Response to the reviewer:

We are grateful to the reviewers for their comprehensive evaluation and constructive feedback on our manuscript. Their insightful comments have led to substantial improvements in our work. Below, we provide point-by-point responses to each reviewer's comments, with the original comments in black and our responses in blue. All revisions made to the manuscript are highlighted accordingly.

Major comments:

1. *The title and introduction suggest that CNN-LSTM analysis supports PM_{10} source apportionment, whereas it is actually used only for replacing anomalous data. It should be clearly emphasized what is the added value of such pre-processing step for the subsequent source apportionment.*

Response: We sincerely appreciate your valuable suggestion. In this study, the CNN-LSTM model was applied exclusively during data preprocessing, where its primary function was to detect and replace PM_{10} outliers, thereby enhancing the overall dataset quality. We clarify that this step does not directly contribute to the PMF source apportionment analysis but serves to ensure the reliability of input data for subsequent modeling. By improving data integrity, the preprocessing step facilitates more robust identification of pollution sources and more accurate quantification of their contributions in the PMF results. To better reflect this methodological distinction, we have revised the manuscript's title to explicitly highlight the preprocessing role of CNN-LSTM. The updated title is: "Measurement Report: Unraveling PM_{10} Sources and Oxidative Potential Across Chinese Regions Based on CNN-LSTM Data Preprocessing and Receptor Model"

2. *The used PMF approach lacks a proper statistical description. Key aspects such as number of factors selection, residuals and error estimation should be presented, at least in the supplement section. Moreover, PMF analysis was conducted for only 4 sites, despite data being available for 12. Performing PMF across all sites could yield more*

robust results, highlighting the contrasts between urban, suburban, rural and remote areas. Otherwise, the authors should clearly demonstrate that the selected sites are fully representative of the different typologies.

Response: Thanks for these important and constructive comments regarding the PMF analysis. We have made substantial improvements to address these concerns.

(1) Statistical description of PMF approach

We completely agree that the PMF analysis requires more comprehensive statistical documentation. In the supplement, we have added in "S3 Source profiles from the PMF" that includes:

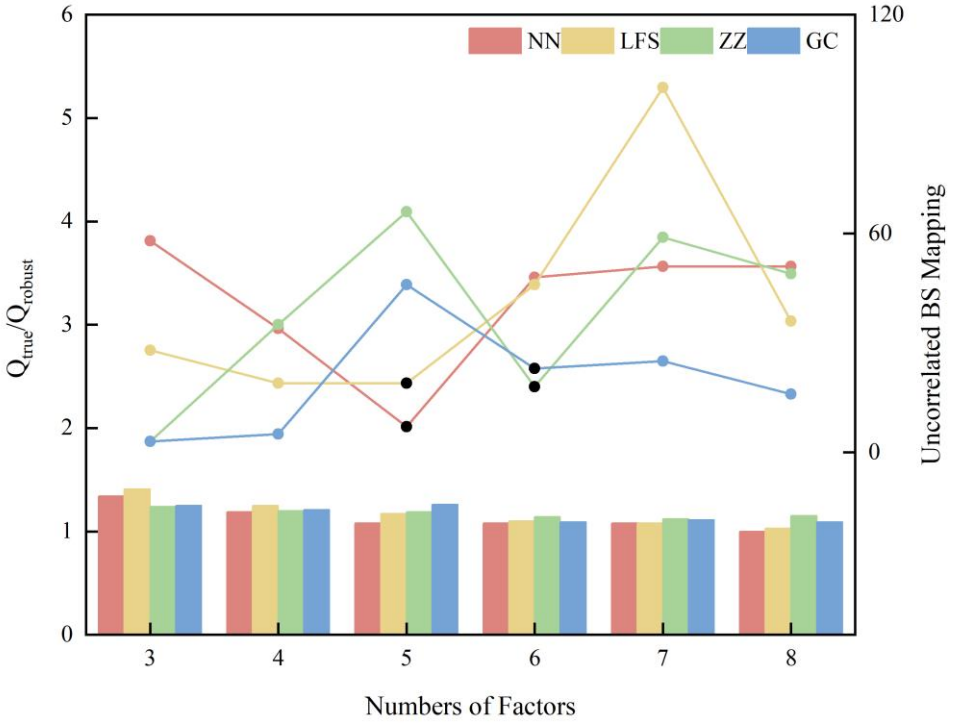

Figure S1. Changes in Q_{true}/Q_{robust} and uncorrelated bootstrap (BS) mapping calculated by the PMF model for varying numbers of factors (3-8). Black dots represent the optimum solution factors at each station.

Table S4. Summary of error estimation diagnostic with PMF at NN, LFS, ZZ and GC station.

Diagnostics		NN	LFS	ZZ	GC
	Number of base run	20	20	20	20
	Q_{robust}	2338	2658.86	2043.86	2066.75
	Q_{true}	2521.1	3160.89	2341.26	2249.16
	Q_{true}/Q_{robust}	1.08	1.18	1.14	1.09
DISP	% dQ	< 0.1 %	< 0.1 %	< 0.1 %	< 0.1 %
	swaps	0	0	0	0
BS	Number of run	100	100	100	100
	Extra modeling uncertainty (%)	0	0	0	0
	Min. Correlation R-Value	0.6	0.6	0.6	0.6
	Mapping	Dust: 94% Biomass burning: 99% Others: 100%	Secondary aerosol: 88% Traffic: 97% Agricultural activities:95% Others: 100%	Coal combustion: 87% Secondary aerosol: 97% Agricultural activities: 98% Others: 100%	Agricultural activities: 90% Coal combustion: 85% Others: 100%
BS-DISP	#of Cases Accepted	89	78	80	71
	% cases with swaps	9%	18%	15%	25%

