The following publication Xu, Y., Song, Y., Cai, J., & Zhu, H. (2021). Population mapping in China with Tencent social user and remote sensing data. Applied Geography, 130, 102450 is available at https://doi.org/10.1016/j.apgeog.2021.102450.

Population mapping in China with Tencent social user and remote sensing data

Highlights

• Population mapping in China using social user and remote sensing data.

• Comparison of different indicators in mapping population.

• Performance of different models in estimating population.

• Implications of social user data and remote sensing data in mapping population.

Abstract

Real-time population data are vital for urban planning and resource management for sustainable development. To complement satellite-based population estimation methods, geospatial social media data provide additional opportunities to estimate the distribution of population with high levels of efficacy and accuracy. Thus, this study attempts to assess the performance of various sensing data to disaggregate population data in China; the tested data include Tencent location-based service (LBS) data (about 0.8 billion users), satellite-derived land use/cover data, and nightlight imagery data. With the use of census data for validation, the experimental results show that Tencent LBS data are much better than satellite-derived land use/cover data and nightlight satellite data for mapping the population distribution. The overall mapping accuracy at the city level using Tencent LBS data was 88.9%, whereas the accuracy using land use/cover data was 87.1% and that using nightlight satellite data was 85.5%. The experimental results also indicate that LBS data and remote sensing data could both be well integrated to map the population distribution in China. Thus, a population spatialization model was further developed using all of the tested indicators; this model allowed the overall population estimation accuracy at the city level to reach 90.4%. This model could help determine the population distribution on various spatial scales quickly and efficiently, and the developed tool and the provided population estimates may be vital for the sustainable development of cities and regions for which population data are lacking.

Keywords

Multisource data; Population estimates; Population distribution

1. Introduction

Demographic data show the aggregation and distribution of people at a specific time and space. Up-to-date urban population data are crucial for urban planning, environmental protection, and resource allocation. Demographic data are traditionally obtained from field surveys. Taking the census data of China as an example, a national household survey is conducted every 10 years, and a 1% sample survey is conducted every 5 years. The census data for the remaining years are therefore obtained by performing extrapolation and smoothing operations based on the census data from the survey years (Cai et al., 2008). Due to the long time interval needed to obtain real demographic data, it is difficult to determine the current population size and its spatial distribution due to human movement. Accurate assessment of the latest population data benefits the implementation of forward-looking urban planning and resource management, particularly for developing countries like China, in which an unequal distribution of population and great mobility have led to considerable public security and resource constraints (Wu et al., 2019).

The development of geospatial technology has led to the extensive application of contactless remote sensing data in population mapping studies due to the rapid and extensive coverage they provide (Chen et al., 2019; Wang et al., 2018; Tan et al., 2018). Two types of remote sensing data are currently used in population estimation studies: nightlight satellite data and conventional optical satellite data. The nightlight data express the richness of people's social and economic activities as the nightlight intensity, from which the population information can be inferred (Yu et al., 2017). Some widely used nightlight satellite data sets include the National Polar-Orbiting Partnership Visible Infrared Imaging Radiometer Suite (NPP/VIIRS) and its predecessor Defense Meteorological Program Operational Line-Scan System (DMSP/OLS) from the United States National Aeronautics and Space Administration (NASA) and the Luojia 1-01 (LJ 1-01) nightlight satellite data from China (Zhang et al., 2019). Due to their coarse spatial resolution and restricted radiometric resolution, the DMSP/OLS nightlight data were mainly used to explore the population size at the country level rather than the city level. Elvidge et al. (1999) investigated the use of DMSP/OLS nightlight data to detect population differences within a city and found that the oversaturated brightness value from the DMSP/OLS data makes it difficult to express spatial differences in the population within the city center. Sutton et al. (2001) explored the use of the same data to estimate the distribution of the world's population and found that countries

can be classified into three categories (high, low, and middle income) according to the statistical relationship between population and luminous intensity.

Advances in luminous technology have endowed the NPP/VIIRS nightlight data currently in wide use with higher spatial and radiometric resolutions than that of their predecessor DMSP/OLS. Based on an experiment with African cities, Chen et al. (2015) confirmed that the NPP/VIIRS nightlight data can achieve more accurate population estimates than the DMSP/OLS nightlight data. Wang et al. (2018) used NPP/VIIRS luminous data and a nonlinear model to explore the relationship between the nightlight data and the populations of some Chinese cities and found that the indicator of total nightlight intensity performs much better than other indicators, such as the luminous coverage area or average nightlight intensity. Zeng et al. (2011) compared nightlight and land cover data to disaggregate population and found that land use/cover data performed slightly better than the nightlight satellite data. Li et al. (2018) used NPP/VIIRS nightlight data to simulate the spatial distribution of the population of Beijing and found that the overall estimation accuracy achieved with nightlight data was about 62% as calculated with reference to the census data.

In addition to nightlight satellite data, intermediate-to high-resolution optical satellite data are also often used to estimate the population distribution by monitoring human settlements on the ground (Lung et al., 2013). In contrast to luminous satellite data, human physical settlements, such as residences and buildings, are the most important features to exploit for estimation of the population distribution (Ye et al., 2019). Stevens et al. (2015) used satellite data and open-source Geographic Information Systems data to estimate the worldwide population, which is also well known as the WorldPop program. Similarly, the Gridded Population of the World program initiated by Columbia University explored the integration of remote sensing and open geographic information data to produce information regarding the global population distribution. Moreover, Freire et al. (2018) explored the use of low- and medium-resolution satellite data to obtain the Global Human Settlement Layer and then combined the spectral and textural information from satellite images with the Global Human Settlement Layer data to generate a global population distribution map.

