Chin. Geogra. Sci. 2021 Vol. 31 No. 6 pp. 966−978
https://doi.org/10.1007/s11769-021-1240-6

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

Population Spatial Distribution Based on Luojia 1-01 Nighttime Light Image: A Case Study of Beijing

SUN Lu[1, 2], WANG Jia[1, 2], CHANG Shuping[1, 2]

(1. *Institute of GIS, RS & GPS, Beijing Forestry University, Beijing* 100083, *China*; 2. *Beijing Key Laboratory of Precise Forestry, Beijing Forestry University, Beijing* 100083, *China*)

Abstract: With the continuous development of urbanization in China, the country's growing population brings great challenges to urban development. By mastering the refined population spatial distribution in administrative units, the quantity and agglomeration of population distribution can be estimated and visualized. It will provide a basis for a more rational urban planning. This paper takes Beijing as the research area and uses a new Luojia1-01 nighttime light image with high resolution, land use type data, Points of Interest (POI) data, and other data to construct the population spatial index system, establishing the index weight based on the principal component analysis. The comprehensive weight value of population distribution in the study area was then used to calculate the street population distribution of Beijing in 2018. Then the population spatial distribution was visualize using GIS technology. After accuracy assessments by comparing the result with the WorldPop data, the accuracy has reached 0.74. The proposed method was validated as a qualified method to generate population spatial maps. By contrast of local areas, Luojia 1-01 data is more suitable for population distribution estimation than the NPP/VIIRS (Net Primary Productivity/Visible infrared Imaging Radiometer) nighttime light data. More geospatial big data and mathematical models can be combined to create more accurate population maps in the future.

Keywords: Luojia1-01 nighttime light image; principal component analysis; points of interest; landuse type data; population spatial distribution

Citation: SUN Lu, WANG Jia, CHANG Shuping, 2021. Population Spatial Distribution Based on Luojia 1-01 Nighttime Light Image: A Case Study of Beijing. *Chinese Geographical Science*, 31(6): 966−978. https://doi.org/10.1007/s11769-021-1240-6

1 Introduction

Population data is one of the most important types of basic geographic data, and is of great significance for various population-related research and applications. With the continuous development of urbanization in China, the growing population brings great challenges to urban development (Alahmadi et al., 2013). Census is the main channel used by many countries to achieve population information statistics and analysis (Hu et al., 2018). In China, population census is carried out every ten years. Although authoritative, systematic and standardized, its renewal cycle is slow, which makes it impossible to grasp the population spatial distribution of the country. Especially in Beijing, often influenced by the national fertility policy and local evacuation and control measures, the population have been dramatically changing in recent years. Statistical data that is currently available does not meet the research need of the community.

Received date: 2021-05-06; accepted date: 2021-08-08
Foundation item: Under the auspices of Natural Science Foundation of China (No. 42071342, 31870713), Beijing Natural Science Foundation Program (No. 8182038), Fundamental Research Funds for the Central Universities (No. 2015ZCQ-LX-01, 2018ZY06)
Corresponding author: WANG Jia. E-mail: wangjia2009@bjfu.edu.cn

Many scholars have carried out in-depth studies based on the spatial distribution of social data and socialization of spatial data based on remote sensing data and GIS technology (Wu et al., 2005). Some scholars commonly used the spatial autocorrelation model (Birch and Young, 2006; Zhang and Zhu, 2011; Krishna, 2017), spatial logistic regression model (Law et al., 2003), and density-independent matrix model (Westerberg and Wennergren, 2003) to simulate the spatial distribution of the population. With the abundance of data sources and the development of GIS technology, the overlay method can be used to easily integrate various types of spatial data. For instance, Monmonier and Schnell (1984) interpreted Landsat image data to obtain residential land data and combined them into the zoning density method. Bakillah et al. (2014) studied the correlation between POI data and population and calculated the quantity and distance information of POIs to construct the weight of the population spatial distribution grid. Thus, it can be said that remote sensing data and GIS technology play a great role in populational studies.

Nighttime Satellite Data records the night light density value, which shows the gathering location of the population, comprehensively reflects human activity. Elvidge et al. (1999) and Sutton et al. (1997) found that DMSP-OLS nighttime satellite data were significantly correlated with population distribution. Subsequently, many scholars conducted a large number of studies and applied the nighttime satellite data to the population distribution simulation. For the areas where the demographic data is missing, this type of data is used to create new ways for regional population estimation and population spatial distribution simulation.

Briggs et al. (2007) conducted regression analysis on the light area, non-light area, and light brightness of each land use type in each region produced a population density map with a spatial resolution of 200 m and 1 km. Bagan and Yamagata (2015) established a common least-squares regression model and a geographically weighted regression model, which took the spatial relationship of population distribution into account to improve its simulation accuracy. POI (Points of Interest) is a kind of social perception data that generally refers to all the geographical objects that can be abstract as points, especially some geographical entities closely related to the daily life, such as schools, banks, restaurants, gas stations, hospitals, and supermarkets (Mckenzie et al., 2015; Liu et al., 2019). The traditional social and economic data with geographical location is only the count of the large facilities above district level. A lot of manpower and material resources are required to obtain it, and updating it is a very time-consuming process. POI has many advantages, such as easy access, abundant data, high positioning accuracy, and the ability to reflect micro details (Yao et al., 2017). As a new type of spatial data source, the distribution pattern and density of POIs are of great significance in urban planning and a population simulation.

In recent years, multi-source data fusion has been used to study population in different regions and this field of research made rapid progress. Some scholars combined POIs and multi-source remote sensing data in the random forest model, and disaggregated 2010 county-level census population data of China to 100 m × 100 m grids (Ye et al., 2019). Other scholars adopted Luojia 1-01 nighttime light image, POIs, and social media check-in data to map Chinese population. Multiple variables were used to train a random forest model, and conduct a fine-scale population mapping in Zhejiang Province, China (Wang et al., 2020). In 2018, the population of 45 Chinese cities at 100 m resolution was calibrated by using Luojia 1-01 data via random forest model and geographically weighted regression model. The NPP/VIIRS (Net Primary Productivity/Visible infrared Imaging Radiometer) data were used for comparison and the outcomes showed that the precision of Luojia 1-01 is better than that of NPP/VIIRS except for some specific cases (Guo et al., 2021). Therefore, it is considered reasonable to combine the data of Luojia 1-01 and POI to visualize the population spatial distribution.

Accurate population distribution data is the basis of better urban management, which can effectively improve the quality of urban development. In previous years, studies about population spatial distribution based on remote sensing data mostly used NPP/VIIRS nighttime light data and landuse type data. It is seldom carried out based on Luojia 1-01 and multi-source data. At present, there are few studies on population distribution at small scale. This study jointly used nighttime satellite Luojia 1-01 and multi-source data to produce an accurate population map of Beijing, which can enrich and expand the methods of population spatial research, and provide the basis for more rational urban planning.

