Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC1
https://doi.org/10.5194/gmd-2021-428-RC1, 2022

Comment on gmd-2021-428

Anonymous Referee #1

Referee comment on "Estimation of missing building height in OpenStreetMap data: a French case study using GeoClimate 0.0.1" by Jérémy Bernard et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-428-RC1, 2022

Thank you for the work put in this study. GeoClimate seems to be a very interesting project, its relevance is clear and adding building heights to it would certainly be very valuable. However, I believe this submission is currently far from being ready for publication in this journal, for several reasons detailed below, and I recommend major revisions.

1. The innovation compared to previous literature is minimal, in particular because the authors missed a study [NMD20] that looked at the very same question, with more data and a larger geographical scale. The authors should rethink the gap in the literature they are trying to address within their study by taking this closest reference point into account.

2. There is a lack of framing regarding under which context the method should be used: e.g. is it to predict in areas with no data available as proposed in [NMD20], which requires to test the spatial generalization of the model, or to fill the gaps in a city where data is available as in [BIL17] and where more over-fitting makes sense? Those different cases require different training and testing approaches. It seems that the authors are primarily interested in the first one, if so, the methods should more robustly assess generalization, see next point.

3. The methodology needs improvements to be more robust. The results seem reasonable e.g. the RMSE and MAE values, but the training and test procedure should be improved. Some examples of points I believe could be addressed: First, the authors should undertake cross-validation to test their model on different folds of the data to account for different urban situations that easier or hard to train/predict on. Second, spatial cross-validation on spatially-distant folds would be particularly relevant to enhance/demonstrate generalization. Third, why choosing train and test cities in the same region e.g. Corbonod and Annecy or Nantes and Saint-Nicolas-.. are close, while there are so many cities in France to separate further spatially the sets?

4. The text could be much clearer. In particular, I found the structure of the introduction confusing (this relates to the previous points on lack clear gap in literature, use case, etc.). The presentation of the results is also perfectible. Some metrics are given without clear indications of the sets they are referring to e.g. are the RMSE line 170 and “all cities” line 175 for the training or test set, or both? Results from training and test should in principle be presented separately, not together as on Fig. 6, as they represent different prediction settings. One/few summary result table(s) would also help the reader.

5. Why using OSM if you have higher quality data from a government source? BDTOPO is great in France, why using data with uncertain coverage when best coverage is available? There is also a lot of great government data across Europe, so why OSM specifically? For scaling globally? Because GeoClimate is specifically built for OSM? I believe this is not explained. Also, one would need to take into account that in areas where OSM building footprint coverage is low, say rural Greece, the model will likely be wrong as the urban form input will be wrong. If the goal of the authors is to specifically investigate prediction from OSM, then an option to differentiate this study from [NMD20] could be to predict for different scenarios of OSM quality and compare the results, which might show that OSM is good enough even with medium-low quality, or not, and then identify where and why?

References

[NMD20] Milojevic-Dupont, Nikola, et al. "Learning from urban form to predict building heights." *PLOS ONE* 15.12 (2020): e0242010.

[BIL17] Biljecki, F., Ledoux, H., and Stoter, J.: Generating 3D city models without elevation data, Computers, Environment and Urban Systems, 64, 1–18, 2017.