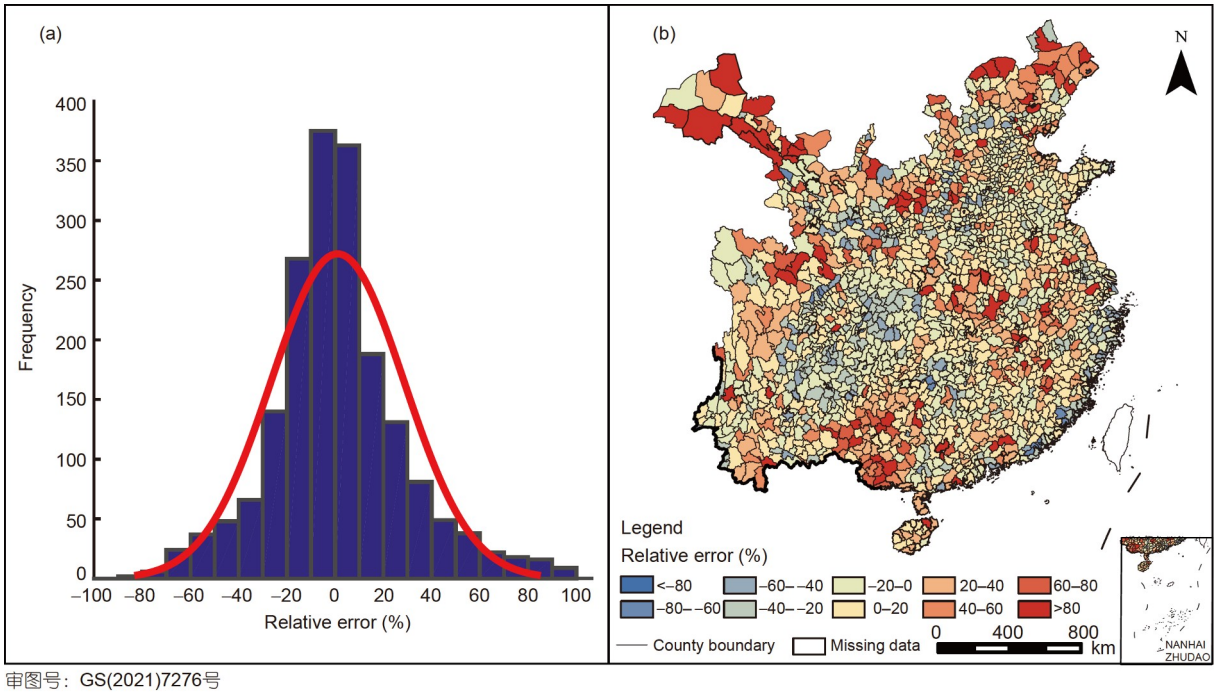


Figure 2 Scatter plot of the random forest regression model predictions against the county-level census data for 2000.

from other statistical models. The stability of the model is thereby reconfirmed by this finding. Figure 3b shows the spatial distribution of the errors. Small errors mainly existed in the North China Plain and lowlands along the Yangtze River, while large errors mainly existed in the mountainous areas of the northwestern and southwestern study areas, the coast of Southeast China, the Guanzhong Plain, the Jiangnan Hills and the Chengdu Plain. Among them, positive errors mainly existed in the northwest and southwest mountainous areas as well as the Jiangnan Hills. This finding indicates that the predictions are greater than the actual population den-



审图号：GS(2021)7276号

Figure 3 Histogram (a) and spatial distributions (b) of the relative errors in the random forest regression model predictions.

sities in these regions. Negative errors mainly existed in the Guanzhong region, along the southeast coast, and on the Chengdu Plain. This finding indicates that the predictions are less than the actual population densities in these regions. The main reason for these errors was that the model did not include the effects of economic location factors on the population spatial distribution. On the Guanzhong Plain and Chengdu Plain, the terrain is flat; hence, agriculture is developed, and transportation is convenient. Both of them are regional economy centers. There are dominantly hills on the southeast coast of the study area, but it is favorable for ocean shipping, and the economic level is higher than that in adjacent areas. Due to the developed economies, there are high populations in these regions. However, following the Geodetector analysis results, the population density would be low in both high-elevation and complex terrain areas. The elevations of the Guanzhong region and the Chengdu Plain are approximately 500–600 m above sea level, which is much higher than the elevation over the North China Plain of 0–50 m above sea level. Therefore, the model predictions are lower than the actual population density.

