Accuracy Assessment of Representative Population Grid Datasets in Arid Regions: A Case Study of the Gansu-Ningxia-Qinghai Region (Postprint)

Authors: Xiao Dongsheng, Wang Ning, Liu Zhicheng, Wang Ning

Date: 2023-04-07T16:57:20+00:00

Abstract

High-precision population grid datasets hold significant value in risk assessment, disaster emergency response, ecological environment protection, and regional development and planning. Differences in input data precision and model selection result in distinct characteristics and advantages, making it crucial to evaluate the accuracy of representative datasets and analyze their applicable conditions. This study assesses the accuracy of the WorldPop dataset and the Gridded Population of the World version 4 (GPWv4) dataset in the arid regions of northwestern China, specifically Gansu Province, Ningxia Hui Autonomous Region, and Qinghai Province. Using township administrative divisions—the finest available unit of Chinese census data—as the study unit, we conduct correlation analysis between the WorldPop and GPWv4 datasets and the 2020 Seventh National Census data, calculate the spatial distribution of statistical errors and relative errors, and quantitatively evaluate the accuracy of each dataset. We qualitatively analyze the mapping performance of the datasets through visual estimation and discuss their error sources. Statistical error results indicate that the WorldPop dataset exhibits higher accuracy, with its correlation coefficient (r), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) reaching 0.76, 23016, 0.73, and 0.60, respectively, while the corresponding statistics for the GPWv4 dataset are 0.70, 22297, 0.75, and 0.58. Additionally, the spatial distribution of relative errors reveals that the WorldPop dataset provides more accurate estimations across a larger number of regions. Visual estimation results demonstrate that both population grid datasets exhibit similar mapping performance, characterized by densely populated eastern regions and sparsely populated western regions. Evaluation studies on the accuracy of population grid datasets in arid regions facilitate analysis of dataset error sources and guide rational dataset utilization.

Future research should leverage ancillary data on human habitation to generate population distribution patterns specific to arid regions, thereby enhancing the accuracy of population datasets for northwestern arid areas.

Full Text

Preamble

ARID LAND GEOGRAPHY
ChinaXiv Partner Journal
Vol. 46 No. 3 Mar. 2023

Accuracy Assessment of "Representative Population Grid Datasets" in Arid Regions: A Case Study of Gansu-Ningxia-Qinghai Region

XIAO Dongsheng[1,2], WANG Ning[1,2], LIU Zhicheng[1,2]
(1. School of Civil Engineering and Surveying, Southwest Petroleum University, Chengdu 610500, Sichuan, China;
2. Disaster Prevention and Emergency Response Research Center of Surveying, Remote Sensing and Geographic Information, Southwest Petroleum University, Chengdu 610500, Sichuan, China)

Abstract: High-precision population grid datasets hold significant value in risk assessment, disaster emergency response, ecological environment protection, and regional development planning. Variations in input data precision and model selection endow these datasets with different characteristics and advantages. Therefore, evaluating the accuracy of representative GPWv4 and WorldPop datasets and analyzing their applicable conditions is of great importance. This study assessed the accuracy of the WorldPop and GPWv4 datasets in the arid regions of Gansu Province, Ningxia Hui Autonomous Region, and Qinghai Province in northwestern China. Using the finest available unit of Chinese census data (township administrative divisions) as the research unit, correlation analysis was conducted between these datasets and the seventh census data from 2020. Statistical and relative error distributions were calculated to quantitatively evaluate each dataset's accuracy. Mapping performance was qualitatively analyzed through visual estimation. Finally, error sources were discussed. Statistical error results demonstrate that the WorldPop dataset achieves higher accuracy than GPWv4, with correlation coefficient (r), root mean square error, mean absolute error, and mean absolute percentage error of 0.76, 23016, 0.73, and 0.60, respectively, compared to 0.70, 22297, 0.75, and 0.58 for GPWv4. Concurrently, the spatial distribution of relative errors indicates that WorldPop accurately estimates population in more areas. Visual estimation results reveal similar mapping performance between the two population grid datasets, both showing dense population in the east and sparse distribution in the west. This accuracy assessment of population grid datasets in arid regions facilitates analysis of error sources and guides rational dataset utilization. Future research should incorporate auxiliary data on human habitation to generate arid-region-specific

population distribution patterns, thereby improving the accuracy of population datasets in northwest China's arid areas.

