Comment on acp-2022-502

Anonymous Referee #2

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-RC2, 2022

Analyzing long-term trends by excluding the effects of meteorological factors is critical in the assessment of anthropogenic air pollution factors. In this paper, the authors have used three different methods to decouple meteorological effects and investigate the trends of different pollutants in South China. I find the comparison of these three methods valuable and novel even though the trends were only consistent in 30% of the conditions between these approaches. The manuscript is well-written and has a proper flow to it. The problem statement and introduction are well-written. The discussion of results is clear. However, I think the method section should be expanded and better explained. Here are some general comments for improvement:

- RF and BRTs Modeling can be explained better. In particular, how the train-test splitting was applied is not explained thoroughly as it is important in model development. Was this random or sequential? For time series with long-term trends, this split should not be applied randomly, as might be customary in most of the random forest models in other fields, and should be applied sequentially. This is due to the fact that random split will bring extra information to the test validation (e.g. seasonal or weekly trends) that should not be available to the test and cause data leakage.
- The modeling work needs a feature importance analysis. This is very important since some of the features might not add anything to the model and can be simply eliminated from the analysis. Also, it shows the most influential meteorological factor on the trends.
- Some additional information can be added to the discussion section about the reasons for observing some of the trends. For example, if authors hypothesize specific regulation(s) as the reason for a specific deweathered trend, that can be added in the discussion section in addition to the introduction.
- The modeling
- The error or confidence intervals should be added to trend figures (e.g. figures 4 and 5).

Minor comments:

- Line 37: change "..two-three year.." to "…two-three years"
- Line 100 and figure 1: "Hourly meteorological data … were obtained from the meteorological observational station at a nearby airport". The meteorology factors, especially wind direction, change rapidly spatially at nearshore sites similar to the ones used in this work. Please mention that the meteorological stations were the closest available to the air quality sites if that is the case. Otherwise, please try to use the closest possible station in your database. Also, this should be mentioned as a source of error in the analysis.
- Line 128: change "predicated" to "predicted"
- Line 138: change "indicates" to "indicate"
- Figure 3: the range of predicted values is considerably smaller than the observed value. This is an inherited issue with RF and BRT models and should be explained in the text.
- Line 193: change "decreases" to "decrease".
- Line 232: change "conducted to" to "conducted on"
- Line 239: change "obtained between" to "obtained by"
- Line 269: change "annul" to "annual"