(2) PMF analysis site representativeness

The selection of these four sites was based on the following considerations: First, these four sites possess good representativeness in terms of geographical location and environmental characteristics. Second, considering the data quality requirements for PMF model analysis, we selected sites with high data completeness and reliability. Specifically, Nanning represents the southern inland urban environment (located relatively close to the coast), Longfengshan is located in the northeastern region and serves as a background site reference, Zhengzhou, as an important transportation hub, exhibits typical suburban characteristics, and Gucheng is situated in the key pollution area of the Beijing-Tianjin-Hebei region, representing the rural environmental characteristics of this area. We have added detailed explanations for the site selection criteria in the revised manuscript (**Lines 451-458**): "This study employed the PMF model to conduct a detailed analysis of PM_{10} sources at four representative sites selected based on distinct geographical and environmental characteristics. Specific tracers used in PM_{10} source apportionment in this study are shown in **Table S3**. The selection criteria considered regional representativeness, pollution characteristics, and geographical diversity across China. The selected sites include: NN, an urban site in southern China with coastal proximity; LFS, a remote site located in the northeastern region of Heilongjiang Province; ZZ, a suburban site serving as a major transportation hub in central China; and GC, a rural site situated in the heavily polluted Beijing-Tianjin-Hebei region. These four sites collectively represent different pollution source characteristics and regional environmental conditions, enabling a comprehensive understanding of PM_{10} source apportionment across diverse geographical and climatic zones in China."

3. *The discussion about OP might be improved. For example, in section 3.3, the link between the chemical composition and OP is not addressed. This could help to understand the elevated OP level observed in Chengdu during summer or in Jinsha during spring. Conducting PMF at these sites might provide further insight. Additionally, the sources contributions to OP were obtained from the PMF analyses,*

but the reported percentages should be clarified. Do they account for the explained variability of the model? What are the contributions to the unexplained part? Moreover, given that OP concentrations are often known to be associated with biomass burning (Daellenbach et al., 2020), the absence of any such contribution for LFS and ZZ is surprising and should be discussed.

Daellenbach, K.R., Uzu, G., Jiang, J., Cassagnes, L.-E., Leni, Z., Vlachou, A., Stefenelli, G., Canonaco, F., Weber, S., Segers, A., Kuenen, J.J.P., Schaap, M., Favez, O., Albinet, A., Aksoyoglu, S., Dommen, J., Baltensperger, U., Geiser, M., El Haddad, I., Jaffrezo, J.-L., Prévôt, A.S.H., 2020. Sources of particulate-matter air pollution and its oxidative potential in Europe. Nature 587, 414–419. https://doi.org/10.1038/s41586-020-2902-8

Response：We appreciate the reviewer's insightful suggestions.

(1) Correlation analysis between OP and chemical composition at CD and JS

Thank you for your valuable comment regarding the discussion of OP in section 3.3. We appreciate your suggestion to strengthen the link between chemical composition and OP. Following your recommendation, we have conducted Spearman correlation analysis between OP and other chemical components at both Chengdu in summer and Jinsha in spring. As shown in **Figure R1**, the results reveal strong positive correlations ($r > 0.5$, $P <= 0.001$) between OP and various chemical constituents at both locations. Based on these correlation findings, we believe that the elevated OP levels observed in Chengdu during summer and in Jinsha during spring can be attributed to the complex mixture of pollutants at these sites. The strong correlations suggest that co-pollution from multiple sources contributes synergistically to the enhanced oxidative potential. The seasonal variations in OP likely reflect the combined effects of meteorological conditions, emission patterns, and chemical transformation processes that influence the concentration and composition of these reactive species.

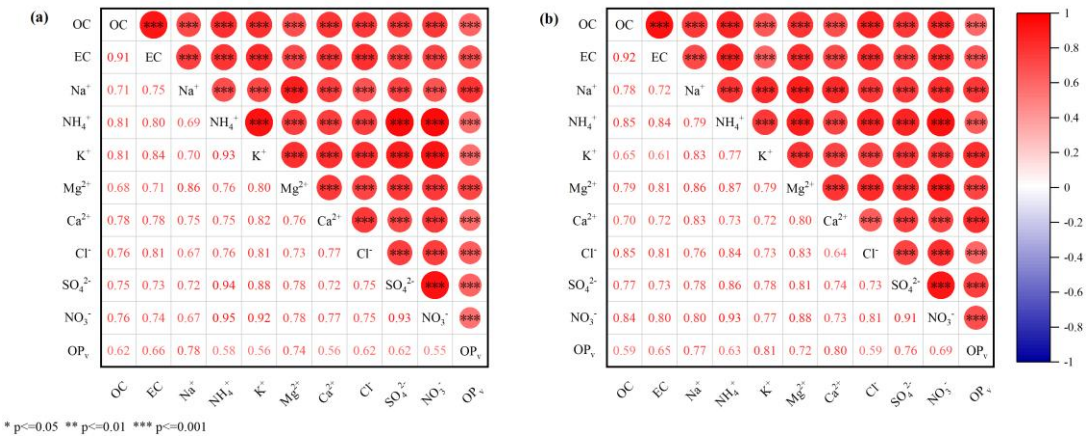

Figure R1. Spearman correlation coefficients between OP_v and the chemical species observed at (a) CD in summer and (b) JS in spring.

(2) Explained variability of the model

Thank you for your insightful comment regarding the PMF analysis and source contributions to OP. We acknowledge that providing detailed metrics for explained variability would enhance the interpretation of our PMF results. However, we would like to clarify our approach following established practices in the field. It should be noted that OP_v was set as "weak" in the PMF input parameters, which resulted in a lower r^2 value compared to other chemical components. As noted in numerous PMF studies, the primary objective is to reproduce the observed mass concentrations with reasonable reconstruction accuracy. The main criterion for model validation is achieving satisfactory mass reconstruction with $r^2 > 0.8$, which indicates that the identified sources can adequately explain the observed variability in the dataset. In our study, the PMF results demonstrate that most chemical components achieved r^2 values above 0.8, suggesting robust source identification and apportionment. The reported percentages for source contributions to OP represent the relative importance of each identified source in explaining the observed OP values based on this validated model framework.