However, the current population estimation accuracy of remote sensing-based methods is generally not high; some studies confirmed that the overall accuracies of four existing population estimation products for China based on satellite data are about 60% (Bai et al., 2018; Li et al., 2018), which limits their wide application. Compared with conventional remote sensing methods, location-based service (LBS) data from mobile applications such as

social media (e.g., Facebook, Twitter), crowd-sourced reviews (e.g., Yelp, TripAdisor, and Meituan-Dianping), and transportation services (e.g., Uber, Didi), can reflect the social activities of individuals within a city from the perspective of social instead of physical evidence in space, which indicates that LBS data may have great potential in mapping the distribution of people within a city (França et al., 2016; Patel et al., 2017). Based on the spatial attribute of mobile phone data, Deville et al. (2014) estimated the population distribution and noted that mobile phone data are more effective in areas with a sparse population and a lack of census data. Patel et al. (2017) used geospatial information from Twitter to successfully simulate the spatial distribution of population in Indonesia. Other than spatial attribute, LBS data also has a high degree of time continuity, thus, some scholars even explored the use of LBS data in detecting spatio-temporal patterns of human activities as well as migration flows (Lai et al., 2019; Santa et al., 2019; Song et al., 2019; Song et al., 2020). These studies indicated that LBS data have great potential in estimating the population distribution within cites.

Based on the rich research about population mapping with the use of geospatial data from multiple sources, this study aims to 1) perform a comparative analysis and evaluation of the performance of various data in mapping population, including Tencent LBS data, nightlight satellite data, and satellite-based land use/cover data; 2) combine remote sensing and LBS data to assess various population spatialization simulation methods and technologies to achieve precise population estimations for China and verify its accuracy. The results and quantitative simulation tools could provide important population data and technical support for the long-term plans for countries and regions for which population data are lacking.

2. Study area and datasets

2.1. Study area

China's vast population of more than 1.3 billion has an uneven distribution within various administrative units. As a result, coarse administrative units (e.g., the province level and prefectural level) may not adequately describe the complex pattern of population in China. Thus, the finest administrative level—the county, of which more than 2000 exist—was selected for this study, in which open datasets from multiple sources were investigated to obtain population estimations for China. Fig. 1(a) shows the spatial distribution of China's entire population based on census data from the statistics bureau of China; red indicates a larger population, and green indicates a smaller population. The population map shows that

more people are crowded in the eastern part of China, especially in the southeast coastal areas. The most densely populated areas include the southern Pearl River Delta area and the eastern Yangtze River Delta area. According to the demographic data, the top five provinces in terms of population are Guangdong, Shandong, Henan, Sichuan, and Jiangsu, with more than 80 million in each province, which is consistent with the distribution map in Fig. 1(a). Over the past 30 years, China's population has increased from 1.15 billion in 1990 to 1.40 billion in 2018, with an average annual growth rate of about 0.4%. Meanwhile, the total urban area increased from 12,855 square kilometers in 1990 to 58,455 square kilometers in 2018, with an average growth rate of about 5.5%. The urbanization rate thus far exceeds the population growth rate, and the rapid urbanization process has greatly altered the living environment of most regions in China. Quick and accurate assessment of the population distribution is conducive to accurate assessment of the human-earth relationship and to the execution of appropriate spatial planning and environmental protection strategies to achieve sustainable development.