3.3 Population distribution pattern in the TCRC from 1776 to 1953

The reconstruction results show that the spatial distribution pattern of the population in the TCRC remained approximately unchanged from 1776 to 1953. It was dominantly characterized by more persons in the east and fewer persons in the west. High population densities always existed in the middle and lower reaches of the Yangtze River Plain, North China Plain, Guanzhong Plain, Sichuan Basin and Pearl River Delta. Low population densities always existed in the Western Sichuan Plateau, Yunnan-Guizhou Plateau, Southeast Hills, Loess Plateau, Inner Mongolia Plateau and north of the Hexi Corridor. At the grid cell scale, the global Moran's I remained greater than 0 (p<0.01) throughout the study period (Table 3). This finding indicates that there was always significant spatial agglomeration of the population in the TCRC. Notably, the dominant characteristics of spatial agglomeration varied temporally.

From 1776 to 1851, along with the opening of commercial ports and the development of canal trade, the Yangtze River Delta region was the area where the greatest population growth occurred (Figure 4a–4c). During this period, the grid cells where the population density was higher than 500 persons per km^2 increased, from 292 to 683, by 134%. The population densities of the grid cell where Suzhou prefectural capital was located were the highest, exceeding 5,000 persons per km^2. The provincial and prefectural capitals in the eastern provinces were the main population agglomeration areas. For instance, in the Shuntian, Kaifeng, Jinan and Nanchang prefectural capitals, the population densities were more than 1,500 persons per km^2. There was spatial variability in the population densities in the hinterland of the eastern plain areas, but the population densities were consistently high. As a result, these areas were characterized by high-density population agglomerations; they were the main high-density population agglomeration areas in this period (Figure 5a–5c). Although the population density in the urban areas of the eastern study area increased prominently in this period, the area with a high-density population concentration remained approximately unexpanded. The main reason for this result is that the population had reached nearly the maximum value that can be supported by production ability and that farmers were reluctant to leave their ancestral homelands due to traditional beliefs (Pei, 2017); hence, there was no immigration. In the western region, the population densities in the urban areas of the Sichuan Basin and the Yunnan-Guizhou Plateau greatly increased, and there was immigration to nearby areas. The urban settlements developed, and population agglomeration strengthened over the Chengdu Plain. Until 1851, a high-density population concentration area existed surrounding the Chengdu Prefecture, with a population density of 1,770 persons per km^2 in the Chengdu Prefecture capital. This was mainly caused by the strong inertia of population growth led by the massive immigration that occurred during the Kangxi-Qianlong empire periods, when nearly one million people immigrated into Sichuan. Therefore, high immigration provided the basis for the high population growth rate in the mid-late Qing Dynasty, and the population spatial agglomeration also intensified.

From 1851 to 1880, the changes in population density in the study area exhibited distinct spatial variability (Figure 4d). On the one hand, due to war and famine, the population declined sharply in the urban areas in southern Jiangsu, northern Zhejiang, southern Anhui, central Shaanxi and southern Gansu. Due to the Taiping Rebellion, there were severe population losses in the middle and lower reaches of the Yangtze River. In 1880, grid cells where the population density was higher than 500 persons per km^2 declined to 145, accounting for approximately 20% of that in 1851. The population density of the grid cell where Suzhou Prefecture was located declined by 64%, i.e., from 5,235 persons per km^2 to

Table 3 Global Moran's I index and Z values of the population in the TCRC from 1776 to 1953 at the grid scale of 10 km by 10 km

Year	Moran's I	Z
1776	0.85	247.86
1820	0.85	245.52
1851	0.85	245.58
1880	0.83	242.25
1910	0.81	236.68
1953	0.82	236.86

