

Atmos. Chem. Phys. Discuss., referee comment RC2
<https://doi.org/10.5194/acp-2021-864-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on acp-2021-864

Anonymous Referee #3

Referee comment on "Model output statistics (MOS) applied to Copernicus Atmospheric Monitoring Service (CAMS) O₃ forecasts: trade-offs between continuous and categorical skill scores" by Hervé Petetin et al., Atmos. Chem. Phys. Discuss., <https://doi.org/10.5194/acp-2021-864-RC2>, 2022

Review of é Petetin et al.

The paper entitled "Model Output Statistics (MOS) applied to CAMS O₃ forecasts: trade-offs between continuous and categorical skill scores" by Petetin et al. is well written and provides a very interesting perspective on different statistical tools and machine learning approaches that can be used to improve air quality forecasts. It falls within the scope of ACP and I truly enjoyed reading it. The analysis is sound and truly comprehensive. I have only a few minor comments, that the authors may want to consider to improve the manuscript further.

- In line 85 the authors say that in this study, daily mean, daily 1-hour maximum and daily 8-hour maximum are computed only when at least 75% of the hourly data are available (i.e. 18 over 24 hours). Theoretically speaking a day during which the data from 9 am to 4 pm is missing could qualify this criterion, yet both the computed d1max and d8max of such a day would be far off. Instrumental interventions such as service visits, purging with zero air to get moisture out of the system and calibrations usually occur during working hours. Hence the authors may want to consider applying a filter directed at daytime rather than night-time observations next time. This may remove a few extra datapoints but would have been preferable considering the target of the paper. The current choice is, however, hardly going to impact the results pertaining to the model evaluation of the various techniques with respect to d8max and d1max. The days analysed are driven mostly by observational stations reporting d8max >60 ppb or D1max >90. Hence most days included in the analysis would have daytime data. Nigh time events with d8max >60 ppb and d1max >90 ppb are relatively rare, although they do occur occasionally at high altitude stations. So, in my opinion the best way out without redoing the analysis would be to run a quick check on the data for the following two parameters

- How many days with large data gaps during the day (9 am to 4 pm) were included in the analysis?
- How many of the observed d8max and d1max events are night time events?
Both numbers would be small and can be reported and discussed as limitation.
- Figure 2: I find it hard to see the colour difference between the purple and black line. In particular where they are not superimposed. The colour contrast in Figure F1 which is similar is much better
- Figure 3: Some people have bad memory for abbreviations or the habit of skipping to the figures. Just like the authors gave the full form for (h: hourly; d: daily mean; d1max/dd1max: daily 1-hour maximum; d8max/dd8max: daily 8-hour maximum) which is much appreciated can they please give the full form of the abbreviations S, H, F, FB, SR, CSI, PSS, AUC, PCC (which people may be more familiar with a R) in the figure caption. It will save a lot of readers from having to scroll back to the method section where these are defined.