Response (in blue) to Reviewer's comments #2

Manuscript number: egusphere-2025-165

Title: Explainable ensemble machine learning revealing enhanced anthropogenic emissions of particulate nitro-aromatic compounds in eastern China

Response to reviewer #2:

The manuscript investigated the sources and drivers of particulate nitro-aromatic compounds (NACs) in eastern China using a combination of machine learning and receptor modelling. The study's main focus is how primary emissions, secondary formation, and meteorological factors contribute to ambient NAC levels across different locations and seasons. The authors proposed an ensemble machine learning (EML) model coupled with SHAP (SHapley Additive exPlanation) values and a PMF (Positive Matrix Factorization) source apportionment to interpret NAC variations. Eleven sampling sites (urban, rural, mountain) over multiple seasons provide a robust dataset of NAC concentrations and related variables. The EML model achieves high predictive performance (as can be expected from statistical modelling). The authors conclude that strengthened control of combustion emissions is necessary to mitigate particulate NAC pollution, as their modelling highlights the outsized role of human sources even in a region with complex meteorological and secondary processes.

Overall, this work is important. It extends existing literature on NAC sources (which previously relied on linear models or standalone PMF) by providing a more interpretable quantification of each factor's contribution. The study is well grounded in current literature and clearly exhibits its novelty by bridging source apportionment with explainable AI. A few methodological clarifications and edits (detailed below) could further strengthen the work before this paper could be submitted.

We sincerely thank the review for the valuable feedback that we have used to improve the quality of our manuscript. According to the comments, we have made extensive modifications to this manuscript to make results convincing. In the revised version, the reviewer comments are laid out below in **bold black** font. Below, we provide a point-to-point response to each comment. Our response is given in blue and changes to our manuscript are all highlighted by using *blue italic* text. We have tried our best to improve the manuscript and we hope the revision would satisfactorily address the comments and concerns of the reviewer.

Furthermore, we would like to show the details as below:

1. I agree that the ensemble machine learning approach is appropriate for capturing complex nonlinear relationships, but some details would benefit from clarification to enhance confidence in the results. The authors note an 80/20 random split with cross-validation, but given data from multiple sites and seasons, it would be helpful to discuss whether any site-specific bias could affect the model. If, for instance, all data from a particular location or season mostly fell into the training set, the reported high R2 might not fully reflect generalizable performance. An ideal approach (if data allow) would be to test the model's predictive skill in a leave-one-site-out or leave-one-season-out manner to ensure it generalizes across different scenarios.

Response: We appreciate the reviewer's insightful comment regarding the potential influence of site-specific or seasonal bias in our model evaluation. To address this concern and rigorously assess the spatial generalizability of our ensemble machine learning (EML) model, we performed a leave-one-site-out cross-validation analysis.

Under this cross-validation scheme, data from each site were systematically withheld from model training in turn, and predictions were made exclusively on the excluded site. The procedure ensured a strict separation between training and testing data, thereby providing an unbiased estimate of model transferability across different scenario.

As shown in the revised manuscript (Figure S7), the leave-one-site-out results demonstrated good predictive performance, with an overall R^2 of 0.84 and a regression slope of 0.92 between observed and predicted NACs concentrations, which further indicates that the model is generalizable rather than a location-specific model. The relevant description has been incorporated into the revised manuscript accordingly.

Revised sentence in manuscript (Line 212–216):

"*To further evaluate the generalizability of the EML model, a leave-one-site-out cross-validation approach was implemented. The data from each site were iteratively excluded from model training and used exclusively for testing, ensuring complete independence between training and testing sets. The results show that this model exhibits robustness and transferability rather than limited to specific scenarios (see Fig. S7).*"

Added Figure S7 in Supporting Information:

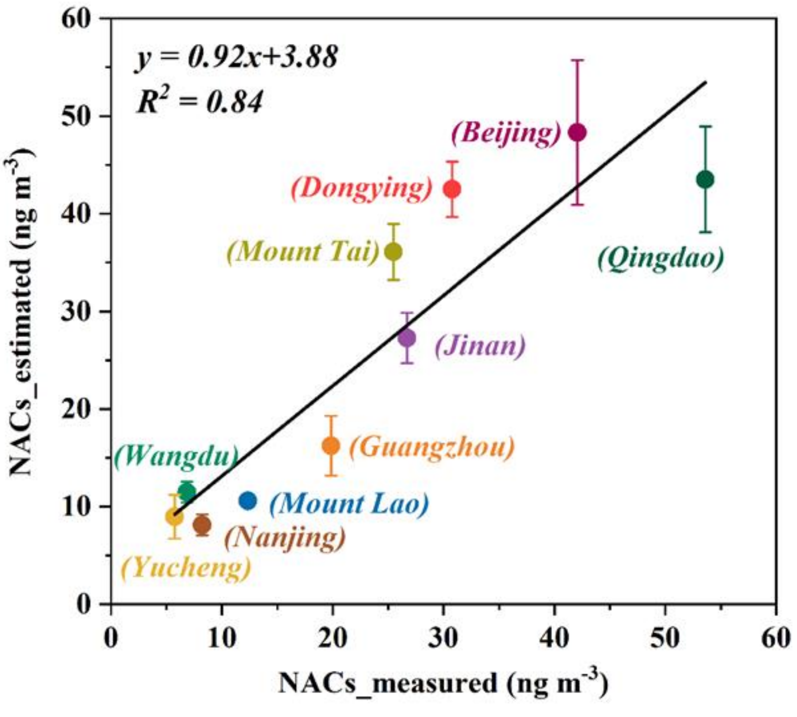

Figure S7. Comparison of observed and simulated NACs at different sites with a leave-one-site-out cross-validation approach.

While the integration of PMF source contributions as input features is innovative, this could introduce circular reasoning if not carefully handled – since NAC concentrations themselves (via their speciation) inform the PMF factors. The authors should reassure that using PMF outputs (four source factor contributions) as predictors does not inadvertently "double count" NAC information. One way to address this would be to emphasize that the ML model's target was the total NAC (or NAC subgroups) concentration and that PMF factors, being based on species patterns, serve as

independent explanatory variables capturing source-type influences. Clarifying these points will help readers understand the modelling strategy and trust that the conclusions (e.g., anthropogenic share of ~49%) are data-driven and not an artifact of the model design.

Response: Thanks for the reviewer's valuable comment regarding the potential risk of circular reasoning due to the integration of PMF source contributions as input features in the ML model. In this study, the target variable of the ML model is the total concentration of NACs, which includes not only the NACs species used in the PMF model (*i.e.*, 4NP, 3M4NP, 2M4NP, 4NC, 4M5NC, 3M6NC, 5NSA, 3NSA), but also other NACs that were not included in the PMF input matrix (as shown in Table S1). Therefore, there is no direct overlap between the target variable and the PMF input.

Additionally, the PMF outputs are composite source-type signatures derived from the covariation of these eight NACs and tracer gases, rather than reconstructions of individual NAC concentrations.

Importantly, in the ML modeling process, we used only the PMF-derived source contributions as input features, and no individual NAC concentrations were directly included. As a result, the ML model avoids any potential data leakage or double-counting of NACs, which further supports the robustness of the conclusion regarding anthropogenic influence.

In the revised manuscript, we have added relevant sentence and provided a comprehensive description of the data-driven modeling framework on PMF and ML.

Revised sentence in manuscript (Line 192–198):

"*The* dataset (613 rows) *used* for the four ML algorithms *consisted of eleven parameters as inputs, including PMF-derived source contributions,* meteorological conditions (T, BLH, RH, SSR, WS_H, and WS_V), and heterogeneous reaction represented by the aerosol surface area *(Sa)*, *all of which influence the sources and sinks of NACs. To avoid circular reasoning, the ML model was constructed to predict the total concentration of NACs as target variable. The four PMF-derived source contribution factors, which serve as independent explanatory variables capturing source-type influences, were used as input features instead of individual NAC species. This approach ensures a clear separation between PMF inputs and the ML target, effectively preventing data leakage or double counting*."

