The manuscript by Qiu et al. investigates the sources and meteorological factors influencing ozone variation over four years in Hangzhou China, using observation-based approaches including machine learning (ML) -based meteorological normalization, PMF, and a Box Model for ozone simulation. Overall, the manuscript is well-organized, clearly written, and presents the results effectively. The application of ML in this study provides a strong example of its potential to enhance our understanding of ozone formation. My comments below are primarily focused on the methodology regarding source apportionment and the ML aspects, which the authors identify as novel points of this work.

General comments

Ozone concentrations are determined by various drivers (e.g., precursor emissions, dilution, transport, deposition, and chemistry). It is easy to relate ambient ozone to its drivers using ML algorithms, while it is important to emphasize the physical interpretation, not just the mathematical relationships, in data-driven approaches. This is why knowledge-guided ML is now highly recommended. Specifically, in the application of ML for explaining ozone formation, emphasis should be placed on feature selection (i.e., variables representing potential drivers) and the interpretation of results.

1. In the Methods section, the meteorological normalization method is applied to decouple the impact of meteorology from emission-driven changes in ozone and source-specific VOCs. ML-based meteorological normalization is essentially an adjustment method that aims to correct meteorologically induced variations in air quality time series. Similar statistical approaches have been used since the 1980s in the USA to estimate emission-driven trends of ozone. It is important to clarify that this technique does not "remove" meteorology from observational data but rather reduces its impact through specific techniques. We cannot have air pollution without meteorology. The term "REMOVE" is used throughout the text, it would be prudent to use quotation marks around "REMOVE" to avoid misunderstanding.

2. A key question here is the physical meaning of meteorologically normalized ozone. The level of normalized air pollutants depends on how normalization is applied according to the research purpose. Section 2.2 focuses heavily on random forest modeling but lacks sufficient detail about the rationale for feature selection and the meteorological normalization processes, making it difficult for readers to fully understand the implications of the results. The authors discuss the relative importance of dispersion +/ transport and chemistry in driving air pollutants, assuming these atmospheric processes are well represented by variables like wind and air mass clusters. This assumption needs clarification to build confidence in the model results—specifically, what features are proxies for specific atmospheric processes?

3. In Section 2.4, the authors state "*In this study, the observed and meteorological normalized VOCs concentrations were fed into US EPA PMF v5.0 to identify and quantify major emission sources of VOCs*." This approach is interesting for PMF modeling, particularly in examining changes in source contributions after meteorological normalization to understand the impact of dispersion (should be the overall impact of meteorology) on VOC sources (a good point to address). However, since PMF is a bilinear model requiring additive input variables, questions arise: are these normalized VOCs still additive? How is the total VOC for normalized

concentrations calculated? Is the normalized VOC comparable to the observed VOC? An alternative approach to achieve the same goal might be to meteorologically normalize the PMF-resolved source-specific VOCs (i.e., run PMF with observed VOCs first, then normalize each source-specific VOC). This work may be of the authors interest: https://doi.org/10.1029/2023JD038696.

4. In Figure 2, all features are ranked with positive values, which describe the magnitude of their impacts without considering the sign of those impacts. However, dispersion can have both positive (enhancing concentration during poor dilution) and negative (reducing pollutant levels) effects. Additionally, using pie charts to illustrate the roles of dispersion and chemistry is problematic because chemistry is not independent of dispersion and transport. Can the authors elaborate more about this?

5. In the Results & Discussion section, the authors demonstrate model performance using only the squared correlation coefficients. It is recommended to also include root mean squared errors, as this is an important metric for describing the accuracy of model predictions.

In summary, I strongly recommend that the authors add more details about feature selection, the adopted meteorological normalization process, and the physical meaning of the normalized air pollutants. One of the existing literature has discussed and reviewed various meteorological normalization strategies based on ML modeling, which may be helpful for this work (https://doi.org/10.1007/s11430-022-1128-1).

Minor Comments

1. Line 381: Clarify what is meant by “After normalizing the effect of dispersion.” Meteorological normalization is not limited to normalizing the effect of dispersion.

2. Figure 8: Regarding the pies for normalized source contributions, are these contributions additive? What is the physical meaning of the sum of normalized source contributions?