Comment on gmd-2022-2

Anonymous Referee #1

Referee comment on "Global, high-resolution mapping of tropospheric ozone – explainable machine learning and impact of uncertainties" by Clara Betancourt et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2022-2-RC1, 2022

I appreciate the effort that the authors have put in assessing spatial uncertainties associated with their predictions. Unfortunately, this is often not a priority in global mapping papers, so I welcome this focus on uncertainty very much. The manuscript is structured well, but overall fairly lengthy and is written in a relaxed, almost conversational writing style - which I like, but from sometimes it's perhaps a bit too much. Another suggestion would be to move most of the tables and figures to the supplemental to improve readability. With 12 figures and 5 tables in the main text, I found it sometimes hard to navigate and find the most relevant results.

I must admit that I know close to nothing about ozone and what variables structure its spatial patterns in the troposphere, but essentially the model is predicting O3 using latitude, altitude and human development (nearby nightlight); collectively explaining the majority of variation in the model. However, latitude and altitude are both merely a proxy for temperature and radiation; which are probably the actual main drivers of O3 levels (L153). Why not use them directly as predictors in the model? There are various high quality global layers available.

With the very strong effect of absolute latitude in the model (~25% global importance), the predicted pattern will strongly reflect absolute latitude – which is clearly visible in the final map (fig 11).

Next, I have a couple of doubts with the current "area of applicability" (AOA) approach, where features are weighted by their respective importances as in the model. By weighting the features, you're essentially using absolute latitude and altitude (twice, including relative altitude) as to define feature space in which you apply the model. But as above; it's not latitude or altitude that define this space, but temperature and radiation (L153). A major drawback, I think, from scaling the features used in the AOA analysis using their respective importances, is that you're now basically assuming you can predict everywhere on the same latitude, save for some places (some of) the fewer important features fall outside the sampled space.

It would be useful if the authors would include a comparison with previously published tropospheric ozone predictions, based on mechanistic models. Of course, the authors did do a substantial test of their model's validity and stability – but these really only hold true for this particular model and dataset; and don't provide insight in how the results compare to other tropospheric ozone predictions.

Further points

- The statement on L440 (*"The effect of 'absolute latitude' on predictions is consistent with what is known about ozone formation – ozone generally increases toward the equator"*) seems to be the opposite of the patterns that are predicted (Fig 11), with lower values near the equator, and higher values in temperate regions?
- The selection of the threshold distance for 'non-predictable' data (ie., the upper whisker of all the cv distances), is seemingly arbitrary. It is in line with the AOA paper by Meyer and Pebesma, but neither that paper provides a statistical reasoning for picking this particular threshold.
- The manuscript includes various subjective interpretations of the results. I believe the manuscript would benefit from a more objective wording. Some examples:
  - L310: *"East Asia is a special case because the ozone value distribution is rather narrow there"*
  - L325: *"Very high 'nightlight in 5km area' values"*
  - L406: *"are considered significant"*
  - Figure C1: *"really high values"*
- The authors do use "not shown" rather often, a total of 8 times in the entire manuscript. I guess this is ok, but if the authors feel that the data/results aren't necessary to show, arguable the entire section can be removed. If not, I would suggest placing the evidence for the statement in the supplemental materials.
- Figure 3: are these 'example data points' points in the AQ-bench dataset or in the raster data you're predicting?