Added Table S1 in Supporting Information:

Table S1. Sampling sites and sampling periods involved in this study.

Sampling site	Site type	Sampling period	Season	Number of samples	Detected species
Jinan	urban	2016.04.12-2016.04.27	spring	9	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
		2014.09.04-2014.09.21	summer	37	1, 2, 3, 5, 6, 7, 8, 9, 10
		2016.06.27-2016.07.11			1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
		2017.10.22-2017.11.01	autumn	20	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
		2013.11.26-2014.01.05	winter	16	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
		2016.02.19-2016.03.07			1, 2, 3, 5, 6, 7, 8, 9, 10
Guangzhou	urban	2017.06.28-2017.07.08	summer	20	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
Nanjing	urban	2017.10.22-2017.10.31	autumn	16	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
Beijing	urban	2018.01.15-2018.01.31	winter	14	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
Yucheng	rural	2014.06.09-2014.06.20	summer	16	1, 2, 3, 5, 6, 7, 8, 9, 10
Wangdu	rural	2014.06.19-2014.06.29	summer	18	1, 2, 3, 5, 6, 7, 8, 9, 10

Dongying	rural	2017.06.04-2017.06.15	summer	10	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
		2017.01.15-2017.01.23	winter	9	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
Qingdao	rural	2019.01.10-2019.02.23	winter	132	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
		2019.11.11-2019.12.25			1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12
Mount Tai	mountain	2018.03.22-2018.04.05	spring	25	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
		2014.07.27-2014.08.06	summer	17	1, 2, 3, 5, 6, 7, 8, 9, 10
		2017.11.28-2017.12.09	winter	157	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
		2019.12.01-2019.12.31			1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
Mount Lao	mountain	2021.04.16-2021.05.19	spring	97	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

NOTE: 1 4-nitrophenol (4NP). 2 3-methyl-4-nitrophenol (3M4NP). 3 2-methyl-4-nitrophenol (2M4NP). 4 2,6-dimethyl-4-nitrophenol (2,6DM4NP). 5 4-nitrocatechol (4NC). 6 4-methyl-5-nitrocatechol (4M5NC). 7 3-methyl-6-nitrocatechol (3M6NC). 8 3-methyl-5-nitrocatechol (3M5NC). 9 5-nitrosalicylic acid (5NSA). 10 3-nitrosalicylic acid (5NSA). 11 2,4-dinitrophenol (2,4DNP). 12 4-methyl-2,6-dinitrophenol (4M2,6DNP).

2. The claim of "enhanced anthropogenic emissions" driving NAC pollution needs to be positioned against existing studies to ensure the manuscript's novelty is clear. Prior works have already pointed to combustion sources (coal, biomass burning, vehicle emissions) as major NAC contributors. My understanding is, the manuscript's novelty is primarily methodological, and this study's added value lies in quantifying the contributions with a new method and revealing nuanced patterns (like seasonal driver shifts and differences between urban/rural/mountain sites). The authors should ensure readers recognize that the significance lies in using an explainable ML approach to confirm and detail known drivers, rather than in discovering an entirely new source of NACs. This steer will prevent any impression that the study is merely repeating known information, instead of providing new insights into the magnitude and context of anthropogenic influence.

Response: Thanks for the constructive comment. As suggested, we have clarified that the key contribution of this study lies not in identifying new NAC sources, but in employing an explainable ensemble machine learning framework to provide high-resolution, quantitative assessment of the relative importance of known sources under complex atmospheric conditions. By applying this approach across urban, rural, and mountain sites and throughout different seasons, we revealed nuanced shifts in drivers, which have not been captured in prior NAC source apportionment studies.

