

CC #1:

1. Tencent density user positioning data belongs to instantaneous data, showing significant differences during the day and night, and even between different time periods. What was the filming time for this paper? However, no matter which time period Tencent density user positioning data is available, it is not suitable for census data and is more suitable for using survey data to predict population density.

Response:

Thanks for your suggestion.

Although Tencent density user positioning data is the instantaneous data, the average of Tencent density user positioning data during a long period has been proved to be a reliable proxy for the human distribution in many previous studies. Therefore, we used the average image of Tencent density user positioning data from January 1 and June 30 in 2019. During this period, we obtained the instantaneous Tencent density user positioning data every five minutes. The description is shown in Line 109-113 in the revised manuscript.

2. Similarly, please explain how to obtain the feature importance by stacking ensemble learning . Is the ranking of feature importance applicable to the population distribution of all regions in China due to its complex landforms?

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. The training set included all counties and most towns across the entire China and the models were fitted by the training set. Therefore, the fitted models and the feature importance are applicable to the entire China.

Added or revised contents:

Line numbers in the revised manuscript: Line 328-330.

[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.]

3. The article claims to have used a extensive dataset of 60 million POIs. What is the specific POI category for the application? The impact of POI categories on the population varies. They should be selected and weighted.

Response:

Thanks for your suggestion.

We added the categories for used POIs.

Added or revised contents:

Line numbers in the revised manuscript: Line 118-120.

[Only POIs associated with human activities were used. These POIs were classified into ten main categories: restaurant, shopping, life service, working, education, medical facility, residence, transportation, recreation, and others.]

4. Road data should also specify the category of use. For example, Expressways are not suitable for introduction, while Expressway toll stations are suitable for introduction.

Response:

Thanks for your suggestion.

The used road data exclude the expressways and railways and we added the categories for used road data.

Added or revised contents:

Line numbers in the revised manuscript: Line 126-126.

[Note that the roads used in this study mainly included city roads, as well as provincial, county, and township-level roads, while excluding railways and expressways.]