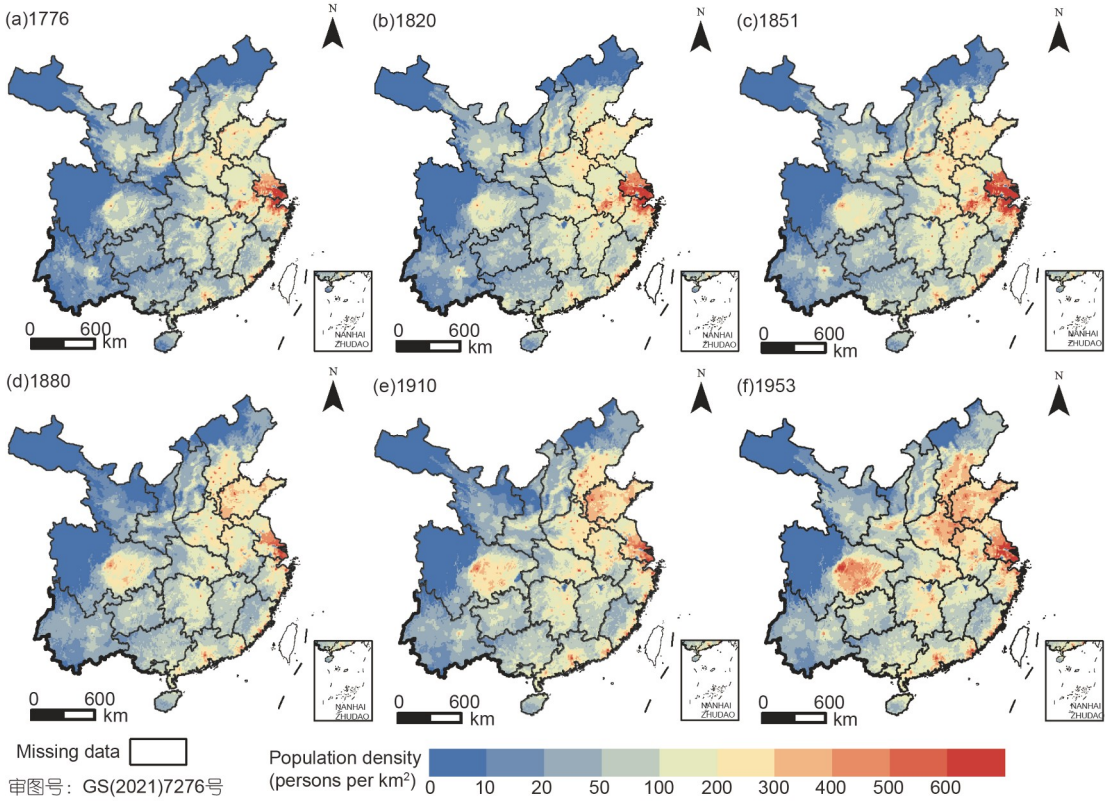


Figure 4 Spatial distribution pattern of the population in the TCRC from 1776 to 1953 (grid size: 10 km by 10 km).

1,894 persons per km^2, from 1851 to 1880. Due to the Ding-Wu severe famine and the northwest clash during the Tongzhi period, the population density in southern Gansu, northern Henan and the Fengxiang-Qianzhou-Xi'an-Datong line in Shanxi and Shaanxi declined by 200-500 persons per km^2. Along with population losses in urban areas, the agglomeration characteristics of the population changed. The population agglomeration areas in the middle and lower reaches of the Yangtze River shrank to the east and expanded to the north, and the population agglomeration areas in the Guanzhong Plain disappeared. On the other hand, with the beginning of modernization, the functions and statuses of cities gradually developed. There was prominent urban development in the North China Plain, the Sichuan Basin and the Pearl River Delta, which were impacted slightly by wars and famines. Their populations grew, and spatial agglomeration characteristics occurred in these areas (Figure 5d). From 1776 to 1880, the population density of the prefectural cities in southern Zhili and Shandong increased by 50–200 persons per km^2. As a result, there was a high-density population-agglomeration area in Shandong, southern Zhili and northeast Henan. Meanwhile, the high-density population agglomeration areas in the Pearl River Delta and the Sichuan Basin also began developing in this period.

From 1880 to 1953, the population increased extensively in the plains and the southeastern coastal areas of the TCRC (Figure 4e to 4f). Up to 1953, the numbers of grid cells where population densities exceeded 500 persons per km^2 in the eastern North China Plain and the Pearl River Delta had increased to 88 and 35 respectively, which were approximately 2.4- and 8.75-fold of those in 1851. At this time, the population densities in the urban areas of Beijing and Shanghai exceeded 4,000 persons per km^2. The high population density agglomeration area in the eastern study area, which was represented by the northern China region and the middle and lower reaches of the Yangtze River, arose at that time. Additionally, due to the expansion of human activities toward the north, the population agglomeration area extended toward the north and, hence, the low-population-density agglomeration area in the northern part of the North China Plain shrunk. In the west, through population growth and expansion over a long period, the population density within the Sichuan Basin greatly increased. In 1953, the grid cells where population densities were greater than 500 persons per km^2 in the Chengdu Plain and Chongqing increased to 60 and 12, respectively. As a result, a high-population-density agglomeration area centered in the central-eastern part of Sichuan and Chongqing arose (Figure 5e–5f). The population in the Yunnan-Guizhou Plateau also grew, but there were no detectable spatial agglomeration characteristics and no high-density population hotspots.