To ensure the manuscript's novelty is clear, we have changed the title and add some sentences in the revised manuscript:

a. Title has been changed into "Explainable ensemble machine learning revealing *spatiotemporal heterogeneity in driving factors of* particulate nitro-aromatic compounds in eastern China"

b. One supplemental clarification is in the Introduction, to position this study with the context of prior source apportionment research and emphasize our methodological innovation.

c. Another supplemental clarification is in Section 3.2, to highlight the added value of the new approach in refining the understanding of known drivers under different environmental conditions.

Revised sentence in manuscript (Line 84–86 and Line 299–303):

"*This study makes a methodological contribution by employing a novel approach to quantify the seasonal shifts in drivers and spatial variations across urban, rural, and mountain regions in a nuanced manner.*"

"*This enhancement in anthropogenic emissions is consistent with the findings reported in previous NAC* *studies (Wang et al., 2018; Yuan et al., 2021). However, the integration of the explainable EML framework constitutes a methodological advancement by enabling quantitative evaluation of source contributions, thereby providing a more nuanced and context-specific understanding of the driving factors across diverse atmospheric conditions.*"

3. The use of SHAP values is a strong point of the study, but some aspects of the SHAP-based findings could be explained more clearly to avoid confusion. One issue is the meaning of negative SHAP contributions for certain factors. For example, the authors mention that at the mountain site, primary emissions had a mean SHAP contribution of –5.7 ng m^{-3}, which initially sounds like primary sources were somehow reducing NAC levels. The intended meaning is presumably that local primary emissions are minimal at the mountain (so their absence corresponds to lower baseline NAC, hence a negative SHAP relative to other sites). Also for discussions regarding "temperature" and "BLH", providing one or two sentences of intuition (e.g., "a negative SHAP value for a factor means that higher values of that factor are associated with lower NAC concentrations") when introducing the SHAP results would make the explanation more accessible, especially for readers new to SHAP analysis.

Response: Thanks for the review's valuable comment. We have revised the manuscript to clarify the interpretation of SHAP values. Specifically, the negative SHAP contribution of primary emissions at the mountain site reflects minimal local emissions, which results in lower NAC concentrations, rather than indicating an actual reduction in NACs due to these sources. Additionally, we also provided concise explanations for the SHAP interpretations of the input variables. These revisions aim to improve the clarity and accessibility of the SHAP-based analysis, especially for readers who are less familiar with the method.

Revised sentence in manuscript (Line 315–317 and Line 423–424):

"*A positive SHAP value indicates that the variable increases the predicted NAC concentration relative to the baseline, whereas a negative SHAP value suggests that higher values of the variable are associated with a decrease in NAC concentrations.*"

"*This negative value reflects the minimal contribution of local anthropogenic emissions in this region, resulting in lower concentrations of NACs compared to other sites.*"

SHAP can sometimes capture pairwise interactions, the authors could discuss on interactions or co-variability among factors if any were observed. For example, did the authors notice if certain meteorological conditions amplify the effect of emissions (high humidity aiding secondary formation of NACs, etc.)? Ensuring the SHAP results are clearly linked back to physical processes (mixing, photochemistry, emissions timing) will make the conclusions more convincing and useful for policy implications.

Response: Thanks for the insightful comment. In response, we conducted a detailed analysis of pairwise SHAP interaction values among key variables. Notably, a significant interaction between temperature (T) and aerosol surface area (Sa) was identified, as shown in the newly added Figure S8. The interaction pattern indicates that high Sa facilitates NACs formation under low-temperature conditions (T < 10°C), suggesting the enhanced gas-particle partitioning and heterogenous reactions. In contrast, high Sa appears to inhibit NACs formation at high temperature (T > 10°C), potentially due to intensified photochemical reactions shifting towards gas-phase products, high temperature promoting to particle-to-gas partitioning, or dilution effects arising from elevated mixing heights in hot seasons. The temperature-dependent behavior highlights the complex role of heterogenous reaction in atmospheric aerosol formation.

However, no other variable pairs exhibited comparable interaction effects across the dataset. Future research incorporating more comprehensive datasets with machine learning or deep learning model is required to better elucidate the synergistic effects on ambient NAC concentrations.