Keywords: population grid dataset; GPWv4 dataset; WorldPop dataset; accuracy evaluation; arid area in northwest China

Introduction

Population data serve as crucial foundational information for risk assessment, disaster emergency response, ecological environment protection, and regional development planning, and constitute the primary data source for population spatialization research. Traditionally, population data are collected through national censuses using administrative regions as basic units, which suffer from low spatial and temporal resolution and fail to adequately reveal intra-administrative population distribution patterns. Moreover, administrative unit-based statistical population data cannot be easily coupled with grid-based geographic unit data, hindering flexible utilization in spatial analysis and statistics. Consequently, establishing population grid datasets that reflect actual population spatial distribution and predict population data and its spatiotemporal patterns holds important theoretical and practical significance.

Currently, numerous population grid datasets have been developed, with widely used examples including LandScan, Gridded Population of the World (GPWv4), Global Rural-Urban Mapping Project (GRUMP), OpenPopGrid, Global Human Settlement Layer (GHSL), and WorldPop. These datasets have been flexibly applied in various research contexts. Presently, only WorldPop and GPWv4 have been updated to 2020, meeting the requirements for contemporary spatial analysis. However, the complexity of human spatial distribution and limitations of dataset generation models inevitably introduce errors into population grid datasets. Evaluating the accuracy of existing datasets not only helps data producers identify shortcomings but also enables users to understand dataset characteristics and select appropriate data for specific research needs.

Accuracy assessment of population grid datasets remains challenging. Two main approaches exist: evaluating the models and datasets that produce the data, or aggregating grid datasets to administrative units for comparison with census data. The former is difficult due to the complexity and mobility of population distribution, making it impossible to obtain accurate grid-level population values. The latter approach has been employed in several studies. For instance, Tatem et al. evaluated GRUMP dataset accuracy by constructing a global malaria prevalence map, revealing substantial differences in estimated population distribution among existing datasets. Wang Xuemei et al. compared GPWv3 and LandScan datasets with statistical data at the watershed scale in the Heihe River Basin, demonstrating that CnPop data exhibited higher accuracy. Bai Zhongqiang et al. assessed GPWv3, GRUMP, and WorldPop datasets using China's fifth census data, finding all three datasets exhibited low accu-

racy in northwest China. Xu Yong et al. evaluated WorldPop and GPWv4 accuracy in Yunnan, Guangxi, and Guizhou provinces in southwest China, while Lin Danchun et al. assessed four gridded population datasets in Guangdong Province. These studies confirm that population grid datasets demonstrate good accuracy in densely populated areas but significantly lower accuracy in northwest China's arid regions. Investigating the accuracy and influencing factors of these datasets in such arid areas represents an important research direction. Furthermore, most previous studies were based on 2010 census data, lacking the most recent findings.

Environmental factors such as water resources profoundly influence population distribution patterns. Northwest China is characterized by arid conditions that result in population aggregation patterns distinct from those in coastal regions with favorable climates. Previous studies have categorized the accuracy of population grid datasets in arid regions as inadequately estimated. Exploring error sources and dataset deficiencies in such areas can help data developers improve dataset accuracy. Therefore, this study utilizes 2020 seventh census data as ground truth to evaluate WorldPop and GPWv4 datasets, analyzing their accuracy and characteristics in northwest China's arid regions to address this research gap.

