Response to Reviewer #2

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.

Response:

We apologize for our imprecise wording and appreciate the reviewer's comments and suggestions. The machine learning-based meteorological normalization method indeed did not remove the influence of meteorological factors. We have taken your advice and placed "remove" in quotes in the manuscript.

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?

Response:

We appreciate the reviewer's comments. We have made the following revisions to clarify which features serve as proxies for specific atmospheric processes:

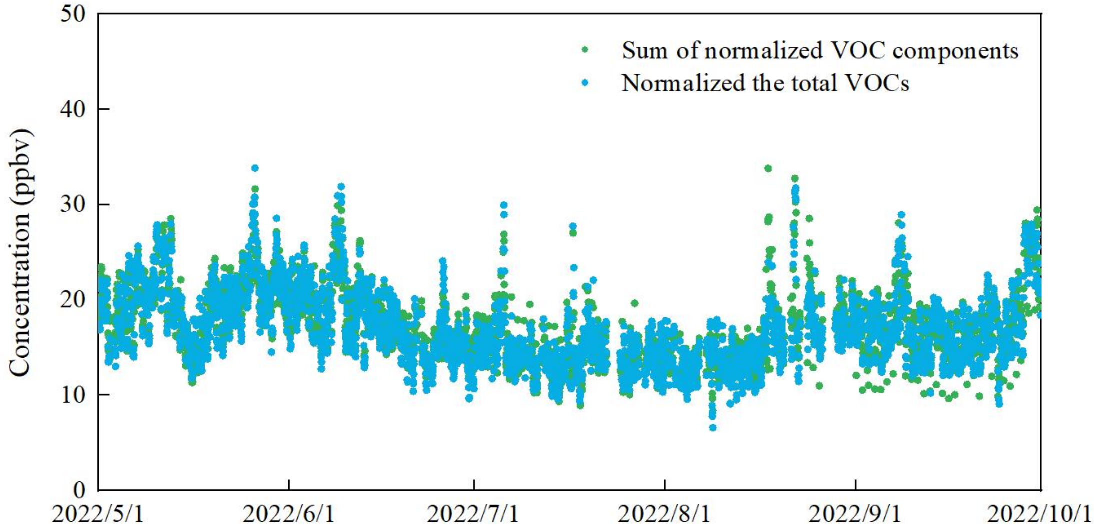
In the RF model, the air pollutants were the response variables, while the explanatory variables included time variables representing source emissions and meteorological variables representing physical and chemical processes. Time variables such as day of year, month, weekday and hour are used to indicate the seasonal, weekly, and daily cycles of emission intensity (Dai et al., 2023; Vu et al., 2019). Temperature was a key factor influencing the rate of chemical reactions, with higher temperatures typically promoting the photochemical reactions that generate O_3. UVB served as the driving force for the photochemical reactions, directly impacting O_3 formation. Additionally, humidity played an important role in the chemical processes involved in O_3 formation. Therefore, T, RH, and UVB were identified as the key features associated with atmospheric photochemical reactions. WS influences the dispersion of atmospheric pollutants. At high wind speeds, air pollutants tended to be dispersed, while low wind speeds resulted in local pollutant accumulation, leading to increased concentrations. WD determined the dispersion path of atmospheric pollutants. BLH was a critical factor affecting the vertical dispersion of pollutants. A higher boundary layer allowed pollutants to disperse more effectively into the upper atmosphere, reducing surface concentrations, whereas a lower boundary layer resulted in pollutant accumulation near the ground. Thus, WS, WD, and BLH were regarded as the features of atmospheric physical dispersion on a local scale. Cluster can serve as a feature of transport from remote regions.

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.