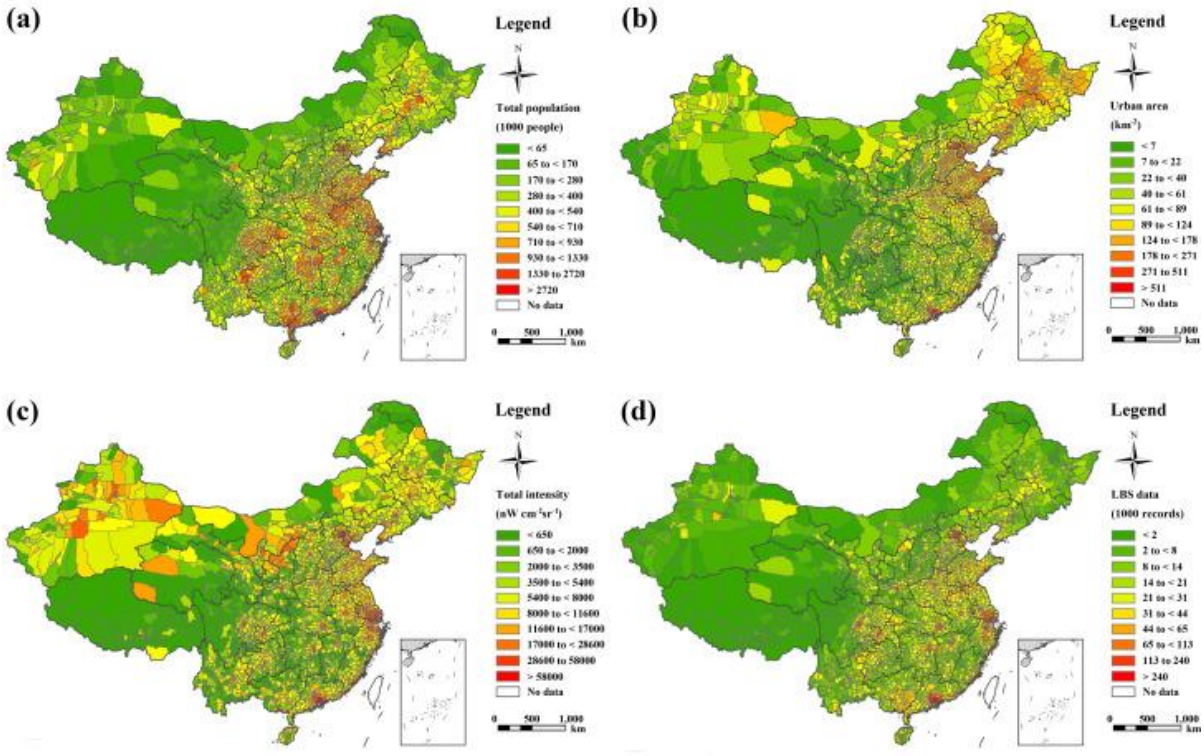

Fig. 1. Study area and collected datasets. (a) Counties/districts of China by population; (b) Satellite-derived built-up urban area at county level; (c) Overall NPP/VIIRS night-light intensity at county level; and (d) Tencent LBS records at county level.

2.2. Data and research idea

The collected data include census data, night-time images, satellite-derived land cover products, and Tencent LBS records (hereafter “LBS data”) at the county and prefecture level; all datasets were collected for 2015. The census data were obtained from the demographic yearbooks of the statistical department of each county. Land use/cover data were obtained from the standardized global land use/cover products provided by the European Space Agency using satellite data. Because the original land use/cover product covers the entire world, the boundary data of China were used to draw its land cover map (see Fig. 1 (b)). Night-time satellite data were obtained from the yearly averaged NPP/VIIRS nightlight data product provided by NASA, in which the effects of clouds have been eliminated (shown in Fig. 1 (c)). Because Tencent is one of the largest Internet service providers in China, the location-based service records from Tencent's active uses were collected through the method introduced in Song et al. (2018) for this study. This dataset includes data from a total of about 0.8 billion online users between January 1, 2015 and December 31, 2015, and it is projected on a map based on the spatial location of each LBS record (shown in Fig. 1(d)). Thus, this map can reflect the spatial distribution of people in China with an online presence. To ensure that the various datasets are consistent and comparable, all datasets were projected into a unified geographic coordinate system with the same spatial resolution.

Herein, the census data at the county and city levels were used as prior information to establish population modeling and to assess accuracy, whereas the obtained remote sensing data and LBS data were used as the prediction factors to simulate the population for each county. The census data are vector-based data based on administrative boundaries, unlike the raster-based indicators from the satellite or social sensor. To ensure that the census data were consistent with the indicators, an area-based statistical method was adopted to aggregate the sum of each indicator on each county. Given that China has 2868 counties, if three different indicators (e.g., land cover data, nightlight data, and LBS data) are used in this study, the predictor size will be 2868×3. Furthermore, the relationship between the population and the indicators can be modeled. To ensure that the obtained model is representative and adaptive, data from all 2868 counties were used to build the population estimation model, whereas the census data for 344 cities were used to verify the model.

3. Methods and results

The tested methods include correlation analysis, stepwise regression analysis, and geographical regression analysis. Correlation analysis was used to detect the association between population and some potential indicators. On this basis, three important indicators—

the total sum of LBS records, the total built-up area, and the sum of the nightlight intensity—were selected from various potential variables, including but not limited to some possible indicators such as the average nightlight intensity and the total sum of agricultural land. The correlation coefficients between population and the obtained indicators were 0.78, 0.72, and 0.65, respectively. Thus, these three indicators—the total sum of LBS records, the total built-up area, and the sum of the nightlight intensity—were selected as predictors and used in this study.

Making use of the selected population indicators, a stepwise regression method was adopted to conduct significance analysis. The results indicated that each of the three indicators was selected in the model. Given the huge spatial differences in population and China's imbalanced development, a geographically weighted regression (GWR) model was adopted in this study, and its effectiveness is addressed in the Discussion.

Referring to Huang et al. (2010), the following geographically weighted model can be constructed by making use of all obtained data for each county:(1)$Pi=\beta 0(ui,vi)+\beta 1(ui,vi)QQi+\beta 2(ui,vi)LCi+\beta 3(ui,vi)NLi+\varepsilon i$where Pi reflects the population of county i; QQi, LCi, and NLi reflect the three indicators used (i.e., the sum of Tencent's users, the total built-up area, and the sum of the nightlight intensity) for county i; and $\beta j=0,1,2,3(ui,vi)$ reflects the jth regression coefficient for county (ui,vi). Herein, each county should have a group of different regression coefficients.

Based on the GWR model in formula (1), the population data and the three corresponding indicators (LBS, land cover, and nightlight data) for 2868 counties were used to solve the model's regression coefficients and obtain a population estimate for each county. The regression results show that the model's goodness of fit (R2) value reaches 0.88, which indicates that the three predictors can adequately explain the population at the county level. Fig. 2 (a) shows the distribution of population at the county level using the proposed model. For comparison, the real population data are also provided in Fig. 2 (b). A comparison of the estimates with the census data shows that the estimates for some eastern counties (prefectures) tended to be overestimated or underestimated, whereas the errors for some western counties were relatively small. This finding indicated that the proposed approach tended to have a larger bias for the eastern parts of China with a highly dense population, whereas the estimates are likely more accurate in the western counties with fewer people. Based on the census data from various administration levels, the population estimation errors at the provincial, city, and county levels were 2.7%, 9.6%, and 26.0%, respectively.

