Using Land Use Data to Estimate the Population Distribution of China in 2000

Yuna Mao, Aizhong Ye,[1] and Jing Xu
College of Global Change and Earth System Science, Beijing Normal University, Beijing, China, 100875

Abstract: To control for the defects found in remote sensing–derived estimates of population distributions, this study created a new method for estimating the population density of China at a 1 × 1 km spatial resolution by combining remote sensing–derived land use with China residence polygon data. As a result, we obtained three sets of land use data (i.e., remote sensing–derived, China residence polygon, and a combination of the two) to estimate population. On the basis of these data, we developed both urban and rural population distribution models. The results demonstrated that this new method could improve the accuracy of population estimation.

INTRODUCTION

Information about the distribution of population plays an important role in a country because it affects social development, resources, and other aspects of society. Large populations have placed great pressures on global resources, the environment, and sustainable development (Lo, 1986; Sutton et al. 1997); therefore, timely and accurate estimations of the spatial distribution of population and its development are essential to protect the environment. Normally, population distributions are obtained by periodic censuses or statistical analysis. This type of data is useful for macro-analyses of the population, resources, the environment, and social and economic development. However, such data are usually not suitable for conducting analyses on micro-aspects of population distribution, such as spatial analysis, because the accuracy of these methods of data collection do not provide sufficient spatial resolution.[2] Furthermore, the data structure is not applicable and has created difficulties for spatial analyses. Moreover, this type of method for estimating population is time-consuming, costly, and difficult to update on a frequent basis.

Due to the urgent demand for high-spatial-resolution population data and the advancement of new technologies, such as geographic information systems (GIS) and remote sensing (RS), many researchers have begun to use digital simulation technology to spatially estimate population distributions. Remotely sensed data have become an important resource in population estimation because of their strengths in data coverage, reasonable accuracy, and low cost (Lo, 1995; Jensen and Cowen, 1999).

[1]Corresponding author; email: Azye@bnu.edu.cn

[2]This type of analysis often regards the population distribution in a region as uniform, which is often far from what actually occurs.

GIScience & Remote Sensing, 2012, **49**, No. 6, p. 822–853. http://dx.doi.org/10.2747/1548-1603.49.6.822

Establishing spatial population distributions requires the generation of a gridded population, and many researchers have explored the potential use of remote sensing data in the estimation of population distributions and have performed many investigations for this purpose. Methods for generating a gridded population can be divided into five types (Zhuo et al., 2005).

The first approach is average (e.g., weight average) allocation. This approach uses the average or weighted average principle to allocate administrative unit population statistics to a grid cell (Jin et al., 2003). This method is the most simple and crudest, and it often results in large mutations on the boundaries among the grid cells.

The second approach involves the analysis of factors that affect population distributions (Liao and Sun, 2003; Tian et al., 2004). The research method underlying this approach analyzes a series of factors that affect population distribution, such as land use, temperature, and climate, and to assigns a weight to each factor. Subsequently, the population density allocation coefficient of every grid can be obtained, and this coefficient is multiplied by the grid area, which yields the population density of each grid cell.

The third approach is grid interpolation (Liu et al., 2003a; Lv et al., 2003). The main research method of this approach is to divide the study area into a grid of a certain resolution and to use a variety of interpolation methods to calculate the grid population density.

The fourth approach is based on the population distribution rule (Paul, 1997; Feng, 2002). In this method, the classic rule of population density is utilized, which states that a distance decay model should be used to simulate the population density of each grid.

The fifth approach uses remotely sensed imagery to extract information on residences, such as types and density, or to extract the factors that affect population distributions to estimate population density (e.g., Joseph et al., 2012). Of the five approaches, the fifth method (i.e., using remote sensor data) has been widely used and has become an important method for simulating population distributions.

Taragi et al. (1994) employed two techniques for obtaining data through remote sensing to estimate the population of India: the Unit Count Technique and the Area Density Technique. Hardin et al. (2007) discussed three broad methodologies (dwelling identification, land type surrogates, and pixel-based estimation) to estimate intra-urban population totals and densities using overhead imagery and focus on the developed, urban world. Li and Weng (2005) explored the potentials of integrating Landsat ETM+ data with census data to estimate the population density of the city of Indianapolis, Indiana, USA and obtained a good accuracy of up to 96.8%. Zhang et al. (2007) proposed a power exponential model that is based on the scale of a district and the distance from the center of the district to each grid of the district, according to the distance decay function. Liu et al. (2003b) simulated population density by fusing remotely sensed data, meteorological data, soil data, and statistical data into a grid-generation model. Tian et al. (2004) used the idea of modeling a population separately as a town or country to construct two different models to simulate the population density of these areas. Tobler et al. (1997) transformed census data into a population grid. Landscan (Dobson et al., 2000; Dobson et al., 2003) can distribute census counts into 30×30 arc-second grid cells that are based on probability coefficients calculated from road proximity, slope, land cover, and nighttime lights.

Although these methods each have individual advantages for estimating populations and focus on the factors that affect population distribution, these methods also have many disadvantages. First, some of these methods rarely investigate the correlations and relative effects of all of the factors. Furthermore, combining the factors makes modeling more complex (Zhang and Yang, 1992). To address this problem, this paper used the most correlative indicator (i.e., land use) (Tian et al., 2005) to simulate the population, which could decrease the effect of redundant information. Second, some of the studies do not take into account the differences between the distribution patterns of urban and rural populations. For this problem, this article separately modeled the population of urban and rural areas. Third, some of the methods do not take into consideration floating populations.[3] For this problem, this article used the Fifth Census of the China population in 2000 instead of the Chinese population by county in 2000 (Chinese Ministry of Public Security, 2001). Fourth, these studies do not take into account the limited accuracy of remotely sensed data. For this problem, this research combined remote sensing data with a topographic map to improve the accuracy of the simulation.

This article consists of six parts. The Study Area section briefly introduces China, and the Data Sources section outlines the necessary data used in the research. The Methodology section introduces models for simulating population, and the Results section presents the simulated population (2000) of China and provides analyses of the results. The Validation and Comparison section validates our results and compares our results with those from the Chinese Academy of Sciences (CAS). Finally, the Discussion and Conclusions section summarizes and states the conclusions from our studies.

STUDY AREA

The study area for this paper is continental China (Fig. 1), which is located in East Asia and spans the the latitudes 3°51′ to 53°33′ North and the longitudes 73°33′ to 135°05′ East. China has a huge population of 1.3 billion and limited land resources (NBSC, 2005).

The overall characteristics of the distribution of the Chinese population can be described as follows: the eastern regions of China are more densely populated than the western; plains and basin areas are more densely populated than mountains and plateaus areas; agricultural areas are more densely populated than forests and pastoral lands; humid regions are more densely populated than the dry, cold regions; the regions that were developed earlier are more densely populated than those that were developed more recently; and regions along rivers, the sea, and other modes of transportation are more densely populated than less accessible areas.

Population data are an important factor that affect many aspects of China. Knowledge of the Chinese population distribution can guide local, regional, and national leaders in planning for resources and supplies, and this knowledge can be quite important for our country's social development and the people's well-being.

---

[3]"Floating populations" consist of individuals registered *de jure* as residents of certain locations, while *de facto* residing in others. An example are migrant workers from rural areas of China who live and work without official registration status (*hukou*) in China's eastern cities.

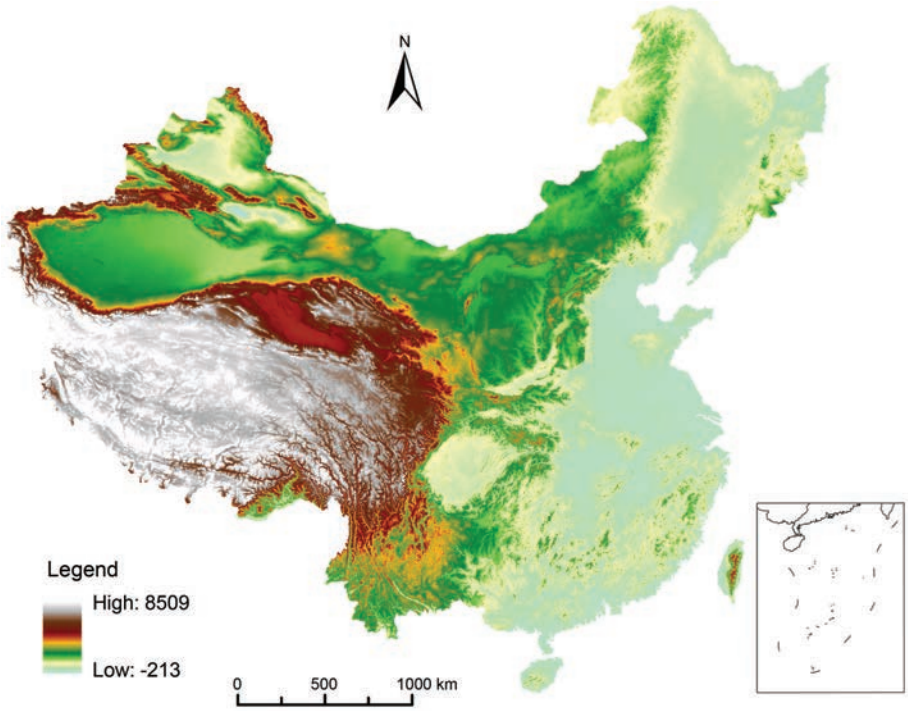

Fig. 1. Digital elevation model (DEM) of continental China.

DATA SOURCES

Population Data

Population data were obtained from the Fifth Census of China in 2000 (China, 2001). These data divide the population of the administration areas of a city into two parts: the city population and county population ((NBSC, 2002). Each part is also divided into two types of populations: urban and rural. In the census, the city population (Chinese: *shi renkou*) is defined as the population in the city proper and exurban districts and does not include the population in counties. For example, the total population of Beijing City is the sum of the following thirteen municipal districts: Dongcheng District, the Xicheng District, Chongwen District, Xuanwu District, Chaoyang District, Fengtai District, Shi Jingshan District, Men Tougou District, Haidian District, Shunyi District, Changping District, Fangshan District, and Tongzhou District. In addition, in the census, the floating population is recorded at their current place of residence rather than at their place of registration if they have lived at the current place for more than six months or a year (Tan et al., 2008). Compared with the data sources of many other papers, which use the original census data of the Chinese population by county in 2000 (Chinese Ministry of Public Security, 2001), our population data are more representative because they include the floating population. Therefore, our data should more closely reflect the actual size of the Chinese population, and our data could be available as an attribute of the administrative polygons at the county level. Other data,

which neglect the floating population who actually work and live in city areas, will lead to considerable distortions of city population in some cities that have a large inflow of migrant workers (Zhou and Yu, 2004).

Land Use Data

The second source is land use data, which is obtained from the Data Center of Resources and Environment at the CAS. These data are extracted from Landsat Thematic Mapper (TM) images from 2000 using land use/land cover classes appropriate for the 1:100,000 scale (Liu, 1996; Liu and Buhe, 2000; Liu et al. 2003a) (these classes are explained in Table 1). The original format of the land use data is similar to ArcInfo coverage. In this study, the data were converted into 25 raster files in the Environment Systems Research Institute (ESRI) Grid form at a 1 × 1 km resolution using the cell-based encoding method of percentage breakdown. Every file represented a land cover type; and the value of each grid cell in the raster file corresponded to the area of the type of land use in the grid cell (Tian et al. 2005). These data have been successfully employed in the study of land use changes in China (Liu et al., 2003a). In this paper, we used the ild51 and ild52 data (for urban and rural residential areas, respectively) to simulate the population.

China County Vector Boundary Data

The third source is the China county vector boundary data (see Fig. 2); these data were obtained from the Resources and Environment Database of China and contain detailed attributes of every county, such as the name and area.