Revised sentence in manuscript (Line 332–337):

"*Notably, at low temperature (approximately < 10°C), the contribution on NACs exhibited an explosive enhancement, accompanied by a pronounced synergistic effect with higher Sa (Fig. S8), indicating enhanced gas-to-particle partitioning and heterogeneous formation. Conversely, at higher temperature, high Sa appears to suppress NAC formation, possibly as a result of intensified photochemical reactions facilitating gas-phase products, high temperature promoting to the partitioning to particle phase, or dilution effects caused by increased mixing heights in hot seasons.*"

Added Figure S7 in Supporting Information:

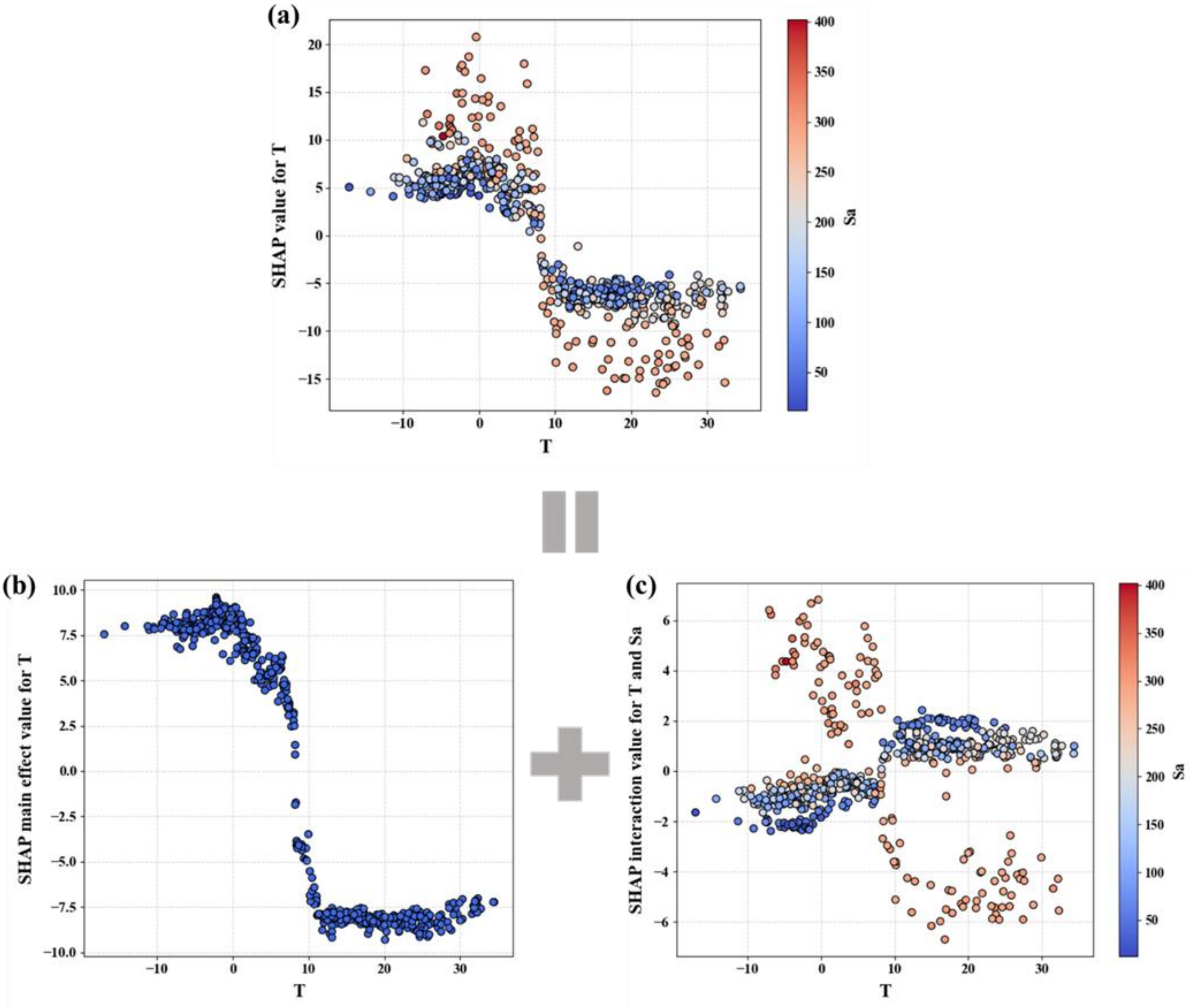

Figure S8. (a) The interaction effect of temperature (T) and aerosol surface area (Sa), (b) the main effects of T on NACS, and (c) the interaction SHAP value between T and Sa shows how the effect of T on NACs varies with Sa.

4. Line 186, the multi-target modelling approach, where NPs, NCs, and NSAs were predicted simultaneously (mentioned in the Methods), is an interesting aspect but is not very prominently discussed in the results. The conclusion hints that different functional groups had different key

drivers (e.g., gas-phase oxidation dominating NSAs). It would strengthen the paper to emphasize these findings a bit more in the Results section 3.3 or 3.4 – for instance, explicitly stating which sources were most important for each NAC subclass. This adds depth to the analysis (showing the model's strength in capturing subtle differences).

Response: Thanks for the comment. As suggested, in the revised manuscript, we have incorporated a summary paragraph at Section 3.3 to explicitly highlight the distinct sources for different NAC subclasses. Specifically, coal combustion was identified as the primary contributor to NPs, biomass burning emerged as the dominant source for NCs, and NSAs were predominantly associated with gas-phase formation. These results demonstrate the model's capacity to resolve nuanced differences in source attribution across functional groups and underscore the importance of implementing targeted control strategies.

Revised sentence in manuscript (Line 395–399):