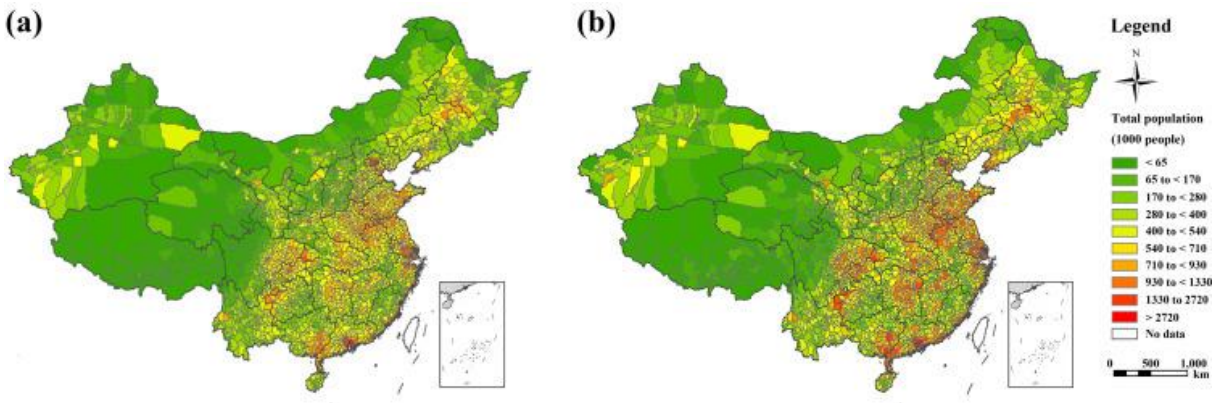

Fig. 2. Distribution of population in China based on: (a) Our proposed population estimation model using both LBS and remote sensing data; (b) Census data at county level (for validation).

Moreover, the population estimations at cities in China using the proposed method was compared with the estimations from four mainstream global population datasets, including GPW, GHS, LandScan and WorldPop (Tatem, 2017; Freire et al., 2018; Xu et al., 2020). Two accuracy indices, including mean average error (MAE) and root mean square error (RMSE) were used to conduct an accuracy assessment by comparing population estimations from different products with the real census data for all 344 cities in China. Table 1 shows the accuracy statistics of the population estimations from different data sets, based on which the results indicated that our proposed model using both LBS and remote sensing data performs better than the four mainstream population datasets over China, as MAE of the proposed method is 9.6%, which is lower than the figures from other data sets, including WorldPop (10%), LandScan (10.9%), GPW (11.6%), and GHS (11.7%). The results also indicated that among four existing products, both WorldPop and LandScan have slightly better performance than GPW and GHS in mapping population in China, as the RMSEs from WorldPop and LandScan were 586850 and 558050, respectively, which are less than the figures from GPW (600050) and GHS (608850). Nevertheless, all the tested population products have high population estimation accuracy (above 88%) for cities in China.

Table 1. Accuracy assessment for the population estimations from different products.

	WorldPop	LandScan	GPW	GHS	Ours
MAE (%)	10.0	10.9	11.6	11.7	9.6
RMSE (people)	586850	558050	600050	608850	473550

4. Discussion

4.1. Impact of various indicators on population estimates

The performance of various types of data, including Tencent LBS data, the nightlight remote sensing data, and the satellite-derived land cover data, on population estimates was assessed via correlation and regression analysis methods.

Fig. 3 shows the performance of various indicators in mapping county level population based on the cumulative percent of real population and different indicators (including Tencent LBS data, nightlight data, and land cover data), in which three curves with different colors correspond to the results obtained using different indicators. For each curve, the more the scattered points are concentrated in a linear pattern, the more effective the use of this indicator to estimate population. These results show that the use of Tencent LBS data (red curve shown in Fig. 3) performed better than the use of other kinds of data, including nightlight data (blue curve shown in Fig. 3) and land cover data (green curve shown in Fig. 3), because the scattered points from the LBS data are more concentrated along a centered line than the points from either the nightlight or land cover data, in which some points tend to deviate greatly from their centered lines. This finding indicates that a larger population estimation error might be obtained with the use of either nightlight or land cover data. The correlation analysis result also indicated the superiority of LBS data to the others; the correlation coefficient between Tencent LBS data and the population was 0.78, whereas those for nightlight data and land cover data were 0.72 and 0.65, respectively.