Response:

We appreciate the reviewer's comments. The reviewer raised a very critical point: the mathematical requirement of the PMF model was that the total concentration was a linear combination of contributions from individual sources. Random forest was a nonlinear machine learning algorithm. We applied the random forest model for meteorological normalization to individual VOC species and total VOCs, and found that the sum of the meteorologically normalized VOC species remained linearly correlated with the total VOCs (Fig. S4). Therefore, we believed that the nonlinear processing did not significantly alter the overall structure of the total VOCs concentrations, the results of the PMF model remained reasonable under the conditions. Additionally, we appreciate the reviewer's suggestion regarding the method for meteorological normalization of source-specific VOCs in PMF analysis. We have carefully read and cited this literature, and we believed that this method directly satisfies the additivity principle of PMF, making it an excellent approach. We will also consider comparing the results of these two methods in the future. We have added the following explanation in Section 2.4 of the Methods:

RF model for meteorological normalization was a nonlinear machine learning algorithm. To satisfy the fundamental mathematical requirement of the PMF model, which stated that the total concentration was a linear combination of contributions from individual sources, the RF model was applied for meteorological normalization of individual VOC species and total VOCs in this study. This ensured that the sum of the meteorologically normalized VOC species remained linearly correlated with total VOCs (Fig. S4), indicating that the nonlinear processing did not significantly alter the overall structure of total VOC concentrations. With this approach, the results obtained by inputting the meteorologically normalized data into the PMF model were reasonable.



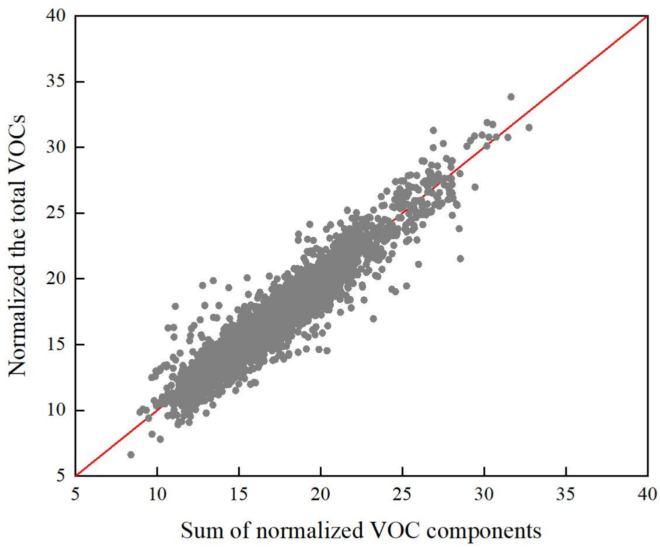


Figure S4: Time series and correlation of the sum of normalized VOC species and normalized total VOCs.

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?*

Response:

We appreciated the reviewer's comments. In this study, we used feature importance to reflect the overall significance of explanatory variables in the RF model. Although these features may have positive or negative effects on O_3 concentrations under different environmental conditions, their importance values in the global model quantified the proportion they occupied, reflecting their contribution to the overall prediction performance. While feature importance did not directly reveal whether the influence was positive or negative, it highlighted the critical factors in the model's predictions (Feng et al., 2019; Liu et al., 2022b; Ye et al., 2022; Yang et al., 2023). We appreciated the reviewer for raising this excellent point, and in future research, we planned to employ SHAP analysis to investigate the positive and negative impacts of individual features. We have added the following explanation about feature importance in the Method:

Feature importance was used to reflect the overall significance of explanatory variables in the RF model. The importance was typically represented as an array, where each value corresponded to the importance score of a specific feature. These scores usually range from 0 to 1. The higher importance score indicated that the feature had a stronger predictive capability for the response variable.

Additionally, we were grateful for the reviewer's comment that "chemistry is not independent of dispersion and transport." The pie chart was intended to simplify the display of the relative contributions of emissions, chemical reactions, local dispersion, and long-distance transport. In the Methods section, we added explanations of the atmospheric processes represented by each meteorological parameter:

Temperature was a key factor influencing the rate of chemical reactions, with higher temperatures typically promoting the photochemical reactions that generate O_3. UVB served as the driving force for the photochemical reactions, directly impacting O_3 formation. Additionally, humidity played an important role in the chemical processes involved in O_3 formation. Therefore, T, RH, and UVB were identified as the key features associated with atmospheric photochemical reactions. WS influences the dispersion of atmospheric pollutants. At high wind speeds, air pollutants tended to be dispersed, while low wind speeds resulted in local pollutant accumulation, leading to increased concentrations. WD determined the dispersion path of atmospheric pollutants. BLH was a critical factor affecting the vertical dispersion of pollutants. A higher boundary layer allowed pollutants to disperse more effectively into the upper atmosphere, reducing surface concentrations, whereas a lower boundary layer resulted in pollutant accumulation near the ground. Thus, WS, WD, and BLH were regarded as the features of atmospheric physical dispersion on a local scale. Cluster can serve as a feature of transport from remote regions.