China Residence Polygon Data

China residence polygon data (see Fig. 3) are derived from a 1:250,000 topographic map of China, which was updated at the end of 1997. It contains over 30,000 towns and 647,705 villages. Because the polygons are small, we used a small part of the Liaoning Province as an example (see Fig. 4). The reasons why we choose this data are twofold: On the one hand, they are obtained by field survey, so have a relatively high reliability; on the other, their integrity, logical consistency, and positional precision all comply with related technical regulations and standard requirements of the State Bureau of Surveying and Mapping. However, there is also a problem caused by the use of data from 1997 instead of from 2000. During 1997–2000, China was in the stage of comprehensive promotion of urbanization, which may affect the estimation result, and could for example lead to underestimation of urban population and overestimation of rural population. Constrained by the lack of data for 2000, we ignore these effects and regard the residence polygon in 2000 same as in 1997.

METHODOLOGY

The entire process for simulating the population distribution of China is shown in Figure 5. All of the information presented in the figure is discussed in detail in the following section.

Table 1. Land Cover and Residential Classification Systems

Primary types (code name)	Secondary types (residential code name types)[a]
1 Farmland	11 Paddy field A
	12 Non-irrigated field A
2 Woodland	21 Forest A
	22 Shrub A
	23 Sparse woodland A
	24 Other woodland A
3 Grassland	31 Dense grassland A
	32 Moderate dense grassland A
	33 Sparse grassland A
4 Waters	41 River N
	42 Lake N
	43 Reservoir and pond N
	44 Glacier and snow N
	45 Beach N
	46 Bottomland N
5 Built-up area	51 Urban area R
	52 Rural residential area R
	53 Other built-up area R
6 Unused land	61 Sandy land N
	62 Gobi N
	63 Saline-alkali land N
	64 Marsh N
	65 Bare land N
	66 Bare rock and gravel land N
	67 Other unused land N

[a]N = exclusive areas; R = residential areas; A = non-residential areas. R and A are habitable areas. In this paper, the land cover types would be represented by "ild" and their codes. For example, ild11 would represent paddy field.

Data Pre-processing

Before undertaking the research, we first needed to pre-process the basic data. From the Chinese county vector boundary map and its corresponding table, we found that there were some improper sections, such as the presence of some null values, which would affect our research. Therefore, we modified these data in several ways.

First was the processing of null values by their deletion. Second, for repeated county names, we distinguished between three cases: if the county did not use its

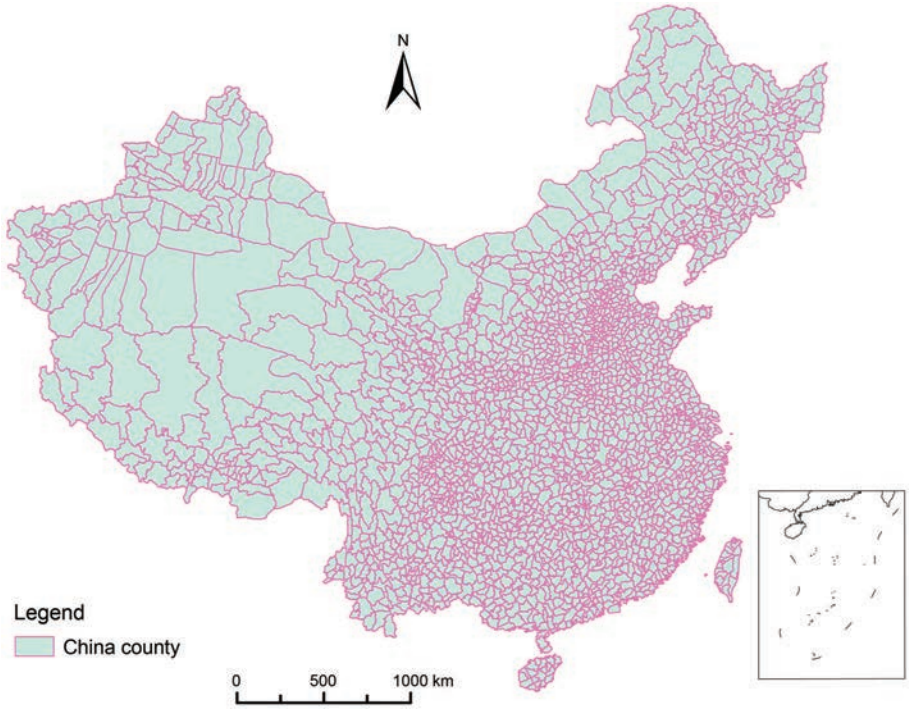

Fig. 2. China county vector boundary data.

land, we deleted it; if the county used its land and it was in close proximity to another county, we combined it with that county. For example, Huaibei City included two sections of the map (see Fig. 6A), but these two sections were not consecutive. If we calculated the population distribution of this region without considering its special location, the simulated result would deviate from the actual situation. We noted that this county was near Xiao County; therefore, we combined the data of Huaibei City with that of Xiao County, and the processed result is shown in Figure 6B. Lastly, if any regions independently used land and were far from any county, we preserved them and distributed their population according to their proportion. After these processes were complete, a proper vector map was constructed, and we began our research.

Modeling on the Basis of Land Use Data

Many factors, such as land use, slope, temperature, and other factors, affect the distribution of population. To decrease the effect of redundant information, we took the most correlative indicator, land use (Tian et al., 2005) to simulate the distribution of China's population.

Modeling at the Scale of County Population

At present, the smallest mapping unit that is available countrywide is the county; therefore, county-level population data are the best data to represent population

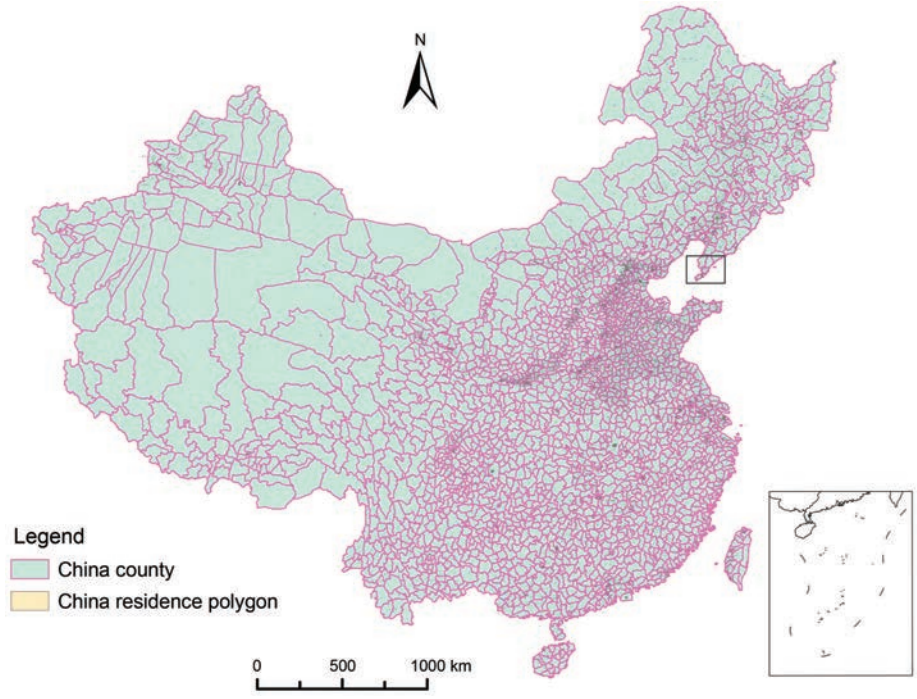

Fig. 3. China residence polygon data. Liaoning Province is shown in the black rectangle.

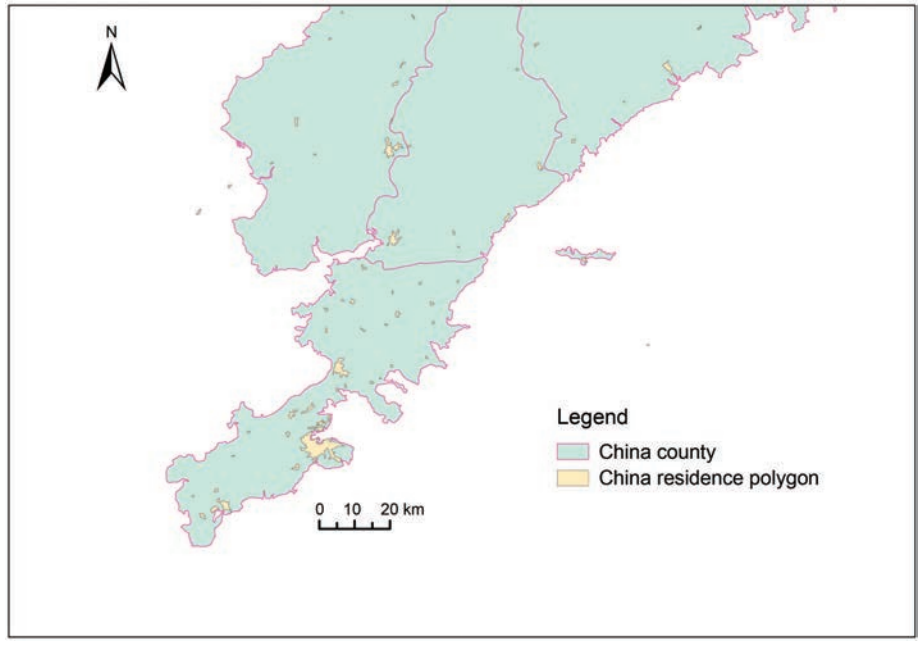

Fig. 4. Liaoning province residence polygon data. Figure 4 is an enlarged view of the black frame in Figure 3.

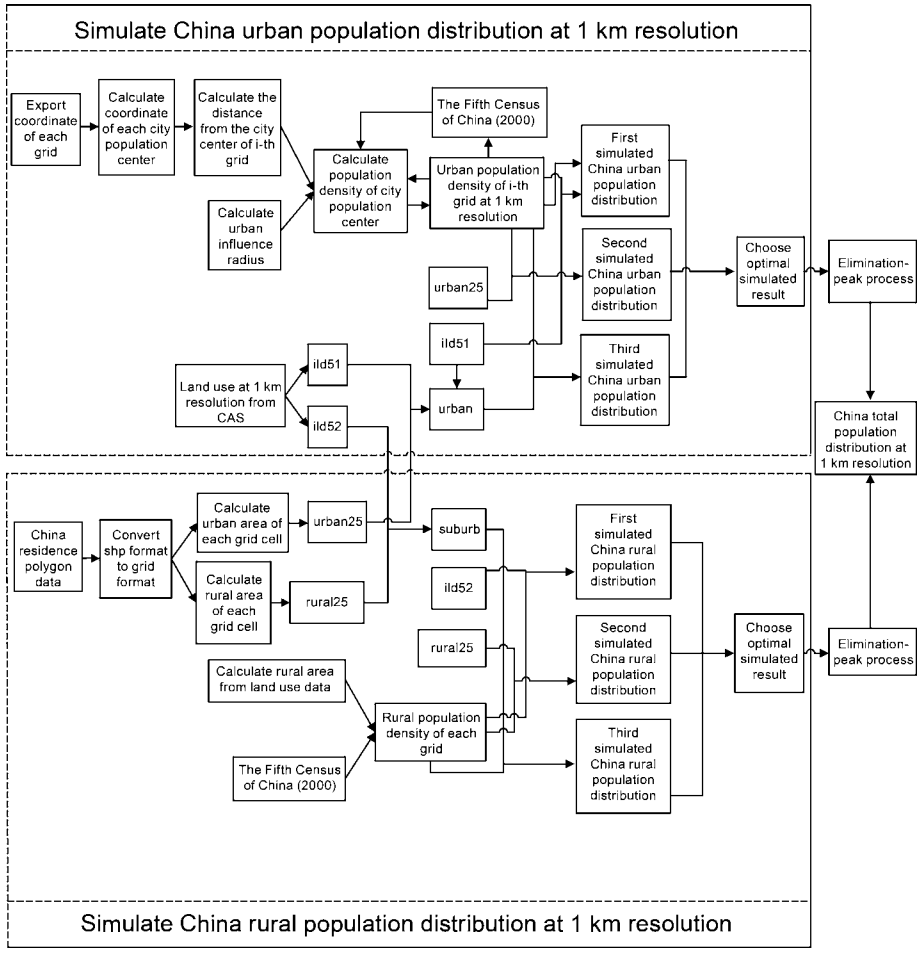

Fig. 5. Processes involved in simulating the population distribution of China.

distribution. These data divide the total county population into two parts: urban and rural. This process can avoid the reallocation of the populations of different counties and between rural and urban areas, which is caused by the improper process of uniformly modeling the parameters.