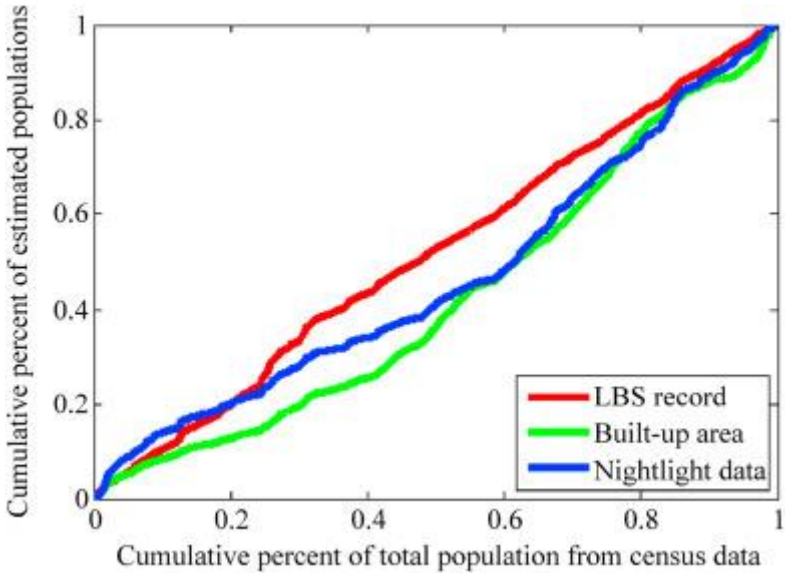

Fig. 3. Comparison of different indicators in mapping population using the cumulative percent of real population and different indicators at all counties in China.

The GWR model was applied to model the relationships between population and various indicators. Given that three kinds of data were available (Tencent LBS data, total built-up area, and nightlight intensity), three regression models were built, with each indicator used to build a single model. Table 2 shows the results of the built model using each of the three indicators. Tencent LBS data performed best (overall accuracy, 88.9%), followed by the land cover data (87.1%) and the nightlight data (85.5%). The corresponding R2 values for the established models based on the three indicators were 0.82, 0.79, and 0.74, respectively. These results indicate that the currently widely popular LBS data (such as Tencent LBS data) perform well in simulating population because their overall accuracy for various cities can be improved by 3–5 percentage points when compared with the use of nightlight or satellite-derived land cover data.

Table 2. Statistical results using spatial regression model with three indicators.

	Model 1 (Tencent LBS data)	Model 2 (Total built-up area)	Model 3 (Nightlight intensity)
Constant	248580	236650	334290
F1: Tencent LBS data	10.3	–	–
F2: Total built-up area	–	4606.8	–
F3: Nightlight intensity	–	–	30.0
R2	0.82	0.79	0.74
Accuracy (%)	88.9	87.1	85.5

4.2. Performance of different models

The performances of the conventional ordinary least squares regression (OLS) method and the GWR method were assessed and compared with the tested indicators. The performances of the indicators in predicting the population of various administrative units were tested with both methods. Fig. 4 shows the distribution of the population estimation error with various combinations of indicators using either the OLS or GWR method; orange indicates that the population is overestimated, and blue indicates that it is underestimated. The simulation results in Fig. 4 show that the GWR method performs much better than the OLS method, because the colors in Fig. 4(d)-4(f) with the GWR method are much lighter than those (Fig.

4(a)–(c)) using the OLS method; this means that a smaller prediction error (better prediction result) was obtained with the GWR method, regardless of which indicator was used.