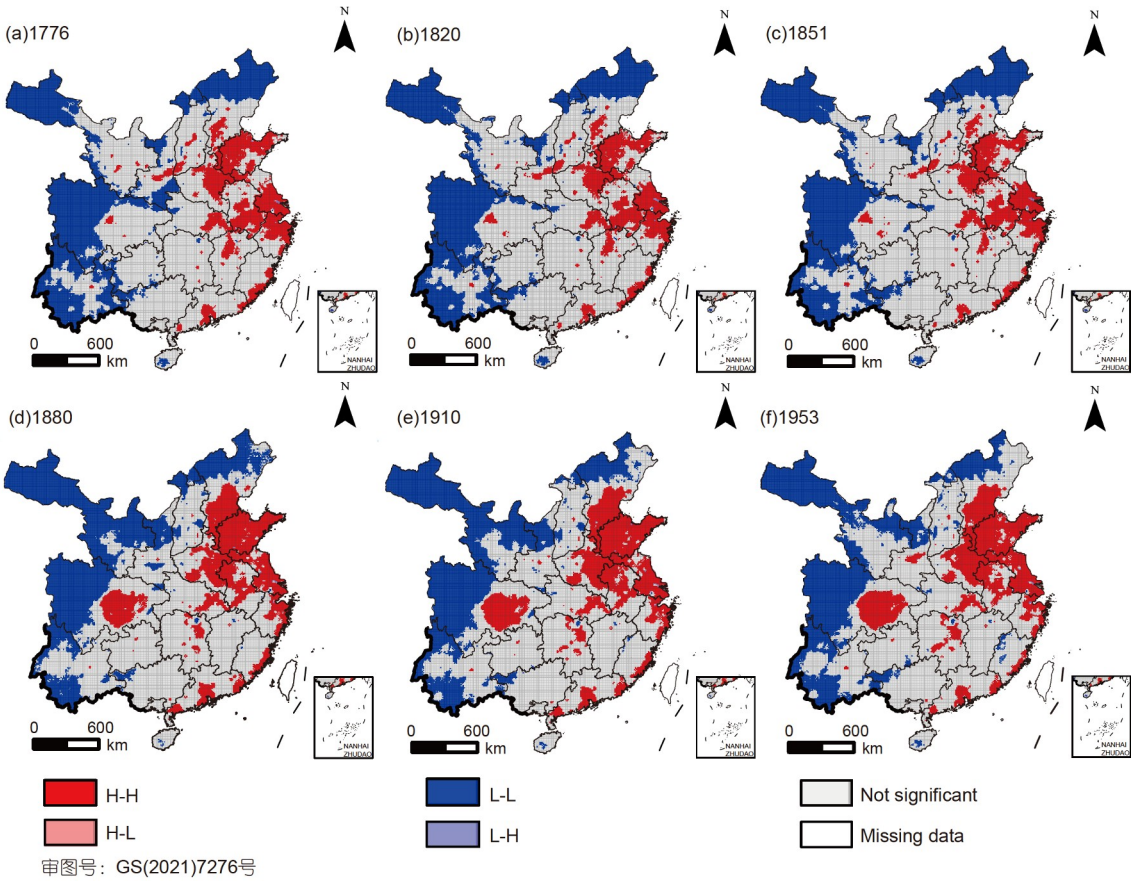


Figure 5 LISA cluster map of the population distribution in the TCRC from 1776 to 1953 (grid size: 10 km by 10 km).

3.4 Comparison with HYDE population dataset

Among the grid cell-based global historical population datasets, the HYDE population dataset v3.2 covers the longest time series (Klein Goldewijk et al., 2017). This dataset partially overlaps with our reconstructions in both the temporal and spatial dimensions. For the years in which both of them do not exactly match each other, we selected the closest year between the HYDE dataset and our reconstructions for the comparison. As shown in Figure 6, the spatial distribution patterns of the population exhibited by both our reconstructions and the HYDE dataset are highly consistent but with differences in some regions. From 1776 to 1851, the population densities in the HYDE dataset were much lower than those from our reconstructions in the eastern plains areas, particularly in the middle and lower reaches of the Yangtze River and urban areas of the provincial and prefectural capitals. Among them, the population densities in some urban areas in the southern region of the Yangtze River in the HYDE dataset were less than those in our reconstruction by exceeding 500 persons per km^2. In the western and northern highland and mountainous areas, the population densities in the HYDE dataset were generally higher than those in our reconstructions. The largest overestimations existed in the Sichuan Basin, where the overestimations were 50–200 persons per km^2, except for the Chengdu Plain. Between 1851 and 1953, in the eastern part of the study region, the population densities in the HYDE dataset were still lower than those in our reconstructions but with smaller differences than those in previous periods. However, the difference between the HYDE dataset and our reconstruction exhibits new characteristics, which are urban and rural discrepancies. The urban population densities in HYDE were higher than those in our reconstructions, while the rural population densities in HYDE were lower than those in our reconstructions. In the western study area, where urban areas developed slowly and rural areas were dominant, the population densities in HYDE were extensively lower than those in our reconstruction. Taken together, we find that HYDE overestimated the urban population and underestimated the rural population during 1851–1953.