2 Materials and Methods

2.1 Study area

Beijing, the capital of China, is located in the North Plain of the country. The central location is 116°20′E, 39°56′N. It has 16 districts, covering 164 10.54 km^2 (Fig. 1). High terrains are mostly located in the northeast and low ones in the southeast. The area is surrounded by mountains to the west, north and northeast, and its southeast region is a plain sloping gently toward the Bohai Sea. Beijing has a warm temperate semi-humid semi-arid monsoon climate and the city's transportation conditions are particularly developed. According to information extracted from the Beijing Area Statistical Yearbook (Beijing Municipal Bureau Statistics, Survey Office of the National Bureau of Statistics in Beijing, 2019), the city had a permanent population of 21.542 million in 2018, with 18.634 million people concentrated in urban areas and 2.908 million in rural areas. In that year, the per capita Gross Domestic Product (GDP) was 140.211 thousand yuan(RMB) and permanent population density was 1313 person/km^2. A high-density population poses challenges to city population management and security planning. Studying population spatial distribution can provide a basis for solving such issues.

2.2 Data and processing

The population data used in this study were from the Beijing Area Statistical Yearbook (Beijing Municipal Bureau Statistics, Survey Office of the National Bureau of Statistics in Beijing, 2019) and slope data were obtained by the digital elevation model (DEM) in the study area, and then processed in ArcGIS 10.5. Its resolution was 30 m × 30 m. DEM data were available in the Geospatial Data Cloud (http://www.gscloud.cn/).

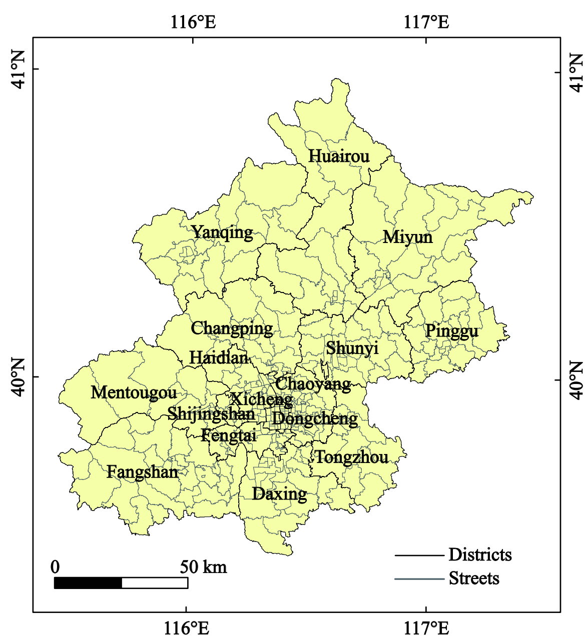

Fig. 1 The position of Beijing in China

Landsat 8 Operational Land Imager (OLI) data were used to classify land use type, road, and river distribution, with a resolution of 30 m × 30 m. According to the expert scoring method, land use type data was normalized to five types. We have assigned 0 to water, 1 to unused land, 2 to forest and grass, 3 to arable land, and 9 to construction land (Wu et al., 2015). Euclidean Metric was used to calculate the closest distance to the road and waters.

POI is a type of social perception data. This paper used 844 624 effective points found within the study area in 2018. These POI records include 16 major categories: food, beauty, tourist, education, medical, realty, hotel, shopping mall, way, company, financial, automobile, culture, sports, import and export, and government. Kernel Density was applied to calculate each POI density in ArcGIS 10.5 and POI density was normalized for subsequent experiments.

Luojia1-01 nighttime satellite data is a new generation of nighttime remote sensing satellites. Compared with traditional nighttime satellite data, its resolution is higher than ever reaching 130 m. The Luojia 1-01 nighttime Satellite Data can be downloaded free of charge at the website (http://www.hbeos.org.cn/). However, it needs radiance conversion before use. The radiance conversion formula is as follows:

$$L = DN^{3/2} \times 10^{-10} \tag{1}$$

where L is the radiation correction value after absolute radiation correction, the unit is W/(m^2·sr·μm), DN is the image gray value. Luojia 1-01 nighttime satellite data has been available since 2018. This paper utilized nighttime satellite data collected in August and September, when the data coverage was relatively more complete than other periods. The specific collection time is 22:00 pm. The mean value of the images was calculated to obtain the luminance value of the night light covering the study area. We have normalized the data after absolute radiation correction.

The WorldPop Mainland China dataset is a relatively new gridded population dataset with a better spatial resolution for the Chinese territory. The WorldPop set for 2018 was obtained from the official WorldPop project website (http://www.worldpop.org.uk/).

2.3 Methods

This method proposed here was divided into three main steps: 1) establishing an index system by integrating the socioeconomic factors and the natural factors, involving the selection of POI categories and normalization of various indexes; 2) calculating the index weight based on the principal component analysis, and comprehensively analyze the relative importance and differences among multiple indicators; and 3) simulating the population spatial distribution to analyze the distribution of the street population in Beijing. The main steps of the methodology are summarized in Fig. 2.

2.3.1 Establishing an index system

According to different scholars in the field, land use type, landform, water resource, infrastructure have a great impact on population distribution, and nighttime satellite images can also reflect the way the distribution goes (Liu and Wang, 2001; Liao and Li, 2004; Cao et al., 2009; Townsend and Bruce, 2010). In this study, POI data were selected to represent the distribution of the infrastructure. After counting the population and the number of POIs in all streets of Beijing, a simple linear regression equation was used to analyze the correlation between the population and the number of POIs (Fig. 3). Six types of POIs have the largest correlation coefficients with population: food, beauty, realty, import and export, medical, and education. Since POI is an architectural object with social attributes, these six types were selected as the socioeconomic factors affecting population distribution.

Land use type data represented the process of urban planning and development. The nighttime satellite data showed the population distribution and reflected the level of urban economic development. All of them were used as socioeconomic factors. Distance to roads, distance to a river, and the slope were considered as natural factors that affect population distribution. The population spatial distribution evaluation index system consists of eleven indicators, as shown in Table 1. After normalizing the range of all indexes from 0-10, each index data had a new distribution map (Fig. 4).