Modeling Urban and Rural Populations

Because of the special natural environment and variation in social economic development among different regions of China, the population distribution is extremely uneven. It can also be divided into several parts, such as town and country. Population in town and country has different associated factors and distribution patterns. For example, urban population distribution may relate to the location of economic centers, traffic conditions, or other factors, whereas rural population may relate to river location, agriculture production, or other factors. Therefore, we built these models

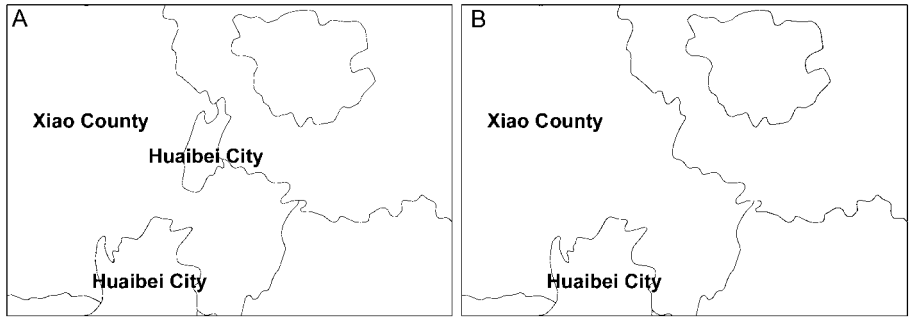

Fig. 6. Pre-processing of Chinese county vector boundary map.

separately so we could avoid over- or underestimation of population density within urban or rural grid-cells, which are caused by the improper combinations of models and parameters.

Modeling on the Basis of Three Types of Land Use Data

Although remote sensor data have been used for population estimation in many studies, these estimates have limited accuracy and cannot fully reflect the population distribution. As a result, we will improve the accuracy by adding the China residence polygon data.

In this paper, three types of land use data were applied to simulate population. The first source of land use data (ild51 and ild52) were obtained from the Data Center of Resources and Environment at CAS and were introduced in detail in the data source section. The second type of land use data (urban25 and rural25) was the raster file that was converted from the China residence polygon data, where each grid cell was given a different value. The value represented the area of the type of land cover in the grid cell. For urban land use, the China residence polygon data were converted to raster data at a 100 × 100 m resolution, then resampled at a 1 × 1 km resolution by averaging the area, which could be explained by the following equation:

$$A_i = \frac{A_{urban}}{A_{total}} \tag{1}$$

where A_i represents the urban area of the grid cell in a town, A_{urban} is the urban area of every town, and A_{total} is the total land area of each town. The resulting urban25 map is shown in Figure 7A.

The China residence polygon data for rural land use was converted to raster data at a 1 × 1 km resolution, and each grid cell is assigned a value of 2500. The 2500 value is obtained by the following equation:

$$A_i = \frac{A_{rural}}{A_{country}} \tag{2}$$

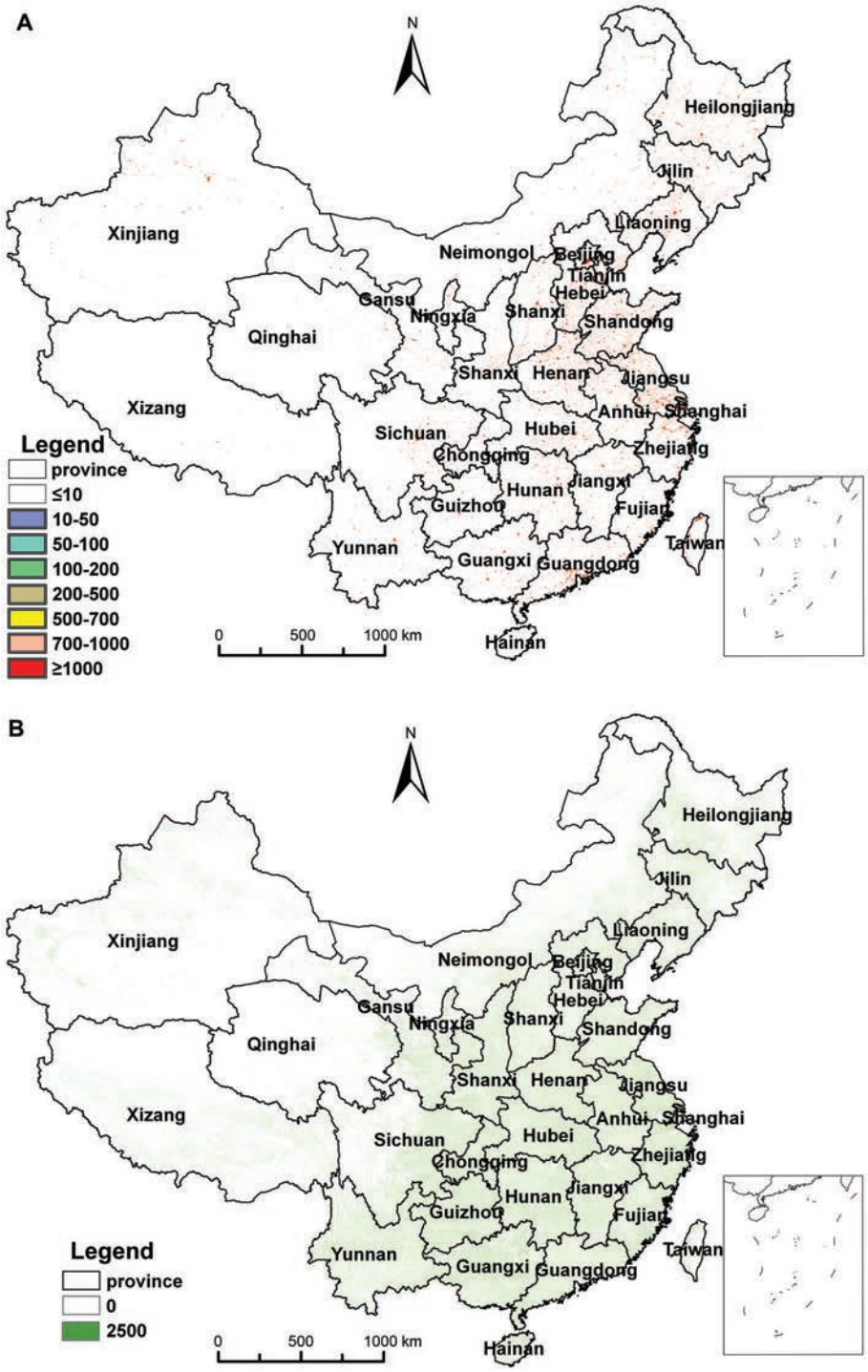

Fig. 7. Processed China urban (A) and rural (B) residence polygon data.

where A_i represents the rural area of a grid-cell in a county, A_{rural} represents the total rural residential area of China in 2000 (Tian et al., 2003), and A_{county} is the total land area of all counties in China. The resulting rural25 map is shown in Figure 7B.

The third type of land use data (i.e., urban and suburban) was a combination of the above two types of data. For these data, we used the merge function in ArcGIS software and obtained the maximum value. Thus, urban and rural land use each had three sets of data.

Development of the Urban Population Model

First, we used regression analysis to separately estimate the populations of rural and urban areas. This analysis was divided into two steps: the effective variables were chosen based on SPSS software, and regression equations were built for population and land use. The result of the first step was not ideal. The correlation coefficient (*R*) between the rural population and the ild51 was 0.67, whereas the correlation coefficient (*R*) between the rural population and the ild52 was 0.61. In theory, the rural population should be more correlative to the ild52, but, on the basis of the linear model, the rural population was more correlative with the ild51, which contradicts this theory. Therefore, we disregarded this method and used a second method, where different models were built according to the rural and urban situations. The process is detailed below.

The urban population is affected by the following factors: geographical position, natural resources, economies of scale in production and consumption, and the usage of automobiles (Fujita ,1989; Alig et al. 2004). Therefore, it is difficult to build a simple, mathematical equation to simulate the population.

Usually, urban population density is proportional to urban scale, i.e., a larger scale corresponds to a greater density (Ye, 2001). In the internal portion of the urban area and counties, a population density difference exists. Generally, the population decreases from the center to the exterior of town. However, because so many factors affect population density, the population does not have a simple, mathematical relationship with scale. Yet, the distribution of population density has some similar rules, and it is mainly affected by the urban scale and the distance from the urban center. The rule can be expressed by the following formula:

$$V_{city} = f(S,L) \tag{3}$$

where V_{city} is the urban population density coefficient, S is the urban scale, and L is the location in the city.

Many simulation models for population distribution have been developed (Yue et al. 2003), such as the Gauss model (Sherratt, 1960; Tanner, 1961). Recently, another type of model, named fractal models for the decay of urban population density (Feng, 2002), was developed and can be described by the following equation:

$$\rho(r) = \rho_0 \times \exp\left[-\left(\frac{r_i}{r_o}\right)\right]^{\sigma} \tag{4}$$

where $\rho(r)$ stands for the population density of the *i*-th grid whose distance from the urban center is r_i, ρ_0 is the city center population density, and r_0 is the urban influence

radius, which can be obtained by determining the radius of a circle of the same area as the town. σ is the constraint parameter, which reflects the information entropy of the spatial change of the urban population. Although the feasibility of this model was doubted, city economists have proven its usefulness (Wang and Guldmann, 1996). Therefore, using this model, we obtained the population distribution coefficient of every grid cell instead of the specific population.

Before we simulated the population density at a 1 × 1 km resolution, we first obtained the city center population density ρ_0 value. Of course what we defined as the city center did not refer to the shape center but to the mass center, and its coordinate was calculated in Eq. (10). As this mass center was from urban land use data, it could decrease the error of the center. Here, we defined the city center as a mono-center in the fractal model. This definition in the above fractal model may not be feasible in some cities, such as some small cities in developed countries, as their population is not concentrated in a single mono-center. However, China is a developing country and most cities had only mono-centers in early stages of development. In China around 2000, this was because few people could drive a car, so in order to work they needed to live near the center. In conclusion, the urban population is still concentrated in urban centers in China. So the fractal model can be used in China's current situation. Some other centers will be calculated for the rural population. Of course, more centers will grow as cities grow. If we use data of 2010, the model error will be larger than the error using data from 2000.