The differences between our reconstructions and the HYDE data mainly resulted from the different data sources and spatial modeling methods (Table 4). In terms of the data sources, HYDE used province-level population data provided by Populstat, whereas our reconstructions used prefecture-level population data taken from historical documents and examined and corrected by local historians.

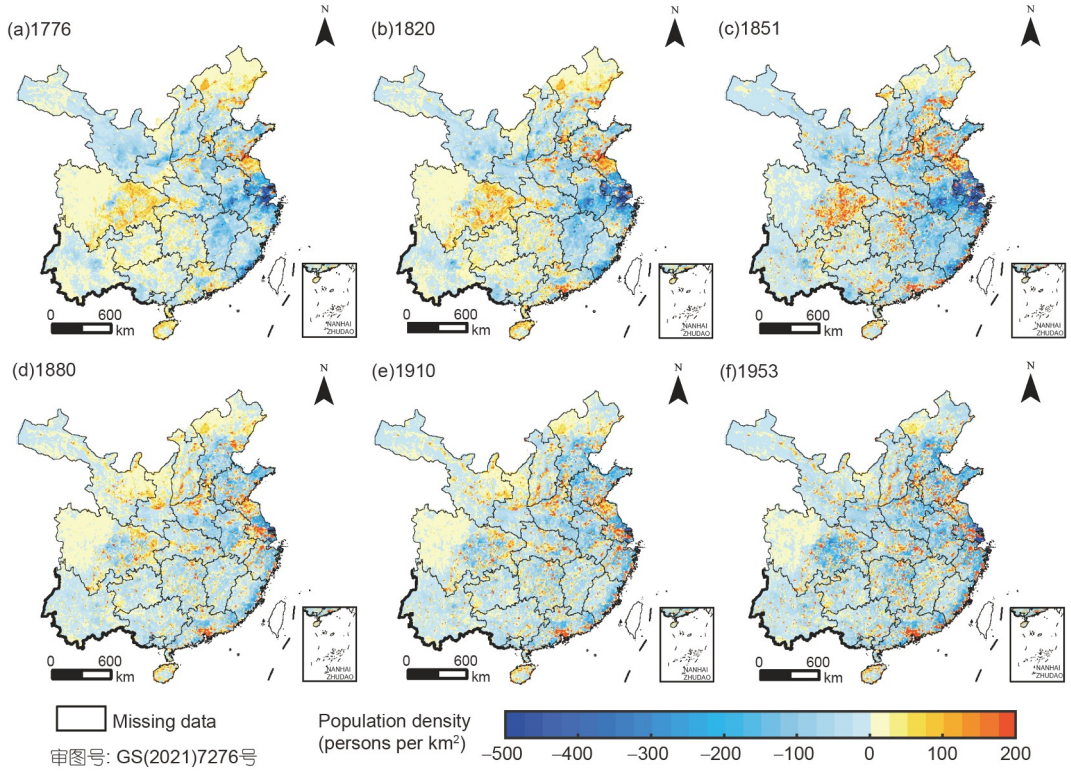


Figure 6 Differences in population densities between the reconstructions obtained in this paper and the HYDE dataset (HYDE dataset minus the reconstructions in this paper).

Table 4 Comparisons of data sources and environmental factors applied by spatial distribution model between the reconstructions from this study and the HYDE population dataset

Dataset	Original population data sources	Environmental factors for spatial distribution modeling
This study	Prefecture-level population data from historical documents	Elevation; slope; relief amplitude; distance to city; climate moisture index; distance to river
HYDE	Province-level population from Populstat	LandScan population distribution pattern; soil productivity; slope; distance to river

Hence, the total populations differed between the two datasets. The overall regional mean population densities were thereby different between the two datasets. This may be the main reason why the populations in the HYDE dataset were generally less than our reconstructions in the eastern areas and higher in the western areas between 1776 and 1851 and why the populations in the HYDE dataset were less than our reconstructions in the western region between 1851 and 1953. In terms of the spatial modeling methods, HYDE used LandScan's modern population spatial distribution pattern as the main control factor and maintained a fixed spatial pattern throughout the historical period. As a result, there are high weights in urban areas, and the population in urban areas would be overestimated. This is because in the preindustrial period, China was dominantly supported by agriculture, and population density differences between urban and rural areas were much smaller than those existing at present. Our reconstructions took into account the environmental factors that were selected through Geodetector analysis. In detail, population spatial modeling took into account the actual historical information on the ranks and locations of cities to indicate the historical spatial agglomeration of the population. As a result, the reconstructed population spatial distribution was dominantly determined by the terrain and city factors as well as by the temporal variability of the cities. Hence, the dominant cause of overestimations in eastern urban areas and underestimations in rural areas in the HYDE dataset may be the maintenance of the unchanged current spatial relative weighting.