We believe that the strong mass reconstruction performance ($r^2 > 0.8$ for most components) provides confidence in the reliability of our source apportionment results and the subsequent OP source contributions. This approach is consistent with standard practices in PMF applications for aerosol source apportionment studies.

(3) Contribution of biomass combustion to no OP at LFS and ZZ

Thank you for this valuable observation and for bringing the Daellenbach et al. (2020) study to our attention, which indeed highlights the typical association between OP and biomass burning. This reference provides important context for interpreting our results.

However, our results do show that the LFS site exhibits no contributions of OP in the biomass burning factor, which may be attributed to specific local combustion conditions. High-temperature complete combustion leads to oxidative decomposition of organic compounds, thereby reducing the oxidative potential of particulate matter despite contributing to PM_{10} levels. We have added in **Lines 592-597**: "The biomass burning at LFS sites, despite contributing to PM_{10}, lacks oxidative activity, which may be related to different combustion conditions or degrees of combustion completeness leading to reduced generation of oxidatively active components. Differences in combustion temperature and oxygen supply conditions significantly affect the formation and transformation of organic compounds, under high-temperature complete combustion conditions, most organic compounds are oxidatively decomposed, thereby reducing the oxidative activity of particulate matter (Tuet et al., 2017)."

For the ZZ site, upon further examination of the PMF results in conjunction with local emission characteristics and considering the reviewer's insight, we have refined our factor identification from biomass burning to industrial emission, which better aligns with the observed chemical profile and local source conditions at this location. The manuscript has been revised accordingly with additional discussion. We have corrected in **Lines 509-512 and Lines 664-669**.

Lines 509-512: "The first factor had high contribution of K^+ (21.7%) and Cl^- (83.9%), but low contribution of OC (4.6%) and EC (2.8%), possibly indicating the influence of industrial emissions, such as food manufacturing, cement manufacturing, salt production, or industrial activities involving potassium chloride compounds (Yin et al., 2019; Seo et al., 2019), with a contribution proportion of approximately 9.5%."

Lines 664-669: "Industry

The PMF results revealed that industrial emissions at ZZ were dominated by Cl^-

and K^+ with negligible contributions from OC and EC. OC typically serves as the primary contributor to particle oxidative activity through redox-active organic species, including quinones and phenolic compounds that can participate in electron transfer reactions and generate reactive oxygen species (Libalova et al., 2018; Jiang and Jang, 2018). The absence of organic carbon compounds provides a mechanistic explanation for the zero oxidative potential observed in this source profile."

4. *The authors are encouraged to carefully revise the text as there is no mention of supplementary materials (Tables S1, S2, S3) in the main text.*

Response: **Thank you for your careful review.** We have now revised the manuscript to include appropriate citations to these supplementary tables at the relevant sections in **Line 271** (Table S1), **Line 437** (Table S2), and **Line 461** (Table S3).

Specific comments:

5. *All acronyms must be clearly defined before being used for the first time, even in the abstract, and not re-defined after it (ex: lines 52, 211, 238, 337, 361). Please revise it carefully.*

Response: Thanks for your suggestion. We have systematically reviewed the manuscript and standardized the use of acronyms. All acronyms are now properly defined upon their first appearance and not re-defined afterwards.

6. *Line 21: Please indicate these "highest OP values".*

Response: We are grateful for your constructive feedback. We have revised **Line 21** to include the precise numerical values. The updated manuscript reads: "Urban sites showed the highest OP values (0.61 ± 0.21 nmol $H_2O_2 \cdot m^{-3}$), with significantly higher PM_{10} concentrations in northern regions compared to southern ones ($p<0.05$)."

7. *Line 48: spaces are missing several times in the text: lines 48, 51, 54, 74, 86, 165, 239, 324, 371.*

Response: Thanks for pointing out these formatting issues. We have corrected all the missing spaces.

8. *Lines 60-61: These health effects are already stated earlier (line 46).*

Response: **Thank you for your careful review.** The health effects mentioned in Lines 60-61 were indeed already stated earlier in Line 46. We have deleted this duplicated information to eliminate redundancy and enhance the clarity of the text.

9. *Lines 66-81: It is mentioned here that to face challenges in source apportionment methods, deep learning technics can be used to deal with anomalous data. However, the CNN/LTSM model was only applied for correcting the total PM_{10} data, and not the data inputs used for PMF. This should be clarified.*

Response: **Thanks for pointing out this important issue.** We agree that our original statement could have been misleading. In our study, the CNN-LSTM model is specifically used to handle missing data and anomalies to improve data quality. We have revised the manuscript to make this clear: (1) the CNN-LSTM approach is applied for data preprocessing to ensure high-quality, complete datasets; (2) this data quality improvement step is essential before conducting source apportionment analyses; and (3) we have removed the misleading reference to traditional source attribution methods in this context. The revised text now clearly distinguishes between data quality improvement (using CNN-LSTM) and subsequent source apportionment analyses, providing a more accurate representation of our methodology. The detailed revisions can be found in **Lines 63-80** of the revised manuscript.