2.3.2 Calculating index weight based on principal component analysis

The scores of each index were calculated by principal

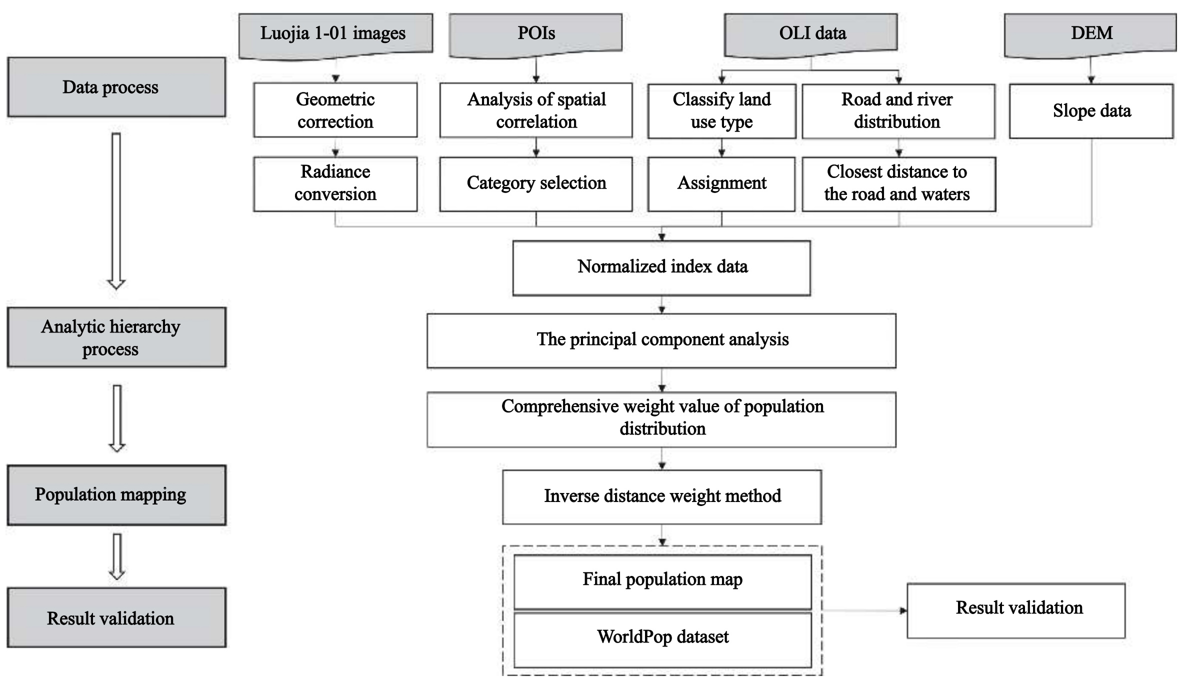

Fig. 2 Flowchart for producing and assessing the accuracy of the population map. POI: Points of Interest; OLI: Operational Land Imager; DEM is Digital Elevation Model

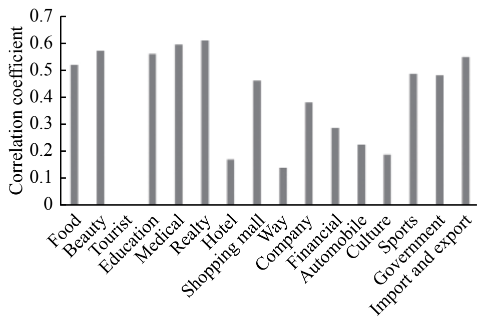

Fig. 3 Correlation between population and the number of POIs (Points of Interest) in 16 categories in Beijing

component analysis (PCA) and then normalized as the weight coefficient of each index. This method can not only reflect the relative importance of each index but also reflect the difference between the indexes. The application of this method to the comprehensive evaluation of multiple indexes can meet the requirements of science, rationality, and objectivity.

(1) Process standardized data by PCA in SPSS, eigenvalue, eigenvectors, and component matrix were shown in Table 2.

Table 1 Population spatial distribution evaluation index system

A Target layer	B Criterion layer	C Index layer
A Population spatial distribution evaluation index system	B1 Socioeconomic factors	C1 Landuse type
		C2 Night light intensity
		C3 Realty
		C4 Medical
		C5 Beauty
		C6 Education
		C7 Import and export
		C8 Food
	B2 Natural factors	C9 Distance to roads
		C10 Distance to river
		C11 Slope

(2) Calculate the coefficients in a linear combination

$$L_{ij} = \frac{x_{ij}}{\sqrt{\lambda_i}} \tag{2}$$

where L_{ij} represents the coefficient of the jth index in

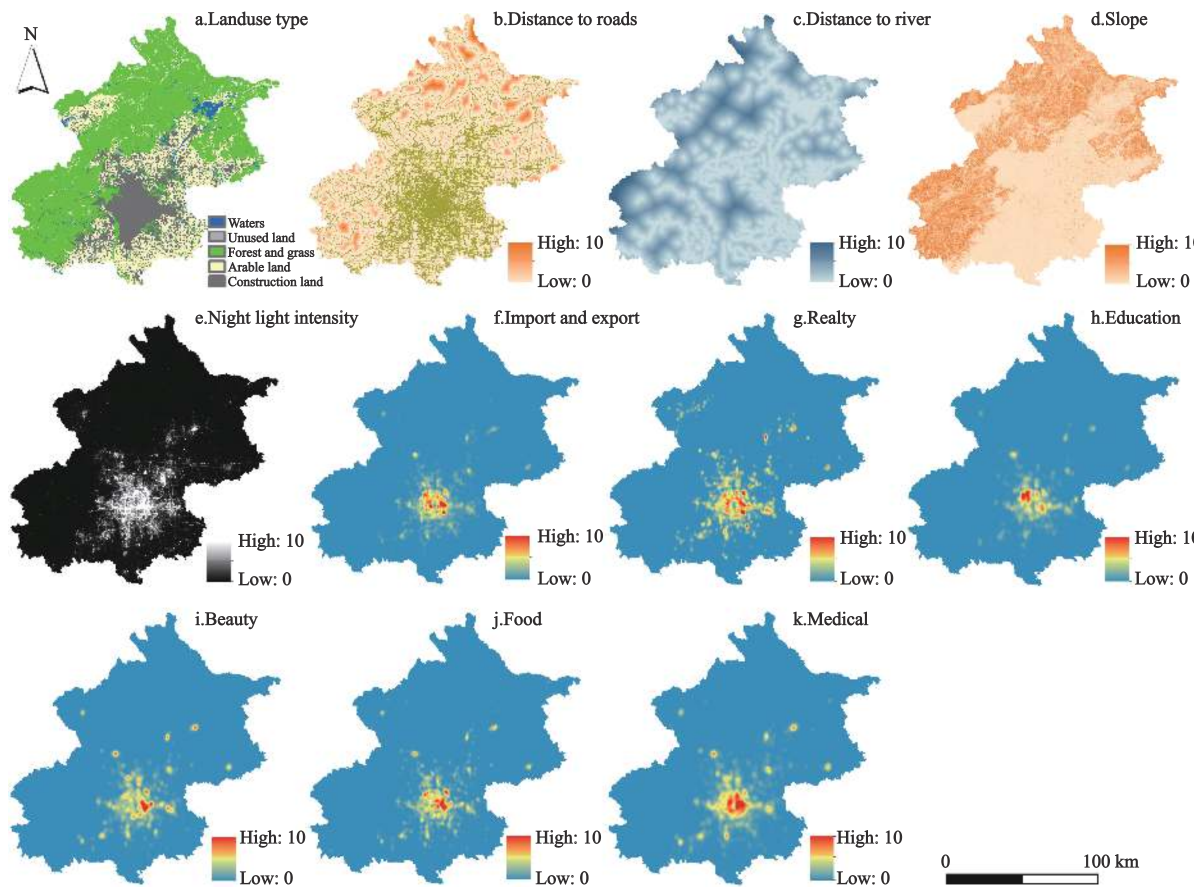

Fig. 4 Normalization of population spatial distribution evaluation index in Beijing

