REVIEWER #1:

Downscaling census data into population grids can address the limitations of census data with irregular units. This paper proposed a new population downscaling method using ensemble learning and geospatial big data. The method is adopted to generate a 100-m gridded population dataset of China's seventh census. The accuracy assessment on the generated population dataset shows that it has higher accuracy than the two existing datasets of WorldPop and LandScan. In general, the paper is well-written, clear, concise and complete in structure. I believe the generated dataset is important to a wide range of geoscience applications. Yet the following comments need to be addressed.

Response:

Many thanks for your valuable suggestions and we deeply improved the manuscript according to your comments.

1. Line 100, what resample method is used for the variable?

Response:

Thanks for your suggestion.

We added the used method (i.e., bilinear resampling method) in the revised manuscript.

Added or revised contents:

Line numbers in the revised manuscript: Line 113-115.

[A projection transformation was conducted and the bilinear resampling process was further implemented to convert it into a 100-m Tencent user density image, as illustrated in Figure 2 (a).]

2. How to get the number of POIs and the road length for each grid in Section 2.2?

Response:

Thanks for your suggestion.

We used the Point Density tool in ArcMap to get the number of POIs within each grid. The Line Density tool in ArcMap was used to calculate the road length within each grid. We revised the descriptions.

Added or revised contents:

Line numbers in the revised manuscript: Line 120-121 and 123-124.

[The Point Density tool in ArcMap summarized the number of POIs within each 100-m grid and this information was used as the POI density in Figure 2 (b).]

[The length of roads within each 100-m grid was computed by the Line Density tool in ArcMap using the road data acquired from the online map of AutoNavi Maps and it was considered as the road density in Figure 2 (c).]

3. Although the authors provided the downscaling procedure (Figure 4), it's still unclear why do the 'population density' used instead of the direct population count.

Response:

Thanks for your suggestion.

It's similar to previous studies, we also used population density instead of population count. Because population density is more suitable for comparing regions of different sizes. We added the reason for this issue.

Added or revised contents:

Line numbers in the revised manuscript: Line 179-181.

[Compared to population count, population density is more suitable for comparing regions of different sizes and is frequently used as the dependent variable in population estimations (Stevens et al., 2015; Cheng et al., 2020; Ye et al., 2019).]

4. In Section 5.2, how do you get the feature importance for stacking? How about the three base models?

Response:

Thanks for your suggestion.

We used the ELI5 Python package to get the feature importance of stacking ensemble learning and we added the description for this issue. For the feature importance of the three base models, we revised Figure 10 and added their feature importance in Figure 10. Meanwhile, we also added their comparisons.

Added or revised contents:

Line numbers in the revised manuscript: Line 328-337.

[To investigate the influence of ten covariates on the fitted PopSE, the feature importance (i.e., weight) of covariates was obtained using the ELI5 Python package. It allows to show the feature importance of various machine learning algorithms, including random forest, XGBoost, LightGB, and stacking ensemble learning. Figure 10 illustrates the feature importance of each covariate for the fitted PopSE and its three base models. Notably, POI density emerges as the most impactful on fitting PopSE and the three base models, with a significantly higher feature importance compared to the other nine covariates. Following closely are the four covariates of latitude, percentage of built area, NTL, and building height and they have similar importance level with relatively equal feature importance. Subsequently, the covariates of Tencent user density, slope, DEM, and longitude exhibit comparable levels of feature importance in PopSE and the three base models. Road density has the lowest contribution to build PopSE with the smallest feature importance.

Except building height, the feature importance of PopSE and its base models is comparable for each covariate in Figure 10.]

