EGUsphere, referee comment RC2
https://doi.org/10.5194/egusphere-2022-396-RC2, 2022

Comment on egusphere-2022-396

Anonymous Referee #2

Referee comment on "Predicting peak daily maximum 8□h ozone and linkages to emissions and meteorology in Southern California using machine learning methods (SoCAB-8HR V1.0)" by Ziqi Gao et al., EGUsphere, https://doi.org/10.5194/egusphere-2022-396-RC2, 2022

General Comments

In the manuscript, four observation-based machine learning models are developed to predict the top 30 and the 4th highest maximum daily 8-hour average (MDA8) ozone (O_3) concentrations as a function of emissions, meteorological factors, and large-scale climate patterns in Southern California, USA. The top O_3 concentrations, especially the extreme statistics of O_3 concentration, are very difficult to accurately predict. The results show that these four models can explain most of the variations of the observed high O_3 concentrations. The study has examined the applicability of these built models in the South Coast Air Basin (SoCAB) and provide alternative methods for predicting top O_3 concentrations in other regions. I would recommend publication in Geoscientific Model Development after consideration of the following comments.

Specific comments

1. As the results shown in Figure 2, compared with the observations, all of the four models tend to slightly overestimate the lower MDA8 O_3 concentrations and to underpredict the higher ones. The four models have very small mean bias (MB, around 1ppbv) when predicting the top30 MDA8 O_3 concentrations (shown in Table S3), but they all have higher MB with the average ~10 ppbv underestimation on the 4th high MDA8 O_3 (shown in Table 2). As shown in Figure 3, more than 90% predicted O_3 concentrations are lower than the observations, which is consistent to the underestimations on the higher MDA8 O_3 shown in Figure 2. It indicates that the relationships between model inputs and predicted ozone are different at different ozone levels even addressing the highest 30 MDA8 O_3 concentrations. I wonder whether lower MB and RMSE for predicting the 4th high MDA8 O_3 would be expected with the empirical models developed using much higher MDA8 O_3 (for example, using the data on the top 15 MDA8 O_3 days).

2. As discussed in the Section 3.3 (Limitations), the precursors' emissions in SoCAB and local meteorological variables have been included in the development of the four models. The structure of the built model equations in the manuscript would be applicable for those regions where top MDA8 O_3 concentrations are mainly affected by local emissions. However, for the regions where the top MDA8 O_3 are significantly influenced by cross-regional O_3 transport, more variables might be considered in developing the predicting models (such as the precursors' emissions in surrounding regions).

3. In the study, the precursors' emissions have been proved to be the most significant factors impacting the peak O_3 levels in SoCAB, and maximum temperature is of relatively high importance among all the meteorological variables. The annual NOx and VOCs emission amounts and maximum temperature from 1990 to 2019 are suggested to be illustrated together with the corresponding 4^{th} high MDA8 O_3 (or the top30 MDA8 O_3 concentrations) in the Supplementary Information.

Technical comments

None.