Table 2 Normalization of population spatial distribution evaluation index

Component	F1	F2
Medical	0.957	0.090
Food	0.953	0.133
Import and export	0.942	0.155
Beauty	0.939	0.101
Realty	0.910	0.071
Education	0.897	0.137
Night light intensity	0.701	–0.018
Landuse type	0.580	–0.452
Slope	–0.328	0.735
Distance to river	0.056	0.725
Distance to roads	–0.262	0.637
Characteristic root	6.231	1.76
Principal component variance	56.647	15.997

principal component i, x_{ij} is the jth normalized data corresponding to the ith characteristic root, λ_i is the ith characteristic root.

(3) Calculate the coefficients in the comprehensive score model

$$\delta_j = \frac{\sum_{i=1}^{m} L_{ij} \times \partial_i}{\sum_{i=1}^{m} \partial_i} \tag{3}$$

where δ_j is the coefficient of index j, L_{ij} is the coefficient of the jth index in principal component i, ∂_i is the variance contribution rate of principal component i.

(4) Weight calculation and normalization

Calculate the weight and normalize all the indexes to make their weight synthesis value 1.

$$W_j = \frac{\delta_j}{\sum_{j=1}^{n} \delta_j} \tag{4}$$

where W_j is the final index weight of index X_j.

The next step is to calculate the weight of each index factor according to Eqs. (2)–(4).

2.3.3 Simulating the population spatial distribution

The weight value of each index was used to calculate the comprehensive weight value of population distribution in the study area. The area with water and zero brightness is uninhabited theoretically. To reduce the error, grids in which the land type was water or the light intensity pixel value was zero were assigned to zero. In this study, the census data of districts and counties in Beijing were used to predict the population distribution of streets in Beijing. The spatial distribution of the population was obtained according to Eqs. (5), (6).

$$F = \sum_{j=1}^{m} w_i P_{ij} \tag{5}$$

$$POP_k = POP \times \left(\frac{F_k}{\sum F_k} \right) \tag{6}$$

where F is the comprehensive weight value of each grid, w_i is the weight of the ith index, P_{ij} is the normalized score value of the jth grid corresponding to the ith index, and m is the number of evaluation indexes. POP_k represents the number of people in the kth grid after spatial distribution, and POP is the demographic value of the county where the grid is located. F_k represents the comprehensive weight value of the kth grid, and $\sum F_k$ is the sum of the comprehensive weight value of all grids in the county where the grid is located.

2.3.4 Result validation

Mean absolute deviation (MAE) was adopted to quantify the estimation errors of the proposed model and the coefficient of determination (R^2) was used to compare accuracy with other models:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (\widehat{y_i} - y_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2} \tag{7}$$

where $\widehat{y_i}$ denotes the population of pixel i derived from the WorldPop data, y_i denotes the estimated population data through the weighted model prediction, $\bar{y}_i$ denotes the average value of the estimated data and n denotes the number of pixels.

The difference between the population grid estimate and the WorldPop data was used as the error value of the population estimate:

$$\varepsilon = E - a \tag{8}$$

where ε is the error value of the population estimate, E is the population estimate, a is the actual value based on the WorldPop data.

3 Results and Analyses

3.1 POIs data selection

To avoid obvious errors, the correlation between POIs and the population was analyzed by a linear regression method. After collating the data (Fig. 5), six types of

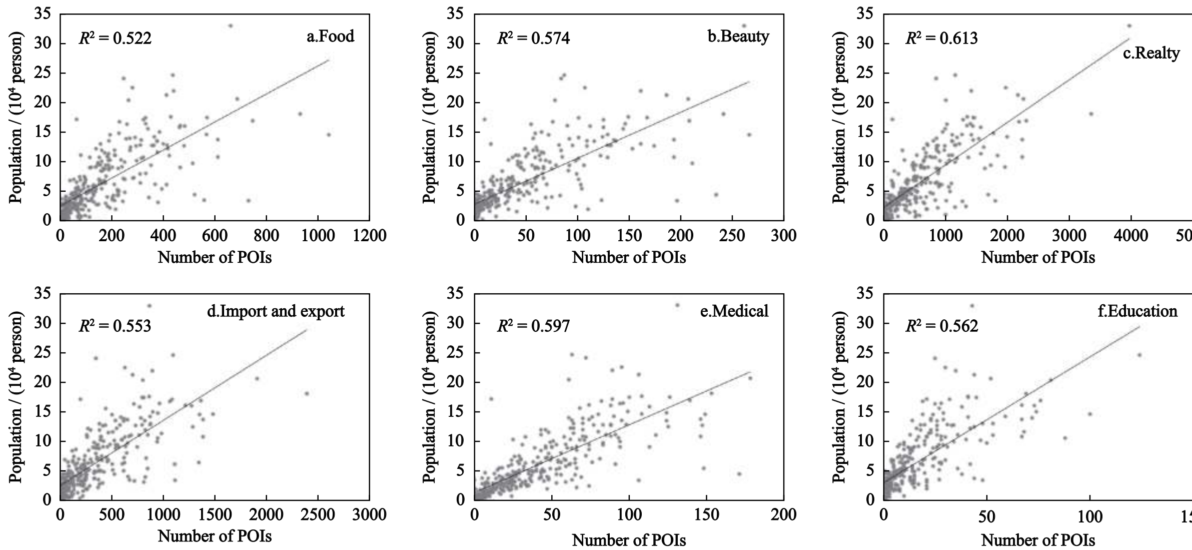

Fig. 5 Linear relationship between POI (Points of Interest) number and population in Beijing in 2018

POIs have larger correlation coefficients with population, such as food, beauty, realty, import and export, medical, and education. In densely populated areas, entertainment and leisure establishment (such as food courts and beauty salons) were bound to increase, and had a positive linear relationship. Schools and hospitals were bound to be set up in areas with more children and elderly people. In addition, the correlation between tourist attractions and the population was minimal. The urban areas of Beijing embrace a myriad of tourist attractions, including the Forbidden City, the Temple of Heaven, and the Beihai Park. However, the residents were mostly distributed in the districts of Haidian and Chaoyang, and other surrounding areas. Many mountainous areas around Beijing have been developed into natural tourist attractions, with unsatisfactory surrounding facilities and few residential areas. We have also verified the existence of a relationship between the number of POIs and the population, but it depended on the type of POIs.