"*Overall, the results demonstrate that the multi-target EML model effectively captured the distinct source contributions and formation pathways associated with different NAC subclasses. Coal combustion was identified as the most important driver for NPs, biomass burning dominated the formation of NCs, and NSAs were primarily linked to gas-phase formation. These findings highlight the strength of this integrated EML approach in differentiating functional group-specific drivers and emphasize the importance of targeted mitigation strategies for various NAC species.*"

Also, given that the data span 2014–2021, there could be a question that if trends over that period were considered – for example, have emission controls in China over the years impacted NAC levels? This may be outside the scope of the current paper's focus on spatial drivers, but a short note in the discussion could acknowledge that temporal trends were not the focus here (assuming no strong trend was observed after accounting for other factors).

Response: We appreciate the reviewer's thoughtful comment. We appreciate the suggestion to explore potential temporal trends in NAC concentrations, particularly considering the emission control measures implemented in China over the years. We have conducted a thorough analysis of the data spanning from 2014 to 2021, and our results indicate that, under consistent seasonal and site-type conditions, NAC concentrations did not exhibit any significant temporal trends, suggesting that the factors influencing NAC levels during this period were primarily spatial rather than temporal. We have added a brief note in the Discussion section to acknowledge that temporal trends were not the focus of this study, and clarify that no strong significant trends were observed after accounting for other variables.

Revised sentence in manuscript (Line 248–249):

"*Moreover, data from 2014 to 2021 revealed no significant trends in NAC concentrations across the same seasonal and site-type conditions, therefore temporal variation was not considered as a primary focus of this study.*"

Minor issues:

The paper is generally well-written, but a few sentences should be edited for clarity or correctness. Here below are examples but the authors need to read through the manuscript for such minor language issues:

Line 73, "Given the complex nonlinear links… it is necessary to establish an effective and reliable evaluation method to comprehensively understand and assess the importance and contribution of each factor…". This could be broken into two sentences to avoid confusion.

Response: Thanks for the helpful suggestion. In response, we have revised the sentence by dividing it into two parts to enhance clarity.

