Improved Population Mapping for China Using the 3D Building, Nighttime Light, Points-of-Interest, and Land Use/Cover Data within a Multiscale Geographically Weighted Regression Model

Zhen Lei [1,2], Shulei Zhou [1,2], Penggen Cheng [1,2,*] and Yijie Xie [1,2]

[1] School of Surveying and Geoinformation Engineering, East China University of Technology, Nanchang 330013, China; 2020120292@ecut.edu.cn (Z.L.); 2021110337@ecut.edu.cn (S.Z.); 2021110155@ecut.edu.cn (Y.X.)

[2] Key Laboratory of Mine Environmental Monitoring and Improving around Poyang Lake of Ministry of Natural Resources, East China University of Technology, Nanchang 330013, China

* Correspondence: pgcheng@ecut.edu.cn

1. Accuracy of the GHSL building height dataset

To assess the accuracy of the GHSL building height dataset, we obtained the 2019 Baidu building footprint vector data. To obtain the building height for each building, the number of building floors was multiplied by 3 m [1,2]. Liu et al. [3] report an accuracy of 86.8%, with a mean height deviation of approx. 1 m, for this dataset. Previous studies [4,5] have typically used this dataset as a sample to train models to generate building height data products. There, we used it as a standard for validating the accuracy of the GHSL building height dataset. In order to match the raster form of the GHSL building height dataset, we used the method (Formula 1) from Frantz et al. [6] to convert the vector building height data of single buildings to 100 m spatial resolution grid data. In addition, we obtained the China Building Height (CNBH) data product generated by Wu et al. to compare the accuracy with the GHSL building height dataset, which is the first 10-meter-resolution building height data product in China and is prominent in the currently open datasets. We also converted it to 100-meter resolution based on the method of Frantz et al. [6].

$$BH_{100m} = \frac{\sum_p^n H_p * A_p}{\sum_p^n A_p} \tag{1}$$

where H_p and A_p denote the height and area of buildings in a small-scale cell (10m×10m grid or vector patch), respectively, and $\sum_p^n A_p$ denotes the total area of all buildings in the corresponding large grid cell (100m×100m).

It is worth noting that BBH data has serious data deficiencies in suburban and rural areas, and the deficiencies tend to be more severe in smaller cities than in larger ones. Therefore, we conducted experiments in the core built-up areas of four large cities: Beijing, Shanghai, Guangzhou, and Chengdu. Figure S1 shows the three building height datasets in the core built-up areas of the four cities. As can be seen from the figure, the BBH dataset has the best-detailed performance, and the GHSL dataset has the worst. This can be attributed to the fact that the original dataset of BBH has the highest resolution (meter), while the GHSL has the lowest resolution (100 meters). However, we also find that the overall distribution of the three datasets is similar, with most of the hotspot and coldspot areas in the same location.

We randomly selected 25,394 grids in the four cities' core built-up areas (areas covered by Baidu buildings) for accuracy verification, accounting for about 10% of the total sample grids. The Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE) were used as accuracy metrics. As seen from Table S1, the overall accuracy of CNBH and GHSL perform roughly the same, with RMSEs of 10.40 and 10.26 and MAEs of 8.36 and 8.27, respectively. Thus, although the details of building heights in GHSL do not perform as well as those in CNBH, the overall accuracy in large cities is relatively good.

Table S1. Accuracy assessment of CNBH and GHSL building height datasets.