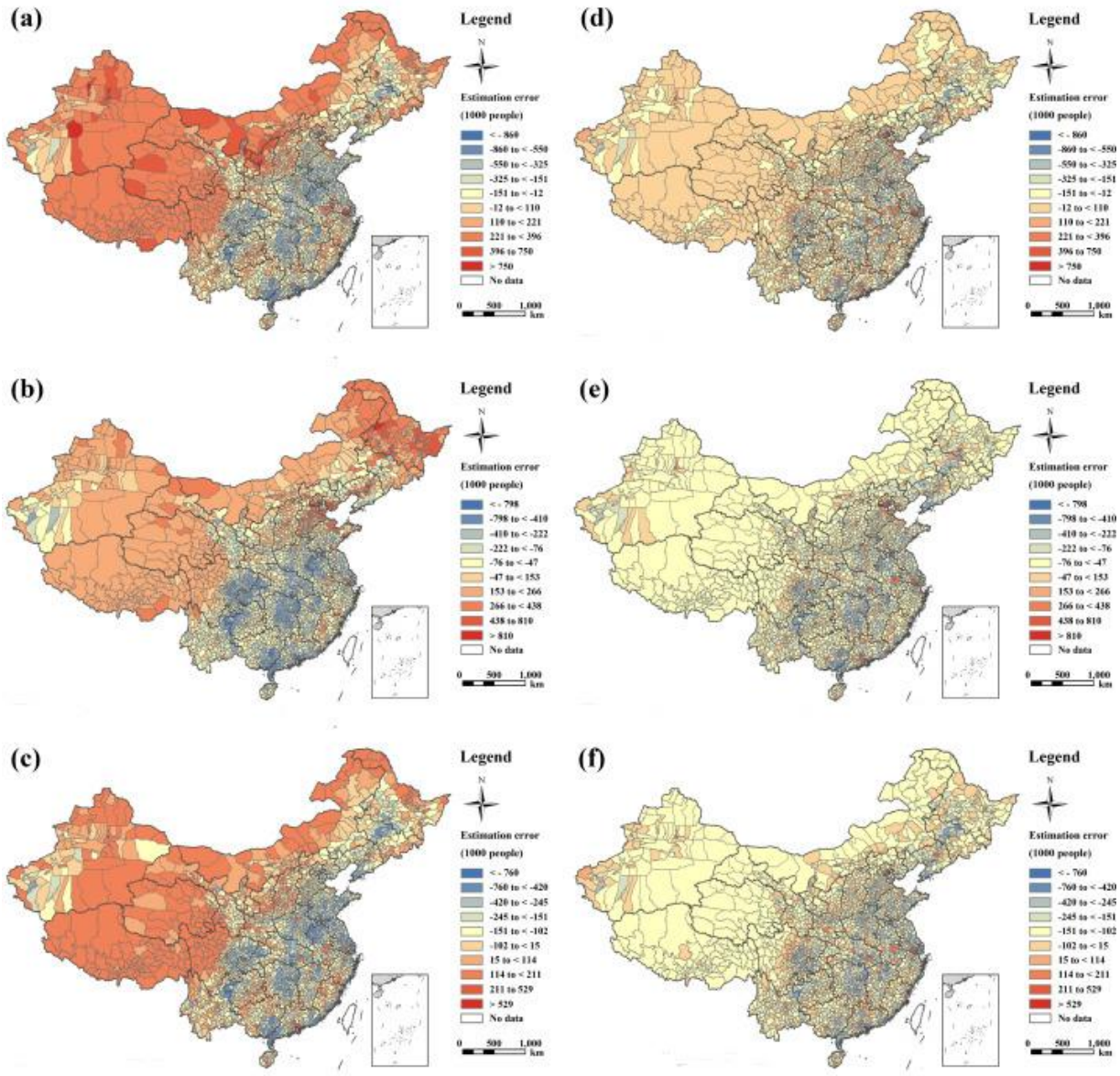

Fig. 4. Distribution of population estimation error with OLS and GWR methods using different indicators (Tencent LBS data, land use/cover data, nightlight data). (a) OLS model with nightlight data; (b) OLS model with land use/cover data; (c) OLS model with Tencent LBS data; (d) GWR model with nightlight data; (e) GWR with land use/cover data; (f) GWR model with Tencent LBS data.

Quantitative indices, such as the average percentage error, can be obtained by comparing the population estimates with the actual census data. Table 3 shows the average percentage error with the various indicators using either the OLS or GWR method. The results indicate that the GWR method always performed much better than the conventional OLS method,

regardless of which indicator was used. The average percentage errors with the GWR method were all below 15%, whereas those with the OLS method were all above 28%; thus, the average improvement with the GWR method can reach to 20 percentage points. For example, if only the nightlight data were used, the overall prediction errors with the OLS and GWR methods were 37.4% and 14.5%, respectively. In this case the improvement is 23 percentage points. The testing results also indicate that the best result can be reached with the use of all of the tested indicators, including land use/cover data, Tencent LBS data, and nightlight data; the overall prediction error was 9.6%, which means that the overall prediction accuracy with all indicators was 90.4%. The results indicate that the use of Tencent LBS data performs much better than the use of nightlight data or land use/cover data and thus these data have great potential for the rapid and efficient estimation of an urban population.

Table 3. Population estimation error based on various strategies combining a method and indicator(s).

Method	Indicator 1 error (Tencent LBS data)	Indicator 2 error (Built-up area data)	Indicator 3 error (Nightlight data)	All indicators error (1, 2, & 3)
OLS	29.4%	35.3%	37.4%	28.1%
GWR	11.1%	12.9%	14.5%	9.6%

4.3. Population mapping implications

Differences in economic levels and the uneven population distribution in China can lead to some uncertainties in its population mapping, including overall accuracy and spatial differences. Improved characterization of the spatial difference in population mapping and improvements in the use of various indicators to simulate the population distribution are discussed in this section. Given the ability of the GWR method to handle spatial differences, the regression coefficient of the GWR method can be used to reflect the spatial characteristics of the population distribution. Fig. 5(a)-(c) show the standardized regression coefficients (Z-Scores) by GWR method using various indicators, including Tencent LBS data, land cover data, and nightlight data; larger regression coefficients are in red, and smaller regression coefficients are in blue. The results show that nearly all of the obtained regression coefficient in the central (e.g., Hunan province) and western (e.g., Tibet) parts of China tend to have a larger regression coefficient (shown in red), which means that these factors might have great impact on population estimation for these areas. In particular, the use of Tencent LBS data tends to show great population estimation effect in western and northern China and in the

border areas of the central provinces, which means that these areas might have scarce signal transmission equipment and low social media usage. For the land use/cover data, large regression coefficients also tend to be obtained for some central (e.g., Hunan province) and western (e.g., Tibet) areas, possibly because these regions have low urbanization rates due to their geographic location (e.g., mountainous areas) and for historical reasons. The results with the use of nightlight imagery are similar to those with LBS data, where large regression coefficients are obtained in the western, northeastern, and some border areas of China's central provinces, possibly due to the low economic levels in these areas; thus, the per capita luminous usage (including industrial and domestic electricity consumption) is lower than in other areas. The results obtained with various indicators indicate that the signal station distribution, economic level, and urbanization rate all contribute to the overall population estimation accuracy. Thus, the appropriate selection of indicators can effectively improve the spatial uncertainties of population estimation in China due to regional differences, such as in the western, central, and northeastern parts of China.