Revised sentence in manuscript (Line 76–80):

"Given the complex nonlinear links between primary emissions, secondary formation, and meteorological conditions and the ambient particulate NACs, *a clear understanding of the separate role of each factor is challenging. Therefore,* it is necessary to establish an effective and reliable evaluation method to comprehensively understand and assess the importance and contribution of each factor on the abundances of NACs under complicated atmospheric conditions."

Line 258, the phrasing "is in coincided with" is grammatically incorrect.

Response: Thanks for the comment. It has been corrected in the revised manuscript.

Revised sentence in manuscript (Line 272–273):

"The dominance of NPs and NCs in this study *coincides* with *the findings from previous studies* in other locations (Cai et al., 2022; Li et al., 2020c; Wang et al., 2019)."

Line 182, a typo "leaner" should be "learner"

Response: Thanks for the comment. It has been corrected.

Line 389, "Jinan ang Beijng" should be "and Beijing"

Response: Thanks for the comment. It has been corrected.

Line 363, "confirmed by a previous observational study"

Response: Thanks for the comment. It has been corrected.

1. As noted, the manuscript uses many abbreviations (NACs, PMF, EML, SHAP, NP, NC, NSA, BLH, SSR, WS_V, WS_H, etc.). It would be very helpful to provide a list of abbreviations early on to improve readability.

Response: Thanks for the valuable comment. As suggested by the reviewer, a comprehensive list of abbreviations was added early in the manuscript to enhance clarity and ensure the accessibility of the main text.

Abbreviation			
NACs	Nitro-aromatic compounds	**BrC**	Brown carbon
NPs	Nitrophenol and its derivatives	**VOCs**	Volatile organic compounds
NCs	Nitrocatechol and its derivatives	**T**	Temperature
NSAs	Nitrosalicylic acids	**RH**	Relative humidity
DNPs	Dinitrophenol and its derivatives	**SSR**	Surface net solar radiation
4NP	4-nitrophenol	**PMF**	Positive matrix factorization
3M4NP	3-methyl-4-nitrophenol	**PCA**	Principal component analysi
2M4NP	2-methyl-4-nitrophenol	**ML**	Machine learning
2,6DM4NP	2.6-dimethyl-4-nitrophenol	**SHAP**	SHapley Additive exPlanation
4NC	4-nitrocatechol	**EML**	Ensemble machine learning
4M5NC	4-methyl-5-nitrocatechol	**BLH**	Boundary layer height
3M6NC	3-methyl-6-nitrocatechol	**WS_H**	Horizontal wind speed

3M5NC	3-methyl-5-nitrocatechol	**WS_V**	Vertical wind speed
5NSA	5-nitrosalicylic acid	**Sa**	Aerosol surface area
3NSA	3-nitrosalicylic acid	**EDTA**	ethylenediaminetetraacetic acid
2,4DNP	2,4-dinitrophenol	**BB**	Biomass burning
4M2,6DNP	4-methyl-2,6-dinitrophenol	**RF**	Random forest
CC	Coal combustion	**XGBoost**	Extreme gradient boosting
TE	Traffic emission	**LightGBM**	Light gradient boosting machine
GR	Gas-phase reaction	**MLP**	Multilayer perceptron
PE	Primary emission	**R^2**	Coefficient of determination
SF	Secondary formation	**MAE**	Mean absolute error
		RMSE	Root mean squared error

2. Line 95, there is a minor point about terminology. Calling Mount Lao a "mountain" site when it's only 166 m altitude is a bit confusing as Mount Tai is at 1534 m a.s.l. It might be worth clarifying that Mount Lao site is at a lower elevation (perhaps a foothill or a coastal mountain location) to avoid readers questioning if it truly represents a clean mountain background.

Response: Thanks for the comment. In response, we have revised the manuscript to characterize Mount Lao more precisely as a lower-elevation site situated in a coastal mountainous region. This clarification aims to provide a more accurate depiction of the site's geographic and environmental context, and to avoid potential ambiguity regarding its representativeness as a clean mountain background location.