---

1 Study Area Overview

Gansu Province, Ningxia Hui Autonomous Region, and Qinghai Province (collectively referred to as Gan-Ning-Qing) are located in the arid region of northwestern China. The area borders Inner Mongolia Autonomous Region to the north, Xinjiang Uygur Autonomous Region to the northwest, Tibet Autonomous Region to the southwest, Sichuan Province to the south and southeast, and Shaanxi Province to the east. The study area comprises 20 prefecture-level administrative units, including 5 ethnic autonomous prefectures, totaling 107 county-level administrative divisions, covering a total area of 2.25$\times10^{6}$ km^2. The Gan-Ning-Qing region serves as a core area of the "Silk Road" economic belt, forming a crucial hub connecting China with Central Asia and Europe. Rich in natural resources, it represents a key region for China's "Western Development" strategy and serves as an energy production and export base for projects such as "West-East Gas Pipeline" and "North Coal South Transport." Despite its vast territory and abundant resources, the region exhibits extremely uncoordinated population distribution. According to 2020 census data, the total population is approximately 3.81$\times10^{7}, with an average population density of about 17 persons \cdot km^{-2}$, far below China's average population density.

Note: This figure was produced based on the standard map downloaded from the Standard Map Service website of the National Administration of Surveying, Mapping and Geoinformation (Approval No. GS(2019)1822), with no modifications to the base map boundaries. The same applies below.

[Figure 1: see original paper] Schematic diagram of the study area

2 Data and Methods

2.1 Census Data

Administrative divisions are hierarchical regions designated for administrative management convenience, also known as administrative areas. China's administrative divisions are structured into four levels: province, prefecture, county, and township. This study evaluates dataset accuracy at the township scale, which represents the finest available unit for Chinese census data. Township administrative boundary data were obtained from the National Geomatics Center of China at a scale of 1:100,000. Due to administrative adjustments, some township boundaries had changed. Using 2020 administrative divisions as reference, inconsistent boundaries were modified in ArcGIS software, resulting in 1,081 township administrative units. Township census data were sourced from district and county census bulletins available on official government websites. For example, census data for various streets and townships in Chengzhong District, Xining City, Qinghai Province, can be queried from the Chengzhong District People's Government website (http://www.xncz.gov.cn/info/2953/121813.htm). The township administrative boundaries and census data are illustrated in Figure 2. Township census data employ the WGS84 geocentric coordinate system and Albers equal-area projection as geographic references.

[Figure 2: see original paper] Number of statistical populations in townships

2.2 Population Grid Datasets

GPWv4 Dataset: The Gridded Population of the World version 4 (GPWv4) dataset is a global population grid dataset released by NASA's Socioeconomic Data and Applications Center (SEDAC). GPWv4 models global population distribution from 2000 to 2020 at 30 arc-second resolution (approximately 1 km at the equator) using basic input data comprising non-spatial population census data and spatially explicit administrative boundary data. Estimated populations are allocated to grid cells proportionally by area within census administrative units, while water bodies are used as a mask to prevent interference from lakes, rivers, and ice-covered areas.

GPWv4_{UNADJ}: The UN-adjusted GPWv4 dataset is modified from the original GPWv4 dataset based on population data provided by UN population agencies.

WorldPop Dataset: The WorldPop dataset provides annual population estimates from 2000 to 2020 at 100 m spatial resolution (3 arc-seconds at the equator). It incorporates spatial ancillary datasets such as settlement locations,

nighttime satellite imagery, roads, vegetation, terrain, and land use to correct residential and built-up area distributions. A random forest regression tree model generates predictive weighting layers to redistribute official census data into grid cells, achieving population spatialization.

WorldPop_{UNADJ}: The UN-adjusted WorldPop dataset is modified from the original WorldPop dataset based on UN population agency data.

[Figure 3: see original paper] Visually estimated population distributions of population grid data sets

2.3 Accuracy Evaluation Methods

To assess the accuracy of population grid datasets, we calculated the correlation coefficient (r), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) between estimated and statistical values. The formulas are as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}}$$

$$MAE = \frac{\sum_{i=1}^{n}|P_i - O_i|}{n}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{P_i - O_i}{O_i}\right|$$

where P_i represents the estimated population within the township administrative area for each grid dataset, O_i represents the corresponding township census population data, and n is the number of administrative units.