4. Discussion

Based on prefecture-level population data that were rigorously examined and corrected by Chinese historians, this paper performed gridded historical population reconstructions and explicitly revealed the evolution of the population spatial distribution in the TCRC from 1776 to 1953. Mod-

eling the population spatial distribution is a crucial task in reconstruction. The model accuracy can exert profound impacts on the reconstructions. To understand the uncertainties of the reconstruction results, the shortages of the modeling and its possible impacts on the reconstruction results are discussed.

First, the selection of environmental factors is limited by the availability of environmental data in the historical period. Due to missing data in the historical period, modern data were used instead of historical data for some environmental factors. For example, due to the absence of high-precision, grid cell-based climate moisture index data for historical periods, the modern climate moisture index was used. According to historical climate reconstructions, the 18th and 19th centuries were in the Little Ice Age, during which the temperatures were lower by 0.3–0.6°C than modern times (Ge et al., 2013). Moreover, in this period, the climate was also wetter, especially in northern China (Zheng et al., 2006). Nevertheless, it is difficult to assess the impact of the substitution of the moisture index between historical and modern times on gridded population reconstruction because most of the available climate reconstruction data are targeted to the entire study region rather than to spatially explicit grid cells. It is generally understood that spatial variabilities in climate change are greater in areas with complex topography than in areas with homogenous topography. It can be inferred that the impacts of substituting the moisture index on the reconstruction results are greater in complex topography areas than in flat areas.

Second, this paper used the distances to cities to measure the impact degrees of the city factors. The industrial structure differs under different social forms. Modern cities are far more attractive to populations than cities in historical periods. In particular, after the reform and opening up, with the acceleration of urbanization, the population agglomeration effect in urban areas strengthens gradually. Thus, the RFRM-based model trained with modern data may overestimate population densities in urban and surrounding areas in historical periods. Finally, economic locations were not taken into account in the model. However, our reconstructions used prefecture-level population data, which are partly modulated by national-level economic locations. Hence, the macroscale population distribution may have been impacted very little.

5. Conclusion

Through the above analysis, it was found that at the county level in the TCRC, terrain factors (i.e., elevation, slope and relief amplitude) were the dominant environmental factors affecting the population distribution. City factors (i.e., the distance to the nearest provincial capitals and prefecture-level cities) were secondary; the climate factor (i.e., moisture index) and the river factor (i.e., the distance to level-1 to level-3 rivers) were in third. Using the above factors as independent variables, the population distribution model based on the random forest regression algorithm had a coefficient of determination of 0.81 (p<0.01), which suggests that this model could reproduce the spatial distribution of the population well. Using the model taking together with the historical demographic data, the gridded spatial distribution patterns of the population at a grid size of 10 km by 10 km were reconstructed for six time slices from 1776 to 1953. The reconstructions show that the population in the TCRC increased overall during this period. Population growth mainly existed in the Yangtze River Delta region before 1851, and it occurred particularly and extensively in the plains and southeast coastal areas from 1851 to 1953. Spatial agglomeration characteristics were observed in the population spatial distribution, but different patterns emerged among different periods. The main reasons for the different spatial patterns included shifting economic development hotspots, traditional beliefs, wars, famines, and migration policies. In comparison to our reconstructions, the HYDE dataset underestimated the population in the eastern region and overestimated the population in the western region from 1776 to 1851, while it overestimated the urban population and underestimated the rural population in the eastern region from 1851 to 1953. The reasons leading to these biases include the data sources, environmental factor selection and spatial modeling methods.

Our reconstructions provide data support to study the issues that are relevant to the resource environment and human society in the TCRC during historical periods, such as the impacts of climate change on human society, the interactions between regional environmental change and human activities, the contexts and impacts of major historical events, epidemic risk analyses, social resource carrying capacity assessments, and human-induced land use and land cover changes and their climatic and ecological effects. However, gridded reconstructions of the population approximate the actual population sizes and distributions. Due to the influences of the model structures and the availability of environmental data in historical periods, the reconstruction results are subject to uncertainty and need to be further refined. Finally, the available gridded population datasets for the post-1950s used different dependent variables, model structures, and grid sizes; thus, systematic differences exist between our reconstructions and other datasets. It is thereby an important issue to eliminate these differences.