We first multiplied both sides of Eq. (4) by the urban area, and then made a summation to obtain the following equation:

$$\sum \rho(r) \times A_i = \sum \rho_0 \times \exp\left[-\left(\frac{r_i}{r_o}\right)\right]^{\sigma} \times A_i \tag{5}$$

where $\rho(r)$ stands for the population density of the i-th grid cell whose distance from the urban center is r_i, A_i is the urban area of the i-th grid-cell, ρ_0 is the city center population density, r_0 is the urban influence radius, and σ is the constraint parameter, which reflects the information entropy of the spatial change of the urban population. Both sides of Eq. (5) referred to the total population of a city or a town and because the total population had been obtained, Eq. (5) could be replaced by the following equation:

$$P_{city} = \sum \rho_0 \times \exp\left[-\left(\frac{r_i}{r_o}\right)\right]^{\sigma} A_i \tag{6}$$

where P_{city} stands for the total population of a city or a town, which can be obtained from the population data we introduced as the population source. The other parameters are defined as above. Therefore, from Eq. (6), we could deduce the center population density by the following equation:

$$\rho_0 = \frac{P_{city}}{\sum \exp\left[-\left(\frac{r_i}{r_0}\right)\right]^{\sigma} A_i} \tag{7}$$

where ρ_0 is the city center population density; P_{city} is the total urban population of a city, which can be obtained from the Fifth Census of China in 2000; r_i is the distance

from the city center of the *i*-th grid cell; r_0 is the urban-influence radius; A_i is the urban area of every grid of a city; and σ is the constraint parameter that reflects the information entropy of the spatial change of the urban population. We can base the value of σ on the basic theory of urban development, which states that over a city's development, it should experience "developing," "developed," and "old" stages. At each of these different stages, the urban population density has different spatial distribution; during urban suburbanization, the urban center's population density decreases and its distribution shape resembles a crater. After years of urban construction, especially after the onset of reform and the open-up policy, China is still in an early developed stage, and some middle-sized or small cities are still in the developing stage. Taking this into consideration, σ is valued at 1, and more detailed information about the acquisition of each parameter will be provided in the following section.

We obtained the value of r_0, by the following formula:

$$r_0 = \sqrt[2]{A_j / \pi} \tag{8}$$

where A_j stands for the total area of the *j*-th city.

We could obtain the value of r_i by the following formula:

$$r_i = \left|(X_i, Y_i) - (X_0, Y_0)\right| \tag{9}$$

where r_i is the distance from the urban center of the *i*-th grid in a city, X_i is the X coordinate of the *i*-th grid, Y_i is the Y coordinate of the *i*-th grid, X_0 is the X coordinate of the urban center, and Y_0 is the Y coordinate of the urban center. We could obtain X_i and Y_i from ild51 with the "Convert Grid theme to XYZ Text file" tool of Arcview software from ESRI. For the coordinate of the urban center, we used the following formula:

$$X_0 = \frac{\sum_1^k X_i \times A_i}{\sum_1^k A_i}, \quad Y_0 = \frac{\sum_1^k Y_i \times A_i}{\sum_1^k A_i} \tag{10}$$

where X_0 is the X coordinate of the urban center, X_i is the X coordinate of the *i*-th grid, A_i is the urban area of the *i*-th grid, K is the number of grids in a city, and Y_0 is the Y coordinate of the urban center. All of the resulting data was obtained using the Matlab computer language.

After obtaining the urban center density ρ_0, we could return to Eq.(1), and using the Matlab computer language, we obtained the population density of each grid ρ(r). These data were in text format, but for use in spatial analysis, they had to be altered. First, we regarded the data as an attribute of the Chinese county vector boundary map and, thus, inputted them into the attribute table of the map. Then, using the conversion function of ArcGIS, we obtained the urban population density raster data.

Development of the Rural Population Model

For the rural population coefficient model, we used the average density model, which could be described by the following equation:

$$P_{rurali} = \rho \times A_i \tag{11}$$

where P_{rurali} stands for the rural population of every grid cell, ρ is the rural population density, and A_i is the rural area of every grid cell. We used the following formula to determine the value of ρ:

$$\rho = \frac{P_{country}}{A_{country}} \tag{12}$$

where P_{county} stands for the rural population of each county in China and A_{county} is the rural area of every county in China. Based on Eq. (11) and using the raster calculator of the ArcGIS software, we obtained the rural population density diagram of China.

RESULTS

Using the above urban and rural population models, we obtained simulated Chinese urban and rural population maps. The total population was 1.24 billion in 2000, the urban population was 0.32 billion, and the rural population was 0.92 billion. As described in the methodology section, this paper used three sets of land use data to simulate population; therefore, we now have three sets of urban and rural population data.

Results from the First Set of Data

The land use data of the first set of data was from the ild51 and ild52 data for 2000, and the simulated population urban and rural results are shown in the following maps. Figure 8A shows the simulated urban population from the first set of data for China (2000), while Figure 8B shows the simulated rural population from the first set of data for China (2000).

Results from the Second Set of Data

The land use data from the second set of data were from the urban25 and rural25 raster file that was converted from the China residence polygon data, and each grid cell was given a different value. Here, the value indicated the area of the type of land cover in the grid cell. As for the urban land use, the China residence polygon data were converted to raster data at a 100 × 100 m resolution, we resampled the data at 1 × 1 km resolution, and finally, we assigned each grid cell a value. As was performed for the rural land use, the China residence polygon data were converted to raster data at a 1 × 1 km resolution, and each grid cell was assigned a value of 2500. The simulated urban and rural population results are shown in the following maps. Figure 9A shows the simulated urban population from the second set of data for China (2000), while Figure 9B shows the simulated rural population from the second set of data for China (2000).

Results from the Third Set of Data

The land use data from the third set of data were urban and rural, which was a combination of the two types of land use data that were mentioned above. The

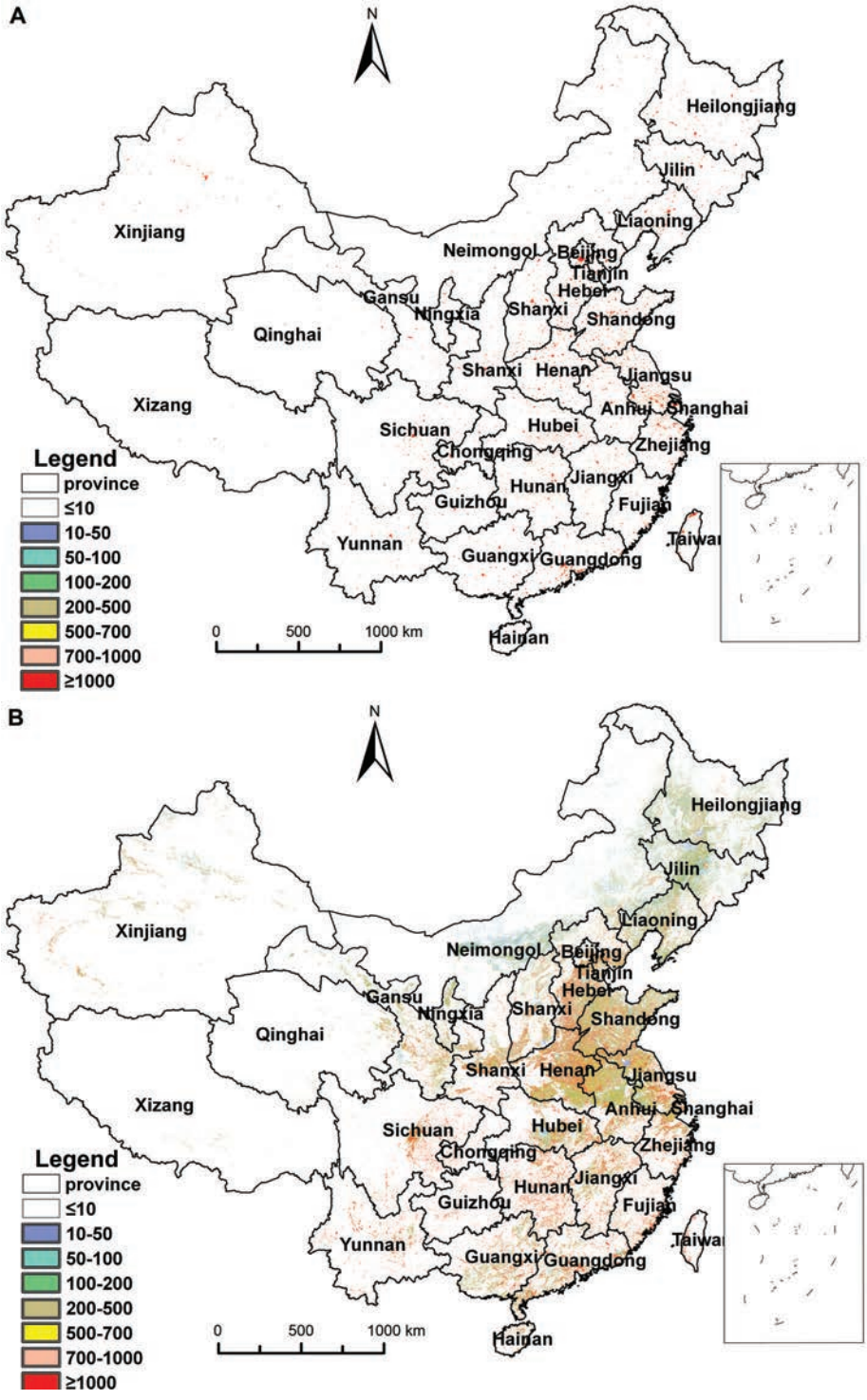

Fig. 8. Simulated urban (A) and rural (B) population from the first set of data for China (2000).

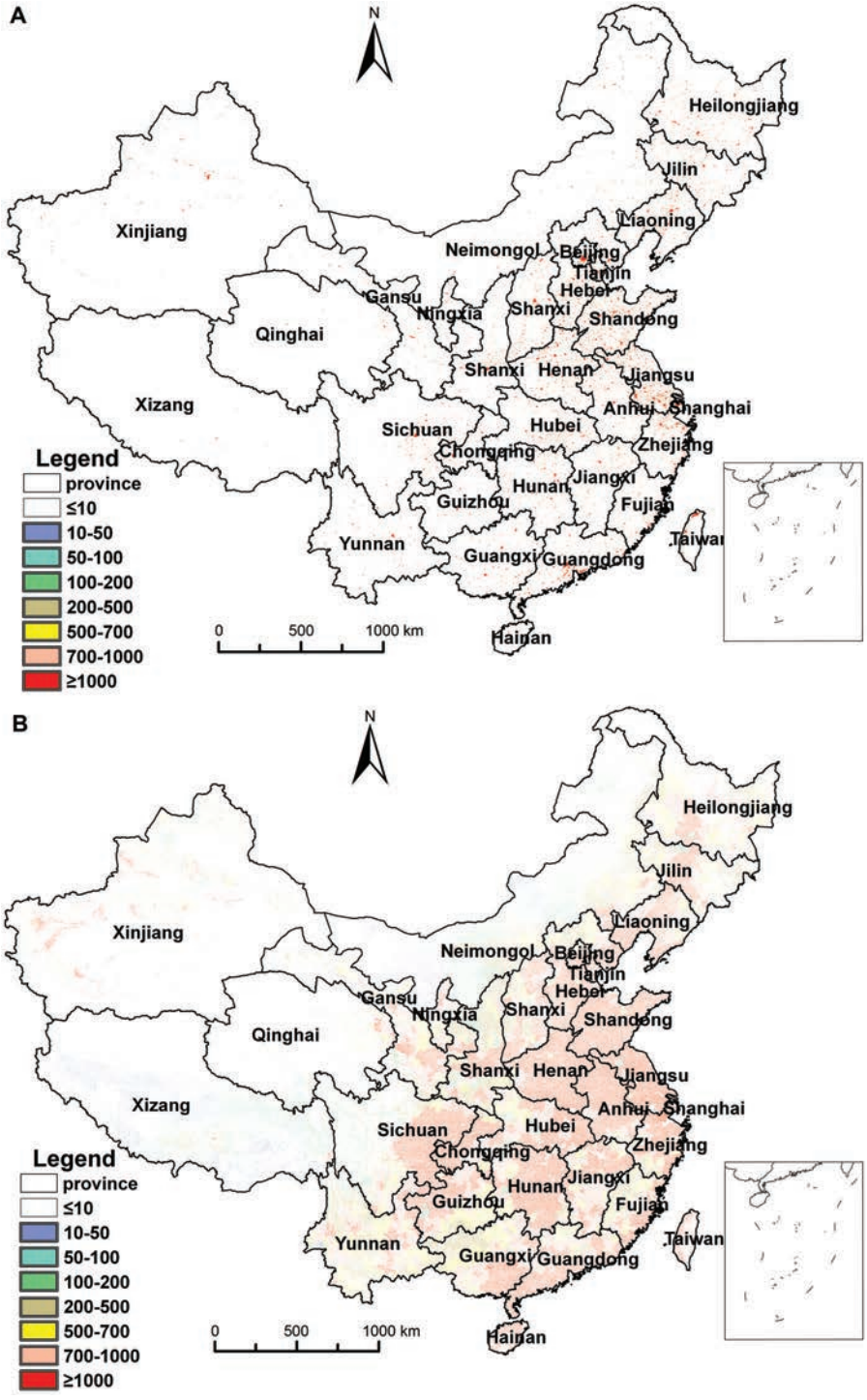

Fig. 9. Simulated urban (A) and rural (B) population from the second set of data for China (2000).

simulated urban and rural population results are shown in the following maps. Figure 10A shows the simulated urban population from the third set of data for China (2000), while Figure 10B shows the simulated rural population from the third set of data for China (2000).