Unit(m)	RMSE	MAE
CNBH	10.40	8.36
GHSL	10.26	8.27

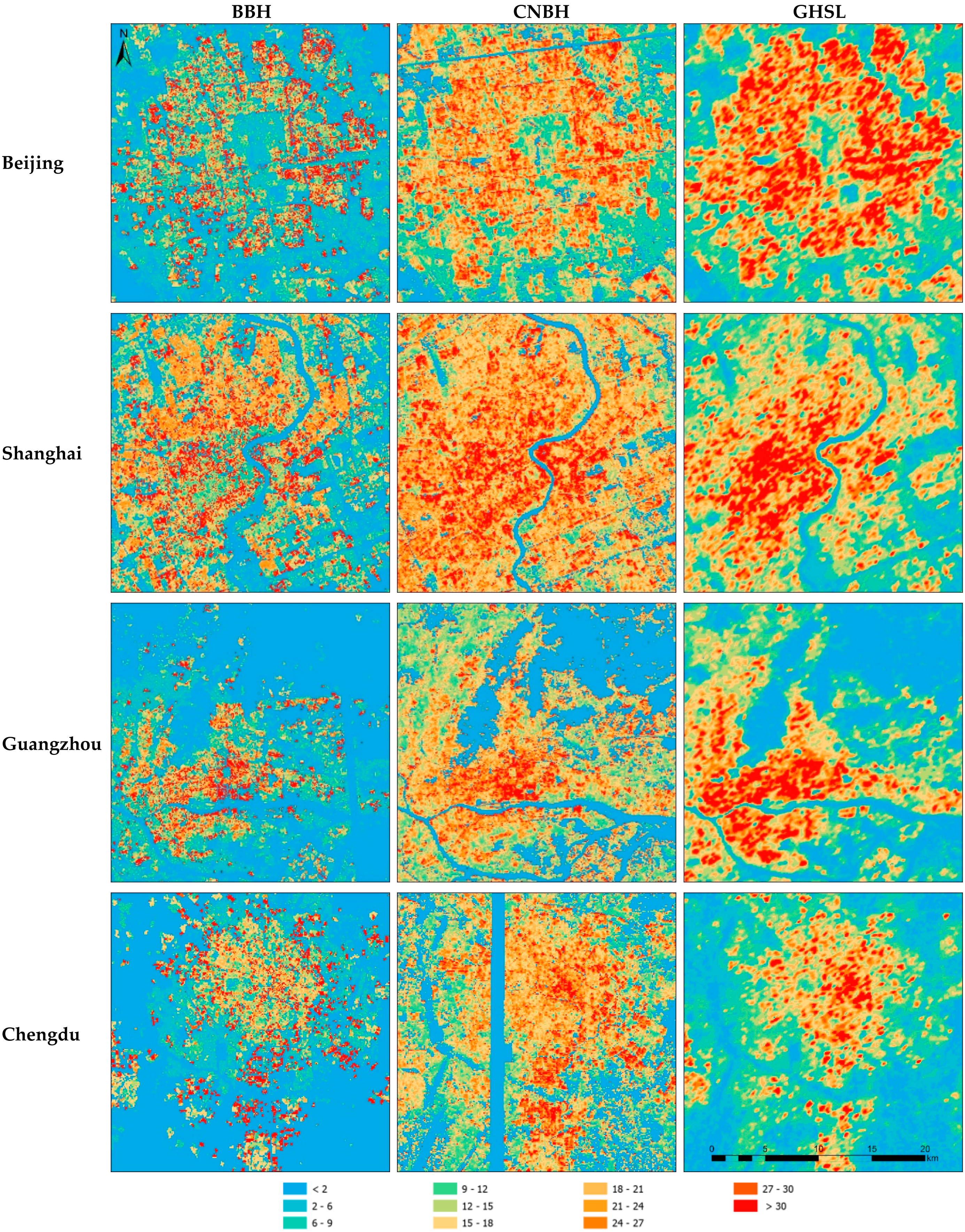

Figure S1. Distribution map of BBH、CNBH and GHSL datasets in the core built-up areas of the four cities.

2. Summary of Pearson's correlation coefficient

Table S2. Pearson's correlation coefficient between population and POI kernel density at various bandwidths.

Bandwidth (m)	Healthcare	Life Services	Car Sales	Auto Repair	Car Service	Motorcycle Service	Science and Education Culture	Company Enterprise	Financial Insurance	Shopping Service	Sports and leisure services	Government agencies and social organizations	Catering Services	Column summation
400	0.776**	**0.780****	0.698**	0.752**	0.718**	0.570**	**0.712****	**0.713****	0.740**	**0.799****	**0.742****	0.758**	0.756**	9.514
800	**0.778****	0.779**	0.700**	**0.753****	**0.719****	0.572**	**0.712****	**0.713****	**0.741****	0.798**	**0.742****	0.760**	**0.757****	**9.524**
1200	0.777**	0.778**	0.702**	**0.753****	**0.719****	0.575**	0.711**	**0.713****	0.740**	0.796**	**0.742****	0.760**	**0.757****	9.523
1600	0.775**	0.777**	0.703**	0.752**	**0.719****	0.577**	0.709**	**0.713****	0.738**	0.795**	**0.742****	0.759**	0.756**	9.515
2000	0.773**	0.775**	0.705**	0.752**	0.718**	0.580**	0.707**	**0.713****	0.736**	0.793**	0.741**	0.757**	0.755**	9.505
2400	0.771**	0.773**	0.707**	0.751**	0.717**	0.583**	0.705**	**0.713****	0.734**	0.791**	0.740**	0.756**	0.754**	9.495
2800	0.769**	0.772**	0.709**	0.750**	0.716**	0.586**	0.703**	**0.713****	0.733**	0.790**	0.740**	0.755**	0.753**	9.489
3200	0.767**	0.770**	0.710**	0.749**	0.715**	0.589**	0.701**	**0.713****	0.731**	0.788**	0.739**	0.754**	0.751**	9.477
3600	0.766**	0.768**	0.711**	0.748**	0.714**	0.591**	0.699**	**0.713****	0.730**	0.787**	0.738**	0.753**	0.750**	9.468
4000	0.764**	0.767**	0.712**	0.747**	0.713**	0.594**	0.698**	**0.713****	0.729**	0.786**	0.737**	0.753**	0.749**	9.462
4400	0.763**	0.765**	0.714**	0.746**	0.713**	0.595**	0.697**	**0.713****	0.728**	0.784**	0.736**	0.752**	0.748**	9.454
4800	0.762**	0.763**	0.715**	0.745**	0.712**	0.587**	0.695**	**0.713****	0.727**	0.783**	0.735**	0.751**	0.747**	9.435
5200	0.760**	0.762**	0.716**	0.745**	0.711**	0.599**	0.694**	**0.713****	0.726**	0.781**	0.734**	0.750**	0.746**	9.437
5600	0.759**	0.760**	0.717**	0.744**	0.711**	0.601**	0.693**	**0.713****	0.725**	0.780**	0.732**	0.750**	0.744**	9.429
6000	0.758**	0.759**	0.719**	0.743**	0.711**	0.602**	0.692**	**0.713****	0.724**	0.779**	0.731**	**0.772****	0.743**	9.446

