Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC2
https://doi.org/10.5194/gmd-2022-2-RC2, 2022

Comment on gmd-2022-2

Anonymous Referee #2

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-RC2, 2022

The authors demonstrate a machine learning approach to generate high-resolution surface ozone concentration products, and evaluate the uncertainties from models and data sources. Many techniques are used in this study and they are generally explained well. The surface ozone products can be potentially used for other studies if the produced ozone mapping is robust. The manuscript is written well in a conversational way and I can feel that the authors try to add the novelty in explaining machine learning results, but overinterpreting should be avoided. There are a few major concerns that I think should be addressed carefully about the motivation of the study and the usage of final ozone products.

Major comments:

- High-resolution ozone mapping is a highlight of this study, but are there any differences between directly using interpolated original ozone products (TOAR) and the products generated here? I think the authors try to extend the application to the regions where measurement sites are not available, but it is clearly that the trained results are limited to the number of measurement sites (mentioned in Sect. 3.2.1).
- High-resolution ozone mapping may introduce extra uncertainties because input features or surface ozone concentrations may have large biases. Since surface ozone is regionally spread, slightly decreasing resolution may reduce uncertainties. This should be discussed to strengthen the motivation of the study.
- The averaged surface ozone concentrations over 2004-2014 are reproduced and the authors also mentioned that the products are static in Sect. 4.4. The geographic variables are used to drive the machine learning model and many of them (e.g. latitude, altitude) instead of physical or chemical variables show high importance to simulated ozone. The relationships between some variables are intuitive, but the issue

is that simulated future ozone concentrations may be quite similar to those simulated over 2004-2014 because the geographic variables with high importance are static – they will not change in the future. The temporal relationships may not be captured by the machine learning model. It may be useful to justify the usage of average ozone earlier in the data description section, and to state the benefits of using final ozone products.

Other comments:

- Line 11: "By inspecting the feature space, …". Not clear in the abstract.
- Line 59: Need to clearly point out the key issues in the current mapping field and what the benefits are by using machine learning approaches.
- Line 79: Please justify the usage of annual mean surface ozone concentrations. I suppose that using monthly data would make the model more robust as more data are involved in the training?
- Line 76: Are there only 5577 data used for machine learning?
- A more specific title is needed for Figure 1 instead of saying 'average ozone values'.
- Line 86: I am not convinced by the association between 'latitude' and ozone photochemistry.
- In Table. 1, many land cover variables are used so they may principally reflect ozone dry deposition? Some discussions are needed here.
- NOx emissions and columns are used. What about other ozone precursor emissions?
- Line 106: It is too confident to state that the random forest is the most suitable; apparently, it is not.
- Figure 3: Does the data points outside the area of applicability simply mean they have extreme high or low values that are not easily to predict? As you scale feature values with SHAP values, it is likely that the threshold used to filter large values is largely dependent on altitude.
- Line 303: I cannot judge if RMSEs in the range of 3.84 to 4.04 ppb are large or small, even though the authors said this is acceptable. I think it will be better to show temporal one standard deviation of surface ozone concentrations along with surface ozone mixing ratios (annual mean) in Fig. 1 for readers.
- Line 327: SHAP value discussions are in Sect. 4.2. I suggest that the authors avoid using many forward references, and merge some discussions in the corresponding sections.
- The evaluation picture (Fig. 10) is important, and I suggest to move it forward.
- Two panels should be indicated in Fig. 11. It would be interesting to show the readers the predicted surface ozone mixing ratios across the globe, even if the authors identify some areas as inapplicable.
- Line 445: I think this is overinterpreted as you are using nightlight conditions to explain monthly or annual mean ozone variation. Ozone chemical production or destruction depends on NOx concentrations and NO titration is one aspect. It is fine if some relationships cannot be explained and I don't expect the relationships derived from SHAP values can explain every feature because machine learning model is not process-based.
- How do authors think of the relative importance of training data number and training strategies (e.g. model types, feature selections) in ozone mapping? The number of

training data may be more important shown in the study, and there is a need to discuss this aspect.