Relative error (RE) was calculated to evaluate dataset accuracy:

$$RE = \frac{P_i - O_i}{O_i} \times 100\%$$

Using the 1,081 township administrative divisions, ArcGIS zonal statistics were employed to aggregate grid dataset values and calculate estimated population densities for each administrative unit. Error statistics were then computed to evaluate each dataset's performance.

Basic information of the population grid data sets

3 Results

3.1 Visual Estimation

Figures 3 and 4 illustrate population distributions from the four population grid datasets across the study area. To enable comparative analysis at uniform resolution, WorldPop data were resampled to 1 km resolution using ArcGIS. Compared with census data, the grid population density data in Figure 3 show similar trends: dense population distribution in the east and sparse distribution in the west. Influenced by climatic and geographic conditions, northwest China exhibits a radial distribution pattern centered on major cities. Key population agglomerations include Yinchuan, Wuzhong, and Zhongwei in Ningxia; Lanzhou, Tianshui, Wuwei, Zhangye, and Jiuquan in Gansu; and Xining and Golmud in Qinghai. In contrast to the highly concentrated distribution in eastern areas, the western region features large areas with sparse, dispersed populations, with most areas exhibiting population densities below 10 persons · km^{-2}.

In terms of mapping performance, the resampled WorldPop dataset shows visual effects similar to census data, indicating that predictive population grid datasets generally reflect actual population distribution with good accuracy. However, significant visual differences exist when comparing GPWv4 and WorldPop datasets. GPWv4 exhibits noticeable mosaic effects due to its simple areal weighting method and low spatial resolution. In contrast, WorldPop demonstrates better spatial continuity and provides more detailed spatial heterogeneity. By integrating widely available remote sensing and geospatial datasets (settlement locations, land cover, building maps, vegetation) to create symmetric weighting layers, the random forest model generates 100 m resolution population density grids, resulting in more pronounced central agglomeration patterns. However, due to coarse classification in ancillary datasets, population density transitions in WorldPop are not entirely smooth. Overall, WorldPop consistently outperforms GPWv4 in describing population distribution differences, whether in densely populated eastern areas or sparsely populated western regions.

3.2 Statistical Analysis

Error statistics of the four population grid data sets

Table 2 presents the township-level population density error statistics for the four datasets. WorldPop exhibits the highest correlation coefficient (r = 0.76), followed by GPWv4 (r = 0.70), WorldPop_{UNADJ} (r = 0.69), and GPWv4_{UNADJ} (r = 0.68). The RMSE values are 23016, 22297, 23571, and 22651, respectively, while MAE values are 0.73, 0.75, 0.77, and 0.77, respectively. MAPE values are 0.60, 0.58, 0.62, and 0.60, respectively. Both GPWv4 and WorldPop show improved accuracy after UN adjustment. Overall, WorldPop outperforms GPWv4 in population grid data accuracy, while UN-adjusted GPWv4 shows improved precision. Notably, all four datasets exhibit

large RMSE values, indicating highly discrete population density estimation errors, while small MAE values suggest high overall accuracy.

[Figure 4: see original paper] Correlation analysis between statistical population at township level and predicted population in grid datasets

Figure 4 displays scatter plots of population density between the four grid datasets and census data at the township level. The overall trend lines approximate the 1:1 line, indicating good consistency between grid datasets and census data. WorldPop samples cluster closest to the 1:1 line with the highest r value (0.76), while GPWv4 samples show the greatest dispersion with the lowest r value (0.70). All datasets contain some outlier samples with large errors, likely caused by non-natural factors such as natural disasters (e.g., mudslides) or residential demolition-induced migration. Discrepancies between township population growth rates and grid dataset growth rates also contribute to prediction biases. Although outlier regions differ across datasets, they share similar characteristics, possibly attributable to different dataset generation methods and input variables.

To ensure experimental accuracy, the 11 samples with largest errors in each dataset (1% of total) were removed and error statistics recalculated. As shown in Table 3, all four datasets show improved accuracy after outlier removal, particularly GPWv4, while WorldPop exhibits minimal r value change. Concurrently, RMSE values improve significantly and MAPE values decrease noticeably. The best overall performance is achieved by GPWv4, followed by WorldPop, GPWv4_{UNADJ}, and WorldPop_{UNADJ}.