Under this assumption, we considered the interactions between dispersion, transport, and chemistry to be insignificant, allowing us to treat the physical and chemical processes independently. Thank you once again for helping us enhance the scientific accuracy of the paper.

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.

Response:

We appreciate the reviewer's comments. We have removed the correlation coefficients (r^2) from the Results and Discussion section of the manuscript and included r^2, along with other performance metrics such as root-mean-square error (RMSE), FAC2 (fraction of predictions with a factor of 2), mean bias (MB), mean gross error (MGE), normalized mean bias (NMB), normalized mean gross error (NMGE), coefficient of efficiency (COE), and index of agreement (IOA), in the Supplement.

Table S2. RF model performance for testing data set .

Pollutants	r^2	RMSE	FAC2	MB	MGE	NMB	NMGE	COE	IOA
O_3	0.88	17.33	0.80	-0.34	12.70	-0.01	0.22	0.68	0.84
NO_2	0.83	9.43	0.97	0.11	6.85	0.00	0.18	0.62	0.81
NMHCs	0.76	6.41	0.99	-0.11	4.60	0.00	0.20	0.54	0.77

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).

Response:

We appreciate the reviewer's comments. We have added the details regarding explanatory variables, model development, model evaluation, and meteorological normalization as follow:

The descriptions of explanatory variables in the response under "General comment 2."

There are approximately 32,856 valid data with a time resolution of 1 hour. The RF model was trained using a forest of 1,000 trees. Training datasets of the RF model was conducted on 80% of the original datasets, and the remaining 20% was selected as testing datasets. Correlation coefficients (r^2), root-mean-square error (RMSE), FAC2 (fraction of predictions with a factor of 2), mean bias (MB), mean gross error (MGE), normalized mean bias (NMB), normalized mean gross error (NMGE), coefficient of efficiency (COE), and index of agreement (IOA) were used to evaluate model performance (Table S2). Based on previous related research, these statistical measures indicated that the model performed well (Emery et al., 2017; Henneman et al., 2017; Vu et al., 2019).

The process of meteorological normalization involved replacing the original meteorological variables with those randomly resampled from the observation dataset, and using the established RF model to predict atmospheric pollutant concentrations under different meteorological conditions. The resampling of meteorological variables was conducted over the two-week period before and after the selected date, with the resampled hours remaining constant. This approach effectively preserved the seasonal and diurnal variations in the response variables (Vu et al., 2019). The resampling and prediction process were repeated 1000 times to generate 1000 predicted pollutants concentrations. The average values were taken as the final meteorologically normalized concentrations. In the meteorological normalization process of O_3 concentration, meteorological variables such as WS, WD, BLH, and cluster, which signify dispersion and transport, were randomly sampled. In the case of O_3 precursors, namely NO_2 and NMHCs, resampling was exclusively applied to WS, WD and BLH. NO_2 and NMHCs have short atmospheric lifetimes, making them less susceptible to the influence of regional transport over large scales (Wang et al., 2023).

Finanlly, We thank the reviewer for providing us with this excellent work, from which we have learned a lot. We have drawn on the analytical methods and statements, and have cited it in our study.

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.

Response:

We appreciate the reviewer's comments. For VOCs, we performed resampling on three meteorological variables: wind speed (WS), wind direction (WD), and boundary layer height (BLH), which we considered mainly to indicate dispersion effects. However, this statement may lead to ambiguity, so we have revised it to the following statement:

"After smoothing out 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?

Response:

We appreciate the reviewer's comments. The normalized source contributions were still additive as the response in General Comment 3. The physical meaning of the sum of normalized source contributions was the relative contribution of all sources to the total VOC concentration after meteorological normalization. We have added the following description in the manuscript:

Fig. 8 showed the proportion of VOCs sources before and after meteorological normalization during the non-pollution periods and pollution periods. The pies for normalized source contributions illustrated the relative contribution of each source to the total VOC concentration after "removing" the effects of dispersion.