"However, an accurate assessment of the health risks associated with PM_{10} requires an accurate analysis of its sources and chemical compositions. High-quality, complete datasets are essential for reliable source apportionment and subsequent risk assessment. Environmental monitoring data often contain missing values and anomalies due to instrument malfunction, maintenance periods, or extreme weather conditions, which can significantly affect the accuracy of subsequent analyses. In recent years, with the rapid development of deep learning technology, its application in handling environmental data quality issues has received increasing attention. Deep learning models, particularly the combination of Convolutional Neural Networks (CNN) and Long Short-Term Memory networks (LSTM), have demonstrated significant

advantages in identifying and correcting anomalies and filling missing values in time series environmental data. CNNs effectively extract spatial features, while LSTMs excel at capturing long-term dependencies in time series (Huang and Kuo, 2018; Li et al., 2020). This hybrid model not only identifies anomalies, but also improves data completeness and reliability by predicting and replacing anomalous or missing values (Lee et al., 2019; Qin et al., 2019). Compared with traditional machine learning methods, CNN-LSTM models show superior performance in several evaluation metrics, such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE) (Huang and Kuo, 2018; Yang et al., 2020a; Li et al., 2020). CNN-LSTM models retain significant value in processing atmospheric particulate matter data for data quality improvement. Their spatial feature extraction capabilities effectively identify and correct anomalies caused by instrument malfunction or local pollution events, thereby improving data quality (Zhang and Zhou, 2023). Through training and learning, CNN-LSTM models can effectively predict and fill missing data, providing a high-quality data foundation for subsequent source apportionment and risk assessment analyses (Li et al., 2020; Yang et al., 2020a)."

10. *Line 95: This is not correct, no MLR model was used in this study for evaluating the PM_{10} sources contributions to OP.*

Response: Thanks for catching this mistake. We apologize for this oversight that was not corrected in time. **Lines 93-95** have been revised to: "Finally, based on the PMF results, we calculated the OP per unit mass of PM_{10} (OP_m) to investigate the intrinsic toxicity of different emission sources."

11. *Lines 190-191: I'm curious how relevant is this criterion. Metals such as Al, Fe, Si, K are not measured and can significantly affect the total PM_{10} mass concentrations during some specific periods, e.g. during desert dust events. I would recommend some more justification.*

Response：Thank you for raising this important point regarding the potential influence of unmeasured metals on PM_{10} mass concentrations.

We acknowledge that metals such as Al, Fe, Si, and K can significantly affect total PM_{10} mass concentrations during desert dust events. However, our analysis shows that such conditions were extremely rare in our study. Among all 1237 samples collected across all monitoring sites, only 20 samples (approximately 1.6%) were collected during blowing dust or haze events. Given this very low frequency of dust-related events, we believe our current approach is appropriate for this particular dataset. The vast majority of samples (98.4%) were collected under normal atmospheric conditions where the measured chemical components adequately represent the PM_{10} composition for our analysis objectives.

Nevertheless, we also suggest that future studies in regions with more frequent dust events should consider including crustal element measurements for more comprehensive mass reconstruction.

12. *Line 216: n, m and p are not described.*

Response: Thank you for the valuable observation. We acknowledge that the variables n, m, and p in the PMF mathematical model were not adequately defined in the original manuscript. We have revised the text to include clear definitions of these parameters. **Lines 215-216** have been revised to: "Where X is the observation data matrix (n×m), G is the factor contribution matrix (n×p), F is the factor profile matrix (p×m), and E is the residual matrix. Here, n represents the number of samples, m represents the number of chemical species, and p represents the number of factors."

13. *Line 231: Is it "PM_{10} concentration measurements" or "reconstructed PM_{10} concentrations"? If PM_{10} were measured at all sites then how?*

Response: Sorry for the confusion. The "PM_{10} concentration measurements" refers to the PM_{10} concentration data available in our dataset. The CNN-LSTM model was trained using PM_{10} concentrations from non-outlier datasets along with their corresponding eleven chemical constituent concentrations. The trained model was then used to predict PM_{10} concentrations for outlier datasets based on their chemical constituent concentrations as input features. **Lines 230-237** have been revised to: "The

CNN-LSTM model was trained using non-outlier datasets consisting of PM_{10} concentration measurements and their corresponding eleven chemical constituents, including OM, EC, Na^+, NH_4^+, K^+, Ca^{2+}, Mg^{2+}, F^-, Cl^-, NO_3^-, and SO_4^{2-}. To ensure the integrity of the data quality, outlier elimination was performed based on the sum of the chemical components. After the outlier screening process described in Section 2.3.3, 471 non-outlier datasets meeting the quality criteria were retained for model training and evaluation, with 85% allocated to the training set and 15% to the test set. The trained CNN-LSTM model was then used to predict PM_{10} concentrations for the 766 outlier datasets by using their eleven chemical constituent concentrations as input features, with the predicted values replacing the original outlier measurements to maintain data completeness."

14. *Line 232: Please justify the use of OM =1.4*OC.*

Response: Thank you for asking for justification of this conversion factor. We sincerely apologize for the confusion in our manuscript. Due to our oversight during the final manuscript preparation, the text incorrectly stated OM =1.4*OC in **Line 336**, but we actually used OM = 1.2*OC throughout all our calculations and data analysis. We have carefully reviewed our analysis code and confirmed that 1.2 was consistently applied. This discrepancy has now been corrected in the revised manuscript.

The conversion factor of 1.2 was specifically chosen based on established literature recommendations for urban atmospheric aerosols. White and Roberts suggested that 1.2-1.4 represents a conservative range commonly used in mass balance analysis (White and Roberts, 1977). We selected 1.2 as the more conservative end of this range to avoid overestimating organic matter contributions in our PMF source apportionment analysis. Therefore, we believe that the use of OM = 1.2*OC represents a reasonable and conservative approach for our urban PM_{10} source apportionment study.

Reference

White, W. H. and Roberts, P. T.: On the nature and origins of visibility-reducing aerosols in the los angeles air basin, Atmos. Environ., 11, 803-812, https://doi.org/10.1016/0004-6981(77)90042-7, 1977.

15. *Lines 232-235: This is already stated earlier, and needs to be better explained.*

Response: Thanks for your kind reminder. We removed the duplicated description of the outlier elimination criteria that was previously mentioned in Section 2.3.3 and enhanced the explanation to clarify that the trained CNN-LSTM model was applied to predict PM_{10} concentrations for the 766 outlier datasets, with these predicted values used to replace the original outlier measurements. **Lines 231-237** have been revised to: "To ensure the integrity of the data quality, outlier elimination was performed based on the sum of the chemical components. After the outlier screening process described in Section 2.3.3, 471 non-outlier datasets meeting the quality criteria were retained for model training and evaluation, with 85% allocated to the training set and 15% to the test set. The trained CNN-LSTM model was then used to predict PM_{10} concentrations for the 766 outlier datasets by using their eleven chemical constituent concentrations as input features, with the predicted values replacing the original outlier measurements to maintain data completeness."