3.2 Population spatial distribution results

Fig. 6 shows the result of the proposed method, the 0.01° population grid in the study area. It was found that the population was concentrated in the centre of Beijing. The population of Beijing presents the distribution characteristic structure of 'central urban agglomeration, periphery multi-core'. Dongcheng, Xicheng, Chaoyang, and Haidian districts are highly populated areas, with a lot busy business districts and universities. Xueyuan, a famous university town in Haidan District, is home for many student apartments and laboratories and showed the highest population density. At the same time, the districts of Shunyi, Changping, and the northwest of the Tongzhou District are also densely populated areas. In recent years, universities in Beijing have set up new campuses in Changping District, which increased the population density of the area. The Beijing Capital Airport in Shunyi District also collaborated to the dense population that formed around this area. Many enterprises had moved to the Tongzhou District and the improvement of infrastructure, the concentration of education, and the development of business brought in-

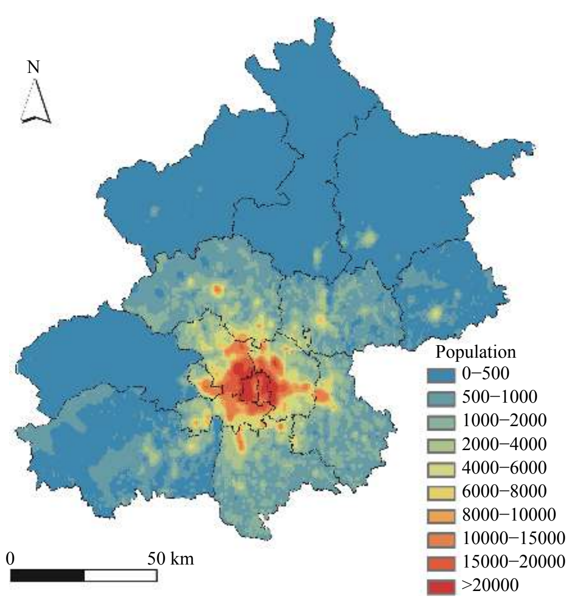

Fig. 6 Population spatial distribution results of Beijing in 2018

creased population concentration. Besides, areas such as Huairou and Miyun districts are mostly farmland, woodland, and mountainous area, generally developed as tourist attractions or rural areas, and not densely populated.

The distribution of population follows the spatial patterns of residential areas. Generally, the geographical conditions in the population concentration distribution area are very favourable. Convenient transportation and urban infrastructure are the most important features for people to live and work. The better these facilities were, the more the population tend to concentrate around them. At the same time, the tendency is that there will be more urban construction in places where people are concentrated, including schools, hospitals, shopping malls, and enterprises. This creates a positive cycle of population concentration.

3.3 Population spatialization and WorldPop for validation

The WorldPop Mainland China dataset is a relatively new gridded population dataset. The accuracy of our newly produced population map was compared with that of the WorldPop data set to verify the accuracy of population spatial distribution using multi-source data, including POIs and Luojia 1-01 nighttime light image. The correlation coefficient between the two was 0.74. Fig. 7 illustrates the relationship between the predicted population density and the WorldPop population density in which each data point corresponds to 1 km^2. The population spatial distribution based on the POIs and other multi-source open data obtained satisfactory results. Therefore, it is believed that a comprehensive consideration of natural geographical and social-economic factors (such as Luojia 1-01 nighttime light image, land use, and POI) was conducive to the accurate realization of population spatial results.

The absolute error value of the population estimate was the difference between the population grid estimate and the WorldPop data. From the geographical distribution (Fig. 8), the error of population estimation in most communities was within 750 people. The accuracy was generally low in the highly or lowly populated areas, and the error was acceptable in regions with medium population densities, such as the Mentougou, Fangshan, Huairou and Miyun Districts. The communities with large errors were distributed in the central area of Beijing, and the main error was that the estimated population value of the community was larger than the actual value. The central area of Beijing is composed of the main urban areas with a high population density. Urban residents' choice of living environment was complex, influenced by economic income, environment, and governmental policies. At the same time, the Forbidden City and other famous scenic spots also had a certain impact on the population distribution in the Second Ring Road of Beijing. More complex factors need to be considered for actual modelling. The overall estimation results were relatively large, which may be correlated with the impact of construction data on population distribution, leading to some built-up areas that can allocated more people.

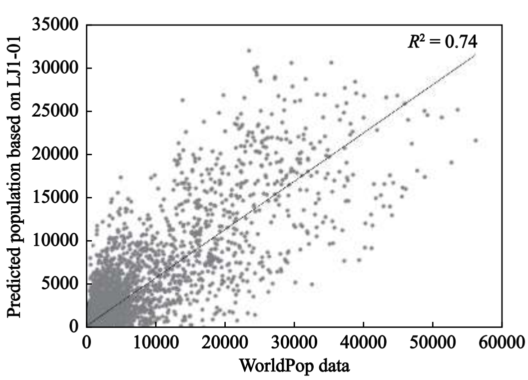

Fig. 7 Correlation coefficient between the predicted population based on Luojia1-01 (LJ1-01) and the WorldPop data

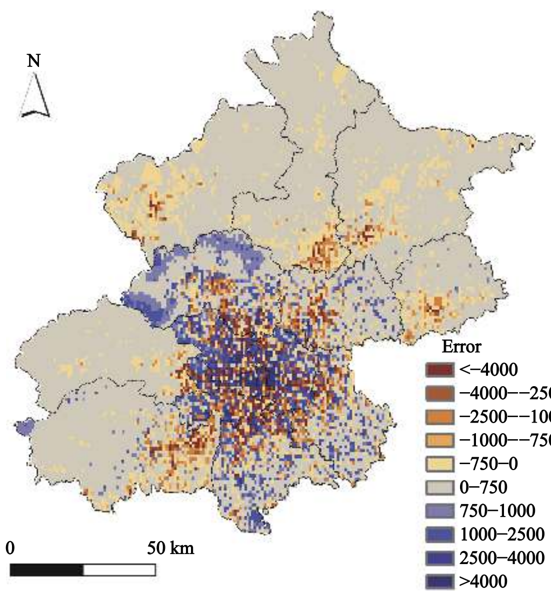

Fig. 8 The absolute error of population estimation in Beijing in 2018

4 Discussion

Fig. 6 shows the overall population distribution of Beijing, and the spatial results presented significant changes. The city centre has a developed economy, complete facilities, and more residential areas, so the population density is higher. Schools, business districts, enterprises are more popular in areas with a denser concentration of people. Most of the urban fringe areas are mountains and farmland, which are sparsely populated.