Acknowledgements *This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA19040101) and the National Key Research and Development Program of China (Grant No. 2017YFA0603301).*

References

Bai Z, Wang J, Yang F. 2013. Research progress in spatialization of population data (in Chinese with English abstract). Prog Geogr, 32: 1692–1702

Breiman L. 2001. Random forests. Mach Learn, 45: 5–32

Cao S. 2001. Population History of China (Vol. 5, Qing Dynasty Period) (in Chinese). Shanghai: Fudan University Press. 691–719

Deville P, Linard C, Martin S, Gilbert M, Stevens F R, Gaughan A E, Blondel V D, Tatem A J. 2014. Dynamic population mapping using mobile phone data. Proc Natl Acad Sci USA, 111: 15888–15893

Fang X, Zhao W, Zhang C, Zhang D, Wei X, Qiu W, Ye Y. 2020. Methodology for credibility assessment of historical global LUCC datasets. Sci China Earth Sci, 63: 1013–1025

Fang Y, Jawitz J W. 2018. High-resolution reconstruction of the United States human population distribution, 1790 to 2010. Sci Data, 5: 180067

Ge Q, Hao Z, Zheng J, Shao X. 2013. Temperature changes over the past 2000 yr in China and comparison with the Northern Hemisphere. Clim Past, 9: 1153–1160

Gou X, Gao L, Deng Y, Chen F, Yang M, Still C. 2015. An 850-year tree-ring-based reconstruction of drought history in the western Qilian Mountains of northwestern China. Int J Climatol, 35: 3308–3319

Han D, Yang X, Cai H, Xu X, Qiao Z, Cheng C, Dong N, Huang D, Liu A. 2019. Modelling spatial distribution of fine-scale populations based on residential properties. Int J Remote Sens, 40: 5287–5300

He F, Li S, Yang F, Li M. 2018. Evaluating the accuracy of Chinese pasture data in global historical land use datasets. Sci China Earth Sci, 61: 1685–1696

Hurtt G C, Frolking S, Fearon M G, Moore B, Shevliakova E, Malyshev S, Pacala S W, Houghton R A. 2006. The underpinnings of land-use history: Three centuries of global gridded land-use transitions, wood-harvest activity, and resulting secondary lands. Glob Change Biol, 12: 1208–1229

IPCC. 2007. Climate Change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge and New York: Cambridge University Press

Kaplan J O, Krumhardt K M, Ellis E C, Ruddiman W F, Lemmen C, Klein Goldewijk K. 2011. Holocene carbon emissions as a result of anthropogenic land cover change. Holocene, 21: 775–791

Klein Goldewijk K, Beusen A, Doelman J, Stehfest E. 2017. Anthropogenic land use estimates for the Holocene: HYDE 3.2. Earth Syst Sci Data, 9: 927–953

Klein Goldewijk K, Beusen A, Janssen P. 2010. Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1. Holocene, 20: 565–573

Legendre P. 1993. Spatial autocorrelation: Trouble or new paradigm? Ecology, 74: 1659–1673

Leyk S, Gaughan A E, Adamo S B, de Sherbinin A, Balk D, Freire S, Rose A, Stevens F R, Blankespoor B, Frye C, Comenetz J, Sorichetta A, MacManus K, Pistolesi L, Levy M, Tatem A J, Pesaresi M. 2019. The spatial allocation of population: A review of large-scale gridded population data products and their fitness for use. Earth Syst Sci Data, 11: 1385–1409

Li K, Chen Y, Li Y. 2018. The random forest-based method of fine-resolution population spatialization by using the international space station nighttime photography and social sensing data. Remote Sens, 10: 1650

Lin S, Zheng J, He F. 2009. Gridding cropland data reconstruction over the agricultural region of China in 1820. J Geogr Sci, 19: 36–48

Linard C, Kabaria C W, Gilbert M, Tatem A J, Gaughan A E, Stevens F R, Sorichetta A, Noor A M, Snow R W. 2017. Modelling changing population distributions: An example of the Kenyan Coast, 1979–2009. Int J Digital Earth, 10: 1017–1029

Liu T, Qi Y, Cao G, Liu H. 2015. Spatial patterns, driving forces, and urbanization effects of China's internal migration: County-level analysis based on the 2000 and 2010 censuses. J Geogr Sci, 25: 236–256

Liu Y, Li J. 2017. Geographic detection and optimizing decision of the differentiation mechanism of rural poverty in China (in Chinese with English abstract). Acta Geogr Sin, 72: 161–173

Lu W. 2014. GIS-supported analysis of regional population change patterns over a long period of time - a case study of population in Shaanxi-Gansu region from 1776 to 1953 (in Chinese). Hist Geogr, (2): 314–324

Luo W, Jasiewicz J, Stepinski T, Wang J, Xu C, Cang X. 2016. Spatial association between dissection density and environmental factors over the entire conterminous United States. Geophys Res Lett, 43: 692–700