Comparison of the Three Sets of Data and the Results

Comparison of the Simulated Urban Population of the Three Sets of Land Use Data. The simulated urban populations from the three sets of land use data are shown in Figure 11. Figure 11A shows the simulated urban population of the first set of land use data, Figure 11B shows the simulated urban population of the second set of land use data, and Figure 11C shows the simulated urban population of the third set of land use data.

Comparison of the Simulated Rural Populations from Three Sets of Land Use Data. The simulated rural populations from three sets of land use data are shown in Figure 12. Figure 12A shows the simulated rural population of the first set of land use data, Figure 12B shows the simulated rural population of the second set of land use data, and Figure 12C shows the simulated rural population of the third set of land use data.

Determination of the Final Results. From the comparisons, we observe that the third set of data was smoother than the other two sets of data, and it more closely resembled the actual population distribution. Therefore, we considered the third set of data as our final population distribution result for the following reasons. First, remote sensor–derived data, which in this paper we refer to as land use data, are at a 1 × 1 km resolution, which is insufficient to show inner details or population distribution. This limited spatial distribution may exaggerate or under-represent the population in places such as mountains and hills. In Southwest China, these regions mainly include Chongqing Province, Sichuan Province, Guizhou Province, Yunnan Province, and the Tibet Autonomous Region and feature more plateaus and mountains. Because of the special terrain of these regions, when remote sensing technologies are used to observe the residential distribution of these areas, only some places will be covered, which could lead to less land use data. Furthermore, these methods will result in a higher population density. The second reason was the introduction of China residence polygon data, which is obtained by field surveys and can more accurately reflect population distributions. Therefore, this method could remedy the defects inherent in remote sensor derivatives.

Post-processing

Once the final urban and rural population distribution data were obtained, we found that some extreme values existed, which were the result of the limited accuracy of the land use data. From the Fifth Census of China, it was determined that Angren County, which is located in the Tibet Autonomous Region, has a rural population of 43,681; however, its rural land use is only 0.8 square kilometers. These data were obtained from the ild52, which does not reflect the real population distribution. Therefore, we used the elimination-peak process and completed this step by programming. The maximum, grid-cell rural population was ruled to be 10,000, and the main

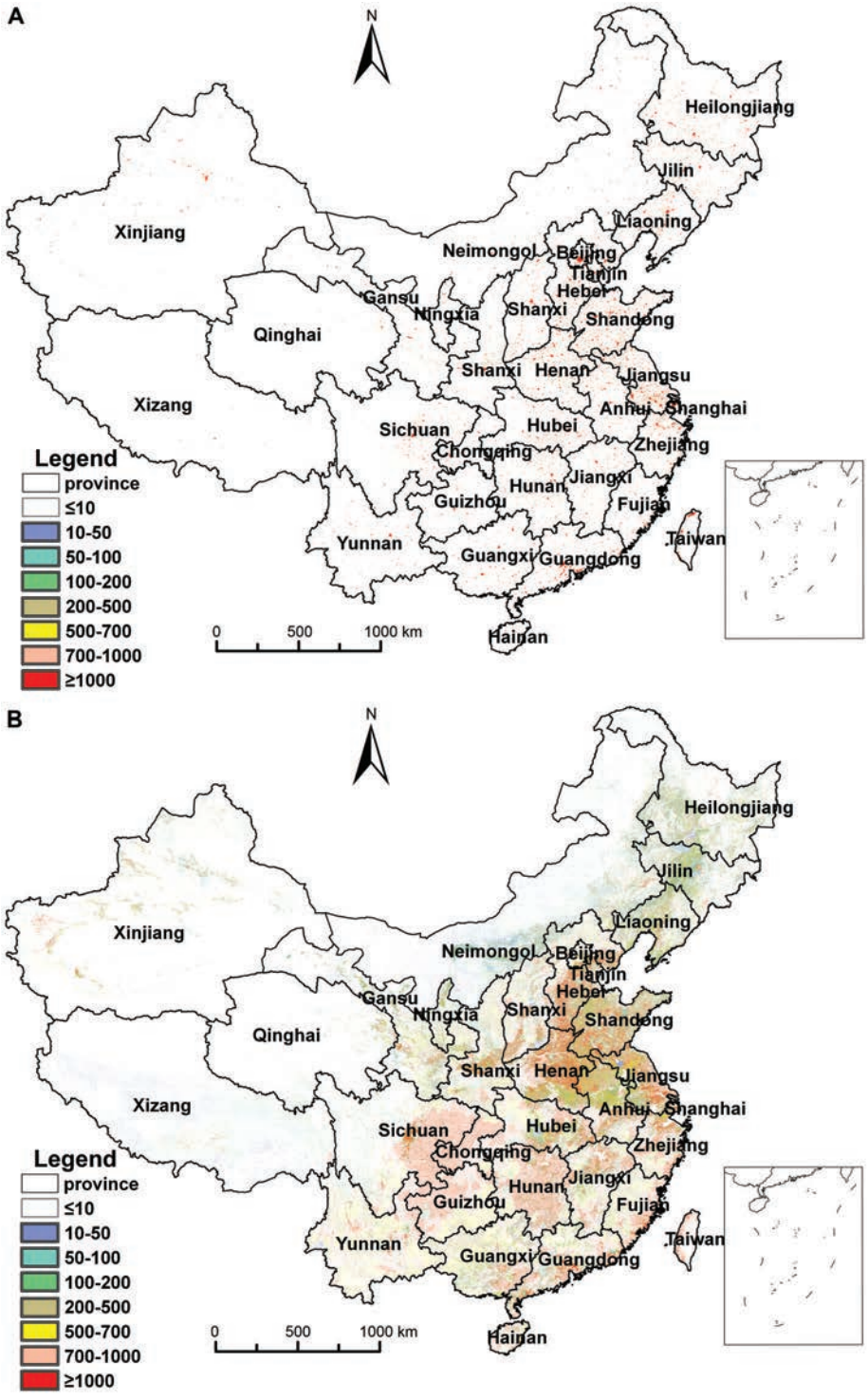

Fig. 10. Simulated urban (A) and rural (B) population from the third set of data for China (2000).

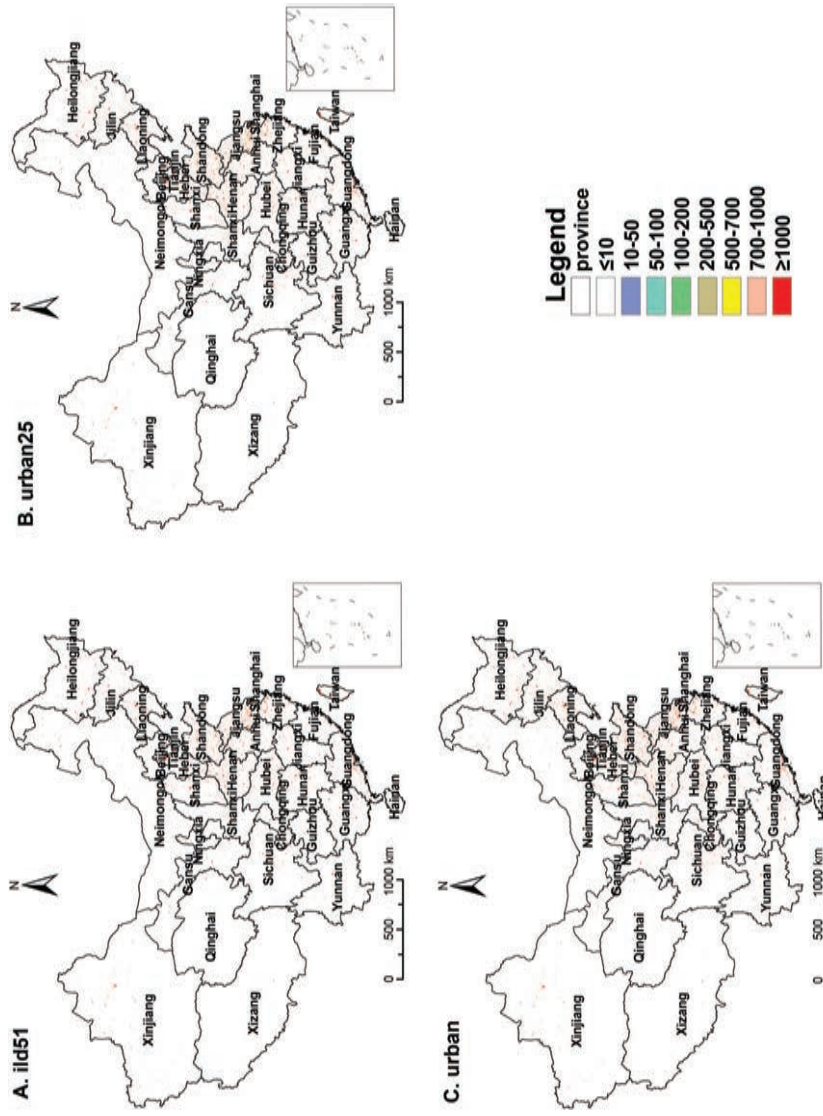

Fig. 11. Simulated urban populations of the three sets of data for China (2000).

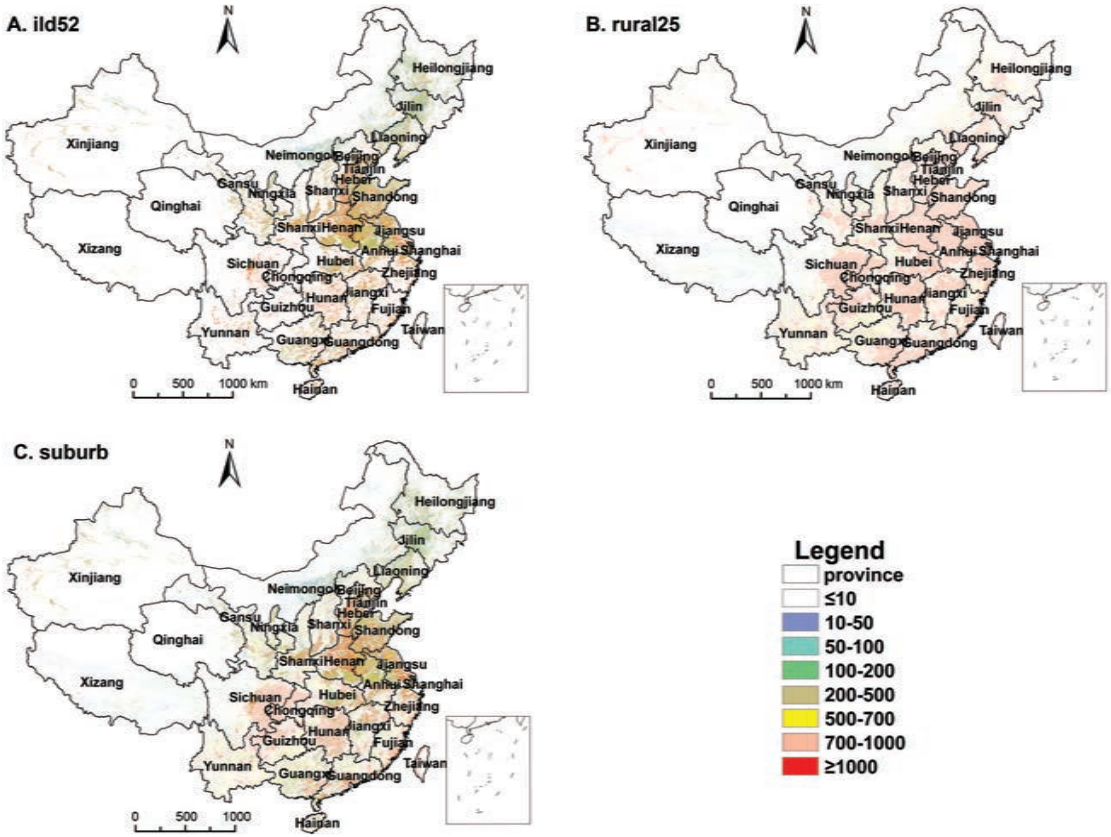

Fig. 12. Simulated rural populations of the three sets of data for China (2000).

