Comment on acp-2022-502

Zongbo Shi (Referee)

Referee comment on "Decoupling impacts of weather conditions on interannual variations in concentrations of criteria air pollutants in South China – constraining analysis uncertainties by using multiple analysis tools" by Yu Lin et al., Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2022-502-RC1, 2022

This paper used two machine learning techniques to remove the meteorological effects on air quality trend in South China. The paper contains new and publishable results.

There are new developments in machine learning, which should be considered (see below). I suggest that the paper may be published after a major revision.
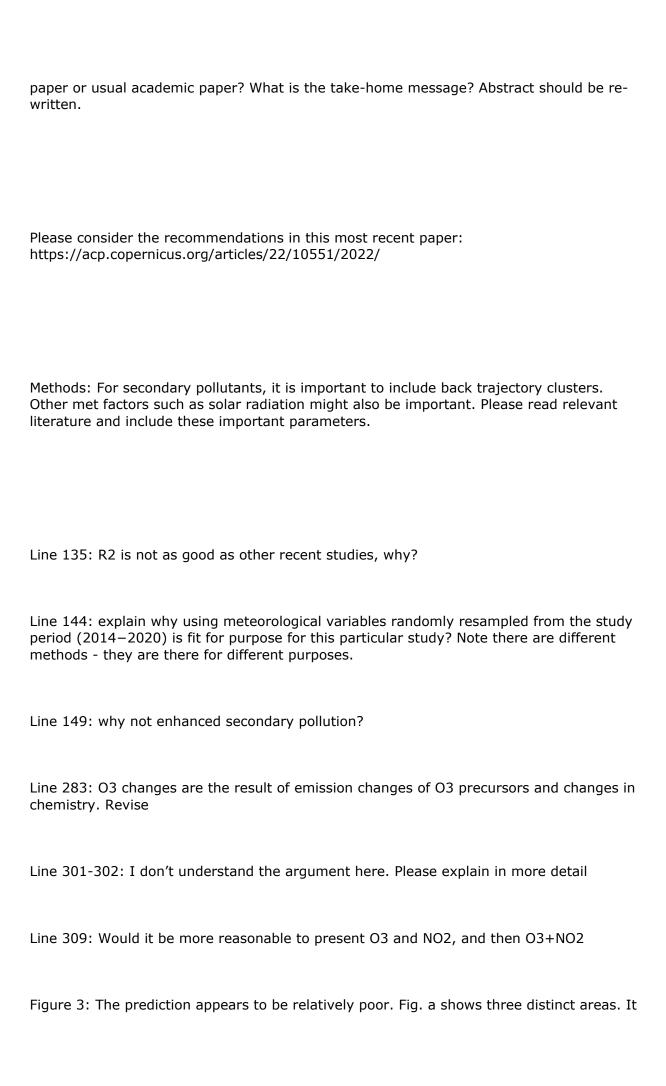
Specific comments

Line 19: define "in annual scale"

Line 21: define "consistency". Explain what does consistency of 70% or 30% means

Line 27: expand this section on results

Abstract focused more on methods but not results. Is this a methodological development

paper or usual academic paper? What is the take-home message? Abstract should be re-written.

Please consider the recommendations in this most recent paper:
https://acp.copernicus.org/articles/22/10551/2022/

Methods: For secondary pollutants, it is important to include back trajectory clusters. Other met factors such as solar radiation might also be important. Please read relevant literature and include these important parameters.

Line 135: R2 is not as good as other recent studies, why?

Line 144: explain why using meteorological variables randomly resampled from the study period (2014−2020) is fit for purpose for this particular study? Note there are different methods - they are there for different purposes.

Line 149: why not enhanced secondary pollution?

Line 283: O3 changes are the result of emission changes of O3 precursors and changes in chemistry. Revise

Line 301-302: I don't understand the argument here. Please explain in more detail

Line 309: Would it be more reasonable to present O3 and NO2, and then O3+NO2

Figure 3: The prediction appears to be relatively poor. Fig. a shows three distinct areas. It

appears to me There is something wrong – I would suggest that the authors check the codes and re-run the results, particularly including other parameters mentioned above. One way to check is to find a recent paper and use their dataset to test the researcers' skills in machine learning modelling.

$NO_2 + O_3$ is often defined as O_x. But you cannot add these two together based on mass concentration. Please turn NO_2 and O_3 into ppb first and then add. This is a mistake the corresponding authors should have picked up.

Discussions are inadequate, more or less repeating the results rather than an in-depth discussion. Two suggestions: interpret the results, in the contexts of literature and clean air policies; examine the implications of the results – e.g., what policies are effective and what are not. Suggest to remove all discussions in the Results, and move to Discussions as needed