Dai, Q., Dai, T., Hou, L., Li, L., Bi, X., Zhang, Y., and Feng, Y.: Quantifying the impacts of emissions and meteorology on the interannual variations of air pollutants in major Chinese cities from 2015 to 2021, Science China Earth Sciences, 66, 1725-1737, 10.1007/s11430-022-1128-1, 2023.

Emery, C., Liu, Z., Russell, A. G., Odman, M. T., Yarwood, G., and Kumar, N.: Recommendations on statistics and benchmarks to assess photochemical model performance, J Air Waste Manag Assoc, 67, 582-598, 10.1080/10962247.2016.1265027, 2017.

Feng, R., Zheng, H.-j., Zhang, A.-r., Huang, C., Gao, H., and Ma, Y.-c.: Unveiling tropospheric ozone by the traditional atmospheric model and machine learning, and their comparison: A case study in hangzhou, China, Environmental Pollution, 252, 366-378, 10.1016/j.envpol.2019.05.101, 2019.

Henneman, L. R. F., Liu, C., Hu, Y., Mulholland, J. A., and Russell, A. G.: Air quality modeling for accountability research: Operational, dynamic, and diagnostic evaluation, Atmospheric Environment, 166, 551-565, 10.1016/j.atmosenv.2017.07.049, 2017.

Liu, B., Wang, Y., Meng, H., Dai, Q., Diao, L., Wu, J., Shi, L., Wang, J., Zhang, Y., and Feng, Y.: Dramatic changes in atmospheric pollution source contributions for a coastal megacity in northern China from 2011 to 2020, Atmospheric Chemistry and Physics, 22, 8597-8615, 10.5194/acp-22-8597-2022, 2022a.

Liu, H., Yue, F., and Xie, Z.: Quantify the role of anthropogenic emission and meteorology on air pollution using machine learning approach: A case study of $PM_{2.5}$ during the COVID-19 outbreak in Hubei Province, China, Environmental Pollution, 300, 10.1016/j.envpol.2022.118932, 2022b.

Liu, Y., Geng, G., Cheng, J., Liu, Y., Xiao, Q., Liu, L., Shi, Q., Tong, D., He, K., and Zhang, Q.: Drivers of Increasing Ozone during the Two Phases of Clean Air Actions in China 2013–2020, Environmental Science & Technology, 57, 8954-8964, 10.1021/acs.est.3c00054, 2023.

Vu, T. V., Shi, Z., Cheng, J., Zhang, Q., He, K., Wang, S., and Harrison, R. M.: Assessing the impact of clean air action on air quality trends in Beijing using a machine learning technique, Atmospheric Chemistry and Physics, 19, 11303-11314, 10.5194/acp-19-11303-2019, 2019.

Wang, Y., Jiang, S., Huang, L., Lu, G., Kasemsan, M., Yaluk, E. A., Liu, H., Liao, J., Bian, J., Zhang, K., Chen, H., and Li, L.: Differences between VOCs and NOx transport contributions, their impacts on O3, and implications for O_3 pollution mitigation based on CMAQ simulation over the Yangtze River Delta, China, Science of the Total Environment, 872, 10.1016/j.scitotenv.2023.162118, 2023.

Yang, C., Dong, H., Chen, Y., Wang, Y., Fan, X., Tham, Y. J., Chen, G., Xu, L., Lin, Z., Li, M., Hong, Y., and Chen, J.: Machine Learning Reveals the Parameters Affecting the Gaseous Sulfuric Acid Distribution in a Coastal City: Model Construction and Interpretation, Environmental Science & Technology Letters, 10, 1045-1051, 10.1021/acs.estlett.3c00170, 2023.

Yang, J., Zeren, Y., Guo, H., Wang, Y., Lyu, X., Zhou, B., Gao, H., Yao, D., Wang, Z., Zhao, S., Li, J., and Zhang, G.: Wintertime ozone surges: The critical role of alkene ozonolysis, Environmental Science and Ecotechnology, 22, 10.1016/j.ese.2024.100477, 2024.

Ye, X., Wang, X., and Zhang, L.: Diagnosing the Model Bias in Simulating Daily Surface Ozone Variability Using a Machine Learning Method: The Effects of Dry Deposition and Cloud Optical Depth, Environmental Science & Technology, 56, 16665-16675, 10.1021/acs.est.2c05712, 2022.