16. *Line 236: What percentage of the total data do the outliers represent?*

Response: Thank you for important question. In our dataset, outliers represent approximately 62% of the total data points. Specifically, we identified 766 outliers out of 1237 total observations.

The relatively high percentage of outliers in our PM_{10} dataset can be attributed to several methodological and technical factors inherent to atmospheric monitoring:

Offline monitoring limitations: Our chemical composition analysis, while strictly following meteorological bureau experimental standards (QXT 70-2007 for carbonaceous components and HJ 799-2016, HJ 800-2016 for water-soluble ions), relies on offline monitoring using quartz filter membranes. These filters are susceptible to damage or scaling-off during transportation and storage processes, and are extremely sensitive to environmental humidity changes, which may lead to abnormally high offline data readings. Therefore, we primarily employed online monitoring equipment for PM_{10} concentration measurements to minimize these offline-related uncertainties.

Instrumental heterogeneity: The 12 monitoring stations employed online

monitoring equipment based on different measurement principles, including light scattering method, beta-ray attenuation method, and tapered element oscillating microbalance method, which cannot be completely standardized. Each method has inherent limitations: (1) light scattering methods are easily affected by particle physicochemical characteristics such as size distribution, chemical composition, and optical properties (Chen et al., 2018; Koestner et al., 2020; Jiang et al., 2021); (2) tapered element oscillating microbalance methods are insufficient in measuring semi-volatile substances (Green and Fuller, 2006) and sensitive to humidity (Li et al., 2012); (3) beta-ray attenuation methods are also sensitive to environmental humidity fluctuations (Shukla and Aggarwal, 2022).

Environmental factors: PM_{10} mass concentration measurements are subject to significant uncertainties due to these combined technical limitations and varying environmental conditions across different monitoring sites, including temporal variations in meteorological parameters and spatial heterogeneity in local emission sources.

Given the higher reliability of our chemical composition data (which follows established national standards and aligns with comprehensive national-scale studies), we employed the CNN-LSTM deep learning model specifically to predict and replace these PM_{10} anomalous values, thereby improving overall data quality and analytical accuracy.

Reference

Chen, D., Liu, X. W., Han, J. K., Jiang, M., Xu, Y. S., and Xu, M. H.: Measurements of particulate matter concentration by the light scattering method: Optimization of the detection angle, Fuel Process. Technol., 179, 124-134, http://dx.doi.org/10.1016/j.fuproc.2018.06.016, 2018.

China Meteorological Administration: Determination of elemental carbon and organic carbon in atmospheric aerosols—thermal-optical analysis method, QXT 70-2007, https://www.cma.gov.cn/zfxxgk/gknr/flfgbz/bz/202209/t20220921_5097811.html (last access: 28 April 2025), 2007.

Green, D. and Fuller, G. W.: The implications of tapered element oscillating microbalance (TEOM) software configuration on particulate matter measurements in the UK and Europe, Atmos.

Environ., 40, 5608-5616, http://dx.doi.org/10.1016/j.atmosenv.2006.04.052, 2006.

Jiang, M., Liu, X. W., Han, J. K., Zhou, Z. J., and Xu, M. H.: Measuring particle size and concentration of non-spherical particles by combined light extinction and scattering method, Measurement, 184, http://dx.doi.org/10.1016/j.measurement.2021.109911, 2021.

Koestner, D., Stramski, D., and Reynolds, R. A.: Assessing the effects of particle size and composition on light scattering through measurements of size-fractionated seawater samples, Limnol. Oceanogr., 65, 173-190, http://dx.doi.org/10.1002/lno.11259, 2020.

Li, Q. F., Lingjuan, W. L., Liu, Z. F., and Heber, A. J.: Field evaluation of particulate matter measurements using tapered element oscillating microbalance in a layer house, J. Air Waste Manage. Assoc., 62, 322-335, http://dx.doi.org/10.1080/10473289.2011.650316, 2012.

Ministry of Ecology and Environment of the People's Republic of China: Ambient air—Determination of water-soluble anions (F^-, Cl^-, Br^-, NO_2^-, NO_3^-, PO_4^{3-}, SO_3^{2-}, SO_4^{2-}) in particulate matter—Ion chromatography, HJ 799-2016, https://www.mee.gov.cn/ywgz/fgbz/bz/bzwb/jcffbz/201605/t20160519_337906.shtml (last access: 28 April 2025), 2016.

Ministry of Ecology and Environment of the People's Republic of China: Ambient air—Determination of water-soluble cations (Li^+, Na^+, NH_4^+, K^+, Ca^{2+}, Mg^{2+}) in particulate matter—Ion chromatography, HJ 800-2016, https://www.mee.gov.cn/ywgz/fgbz/bz/bzwb/jcffbz/201605/t20160519_337907.htm (last access: 28 April 2025), 2016.

Shukla, K. and Aggarwal, S. G.: A Technical Overview on Beta-Attenuation Method for the Monitoring of Particulate Matter in Ambient Air, Aerosol Air Qual. Res., 22, http://dx.doi.org/10.4209/aaqr.220195, 2022.

17.*Line 246: Please remove "W".*

Response：Thanks for your attention to detail. We have eliminated "W" from **Line 247**.