Night light intensity can reflect the vitality of a region to a certain extent and many scholars have already explored the relationship between night light data and population. It has been proved that, at the level of administrative units (e.g., cities) there is a significant correlation between the total population and the total value of night light intensity (Sun et al., 2017). There are 16 districts/counties in Beijing. As shown in Fig. 9, the correlation between the total night light intensity value and the population of urban counties in Beijing is relatively high, reaching 0.81. POI data has been used in population calculations, reflecting the distribution of social buildings and people's demands for work and general life. Where there were more people, there would be more schools, hospitals, shopping malls, and enterprises. On the contrary, the improvement of infrastructure would attract people to gather together. A correlation analysis was conducted between the total number of regional POIs and the population, and the result was up to 0.91 (Fig. 9). The examination showed that it was reasonable to use satellite remote sensing data and POI data to estimate the population spatial distribution.

By comparing the results of the precision analysis (Figs. 6–8), it was found that the spatial accuracy results were reduced in the high or low populated areas. The errors were mainly derived from underestimations in areas with large populations, whereas overestimations were common in those with high populations. When the night light intensity was used as the sole or primary variable in the analysis of population distribution, large numbers of people were distributed to underdeveloped areas, resulting in an over-distribution of the population in rural and suburban areas and uneven distribution in urban areas. Besides, the night lights on a small land area in a city can illuminate a vast surrounding area. This could result in satellite images collecting night lights that cannot accurately reflect population density over relatively small geographical areas. Since urban fringe areas are far away from the city centre, there are few public facilities and a lack of POI data. Transportation is not convenient, and the economy is also relatively under developed. These regions were generally mountainous and cultivated, with fewer residential areas. This indicates that multi-source data (such as POIs) and night remote sensing are not suitable for population estimation in low-density population areas. In today's society, people have an increasing demand for living facilities, e.g., shopping malls, restaurants, tourist attractions, or places necessary for life (schools, hospitals, supermarkets). A more densely populated area, evidently require greater needs of this realm. Therefore, a region with more POIs has a larger population than its counterparts. Although the area is illuminated in the night light image, those without POIs should be allocated with a smaller population. Besides, POI related variables can correctly capture the structure and func-

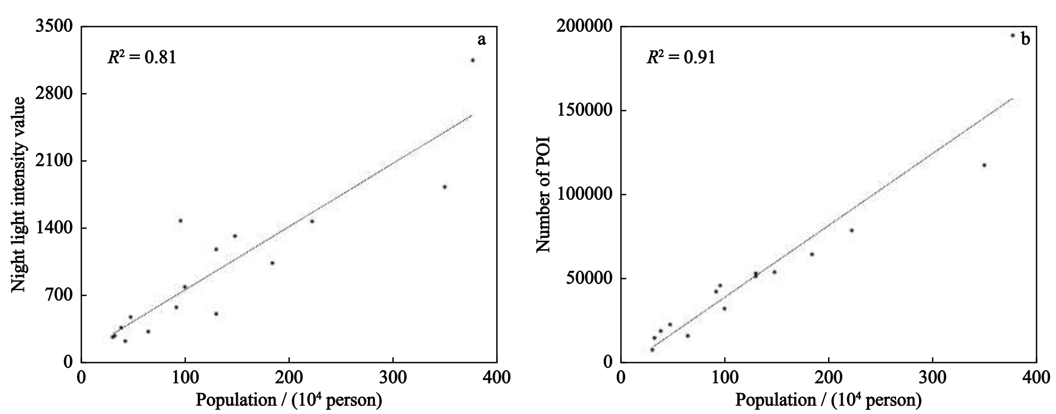

Fig. 9 Correlation between population and the total night light intensity value (a) and the total number of regional POIs (Points of Interest) (b) in Beijing

tion within the urban interior (Jiang et al., 2015; Gao et al., 2017). Thus, adding POI-related variables can greatly enhance the variation of the distributed population and reduce the underestimation of the population in urban areas, since a significant proportion of it was not allocated to suburban and rural areas with pixel lighting, corresponding to a smaller POI density.

Some scholars selected Beijing-Tianjin-Hebei region as a study area to conduct population spatialization with the data of DMSP/OLS, land use and other factors related to population distribution, the accuracy rate is 0.74 in the urban scale (Wu et al., 2015). At present, DMSP/OLS data has been out of service. Therefore, the 2018 NPP/VIIRS data and Luojia 1-01 data were used for comparison. By comparing the results of the precision analysis (Fig. 10, Fig. 11) under the condition that other variables are the same, the population distribution retrieved by Luojia 1-01 data is 0.74, and the population distribution retrieved by NPP/VIIRS data is 0.66. The result showed that the accuracy rate of the population spatialization based on Luojia 1-01 data is more than 74% in the street scale. Therefore, the Luojia 1-01 nighttime light image is more suitable for population spatial distribution. To match the purposes of this study, two typical regions were selected. The first was the main urban area (Dongcheng and Xicheng), which has a high density of population. There were many attractions. The light was strong at night, but there were not so many people. The result of population spatial distribution using the high-resolution Luojia 1-01 nighttime light image was obviously more detailed and realistic in this area. The second region is the Capital Airport. Due to the remote location of the airport, there were not many shopping malls and entertainment facilities nearby, so the light brightness value of the airport at night was relatively high. NPP/VIIRS nighttime light image has low resolution, which was not conducive to distinguish the difference of light brightness values near the airport. Inversion from NPP/VIIRS data clearly overestimated the population of the area, and Luojia 1-01 data showed a clear advantage. Therefore, Luojia 1-01 nighttime light image is a more suitable remote sensing data for population spatial distribution.