[Figure 5: see original paper] Spatial distributions of relative errors of the four data sets

The spatial distribution of relative errors across the four datasets shows similar patterns: severe overestimation dominates in western areas, while eastern areas show relatively balanced error distribution. Relative errors between -25% and 25% are considered accurate estimation; -50% to -25% and 25% to 50% are considered underestimation and overestimation, respectively; -100% to -50% and 50% to 100% are considered severe underestimation and severe overestimation. Figure 5 reveals that GPWv4 contains significantly more severe overestimation areas than other datasets, particularly in large township administrative units in western and northern regions, with few underestimated areas. Overestimation areas are mainly distributed in hilly, Gobi, and plateau regions, while accurate estimation concentrates in plains and urban agglomerations such as Lanzhou. This indicates that GPWv4's areal weighting method achieves better accuracy in plain regions.

[Figure 6: see original paper] Error Taylor diagram of the four data sets

[Figure 7: see original paper] Stacked histogram of relative error percentage of the four data sets

Figure 7 presents stacked histograms of relative error percentages. GPWv4_{UNADJ}

and GPWv4 exhibit larger overestimation areas than WorldPop datasets. Accurately estimated areas constitute the largest proportion in WorldPop datasets, with WorldPop showing the highest proportion of accurately estimated areas among all datasets. Overall, GPWv4 datasets overestimate total township populations, while WorldPop datasets underestimate them. After removing outliers, accurately estimated areas become dominant, overestimation areas decrease substantially, but underestimation areas increase. Comparative analysis indicates that UN-adjusted grid datasets perform better, reducing overestimation at the township level, particularly for WorldPop, where the adjusted dataset shows dominant accurate estimation areas. Overall, WorldPop demonstrates higher accuracy, with its proportion of accurately estimated areas far exceeding those of overestimation and underestimation areas in GPWv4.

4 Discussion

This study evaluated the accuracy differences between WorldPop and GPWv4 datasets in Gansu, Ningxia, and Qinghai provinces in northwest China based on census data. Both datasets exhibit overestimation in northwestern areas and underestimation in eastern regions. At the township scale, WorldPop performs well in densely populated eastern areas, while GPWv4 shows severe overestimation in sparsely populated northwestern areas and moderate population density regions. WorldPop's 100 m spatial resolution (compared to GPWv4's 1 km) provides more options and improved accuracy, enabling more precise description of population spatial distribution and reflecting finer details, with good performance in high-density areas. In contrast, GPWv4's low spatial resolution and areal weighting method limitations prevent it from reflecting true intra-administrative population distribution, though its unique production method yields more accurate total population counts within administrative units, performing well in low-density areas and studies where the minimum research unit exceeds administrative boundaries.

We hypothesize three main factors affect population distribution accuracy in this region. First, the interspersed distribution of farmland and small rural areas in northwest China makes it difficult to extract useful information from land use data to determine human settlements. Second, the fragile ecological environment in northwest China is highly sensitive to human activities, resulting in strong population mobility. Third, population within administrative units is assumed to be fixed, failing to reflect true spatial heterogeneity. Improving population grid dataset accuracy in northwest China requires incorporating auxiliary data on geographic environment and population aggregation patterns that influence human habitation.

Another critical factor affecting dataset accuracy is the data sources employed. Primary input data are census-based population figures adjusted by UN agencies to match global population estimates for target years. However, spatial

variations in natural and socioeconomic conditions cause township-level population growth rates in the study area to differ from national averages, leading to discrepancies between estimated and statistical populations. Additionally, frequent recent administrative boundary adjustments in China may introduce further differences between grid datasets and township-level census data. Ancillary data inputs also affect accuracy: WorldPop incorporates settlement locations, nighttime satellite imagery, roads, vegetation, terrain, and land use to correct residential distribution, with nighttime light data further improving precision. In contrast, GPWv4 relies solely on non-spatial population data and administrative boundaries.