Pei Q. 2017. Migration for survival under natural disasters: A reluctant and passive choice for agriculturalists in historical China. Sci China Earth Sci, 60: 2089–2096

Qin Y, Mueller N D, Siebert S, Jackson R B, AghaKouchak A, Zimmerman J B, Tong D, Hong C, Davis S J. 2019. Flexibility and intensity of global water use. Nat Sustain, 2: 515–523

Rodriguez-Galiano V, Mendes M P, Garcia-Soldado M J, Chica-Olmo M, Ribeiro L. 2014. Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (Southern Spain). Sci Total Environ, 476-477: 189–206

Searchinger T D, Wirsenius S, Beringer T, Dumas P. 2018. Assessing the efficiency of changes in land use for mitigating climate change. Nature, 564: 249–253

Song J, Gao Q, Li Z. 2016. Application of random forests for regression to seismic reservoir prediction (in Chinese with English abstract). Oil Geophys Prospect, 51: 1202–1211, 1052–1053

Sorichetta A, Hornby G M, Stevens F R, Gaughan A E, Linard C, Tatem A J. 2015. High-resolution gridded population datasets for Latin America and the Caribbean in 2010, 2015, and 2020. Sci Data, 2: 1–2

Stevens F R, Gaughan A E, Linard C, Tatem A J. 2015. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. PLoS ONE, 10: e0107042

Tan M, Li X, Li S, Xin L, Wang X, Li Q, Li W, Li Y, Xiang W. 2018. Modeling population density based on nighttime light images and land use data in China. Appl Geogr, 90: 239–247

Tatem A J. 2017. WorldPop, open data for spatial demography. Sci Data, 4: 170004

Vehtari A, Gelman A, Gabry J. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat Comput, 27: 1413–1432

Wang C, Kan A, Zeng Y, Li G, Wang M, Ci R. 2019. Population distribution pattern and influencing factors in Tibet based on random forest model (in Chinese with English abstract). Acta Geogr Sin, 74: 664–680

Wang F, Lu W, Zheng J, Li S, Zhang X. 2020. Spatially explicit mapping of historical population density with random forest regression: A case study of Gansu Province, China, in 1820 and 2000. Sustainability, 12: 1231

Wang F. 2020. Modelling of historical population spatial distribution through fusing multiple data sources and its applications: A case study for the traditional cultivated region of China during 18th to mid-20th century (in Chinese with English abstract). Dissertation for Master's Degree. Jinan: Shandong Normal University. 68

Wang J, Xu C. 2017. Geodetector: Principle and prospective (in Chinese with English abstract). Acta Geogr Sin, 72: 116–134

Wang P, Wang Z, Zhang X, Li C, Wang X, Feng Q, Chen Q. 2010. The spatial patterns of China's population and their cause of format ion in Western Han Dynasty (in Chinese with English abstract). Northwest Popul J, 31: 88–90, 96

Wang S, Yang H. 2019. Study on the Evolution of population and economic spatial distribution in China's Central Plains Economic Zone (in Chinese with English abstract). Popul J, 41: 35–44

Xu X, Zhang Y. 2017. China Meteorological Background Data Set. Data Registration and Publishing System of Resource and Environmental Science Data Center, Chinese Academy of Sciences

Xue Q, Jin X, Han J, Yang X, Zhou Y. 2019. l. Refinement reconstruction of the spatial pattern of regional historical population: Method and

demonstration (in Chinese with English abstract). Sci Geogr Sin, 39: 1857–1865
Yang X, Ye T, Zhao N, Chen Q, Yue W, Qi J, Zeng B, Jia P. 2019. Population mapping with multisensor remote sensing images and point-of-interest data. Remote Sens, 11: 574
Yang X, Yue W, Gao D. 2013. Spatial improvement of human population distribution based on multi-sensor remote-sensing data: An input for exposure assessment. Int J Remote Sens, 34: 5569–5583
Yang Y. 2016. Standardized Database of Chinese Population Distribution. Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences
Ye T, Zhao N, Yang X, Ouyang Z, Liu X, Chen Q, Hu K, Yue W, Qi J, Li Z, Jia P. 2019. Improved population mapping for China using remotely sensed and points-of-interest data within a random forests model. Sci Total Environ, 658: 936–946
Zhang Z. 1987. Evolution Table of China's Modern Political Divisions (in Chinese). Fujian: Fujian Map Publishing House. 13–266
Zheng J, Wang W C, Ge Q, Man Z, Zhang P. 2006. Precipitation variability and extreme events in Eastern China during the past 1500 years. Terr Atmos Ocean Sci, 17: 579–592

(Responsible editor: Jianhui CHEN)