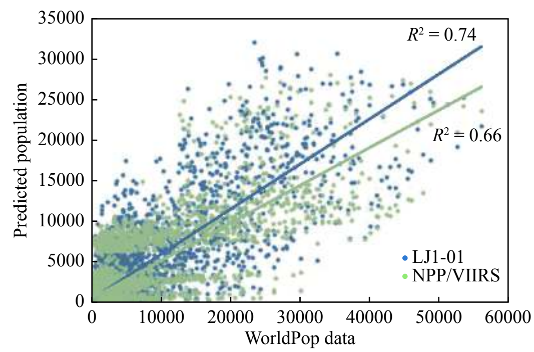

Fig. 10 Correlation coefficient between the predicted population based on LJ1-01, NPP/VIIRS (Net Primary Productivity/Visible infrared Imaging Radiometer) and the WorldPop data

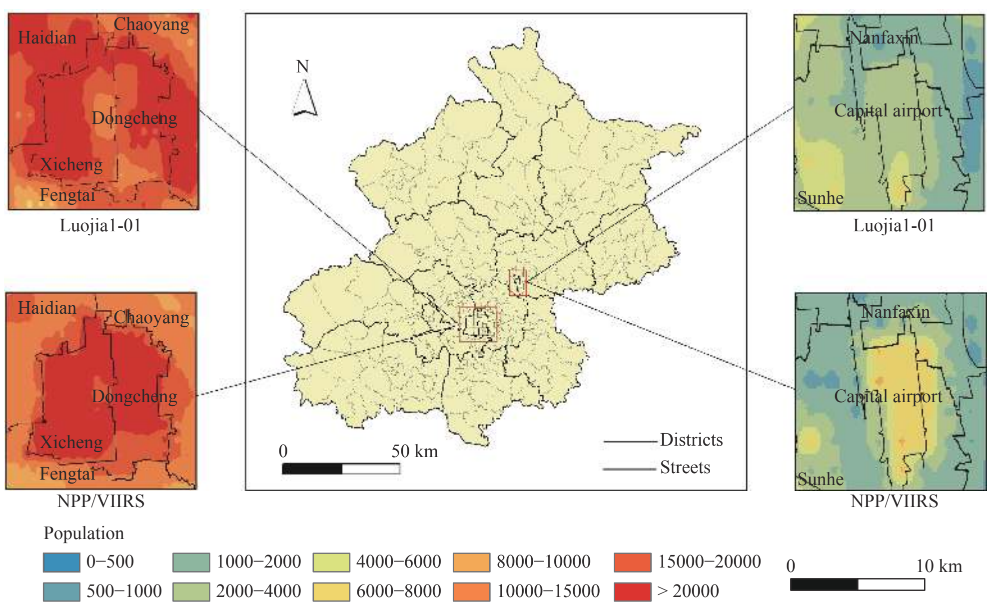

Fig. 11 Contrast of population at different local areas based on different data. NPP/VIIRS: Net Primary Productivity/Visible infrared Imaging Radiometer

This study has made some progress in the study of population spatial distribution, but the processes here presented have some issues. Luojia 1-01 nighttime light image is a new generation of nighttime remote sensing satellite, which resolution is reaching 130 m. Although its resolution, its data is still limited. In the process of data normalization, nighttime remote sensing was resampled to match other data. Therefore, nighttime remote sensing data was still insufficient to guarantee the accuracy of small-scale population spatial distribution. There are already many high-resolution data that can be added to the study, such as Tencent's big location data and social media data. Therefore, exploring the grid-scale under different conditions and forming a complete appropriate scale selection and evaluation method are the key research aims for the future of this study field. Improvements in land use type data can also improve the accuracy of experiments. Cui et al. (2020) used NPP/VIIRS nighttime lights data, land use type data and other auxiliary data to generate the spatial distribution of population at 100 m resolution. The model simulation error against the verified data was less than 10%. They used land use type data which was manually interpreted.

This paper relies on land use type data classified by supervision, so its accuracy is low. However, the final accuracy results were similar. Therefore, combined with high-precision nighttime remote sensing data, improving the accuracy of land use type data will most likely lead to better results. The linear relationship between each type of POI and population was discussed and six types of POI with high correlation coefficients were selected in this paper, food, beauty, realty, import and export, medical, and education. The correlation coefficients are respectively 0.522, 0.574, 0.613, 0.553, 0.597, 0.562. However, the population spatial distribution may be influenced by multiple POI distributions. The demand for public facilities in high-density areas was too high, and the distribution of POIs in such regions is usually dense, leading to an over-population production. In the areas with a low-density population, the lack of demand for public facilities resulted in the neglect of POIs, which would lead to a low estimated population (Antoniou and Schlieder, 2014). Therefore, the number and distribution of POI as well as the influence distance between them will affect the accuracy of population prediction. Accuracy verification has always been a difficult point in population distribution research. At present, it is usual and generally compared with existing research results or field sampling surveys. However, an accuracy verification method with a complete system, high reliability, and wide applicability could not been fully developed. The population distribution was calculated by using Luojia 1-01 nighttime light data, which is the highest accurate noctilucent remote sensing data at present. By comparing the accuracy of population distribution based on the NPP/VIIRS nighttime light data, we have concluded that Luojia 1-01 data is more suitable for population distribution estimation. The seventh census is currently being conducted in China. When the results are published, more accurate population data will be added to the study to assess the accuracy and reliability of the methodology. In future research, it is equally important to study the verification method, which is easily accessible and very precise.

5 Conclusions

Taking Beijing as the research area, this paper used landuse type data, POI (Points of Interest) data, Luojia 1-01 nighttime light image, and other relevant sources to construct a population spatial index system, establishing a new index weight based on the principal component analysis. Utilize GIS technology to visualize the population spatial distribution. Based on accuracy assessments, this method was validated as a promising way of generating population spatial maps. POI data were proved as important indicators to calculate population distribution. Under the condition that other variables are the same, the population distribution retrieved by Luojia 1-01 data is higher than NPP/VIIRS data significantly. On the street scale, Luojia 1-01 nighttime light image is a more suitable remote sensing data for population spatial distribution. Combining high-precision landuse type data with night light image, the result of population spatialization will be better.

Population spatial distribution is a complicated process that can be affected by various natural, economic, and social factors. The method proposed in this study needs further thinking and improvement in the population spatial distribution. With the gradual enrichment of geographic big data and more convenient means of acquisition, the resolution and accuracy of remote sensing data are gradually improved. Using multi-source data to explore the distribution of population composition, including different age groups, ethnic groups, and occupational attributes may be a good idea for future research approaches. Tapping into the rich information contained in the population distribution can help promoting the change from fuzzy population distribution to fine population distribution and better serve the social development.

Acknowledgments

We are grateful to the undergraduate students and staff of the Laboratory of Forest Management and '3S' technology from Beijing Forestry University.

References

Alahmadi M, Atkinson P, Martin D, 2013. Estimating the spatial distribution of the population of Riyadh, Saudi Arabia using remotely sensed built land cover and height data. *Computers Environment & Urban Systems*, 41: 167–176. doi: 10.1016/j.compenvurbsys.2013.06.002

Antoniou V, Schlieder C, 2014. Participation Patterns, VGI and Gamification. 17th AGILE Conference on Geographic Information Science: Geogames and Geoplay Workshop.