idea was to assign a grid-cell population of over 10,000 to neighboring grid-cells. The maximum grid-cell urban population is ruled at 40,000, and it is based on the Fifth Census of China in 2000. According to the Fifth Census of China in 2000, Futian district in Shenzhen City has 909,571 people with the total area of 78.8 km^2. It contains almost 11,543 people within one square kilometer; however, this number is just the average value, and in some dense places, this number may be larger. Furthermore, this number does not include floating population. If we include the floating population within this one square kilometer, the population number will be much larger. As a result, we set the maximum value as 40,000, and we obtained the processed data. For the post-processing of total urban population, we took Kunming City as an example, with the comparisons between before elimination-peak process and after elimination-peak process being shown in Figure 13. Figure 13A shows the simulated urban population before elimination-peak process, while Figure 13B shows the simulated urban population after the elimination-peak process. For the post-processing of total rural population, we took Nima County and Angren County as examples, and the comparisons between before elimination-peak process and after elimination-peak process are shown in Figure 14. Figure14A shows the simulated rural population before elimination-peak process, while Figure 14B shows the simulated rural population after the elimination-peak process.

Total Population Distribution in China

At present, we have obtained the simulated urban and rural populations of 1 × 1 km grid-cells, conducted a summation, and obtained the simulated, total population of the 1 – 1 km grid-cells in China (2000). The result is shown in Figure 15.

From Figure 15, we could see that the majority of the Chinese population was distributed in the eastern region, with fewer people residing in the western region. A geographical dividing line, which was proposed by Hu Huanyong (1983), extends from Mohe County in Heilongjiang Province and to Tengchong County in Yunnan Province. The population is mainly aggregated in the Yellow River, Huaihe River, and Haihe River regions and in the Szechwan Basin, the middle and lower Yangtze River, the Northeast plain, and coastal areas. The densest population was located in the Yellow River, Huaihe River, and Haihe River regions. This phenomenon might be closely linked to the natural and social conditions of eastern China. The topography of China is high in the west and low in the east, and plains characterize eastern China while plateaus and deserts are typical of western China; therefore, eastern China can support more people. Eastern China has been the economic center for a long period, and it has convenient transportation. Additionally, other social factors determine this population gap, such as the reform and open policy.

VALIDATION AND COMPARISON

Validation

To verify the effectiveness of our results, we compared our results to the simulated, total population of the 1 × 1 km square grid-cells in China (2000) from CAS because their results are publically available and have been validated (Tian et al., 2005). More

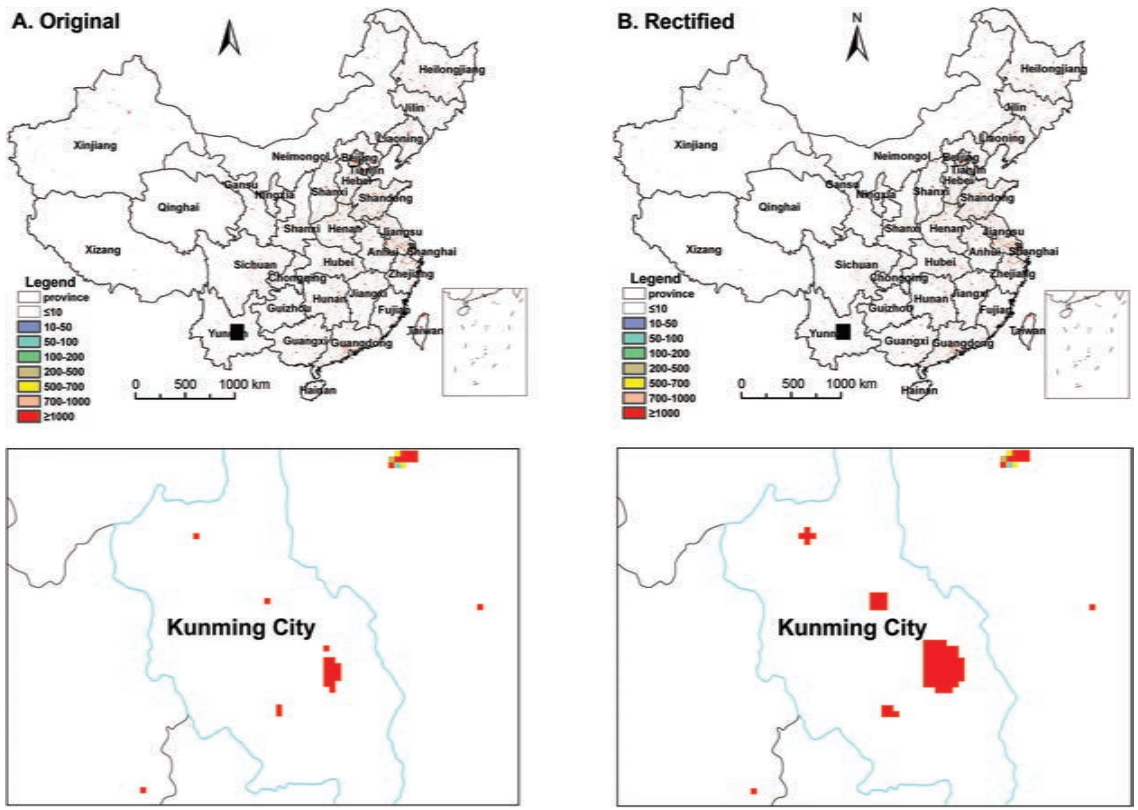

Fig. 13. Original and rectified urban populations of 1 × 1 km grid cells in China (2000). Dark rectangles in upper two images delineate the areas enlarged in the lower two images.

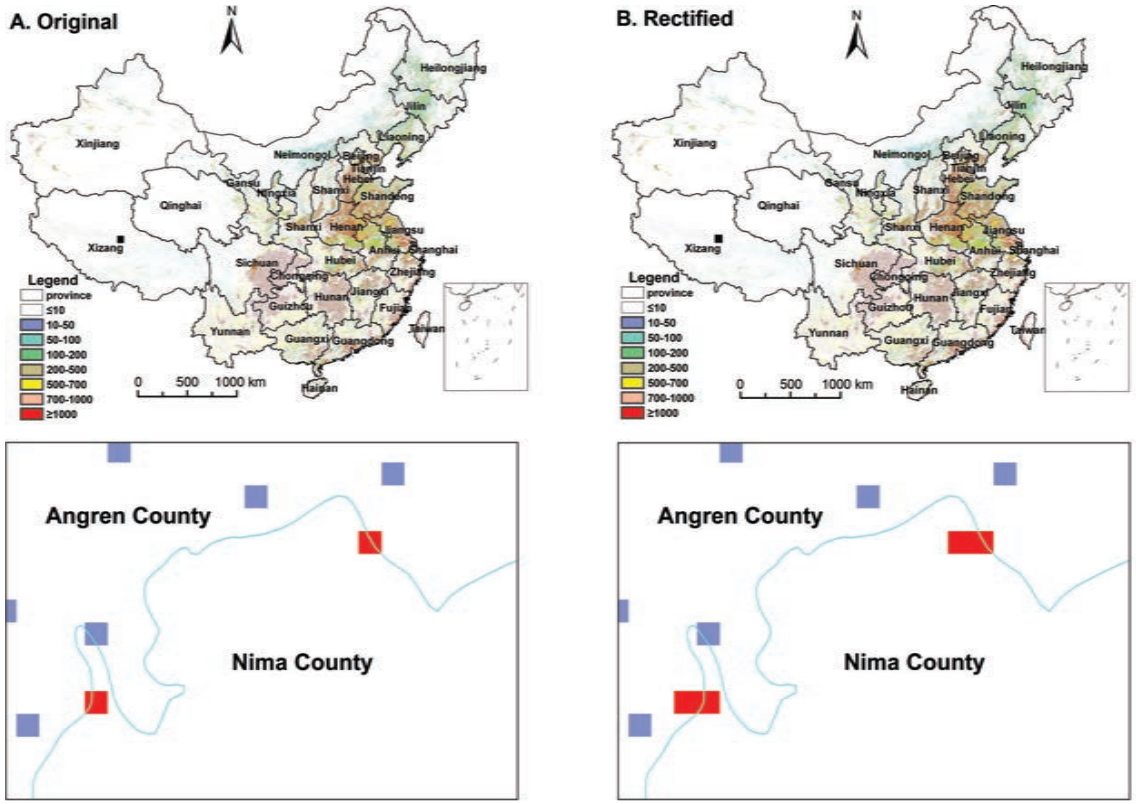

Fig. 14. Original and rectified rural populations of 1 × 1 km grid cells in China (2000). Dark rectangles in upper two images delineate the areas enlarged in the lower two images.

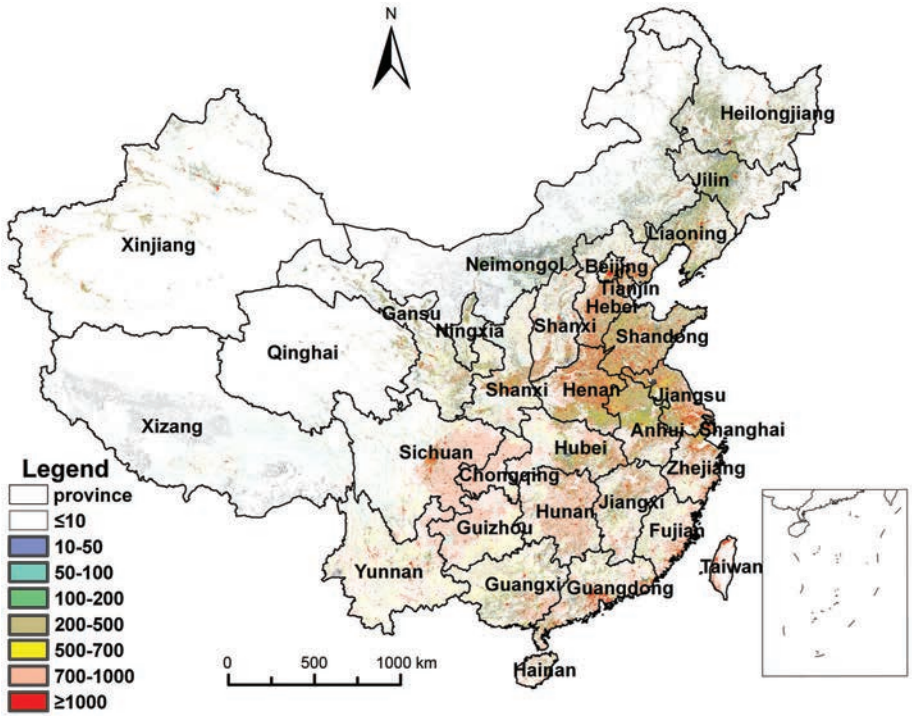

Fig. 15. Our simulated total population of 1 × 1 km grid cells in China (2000).

details on this comparison are provided in the following sections. Of course, this kind of validation may be limited. It is very hard to get finer spatial resolution population data and, as constrained by budget and current capacity, what we can do now is to compare our results with those of the CAS. Their results are also compared with some typical cities to verify the efficiency of estimation results. In the near future, with higher-spatial-resolution GIS and remote sensing data, it should be advantageous to conduct additional validation.