18.*Line 247: Be consistent with significative digits of Table 2.*

Response：Thank you for pointing out the inconsistency in significant digits in

Table 2. We have carefully revised the table to ensure all numerical values are presented with consistent significant digits. **Lines 244-248** have been revised to: "The model was evaluated on both the training and test sets after completion of training, with results presented in Table 2 and Figure 5. As shown in Figure 5 (a), the training process converged effectively, with the loss function decreasing steadily and stabilizing at approximately 0.0007, indicating successful model optimization without overfitting. For the training set, the CNN-LSTM model achieved a MAE of 6.6614 μg·m^{-3}, a RMSE of 8.7162 μg·m^{-3}, and a R^2 of 0.9670. When evaluated on the test set, the model demonstrated an MAE of 12.6705 μg·m^{-3}, a RMSE of 17.4965 μg·m^{-3}, and an R^2 of 0.8840."

Table 2. Comparison of MAE, RMSE, and R^2 among different models.

Model Type	MAE（μg·m^{-3}）	RMSE（μg·m^{-3}）	R^2
Linear Regression	12.6852	17.8804	0.8028
RF	14.6494	20.0135	0.8482
KNN	15.6263	24.2398	0.8135
CNN-LSTM	12.6705	17.4935	0.8840

Additionally, to strengthen the comprehensiveness of our evaluation, we have added comparisons with conventional gap-filling techniques in the revised manuscript. Specifically, we have included a new subsection (Section 3.1.1) titled "Comparison with Conventional Gap-filling Techniques" in **Lines 243-273**, where we compare our CNN-LSTM model with Random Forest (RF), K-Nearest Neighbors (KNN), and Linear Regression methods.

19.*Line 260: Results of Figure 5 would need more discussion. Does the slight deviation occur for the same sites?*

Response: Thanks for your valuable suggestions. We have revised the manuscript to provide a more comprehensive analysis of Figure 5. Specifically, the updated version includes a detailed discussion on the convergence of the loss function (Figure 5a), noting that it stabilizes around 0.0007, indicating successful model optimization without overfitting. The relevant modifications can be found in **Lines 244–248**. "The

model was evaluated on both the training and test sets after completion of training, with results presented in **Table 2** and **Figure 5**. As shown in **Figure 5 (a)**, the training process converged effectively, with the loss function decreasing steadily and stabilizing at approximately 0.0007, indicating successful model optimization without overfitting. For the training set, the CNN-LSTM model achieved a MAE of 6.6614 $\mu g \cdot m^{-3}$, a RMSE of 8.7162 $\mu g \cdot m^{-3}$, and a coefficient of determination (R^2) of 0.9670. When evaluated on the test set, the model demonstrated an MAE of 12.6705 $\mu g \cdot m^{-3}$, a RMSE of 17.4965 $\mu g \cdot m^{-3}$, and an R^2 of 0.8840."

And we also thanks for insightful question regarding site-specific deviation patterns. Unfortunately, due to the limited size of our test dataset, when we attempt to analyze deviations by individual sites, each site contains fewer than 10 data points in the test set, making it difficult to conduct a reliable site-specific evaluation of model performance. However, to address concerns about the model's generalization capability and site-specific performance, we have added Section 3.1.2 "Leave-One-Site-Out Cross-Validation" to the CNN-LSTM model results discussion.

20.*Line 270: Change with "Due to its location".*

Response: Thank you for pointing this out. We have corrected it in **Line 311** as follows: "Due to its location in an arid region, Dunhuang is likely influenced by dust storm events, as evidenced by higher concentrations of crustal elements such as Ca^{2+} (Yu et al., 2020)."

21.*Line 305: Figure 6 is not referenced in the main text, and doesn't provide any additional information beyond what is presented in Table 3. I would suggest to remove it. How was the "unknown components" part determined?*

Response: Thank you for pointing out this problem in the manuscript. We have removed Figure 6 from the revised manuscript and renumbered the subsequent figures accordingly.

The determination of "unknown components" was calculated as the difference between the total PM_{10} mass concentration, as measured by online monitoring

equipment, and the sum of all identified chemical components (OM, EC, NH_4^+, K^+, Na^+, Ca^{2+}, Mg^{2+}, F^-, Cl^-, NO_3^-, SO_4^{2-}). Specifically:

Unknown components($\mu g \cdot m^{-3}$)=PM_{10} total mass - Σ identified chemical components

22.*Line 307: Providing the standard deviations to these values would help assess the significance of the observed differences.*

Response： **Thank you for this valuable suggestion.** We agree that including standard deviations would help readers to better assess the significance of the observed differences. We have revised the manuscript to include standard deviations for all relevant values mentioned in Line 307 (now **Lines 344-345** in the revised manuscript). The specific changes are as follows: "The annual mean PM_{10} concentrations for urban, rural, suburban, and remote sites were 59.99 ± 29.38 $\mu g \cdot m^{-3}$, 62.88 ± 27.58 $\mu g \cdot m^{-3}$, 85.43 ± 39.43 $\mu g \cdot m^{-3}$, and 45.12 ± 14.67 $\mu g \cdot m^{-3}$, respectively."

23. *Figure 7: the axis labels are difficult to read.*

Response： **We are extremely grateful for pointing out this problem.** We agree that the axis labels need improvement for better readability. We have revised Figure 6 (originally Figure 7) with larger, clearer axis labels.