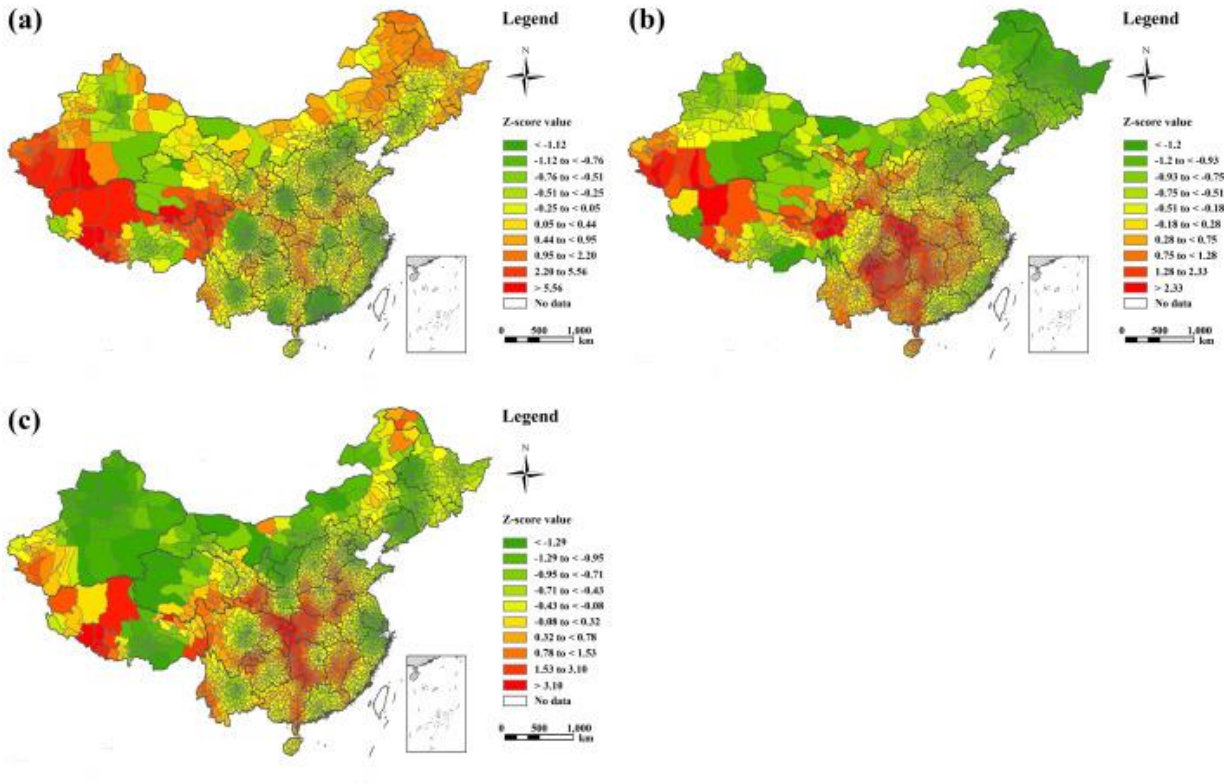

Fig. 5. Z-Scores of regression coefficient with GWR method using different population indicators: (a) Tencent LBS data; (b) land use/cover data; (c) nightlight image data.

5. Conclusions

This study investigated the use of up-to-date LBS data and conventional multisource remote sensing data to map the distribution of China's population to facilitate intelligent urban resource management and spatial planning. The study compares and evaluates the performance of various data sources in population estimation and shows that LBS data have great potential for mapping the distribution of China's population. The study compares and analyzes the performance of various sources of geospatial data on population disaggregation and shows that LBS data can effectively reflect the spatial distribution of China's population and have good population prediction ability. An experiment using census data from 344 prefectures in China found that the overall prediction accuracy using Tencent LBS data can reach 89%, which is much better than the results using either satellite-derived land cover data (87%) or nightlight imagery data (85%). The results indicate that Tencent LBS data are superior to both conventional land use/cover data and nightlight imagery data. It has high population modeling and prediction capabilities, which can help to achieve rapid monitoring and management of human resources in China.

Based on a comparative analysis, this study further investigates the use of multiple sources of sensor data (e.g., LBS and remote sensing data) to map China's population via various methods. The experimental results indicate that the nonlinear GWR method performs much better than the conventional OLS method at population mapping, regardless of whether a single factor or multiple factors are used. The overall prediction accuracy using the GWR method can achieve an improvement of more than 20 percentage points over the use of the conventional OLS method. The results also indicate that multiple factors can achieve a better result than a single factor, regardless of which method is used, and the overall improvement can reach 3 to 5 percentage points. This finding indicates that the use of multiple data sources with the GWR model is much preferred in the task of population mapping in China.

Due to the restriction of available datasets, this study focuses on static population simulation and estimation, so this method has not been verified or applied to dynamic tasks with rapid population movement. The possible integration of real-time LBS data and satellite data to obtain up-to-date population estimations will be among our further pursuits. Moreover, this study focuses mainly on population estimates at the city and county level due to the limited spatial resolution of the datasets; thus, the population at the community and street levels within a city is not considered. Because future LBS data are projected to be more precise, rich, and open (e.g., geotagged position information, mobile communication data, social media), the use of these highly precise open data sources to monitor urban population

mobility and to predict the behavior of various citizens at the street level will be beneficial for intelligent urban planning and community management, which requires further investigation.