Comparison

At present, the dataset that we obtained simulated the Chinese population at a resolution of 1 × 1 km and was simulated by CAS (Tian et al., 2005). Based on the different distribution pattern of towns, counties, and zones, Tian et al. (2005) firstly built a China population simulation model at 1 × 1 km resolution, then calculated urban and rural population coefficients, respectively. As for the urban population coefficient, they built the urban population density coefficient model, which was based on urban area. As for the rural population coefficient, they firstly divided China into twelve ecological-agricultural regions. Then in each region, they built a one-element linear regression model between rural population and all kinds of agricultural lands to choose the model indicators. Thirdly, based on these selected indicators, they built multiple-element linear regression models to obtain the regression coefficient. Lastly,

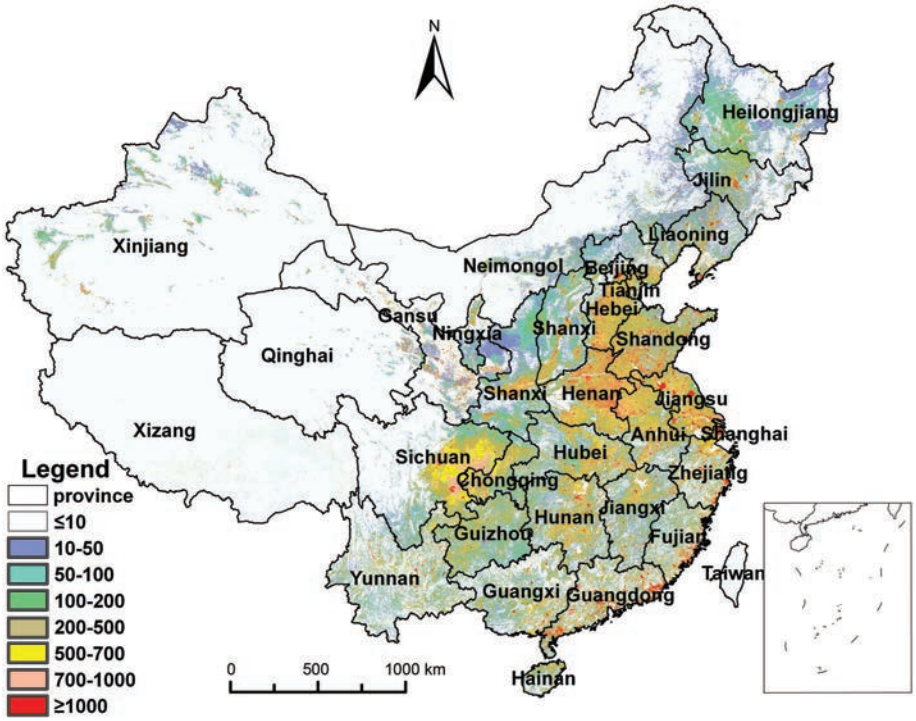

Fig. 16. CAS-simulated total population of 1 × 1 km grid-cells in China (2000).

based on the area of each land type, they obtained the rural population coefficients. The simulation results of those data and ours are shown in two figures (see Figs. 15 and Fig. 16). Figure 15 shows the simulated total population of 1 × 1 km grid-cells in China (2000) from our work, while Figure 16 shows the simulated total population of 1 × 1 km grid cells in China (2000) from the CAS.

First, we compared these two figures macroscopically, and as a whole, the results are in accord. The population distributions of the coastal region are dense. The red areas represent a population of over 1000 and from the comparison, it was evident that the red areas in these two figures were generally overlapping, such as in Beijing City (see Fig. 17). Figure 17A shows the simulated total population of our work, whereas Figure 17B shows the simulated total population of CAS.

When comparing Figures 15 and 16, we also concluded that the simulated population from our work was generally higher than that derived from the CAS, which is a result of using different sources of population data. Because we obtained population data from the Fifth Census of China, which includes fixed and floating population data, our population data are larger than that of CAS. However, our data may reflect the population distribution more accurately.

The most distinct advantage of our work is that we calculated the population distribution of Taiwan, allowing our leaders to obtain knowledge of its population distribution. A second advantage was that our result of the simulated population in China (2000) covers more areas, such as the Geji County, which is located in the

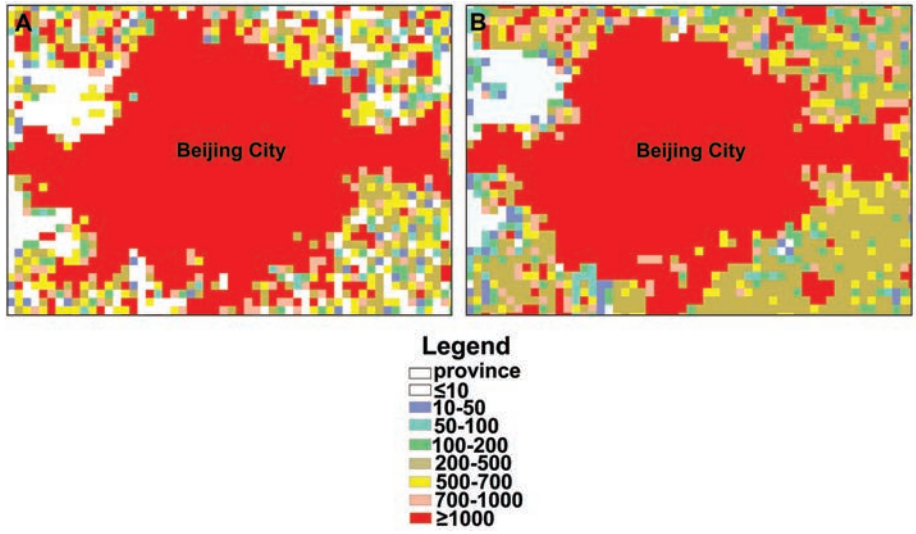

Fig. 17. Simulated total population of Beijing City through our work (A) and CAS (B).

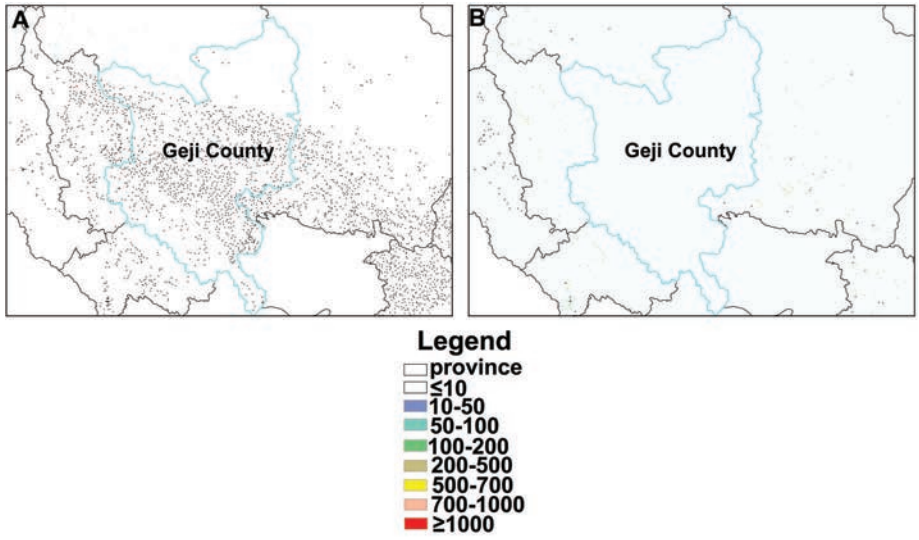

Fig. 18. Simulated total population of Geji County through our work (A)and CAS (B).

Tibet Autonomous Region (CAS did not simulate the population in Geji County). Figure 18 shows the difference between the simulation of the population distribution of Geji County in our work and that conducted by the CAS. Figure 18A shows the simulated total population of Geji County from our work, while Figure 18B shows the simulated total population of Geji County from the CAS. However, from our studies (Nets, 2008), we deduce that this region contains a significant population distribution.

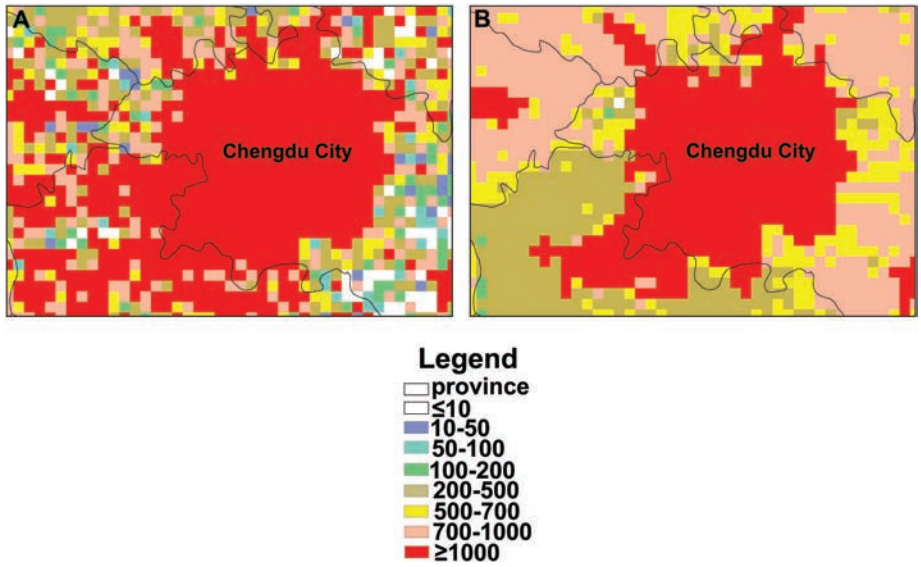

Fig. 19. Simulated distribution of total population through our work (A) and CAS (B).

We concluded that our simulation result was more persuasive and better reflected the actual population distribution.

A third advantage was that our result reflected the actual population distribution because it was based on the China residence polygon data. The population distribution should not exhibit a flaked configuration because the population is normally distributed in a pattern that is concentrated in one place where either suitable natural conditions or convenient transportation exists. The population distribution should thus exhibit a scattered distribution. Our scattered, simulated population result and the flaked, simulated population result from the CAS are shown in Figure 19. Figure 19A shows the scattered simulated population result of our work, whereas Figure 19B shows the flake-like simulated population result of the CAS.

To further compare these maps, we subtracted Figure 15 from Figure 16 using the ArcGIS software to obtain a difference map (Fig. 20). In Figure 20, a positive value indicates that the simulated population at a 1 × 1 km resolution from our work was less than that derived by the CAS, a zero value indicates that the simulated population from our work was equal to the result of the CAS, and a negative value indicated that the simulated population from our work was greater than that of the CAS. From this difference map, it is evident that, in eastern China, our simulation result was greater than that from CAS, and this finding was the result of different population sources. As theorized by Guo (2010), eastern China contains more floating people than western China. Therefore, when we utilize the census data as our population source, the simulated population in the east will be larger than that determined by the CAS. From Figure 20, it is also evident the simulated population in the west derived by our method is less than that determined by the CAS, and this is a result of the smoothness of CAS results. Additionally, given the third advantage of our simulated results, we believe that the CAS results for the west are less reasonable.