6400	0.757**	0.758**	0.720**	0.742**	0.710**	0.604**	0.691**	**0.713****	0.723**	0.778**	0.730**	0.771**	0.742**	9.439
6800	0.756**	0.756**	0.721**	0.742**	0.710**	0.605**	0.690**	0.712**	0.723**	0.776**	0.729**	0.771**	0.741**	9.432
7200	0.755**	0.755**	0.723**	0.741**	0.710**	0.607**	0.690**	0.712**	0.722**	0.775**	0.728**	0.770**	0.740**	9.428
7600	0.755**	0.754**	0.724**	0.740**	0.710**	0.608**	0.689**	0.712**	0.722**	0.774**	0.727**	0.769**	0.739**	9.423
8000	0.754**	0.752**	**0.726****	0.740**	0.710**	**0.610****	0.688**	0.712**	0.721**	0.773**	0.726**	0.768**	0.739**	9.419
Row summation	15.295	**15.323**	14.252	14.935	14.276	11.835	13.976	14.256	14.593	15.706	14.711	15.189	14.967	

**. Significant at 0.01 level

The bolded text in Table S2 indicates the maximum correlation coefficients within the bandwidth series. At the 800-meter bandwidth, the POI correlation coefficients are highest for most categories (8 categories), and the total sum of correlation coefficients across all POI categories (i.e., column sum) is also the highest.

References

1. Tripathy, P.; Balakrishnan, K.; de Franchis, C.; Kumar, A. Generating megacity-scale building height maps without DGNSS surveyed GCPs: An open-source approach. Environ. Plan. B Urban Anal. City Sci. 2022, 49, 2312–2330.
2. Wu, W.-B.; Yu, Z.-W.; Ma, J.; Zhao, B. Quantifying the influence of 2D and 3D urban morphology on the thermal environment across climatic zones. Landsc. Urban Plan. 2022, 226, 104499.
3. Liu, M.; Ma, J.; Zhou, R.; Li, C.L.; Li, D.K.; Hu, Y.M. High-resolution mapping of mainland China's urban floor area. Landsc. Urban Plan. 2021, 214, 104187.
4. Wu, W.-B.; Ma, J.; Banzhaf, E.; Meadows, M.E.; Yu, Z.-W.; Guo, F.-X.; Sengupta, D.; Cai, X.-X.; Zhao, B. A first Chinese building height estimate at 10 m resolution (CNBH-10 m) using multi-source earth observations and machine learning. Remote Sens. Environ. 2023, 291, 113578.
5. Che, Y.; Li, X.; Liu, X.; Wang, Y.; Liao, W.; Zheng, X.; ... Dai, Y. 3D-GloBFP: the first global three-dimensional building footprint dataset. Earth Syst. Sci. Data Discuss. 2024, 2024, 1-28.
6. Frantz, D.; Schug, F.; Okujeni, A.; Navacchi, C.; Wagner, W.; van der Linden, S.; Hostert, P. National-Scale Mapping of Building Height Using Sentinel-1 and Sentinel-2 Time Series. Remote Sens. Environ. 2021, 252, 112128.