Bagan H, Yamagata Y, 2015. Analysis of urban growth and estimating population density using satellite images of nighttime lights and land-use and population data. *Giscience & Remote Sensing*, 52(6): 765–780. doi: 10.1080/15481603.2015.1072400

Bakillah M, Liang, Mobasheri et al., 2014. Fine-resolution population mapping using openstreetmap points-of-interest. *International Journal of Geographical Information Science*, 28(9): 1940–1963. doi: 10.1080/13658816.2014.909045

Beijing Municipal Bureau Statistics, Survey Office of the National Bureau of Statistics in Beijing, 2019. BEIJING AREA STATISTICS YEARBOOK. Publish: China Statistics Press.

Birch D A, Young W R, 2006. A master equation for a spatial population model with pair interactions. *Theoretical Population Biology*, 70(1): 26–42. doi: 10.1016/j.tpb.2005.11.007

Briggs D J, Gulliver J, Fecht D et al., 2007. Dasymetric modelling of small-area population distribution using land cover and light emissions data. *Remote Sensing of Environment*, 108(4): 451–466. doi: 10.1016/j.rse.2006.11.020

Cao Liqin, Li Pingxiang, Zhang Peiliang, 2009. Urban population estimation based on the DMSP/OLS night-time satellite data: a case of Hubei Province. *Remote Sensing Information*, (1): 83–87. (in Chinese)

Cui Xiaolin, Zhang Jiabei, Wu Feng et al., 2020. Spatio-temporal analysis of population dynamics based on multi-source data integration for Beijing municipal city. *Journal of Geo-information Science*, 22(11): 2199–2211. (in Chinese)

Elvidge C D, Baugh K E, Dietz J B et al., 1999. Radiance calibration of DMSP-OLS low-light imaging data of human settlements. *Remote Sensing & Environment*, 68(1): 77–88. doi: 10.1016/S0034-4257(98)00098-4

Gao S, Janowicz K, Couclelis H, 2017. Extracting urban functional regions from points of interest and human activities on location-based social networks. *Transactions in GIS*, 21(3): 446–467. doi: 10.1111/tgis.12289

Guo B, Bian Y, Zhang D et al., 2021. Estimating socio-economic parameters via machine learning methods using luojia1-01 nighttime light remotely sensed images at multiple scales of china in 2018. *IEEE Access*, (99): 1–1. doi: 10.1109/ACCESS.2021.3059865

Hu Yunfeng, Zhao Guanhua, Zhang Qianli, 2018. Spatial distribution of population data based on nighttime light and LUC data in the Sichuan-Chongqin Region. *Journal of Geo-Information Science*, 20(1): 68–78. (in Chinese)

Jiang S, Alves A, Rodrigues F et al., 2015. Mining point-of-interest data from social networks for urban land use classification and disaggregation. *Computers Environment & Urban Systems*, 53: 36–46. doi: 10.1016/j.compenvurbsys.2014.12.001

Krishna A, 2017. Spatial analysis model for estimation of population and other censusdata in india for forecasts in demographic, social and economic arena. *International Journal of Advancements in Computing Technology*, 08(2): . doi: 10.4172/0976-4860.1000184

Law, Richard, Murrell et al., 2002. Population growth in space and time: spatial logistic equations. *Ecology*, 84(1): 252–262. doi: 10.1890/0012-9658(2003)084[0252:PGISAT]2.0.CO;2

Liao Shunbao, Li Zehui, 2004. Relationship between population distribution and land use and spatialization of population census data. *Resources and Environment in the Yangtze Basin*, 13(6): 557–561. (in Chinese)

Liu Yanhua, Wang Qiang, 2001. Optimized population distribution in China: on provincial sustainable development. *China Population, Resources and Environment*, 11(1): 35–38. (in Chinese)

Liu Yilun, Chen Yimin, Liu Ying et al., 2019. Unit rent appraisal in community-scale and spatial pattern mapping in a Metropolitan Area using online real estate data: a case study of Shenzhen. *Tropical Geography*, 39(2): 188–195. (in Chinese)

McKenzie, Grant, Janowicz et al., 2015. Poi pulse: a multi-granular, semantic signature-based information observatory for the interactive visualization of big geosocial data. *Cartographica*, 50(2): 71–85. doi: 10.3138/carto.50.2.2662

Monmonier M S, Schnell G A, 1984. Land use and land cover data and the mapping of population density. *The International Yearbook of Cartography*, 24: 115–121. doi: 10.2307/214769

Sun W, Xia Z, Nan W et al., 2017. Estimating population density using DMSP-OLS night-time imagery and land cover data. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, (6): 1–11. doi: 10.1109/JSTARS.

2017.2703878

Sutton P, 1997. Modeling population density with nighttime satellite imagery and GIS. *Computers Environment & Urban Systems*, 21(3-4): 227–244. doi: 10.1016/S0198-9715(97)01005-3

Townsend A C, Bruce D A, 2010. The use of night-time lights satellite imagery as a measure of Australia's regional electricity consumption and population distribution. *International Journal of Remote Sensing*, 31(16): 4459–4480. doi: 10.1080/01431160903261005

Wang L, Fan H, Wang Y, 2020. Improving population mapping using Luojia 1-01 nighttime light image and location-based social media data. *Science of The Total Environment*, 730: 139148. doi: 10.1016/j.scitotenv.2020.139148

Westerberg L, Wennergren U, 2003. Predicting the spatial distribution of a population in a heterogeneous landscape. *Ecological Modelling*, 166(1-2): 53–65. doi: 10.1016/S0304-3800(03)00118-2

Wu Jiansheng, Xu Duo, Xie Wudan, Peng Jian, 2015. Spatialization of demographic data at medium scale based on remote sensing images: regarding Beijing-Tianjin-Hebei as an example. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 51(04): 707–717. (in Chinese)

Wu S, Qiu X, Wang L, 2005. Population estimation methods in GIS and remote sensing: a review. *Giscience & Remote Sensing*, 42(1): 80–96. doi: 10.2747/1548-1603.42.1.80

Yao Y, Liu X, Li X et al., 2017. Mapping fine-scale population distributions at the building level by integrating multisource geospatial big data. *International Journal of Geographical Information Systems*, 31(5-6): 1220–1244. doi: 10.1080/13658816.2017.1290252

Ye T, Zhao N, Yang X et al., 2019. Improved population mapping for china using remotely sensed and points-of-interest data within a random forests model. *Science of The Total Environment*, 658: 936–946. doi: 10.1016/j.scitotenv.2018.12.276

Zhang J, Zhu Y, 2011. The population spatial distribution model based on the spatial statistics in Shandong Province, China. *International Conference on Geoinformatics*, 1–4. doi: 10.1109/GeoInformatics.2011.5981117