References

Z. Bai, J. Wang, M. Wang, M. Gao, J. Sun. Accuracy assessment of multi-source gridded population distribution datasets in China. Sustainability, 10 (5) (2018), p. 1363

Y. Cai. An assessment of China's fertility level using the variable-r method. Demography, 45 (2) (2008), pp. 271-281

J. Chen, W. Fan, K. Li, X. Liu, M. Song. Fitting Chinese cities' population distributions using remote sensing satellite data. Ecological Indicators, 98 (2019), pp. 327-333

X. Chen, W. Nordhaus. A test of the new VIIRS lights data set: Population and economic output in Africa. Remote Sensing, 7 (4) (2015), pp. 4937-4947

P. Deville, C. Linard, S. Martin, M. Gilbert, F.R. Stevens, A.E. Gaughan, ..., A.J. Tatem. Dynamic population mapping using mobile phone data. Proceedings of the National Academy of Sciences, 111 (45) (2014), pp. 15888-15893

C.D. Elvidge, K.E. Baugh, J.B. Dietz, T. Bland, P.C. Sutton, H.W. Kroehl. Radiance calibration of DMSP-OLS low-light imaging data of human settlements. Remote Sensing of Environment, 68 (1) (1999), pp. 77-88

U. França, H. Sayama, C. McSwiggen, R. Daneshvar, Y. Bar-Yam. Visualizing the "heartbeat" of a city with tweets. Complexity, 21 (6) (2016), pp. 280-287

S. Freire, M. Schiavina, A.J. Florczyk, K. MacManus, M. Pesaresi, C. Corbane, ..., R. Sliuzas . Enhanced data and methods for improving open and free global population grids: Putting 'leaving no one behind' into practice. International journal of digital earth (2018), pp. 1-17

B. Huang, B. Wu, M. Barry. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. International Journal of Geographical Information Science, 24 (3) (2010), pp. 383-401

S. Lai, E. zu Erbach-Schoenberg, C. Pezzulo, N.W. Ruktanonchai, A. Sorichetta, J. Steele, ..., A.J. Tatem. Exploring the use of mobile phone data for national migration statistics. Palgrave communications, 5 (1) (2019), pp. 1-10

X. Li, W. Zhou. Dasymetric mapping of urban population in China based on radiance corrected DMSP-OLS nighttime light and land cover data. The Science of the Total Environment, 643 (2018), pp. 1248-1256

T. Lung, *et al.* Human population distribution modelling at regional level using very high resolution satellite imagery. Applied Geography, 41 (2013), pp. 36-45 2013

N.N. Patel, F.R. Stevens, Z. Huang, A.E. Gaughan, I. Elyazar, A.J. Tatem. Improving large area population mapping using geotweet densities. Transactions in GIS, 21 (2) (2017), pp. 317-331

F. Santa, R. Henriques, J. Torres-Sospedra, E. Pebesma. A statistical approach for studying the spatio-temporal distribution of geolocated tweets in urban environments. Sustainability, 11 (3) (2019), p. 595

Y. Song, B. Chen, M.P. Kwan. How does urban expansion impact people's exposure to green environments? A comparative study of 290 Chinese cities. Journal of Cleaner Production, 246 (2020), p. 119018

Y. Song, B. Huang, Q. He, B. Chen, J. Wei, R. Mahmood. Dynamic assessment of PM2. 5 exposure and health risk using remote sensing and geo-spatial big data. Environmental Pollution, 253 (2019), pp. 288-296

Y. Song, B. Huang, J. Cai, B. Chen. Dynamic assessments of population exposure to urban greenspace using multi-source big data. Science of the Total Environment, 634 (2018), pp. 1315-1325

F.R. Stevens, A.E. Gaughan, C. Linard, A.J. Tatem. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. PloS One, 10 (2) (2015), Article e0107042

P. Sutton, D. Roberts, C. Elvidge, K. Baugh. Census from Heaven: An estimate of the global human population using night-time satellite imagery. International Journal of Remote Sensing, 22 (16) (2001), pp. 3061-3076

M. Tan, X. Li, S. Li, L. Xin, X. Wang, Q. Li, ..., W. Xiang. Modeling population density based on nighttime light images and land use data in China. Applied Geography, 90 (2018), pp. 239-247

A.J. Tatem. WorldPop, open data for spatial demography. Scientific Data, 4 (1) (2017), pp. 1-4

L. Wang, S. Wang, Y. Zhou, W. Liu, Y. Hou, J. Zhu, F. Wang. Mapping population density in China between 1990 and 2010 using remote sensing. Remote Sensing of Environment, 210 (2018), pp. 269-281

J. Wu, Z. Yu, Y.D. Wei, L. Yang. Changing distribution of migrant population and its influencing factors in urban China: Economic transition, public policy, and amenities. Habitat International, 94 (2019), p. 102063

Y. Xu, H.C. Ho, A. Knudby, M. He. Comparative assessment of gridded population data sets for complex topography: A study of Southwest China. Population and Environment (2020), pp. 1-19

T. Ye, N. Zhao, X. Yang, Z. Ouyang, X. Liu, Q. Chen, ..., P. Jia. Improved population mapping for China using remotely sensed and points-of-interest data within a random forests model. The Science of the Total Environment, 658 (2019), pp. 936-946

S. Yu, Z. Zhang, F. Liu. Monitoring population evolution in China using time-series DMSP/OLS nightlight imagery. Remote Sensing, 194 (10) (2017), pp. 1-27

C. Zeng, Y. Zhou, S. Wang, F. Yan, Q. Zhao. Population spatialization in China based on night-time imagery and land use data. International Journal of Remote Sensing, 32 (24) (2011), pp. 9599-9620

G. Zhang, X. Guo, D. Li, B. Jiang. Evaluating the potential of LJ1-01 nighttime light data for modeling Socio-economic parameters. Sensors, 19 (6) (2019), p. 1465