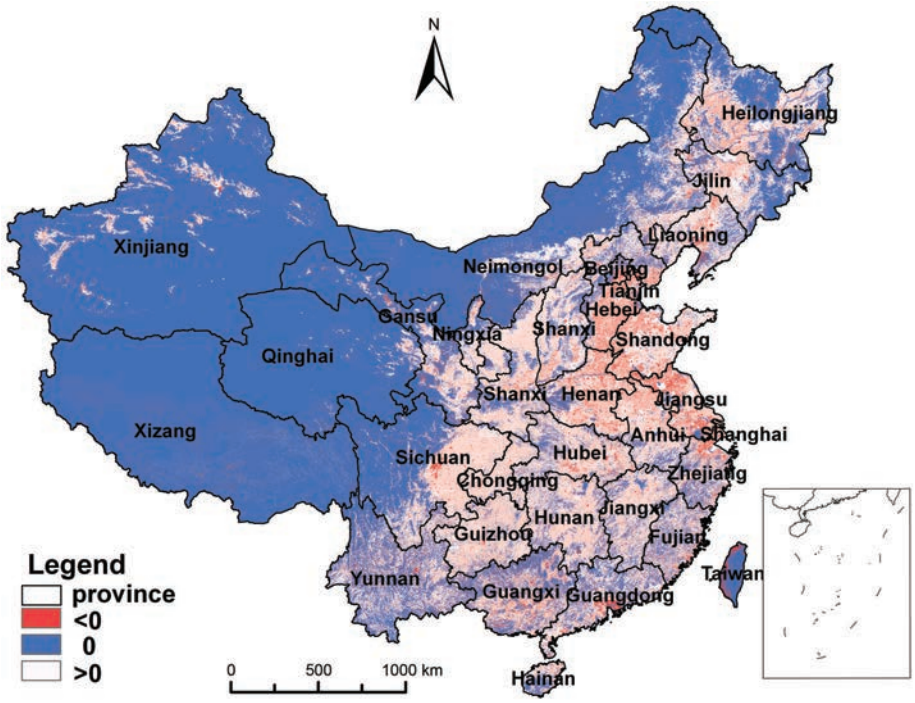

Fig. 20. Difference map between our work and that of CAS.

DISCUSSION AND CONCLUSIONS

Discussion

Using remotely sensed data to estimate population adds a new aspect to population estimation. Some researchers have adopted this method; however, it is still a challenging task in theory and methodology because of the complexity of urban and rural landscapes and of population distributions. The accuracy of population estimates derived from remotely sensed data has not heretofore been particularly high, although these data are linked to surface features and are not directly associated with population distribution. Therefore, a portion of such data can be readily identified as false. For example, areas with a low population maybe located in forest-dominated areas, but the spectral characteristics of the landscape of these areas are fundamentally different. Because of the complexity of population distributions, the use of land use data alone is insufficient. In this study, we only used a single model to estimate population, and in so doing it was not wholly adequate. Therefore, we must combine our model with others or create more powerful models.

Conclusions

This research has constructed models of population estimation that have integrated satellite imagery and census data and have provided relatively high precision,

which is essential in urban and rural planning, natural hazard risk assessment, disaster prevention and response, environment impact assessment, economic decision-making, and evaluation of the quality of life. As a result, we obtained three sets of land use data (i.e., remote sensing–derived, China residence polygon, and a combination of the two) to estimate population, and then we obtained a set of Chinese population data at 1 × 1 km resolution for 2000.

More studies are needed to improve the accuracy of population estimation by developing suitable models and improving land use classification accuracy from higher-spatial-resolution imagery and using multi-source data, such as increasing the quantity of source data or adding road data, which is also an important factor that affects population distribution. In summary, although remotely sensed data have some disadvantages, they may provide an important tool for social scientists and policymakers who seek population knowledge and social justice.

ACKNOWLEDGMENTS

This study was supported by the National Basic Research Program of China (the 973 Program: No.2010CB428402 and No. 2010CB428403) and the Open Foundation of the Changjiang River Scientific Research Institute (CKWV2012324/KY). We would like to thank two anonymous reviewers, Associate Editor Dr. Jason A. Tullis, and the Editor-in-Chief for their helpful comments.

REFERENCES

Alig, R. J., Kline, J. D., and M. Lichtenstein, 2004, "Urbanization on the US landscape: Looking Ahead in the 21st century," *Landscape and Urban Planning*, 69(2–3):219–234.

China, T. F. C., 2001, "The Fifth Census of China," China population information nets [http://www.stats.gov.cn/tjsj/ndsj/renkoupucha/2000pucha/pucha.htm].

Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C. and B. A. Worley, 2000, "LandScan: A Global Population Database for Estimating Populations at Risk," *Photogrammetric Engineering and Remote Sensing*, 66(7):849–857.

Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C., and B. A. Worley, 2003, "LandScan2000: A New Global Population Geography," in *Remotely-Sensed Cities*, Mesev, V. (Ed.), London, UK: Taylor & Francis, 267–279.

Feng, J., 2002, "Modeling the Spatial Distribution of Urban Population Density and its Evolution in Hangzhou," *Geographical Research*, 21(5):635–646.

Fujita, M., 1989, *Urban Economic Theory: Land Use and City Size*, New York, NY: Cambridge University Press, 380 p.

Guo, Z. X., 2010, "Research of Influence on Population Flow of FDI Difference Between West and East China," *Economic Research Guide*, 35:15–17.

Hardin, P. J., Jackson, M. W., and J. M., Shumway, 2007, "Intraurban Population Estimation Using Remotely Sensed Imagery,"in *Geo-spatial Technologies in Urban Environments: Policy, Practice, and Pixels*, Jensen, J. R. and J. Gattrel (Eds.), Berlin, Germany: Springer, 47–92.

Hu, H., 1983, *Study on the Distribution of Population in China*, Shanghai, China: East China Normal University Press, 256 p.

Jensen, J. R. and D. C. Cowen, 1999, "Remote Sensing of Urban Suburban Infrastructure and Socio-economic Attributes," *Photogrammetric Engineering and Remote Sensing*, 65(5):611–622.

Jin, J., Li, C. M., Yin, J., and Z. J. Lin, 2003, "Investigation on the Model for Spatial Distribution of Population Data," *Acta Geodaetica et Cartographica Sinica*, 32(3):278–282.

Joseph, M., Wang, L., and F. Wang, 2012, "Using Landsat Imagery and Census Data for Urban Population Density Modeling in Port-au-Prince, Haiti," *GIScience & Remote Sensing*, 49(2):228–250.

Li, G. Y. and Q. H. Weng, 2005, "Using landsat ETM plus Imagery to Measure Population Density in Indianapolis, Indiana, USA," *Photogrammetric Engineering and Remote Sensing*, 71(8):947–958.

Liao, S. and J. Sun, 2003, "GIS Based Spatialization of Population Census Data in Qinghai-Tibet Plateau," *Acta Geographica Sinica*, 58(1):25–33.

Liu, J. Y., 1996, *Study on the Macro Survey of Chinese Resources and Environment by Remote Sensing and its Dynamics*, Beijing, China: Chinese Science and Technology Press, 353 p.

Liu, J. Y. and A. S. Buhe, 2000, "Study of Spatial-Temporal Feature of Modern Land Use Change in China: Using Remote Sensing Technique," *Quaternary Science*, 3:229–239 (in chinese).

Liu, J. Y., Liu, M. L., Zhuang, D. F., Zhang, Z. X., and X. Z. Deng, 2003, "Study on Spatial Pattern of Land Use Change in China during 1995–2000," *Science in China Series D: Earth Sciences*, 46(4):373–384.

Liu, J. Y., Yue, T. X., Wang, Y. A., Qiu, D. S., Liu, M. L., Deng, X. Z., Yang, X. H., and Y. J. Huang, 2003, "Digital Simulation of Population Density in China," *Acta Geographica Sinica*, 58(1):17–24.

Lo, C. P., 1986, "Applied Remote Sensing," *Geocarto International*, 1(4).

Lo, C. P., 1995, "Automated Population and Dwelling Unit Estimation from High-Resolution Satellite Images: A GIS Approach," *International Journal of Remote Sensing*, 16(1):17–34.

Lv, A. M., Li, C. M., Lin, Z. J., and W. Z. Shi, 2003, "Spatial Continuous Surface Model of Population Density," *Acta Geodaetica et Cartographica Sinica*, 32(4):344–348.

Ministry of Public Security of the People's Republic of China, 2001 [http://www.mps.cn/English/index.htm].

NBSC (National Bureau of Statistics of China), 2002, *China Statistical Yearbook*, Beijing, China: China Statistical Press.

NBSC (National Bureau of Statistics of China), 2005, *China Statistical Yearbook*, Beijing, China: China Statistical Press.

Nets, A. D., 2008, Administrative divisions network [http://www.xzqh.org/html/2008/0606/20255.html].

Paul, S., 1997, "Modeling Population Density with Night-Time Satellite Imagery and GIS," *Computers, Environment and Urban Systems*, 21(3–4): 227–244.

Sherratt, G. G., 1960, "A Model for General Urban Growth," in *Management Sciences, Model and Techniques? Proceedings of the Sixth International Meeting of Institute of Management Sciences*, New York, NY: Pergamon Press, 147–159.

Sutton, P. D., Roberts, D. , Elvidge, C., and H. Melj, 1997, "A Comparison of Nighttime Satellite Imagery and Population Density for the Continental United States," *Photogrammetric Engineering & Remote Sensing*, 63(11):1303–1313.

Tan, M. H., Li, X. B., Lu, C. H., Luo, W., Kong, X. B., and S. H. Ma, 2008, "Urban Population Densities and Their Policy Implications in China," *Habitat International*, 32(4):471–484.

Tanner, J. C., 1961, *Factors Affecting the Amount of Travel*, London, UK: HLMSO, Road Research Technical Report No. 51, 15–220.

Taragi, R., Bisht, K., and B. Sokhi, 1994, "Generating Population Census Data Through Aerial Remote Sensing," *Journal of the Indian Society of Remote Sensing*, 22(3):131–138.

Tian, G. J., Liu, J. Y., and D. F. Zhuang, 2003, "The Temporal-Spatial Characteristics of Rural Residential Land in China in the 1990s," *Acta Geographica Sinica*, 58(5):651–657 (in Chinese).

Tian, Y. Z. , Chen, S. P. , Yue, T. X., Zhu, L. F., Wang, Y. G., Fan, Z. M., and S. N. Ma, 2004, "Simulation of Chinese Population Density Based on Land Use," *Acta Geographica Sinica*, 59(2):283–292.

Tian, Y. Z., Yue, T. X., Zhu, L. F., and C. Nicholas, 2005, "Modeling Population Density Using Land Cover Data," *Ecological Modelling*, 189(1–2):72–88.

Tobler, W., Deichmann, U., Gottsegen, J., and K. Maloy, 1997, "World Population in a Grid of Spherical Quadrilaterals," International Journal of Population Geography, 3:203–225.

Wang, F. and J.-M. Guldmann, 1996, "Simulating Urban Population Density with a Gravity-Based Model," *Socio-economic Planning Sciences*, 30(4):245–256.

Ye, Y. X., 2001, "City and Optimization of Land Use in 21st Century," *China Land Science*, 15:10–13 (in Chinese).

Yue, T., Wang, Y., Chen, S., Liu, J., Qiu, D., Deng, X., Liu, M., and Y. Tian, 2003, "Numerical Simulation of Population Distribution in China," *Population and Environment*, 25(2):141–163.

Zhang, C. and B. Yang, 1992, *Basics of Quantitative Geography* (2nd ed.), Beijing, China: Higher Education Press.

Zhang, G. F. , Shan, X. J., and J. Y. Yin, 2007, "Simulating Population Density Based on TM Images —Taking Shanghai as an Example," *Earthquake*, 27(1):47–54.

Zhuo, L. , Chen, J., Shi P. J., Gu, Z. H., Fan, Y. D., and I. Toshiaki, 2005, "Modeling Population Density of China in 1998 Based on DMSP/OLS Nighttime Light Image," *Acta Geographica Sinica*, 60(2):266–276.

Zhou, Y. and H. Yu, 2004, "Reconstructing City Population Size Hierarchy of China Based on the Fifth Population Census (1)," *Planning*, 28(6):49–55.