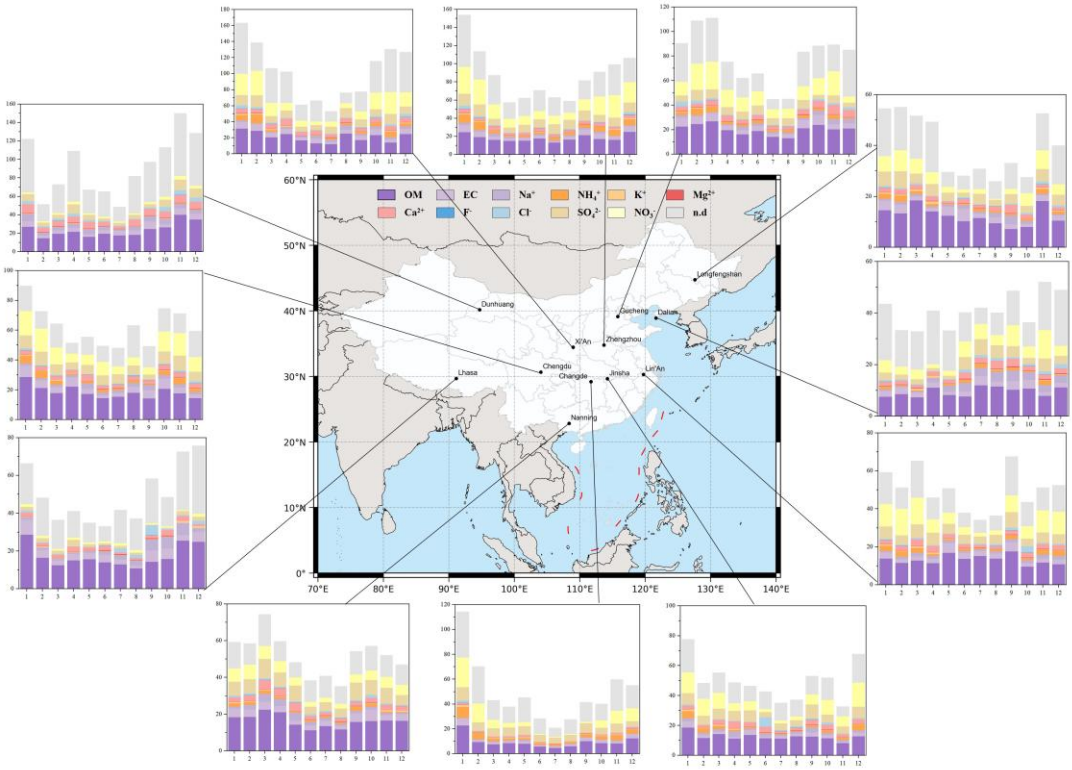

Figure 6. Stacked Representation of Monthly averaged PM_{10} Concentrations and Chemical Composition（$\mu g \cdot m^{-3}$）across Chinese Regions, Including Unknown Components from June 2022 to May 2023. (n.d: Unknown Components). The map base is from the Ministry of Natural Resources' Standard Map Service, review number GS (2019)1822.

24.*Line 321: Change "Seasonal" by "Monthly".*

Response： **Thanks for your kind reminder.** We have replaced "Seasonal" with "Monthly" in **Line 358** according to your suggestion.

"Monthly variations in PM_{10} concentrations are shown in Figure 6."

25.*Line 340: What does the "Chinese atmospheric particulates" term refer to?*

Response: **Thank you for your careful review.** We have revised the sentence to be more specific about the type of atmospheric particulates being referenced. The term "Chinese atmospheric particulates" has been changed to "PM_{10} across China" to clearly indicate that we are referring to carbonaceous components in PM_{10} particles specifically. The revised sentence in **Lines 376-378**: "All three functional site types showed the lowest concentrations in summer and the highest in winter, consistent with previous studies confirming the widespread winter-high and summer-low seasonal pattern of carbonaceous components in PM_{10} across China (Tian et al., 2013)."

26.*Figure 8: Y-Axis units are unclear and not displayed in the figure.*

Response: Thank you for the important observation. We acknowledge that the Y-axis units in Figure 7 (originally Figure 8) were not properly displayed. We have corrected the figure to include clear Y-axis unit labels: (μg·m^{-3}) for Figure 7 (a) and (nmol H_2O_2·m^{-3}) for Figure 7 (b).

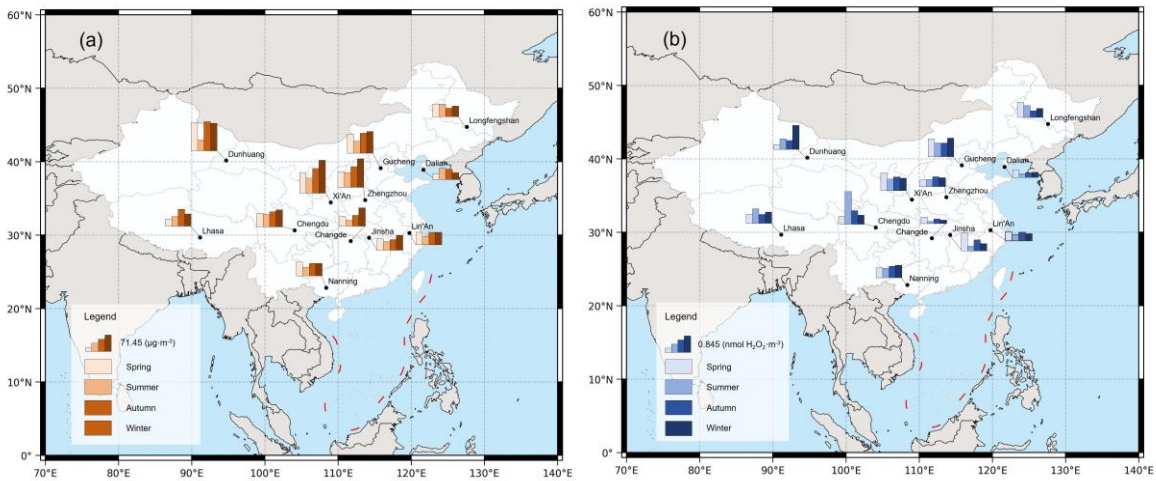

Figure 7. Seasonal variations of (a) PM_{10} concentrations（μg·m^{-3}）and (b) OP_v (nmol H_2O_2·m^{-3}) across different regions of China. The map bases are from the Ministry of Natural Resources' Standard Map Service, review number GS (2019)1822.

27.*Line 375: Change "northern Chinese sites" by "northern China".*

Response: Thanks for your kind reminder. We have changed **Lines 413-414** as follows: "As shown in Figure 7(a) and (b), sites located in northern China exhibited significantly elevated PM_{10} concentrations and OP_v levels during the autumn and winter seasons."