WorldPop employs a random forest model with log-transformed population density, and eliminating zero-count units helps the algorithm achieve good segmentation while making population density more uniform in most cases. However, this creates a dasymetric density weighting layer without zero cells, causing overestimation in low-density areas and underestimation in high-density areas. GPWv4 uses simple areal weighting to distribute population within township administrative units, assuming uniform population distribution across administrative units, with accuracy dependent on input data granularity. Consequently, GPWv4 is suitable for long-term studies where the minimum research unit exceeds the township level, demonstrating high accuracy in coastal regions with relatively uniform population distribution.

---

5 Conclusions

This study compared the accuracy of four population grid datasets (GPWv4, GPWv4_{UNADJ}, WorldPop, WorldPop_{UNADJ}) in the spatially unique "large agglomeration, small settlement" pattern of northwest China's arid regions. The main conclusions are as follows:

1) Both population grid datasets show visual similarity to census data, characterized by dense population in the east and sparse distribution in the west. Through qualitative and quantitative analysis, WorldPop demonstrates superior overall performance in northwest China, with clear classification features, good spatial continuity, and ability to reflect high-precision actual population spatial distribution.

2) Dataset accuracy is primarily influenced by data sources and modeling approaches. WorldPop's random forest regression model generates a dasymetric weighting layer without zero cells, where mean value estimation leads to underestimation in high-density areas and overestimation in low-density areas. GPWv4's areal weighting method assumes uniform population distribution across administrative units, with accuracy dependent on input data granularity. WorldPop is more suitable for refined studies in medium and high population density areas, capable of depicting intra-administrative population heterogeneity, while GPWv4 is appropriate for

studies where the minimum research unit exceeds township administrative divisions.

3) In northwest China's arid regions, population distribution is closely linked to water resources and other living resources, as well as economic, cultural, and historical factors. Fundamentally, accuracy assessment errors stem from the complexity of China's population distribution patterns. Future research should generate arid-region-specific population distribution patterns based on water resource distribution, geographic environment, and population aggregation patterns to improve population distribution accuracy in northwest China's arid areas.

References

[1] Fang J Y, Sun S, Shi P J, et al. Assessment and mapping of potential storm surge impacts on global population and economy[J]. International Journal of Disaster Risk Science, 2014, 5(4): 323-331.

[2] Wu X, Yang J, Zhang H. Analyzing spatial autocorrelation of population distribution in different spatial weights: A case of China[J]. Geomatics World, 2017, 24(2): 32-38.

[3] Gao Yi, Wang Hui, Wang Peitao, et al. Population spatial processing for Chinese coastal zones based on census and multiple night light data[J]. Resources Science, 2013, 35(12): 2517-2523.

[4] Yang Xiaohuan, Wang Naibin, Jiang Dong, et al. Regionalization of population distribution based on spatial analysis[J]. Acta Geographica Sinica, 2002, 57(Suppl.): 76-81.

[5] Guo Hongxu, Huang Ying, Zhao Daiqing. A simulation study on the spatial distribution of urban residents: A case study of Tianhe district, Guangzhou[J]. Tropical Geography, 2013, 33(1): 81-87.

[6] Bhaduri B, Brigh E, Coleman P, et al. LandScan USA: A high resolution geospatial and temporal modeling approach for population distribution and dynamics[J]. GeoJournal, 2007, 69: 103-117.

[7] Jiang Dong, Yang Xiaohuan, Wang Naibin, et al. Study on spatial distribution of population based on remote sensing and GIS[J]. Advances in Earth Science, 2002, 17(5): 734-738.

[8] UNEP. Global resource information database[DB/OL]. [2022-09-24]. http://na.unep.net/siouxfalls/datasets/datalist.php.

[9] Center for International Earth Science Information Network. Global Rural Urban Mapping Project (GRUMP), alpha version: Urban extents[R]. New York: Center for International Earth Science Information Network (CIESIN), Columbia University, 2004.