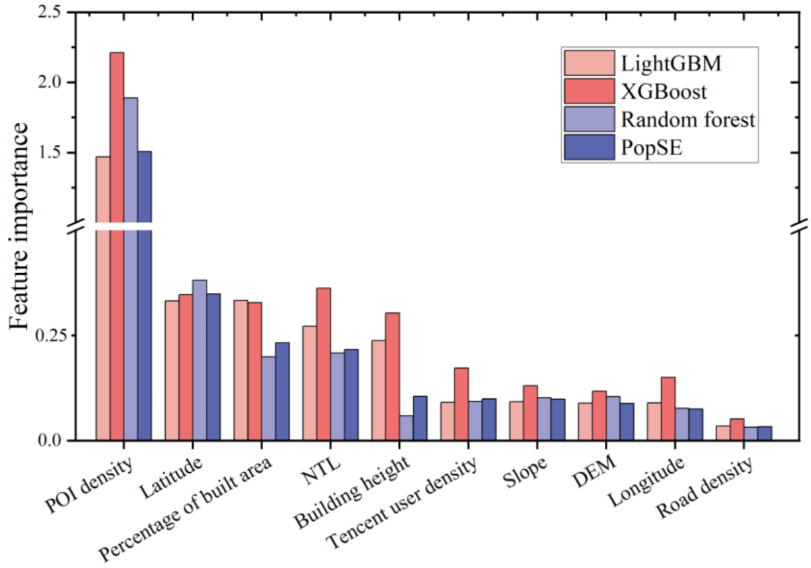

Figure 10: Feature importance of the proposed PopSE and its three base models.

5. Section 5.3 presents the parameter selection, what's the search interval within the search space?

Response:

Thanks for your suggestion.

We added the search intervals for the hyperparameter tuning.

Added or revised contents:

Line numbers in the revised manuscript: Line 344-345.

[The search interval was set to 1 for both the number of trees and the maximum depth of a tree.]

6. Please add the caption of Figure 1c.

Response:

Thanks for your suggestion.

We added the caption for Figure 1c.

Added or revised contents:

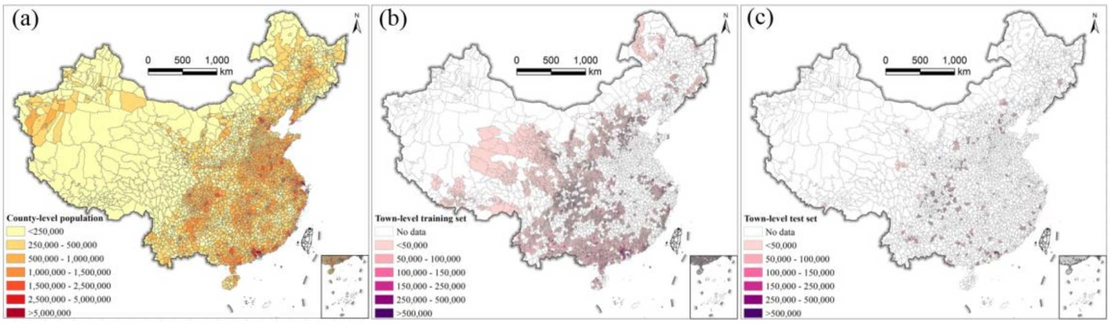

Figure 1: The seventh census data of China. (a) The county-level census data, (b) The town-level census data for training, (c) The town-level census data for test.

7. Please enlarge the font size of Figure 1-3.

Response:

Thanks for your suggestion.

We revised the three figures.

Added or revised contents:

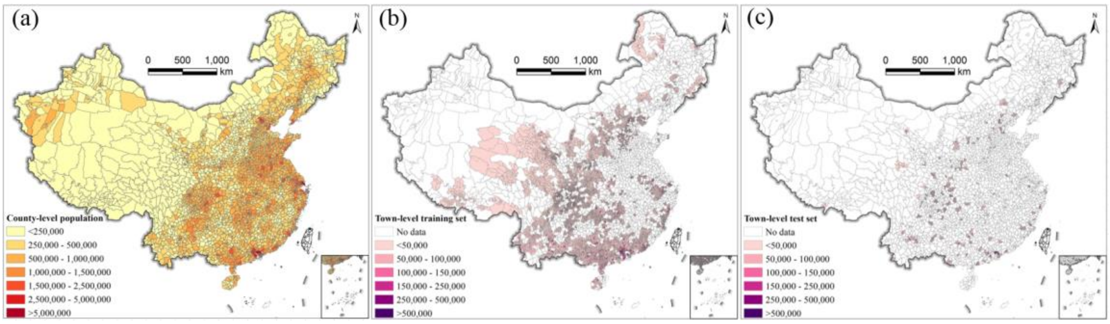

Figure 1: The seventh census data of China. (a) The county-level census data, (b) The town-level census data for training, (c) The town-level census data for test.

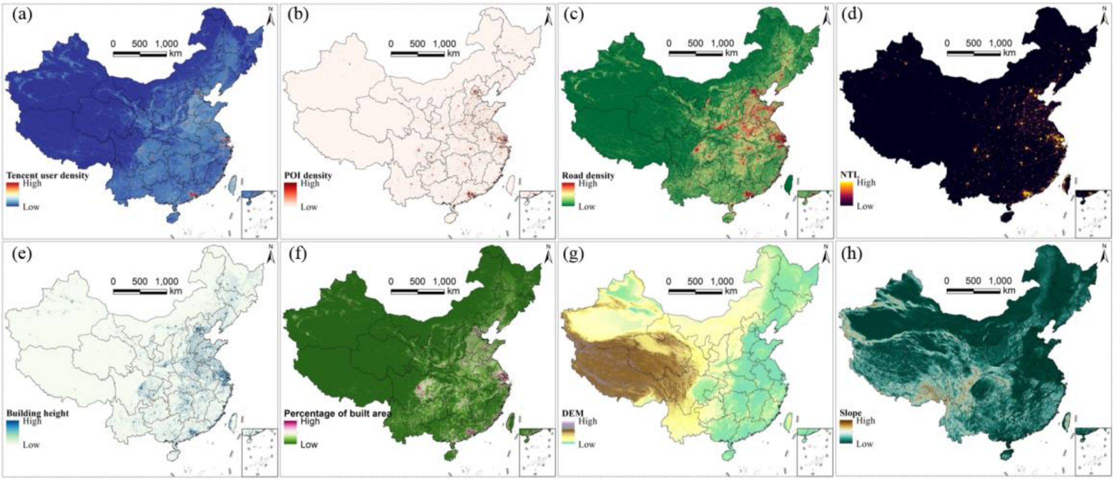

Figure 2: The 100-m gridded covariates of China. (a) Tencent user density, (b) POI density, (c) Road density, (d) NTL image, (e) Building height, (f) Percentage of built area, (g) DEM, (h) Slope.

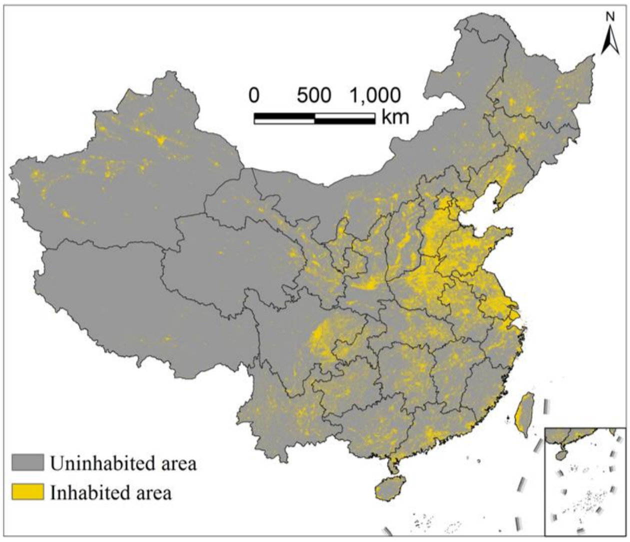

Figure 3: The 100-m inhabited areas of China.