Reply on AC1
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-RC2, 2022

Thank you for the response and revised manuscript. I believe both go a good direction, but I have a few more comments/answers to the authors points. Overall, I think this is valuable contribution and would not oppose strongly to its publication, although ideally I would still think this paper would benefit from some additional improvements.

1. I agree that the aspect of replication of previous research in another geographical context + methods easier to use for practitioners are valuable. I am not sure I understand the third point.

2. I still believe the authors could do a better job explaining how different use cases of missing data (e.g. partial unavailability within a city vs no data available in a region) require different training approaches, e.g. see:


3. My point was not that you may have poorly designed the training set resulting in too pessimistic prediction, but rather than there are ways to more robustly ensure that the choice of training and test set does not generate over-optimistic results (as in, depending on what you are using the results for, e.g. to say that one can expect that the accuracy of the inferred data across the French territory, where there is no ground truth, will be X, Y, Z). "We have actually tested a previous version of the model on more remote cities and the performance was similar" -> why not including this?. Thanks for the interesting discussion on topography etc. The results you mention from [NMD20] might be case-specific and not generalizable, so I would be careful with basing your intuition on these here.

4. Thanks for adding the table.

5. Thanks for the explanations. I believe it would still be important to explain the issues
when using OSM for predictive features, in particular missing building footprints / land use polygons / etc. in many areas that result in biased urban form and consequently wrong input feature values. Ideally this would be something the authors could investigate as previously suggested, by looking by artifically creating missing data situations, but I am happy with at least a mention of this issue. I believe this is important for the users of your model to have in mind the limitations of OSM so that they make the appropriate analysis beforehand. By the way, I do not see any completeness analysis of OSM in your manuscript. Quickly checking simple metrics like accordance between total footprint area in OSM and BDTOPO for the selected cities would be good.

PS: the anonymous reviewer's pronoun could very well be "she" or "they", and the authors may want to mention the anonymous reviewers generically as "they" in the future ;)}