Atmospheric Chemistry and Physics Discussions

Atmos. Chem. Phys. Discuss., referee comment RC1
https://doi.org/10.5194/acp-2022-3-RC1, 2022

Comment on acp-2022-3

Anonymous Referee #1

Referee comment on "Simulating the radiative forcing of oceanic dimethylsulfide (DMS) in Asia based on machine learning estimates" by Junri Zhao et al., Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2022-3-RC1, 2022

The authors present a novel statistical methodology for predicting DMS concentration from a combination of predictors in place of traditionally interpolated data. They use this to estimate the radiative effect of DMS through its contribution to atmospheric sulphate and CCN in Asia and the surrounding Ocean. While a similar approach to estimating DMS has recently been shown I am not aware of such a dataset being used to estimate the atmospheric and radiative differences as compared to a traditional climatology. This should be suitable for publication in ACP after some concerns and clarifications regarding the methodology are addressed.

My main concern is regarding the training and validation data split. The authors state that "we selected the data from 2° latitude bands between 11°S and 30°N as validation datasets (809 points), while the rest of the data was all used as training 175 data (2939 points)." I am unsure exactly what this means but given the large latitudinal variation in DMS in the region (e.g., Fig. S2) this could introduce unnecessary bias in the model. Why not just use a random split?

While the authors provide a good overall evaluation of their model, given the large seasonal variation in DMS (and it's contributing factors), I would also like to see a validation of the RMSE in each season, even if only in the supplemental. This would provide confidence that the model is providing robust predictions in different regimes.

On this point, the authors state that "the training process is not interpretable and not transparent". I would dispute that. One of the benefits of a tree-based model like XGBoost is that it is quite efficient to investigate the sensitivity of the output to each of the predictors. For example, the authors could provide Shapley values for the different predictors, perhaps in different seasons. I don't believe this is too much work and would help demonstrate the robustness of the model and support the interpretation of their model.

On line 244 the authors state that "Overall, the simulation results of XG in other periods were closer to the observations than those of LANA simulation results." but this doesn't look true by eye. Could the authors report the respective RMSE of each result against the observations for this time series? It does seem that XG represents the variability somewhat better though.

Other, minor suggestions:

- L24: 'lack of' -> 'lacking'
- L25: The authors state that the DMS emissions flux 'accounts for 15.4% of anthropogenic sulfur emissions', but DMS isn't an anthropogenic source so this doesn't make sense. Perhaps they mean 'equivalent to 15.4% of …'?
- L30: 'of all sources' -> 'of all sources, respectively'.
- L81-84: The sentence which begins "In this study…" is quite long and doesn't seem to make sense, consider rephrasing.
- L90: The whole manuscript has inappropriate line breaks in the middle of words from this point on, please check.
- L94: " allows modeler" -> "allows modelers"?
- L151: "trained XGBoost" -> "trained an XGBoost"
- L153: "concentrations where without the observations" -> "concentrations in the place of missing observations"?
- L215: "which is 15.4%" -> "which corresponds to 15.4%"?
- L394-396: This is true and the authors might like to cite Schutgens et al. 2016 (https://acp.copernicus.org/articles/16/1065/2016/) in support of the claim.
- Could the authors use the same colour scale for IRF and DRF in Figures 7 and 8 to aid comparison?