Atmospheric Chemistry and Physics Discussions

Atmos. Chem. Phys. Discuss., referee comment RC2
https://doi.org/10.5194/acp-2022-196-RC2, 2022

Comment on acp-2022-196

Anonymous Referee #1

Referee comment on "Correcting ozone biases in a global chemistry–climate model: implications for future ozone" by Zhenze Liu et al., Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2022-196-RC2, 2022

Overview:

The paper presents a multilayer perceptron ML approach to predict the differences between monthly-mean simulations of ozone surface concentrations simulated by the UKESM1 and the EAC4 CAMS Reanalysis. The ML approach manages to predict the differences to a good degree, which is exploited by the authors to ML predict the differences for UKESMI1 scenarios simulation of future ozone surface concentrations. The ML approach is trained by set of 20-30 input variables, such as temperature, photolysis rates and boundary layer height. While not a direct outcome of the ML approach, the importance of the different input variables is quantified using the SHAP framework. The paper is well written.

General review:

The fundamental problem of the paper is the use of the ozone surface concentration fields of the CAMSRA data as the "truth". The ozone surface analysis is corrected too little by the assimilated ozone satellite retrievals, which are dominated by the stratospheric signal, to be considered a representation of the "truth" or the observations; they are just another model run. https://atmosphere.copernicus.eu/sites/default/files/2021-06/CAMS84_2018SC3_D5.1.1-2020_reanalysis_validation.pdf. CAMS84_2018SC3_D5.1.1-2020_reanalysis_validation (copernicus.eu) There is no continuous gridded data set, which would make the identification of biases of simulated ozone surface values for all regions of the earth surface a straightforward task. This is in contrast to meteorological simulations for which meteorological re-analysis (i.e. ERA5) are considered a continuous global reference data set.

For this reason, I suggest rejecting the paper because I can not think of an easy fix of that dilemma. It seems necessary to introduce independent surface in-situ observations such

as the TOAR data sets. Using the TOAR data set as reference make sense but limits strongly the area for which biases can be determined. The TOAR data set could also be used to demonstrate that the biases of UKSEM1 are indeed much larger than the biases of CAMS RA. An alternative approach could be to use the applied method to transition results between different model types, but I am not sure if there are useful applications for that.

The conclusion from the SHAP based ranking of feature importance remains too general and I am not sure what can be learned from that. All the mentioned factors contribute to ozone variability, and it is likely that that is also reflected in their correlation with the inter-model differences. The often-mentioned biases of the emissions were seemingly not identified as a large contributor to ozone biases, which is surprising as the relatively small impact of the NO concentrations. I also wonder why ozone concentrations itself were not used as an input variable for the training. I am not saying the results are wrong, I would only suggest avoiding the over-interpretation of these correlations into actual causes of the biases. But, if the authors can demonstrate that so far unknown deficiencies of UKSEM1 were identified and potentially even fixed based on the finding of the feature importance, this should be mentioned more convincingly.