[10] Balk D L, Deichmann U, Yetman G, et al. Determining global population distribution: Methods, applications and data[J]. Advances in Parasitology, 2006, 62: 119-156.

[11] OpenGMS. Open data[DB/OL]. [2022-09-24]. https://geomodeling.njnu.edu.cn/dataItem/5cd547056af456

[12] Nieves J J. Modelling global human settlement to better inform annual population modelling[D]. Southampton, United Kingdom: University of Southampton, 2020.

[13] Linard C, Alegana V A, Noor A M, et al. A high resolution spatial population database of Somalia for disease risk mapping[J]. International Journal of Health Geographics, 2010, 9(1): 1-13.

[14] Bai Zhongqiang, Wang Juanle, Yang Fei. A summary of the research on population data spatialization[J]. Progress in Geography, 2013, 32(11): 1692-1702.

[15] Dong Nan, Yang Xiaohuan, Cai Hongyan. Research progress and perspective on the spatialization of population data[J]. Journal of Geo-information Science, 2016, 18(10): 1295-1304.

[16] Lin Lijie, Lin Guangfa, Yan Xiaoxia, et al. Spatialization models of census data: A review[J]. Journal of Subtropical Resources and Environment, 2010, 5(4): 10-16.

[17] Bai Z Q, Wang J L, Wang M M, et al. Accuracy assessment of multi-source gridded population distribution datasets in China[J]. Sustainability, 2018, 10(5): 1363, doi: 10.3390/su10051363.

[18] Hall O, Stroh E, Paya F. From census to grids: Comparing gridded population of the world with Swedish census records[J]. The Open Geography Journal, 2012, 5(1): 1-5.

[19] Stevens F R, Gaughan A E, Linard C, et al. Disaggregating census data for population mapping using random forests with remotely sensed and ancillary data[J]. PLoS One, 2015, 10(2): e0107042, doi: 10.1371/journal.pone.0107042.

[20] Tatem A J, Campiz N, Gething P W, et al. The effects of spatial population dataset choice on estimates of population at risk of disease[J]. Population Health Metrics, 2011, 9(1): 1-14.

[21] Wang Xuemei, Li Xin, Ma Mingguo. Research progress and case analysis of population data spatialization based on remote sensing and GIS[J]. Remote Sensing Technology and Application, 2004, 19(5): 320-327.

[22] Xu Y, Ho H C, Knudby A, et al. Comparative assessment of gridded population data sets for complex topography: A study of southwest China[J]. Population and Environment, 2021, 42(3): 360-378.

[23] Lin Danchun, Tan Min, Liu Kai, et al. Accuracy comparison of four gridded population datasets in Guangdong Province[J]. China Tropical Geography, 2020,

40(2): 346-356.

[24] Shi Ying, Mi Ruihua. Differentiation of population spatial distribution in Shaanxi Province[J]. Arid Land Geography, 2015, 38(2): 368-376.

[25] Wang Fenglong, Liu Yungang. Theoretical framework and prospect on quasi-administrative division[J]. Scientia Geographica Sinica, 2021, 41(7): 1149-1157.

[26] Mi Ruihua, Gao Xiangdong. Factors influencing population distribution in Shaanxi Province using spatial econometric analysis[J]. Arid Land Geography, 2020, 43(2): 491-498.

[27] Jiao J Y, Zhang Z G, Bai W J, et al. Assessing the ecological success of restoration by afforestation on the Chinese Loess Plateau[J]. Restoration Ecology, 2012, 20(2): 240-249.

[28] Long H L, Li Y R, Liu Y S, et al. Accelerated restructuring in rural China fueled by increasing vs. decreasing balance policy for dealing with hollowed villages[J]. Land Use Policy, 2012, 29(1): 11-22.

[29] Briggs D J, Gulliver J, Fecht D, et al. Dasymetric modelling of small area population distribution using land cover and light emissions data[J]. Remote Sensing of Environment, 2007, 108(4): 451-466.

Note: Figure translations are in progress. See original paper for figures.