Revised sentence in manuscript (Line 100–102):

"…and two mountain sites: Mount Tai (36.27° N, 117.10° E, 1,534 m a.s.l.), *a typical high-elevation background site;* and Mount Lao (36.15° N, 120.68° E, 166 m a.s.l.), *a lower-elevation site situated in a coastal mountainous region*."

3. Line 184, when talking about the performance of a model, it cannot be validated or verified as natural systems are never closed, it can only be evaluated.

Response: Thanks for the careful comment. Accordingly, we have revised the manuscript to replace "validation" to "evaluation" to ensure accuracy.

4. Linked to point 2, it's needed to ensure Figure (e.g., Figures 4-7) legends and captions fully describe what the plots represent. The caption may list the variables by name (or refer to a legend) so readers don't have to infer abbreviations (e.g., PE, SF, etc).

Response: Thanks for the comment. In accordance with your suggestion, we have revised the captions of Figure 4-7 to explicitly list the variables by name and, where applicable, refer to the corresponding legends to avoid the need for reader to infer abbreviations.

Revised caption of Figure 4 in manuscript (Line 305–313):

"Figure 4: (a) The ranking of the importance for all input variables *("CC": coal combustion; "TE": traffic emission; "T": temperature; "BB": biomass burning; "Sa": aerosol surface area; "GR": gas-phase reaction; "BLH": boundary layer height; "SSR": surface net solar radiation; "RH": relative humidity; "WS_H": horizontal wind speed; "WS_V": vertical wind speed)* calculated via SHAP algorithm (average absolute contribution), (b) the impacts of driving factors on variations of NACs from SHAP analysis during the whole sampling periods *("PE" and "SF" represent primary emissions and* *secondary formation, respectively)*, (c) SHAP summary plots for all samples with the shift in colour of the scatter plot from blue to red indicating an increase in driving factor values, and the relationships between the SHAP values and parameter values for (d) temperature *(T)*, (e) aerosol surface area *(Sa)*, (f) boundary layer height *(BLH)*, and (g) surface net solar radiation *(SSR)* with the right y axis corresponding to the frequency distribution of the measured variables."

Revised caption of Figure 5 in manuscript (Line 368–370):

"Figure 5: The absolute contributions of (a) meteorological conditions, (b) all factors, and (c) primary emissions on the variations of NACs in four seasons from SHAP analysis and box plots with the order of SHAP values for each driving factor in (d) spring, (e) summer, (f) autumn, and (g) winter. *"PE" and "SF" refer to primary emissions and secondary formation, respectively*."

Revised caption of Figure 6 in manuscript (Line 401–406):

"Figure 6: The impacts of primary emissions (PE), meteorological conditions, and secondary formation (SF) on the variations of (a) NPs, (b) NCs, and (c) NSAs from SHAP analysis and relative importance of (d) primary emissions *(including coal combustion (CC), traffic emission (TE), and biomass burning (BB))*, (e) meteorological conditions *(including temperature (T), boundary layer height (BLH), surface net solar radiation (SSR), relative humidity (RH), horizontal wind speed (WS_H), and vertical wind speed (WS_V)*), and (f) secondary formation (*including gas-phase reaction (GR) and heterogeneous reaction represented by aerosol surface area (Sa)*)."

Revised caption of Figure 7 in manuscript (Line 444–448):

"Figure 7: The combined contributions of sources (*i.e.*, coal combustion *(CC)*, traffic emission *(TE)*, and biomass burning *(BB)*), meteorological conditions (*i.e.*, *temperature (*T*)*, *boundary layer height (*BLH*)*, *surface net solar radiation (*SSR*)*, *vertical wind speed (*WS_V*)*, *horizontal wind speed (*WS_H*)*, and *relative humidity (*RH*)*), and secondary formation (*i.e.*, gas-phase reaction *(GR*) and heterogeneous reaction represented by the aerosol surface area *(Sa)*) to the variations of NACs in (a) urban, (b) rural, and (c) mountain areas."