28.*Lines 375-380: Unclear, are the factors presented to explain the high PM_{10} and OP levels in winter and autumn applicable only to the northern Chinese sites?*

Response: We are sorry for confusion. We agree that the original manuscript did not adequately clarify whether these factors apply only to northern China. We have revised this paragraph from two perspectives: climate differences between northern and southern China and distinct heating practices, to illustrate that these factors are unique

regional characteristics of northern China, forming a sharp contrast with southern China. We have revised **Lines 413-421** as follows: "As shown in Figure 7(a) and (b), sites located in northern China exhibited significantly elevated PM_{10} concentrations and OP_v levels during the autumn and winter seasons. This phenomenon in northern Chinese sites can be attributed to several factors unique to northern China's regional characteristics. Firstly, the widespread reliance on coal-based central heating systems and biomass burning for residential heating in northern China during the heating season (typically from November to March) (Liu et al., 2017b; Li et al., 2017) sharply contrasts with southern China where heating demand is minimal due to milder winter temperatures. In addition, northern China's continental climate creates more severe winter meteorological conditions, including prolonged periods of low wind speeds, frequent temperature inversions, and significantly reduced atmospheric boundary layer heights compared to the more temperate conditions in southern regions, which severely inhibited pollutant dispersion (Li et al., 2017)."

29.*Line 396: How are the northern/southern regions defined?*

Response: Thank you for important question regarding our regional classification. We have defined the northern and southern regions of China based on the traditional geographical boundary of the Qinling Mountains-Huaihe River line, which is widely recognized in Chinese geographical and environmental studies as the natural dividing line between northern and southern China. Specifically, monitoring sites located north of this boundary are classified as northern regions, while sites south of this line are classified as southern regions.

It should be noted that Lhasa, due to its unique location on the Tibetan Plateau with distinct high-altitude characteristics, was not included in either the northern or southern regional analysis and was treated separately as a plateau region. This classification approach ensures that our regional comparisons reflect the major climatic and geographical differences across China while appropriately accounting for the unique environmental conditions of the Tibetan Plateau.

And we have added a detailed explanation of our regional classification in the

Supplement S2. The detailed site classification is provided in **Supplementary Table S2**, where we have included the following clarification: "China was divided into northern and southern regions using the Qinling Mountains-Huaihe River line as the boundary. This line represents the traditional geographical and climatic divide in China. Sites north of this boundary were classified as northern regions, while sites south were classified as southern regions. Lhasa, located on the Tibetan Plateau, was treated as a separate plateau region due to its distinct high-altitude environmental characteristics. The detailed site classification is provided in Table S2."

Table S2. The geographical division corresponding to the station.

Geographic region	Station name
Northern sties	LFS, DL, GC, DH, XA, ZZ
Southern sites	LA, CD, JS, NN, CHD

30. *Line 412: What does "more complex" mean?*

Response: We feel sorry for the confusion. We have revised the sentence in **Lines 463-465**: "The ZZ site showed industry, agricultural activities, traffic, coal combustion, dust, and secondary aerosols as the main sources. PM_{10} sources in GC are biomass burning, traffic, dust, agricultural activity emissions, secondary aerosols, and coal combustion."

31.*Line 415: Figure 11 reference is mentioned before the Figure 10.*

Response: Thank you for pointing out the issue with the figure numbering. We have reordered the figures to ensure they are cited in the correct order. Figure 11 is now Figure 9, while Figure 10 remains unchanged. The figures now follow the correct citation order in the text.

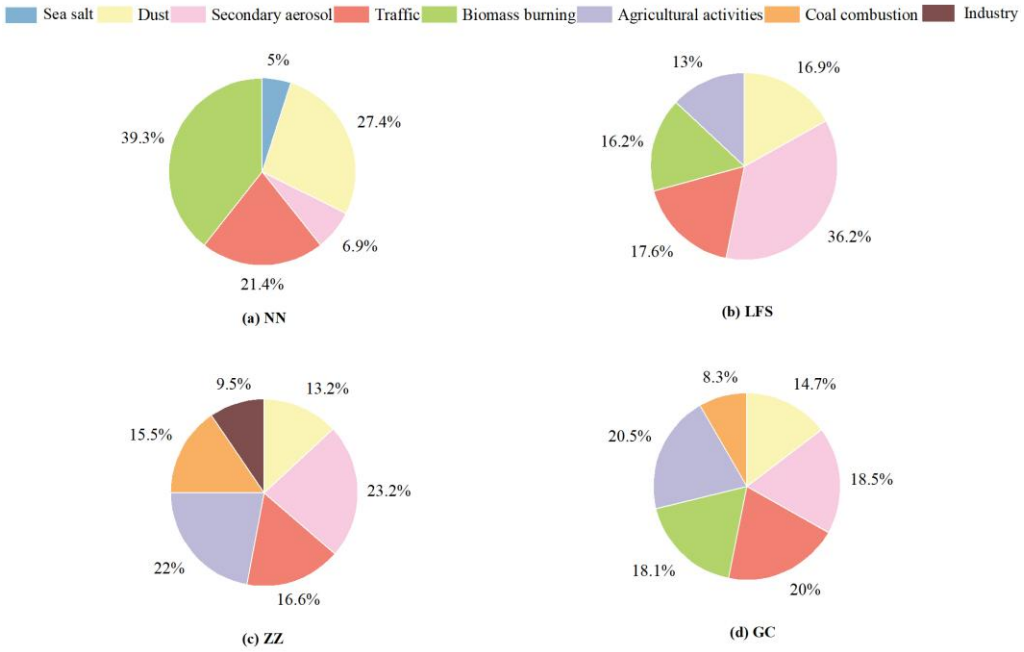

Figure 9. The contributions of Biomass burning, Traffic, Dust, Secondary aerosol, Sea salt, Agricultural activities, Coal combustion, and Industry to the atmospheric concentration of PM_{10} mass (%) as derived by PMF modelling at NN, LFS, ZZ, and GC.

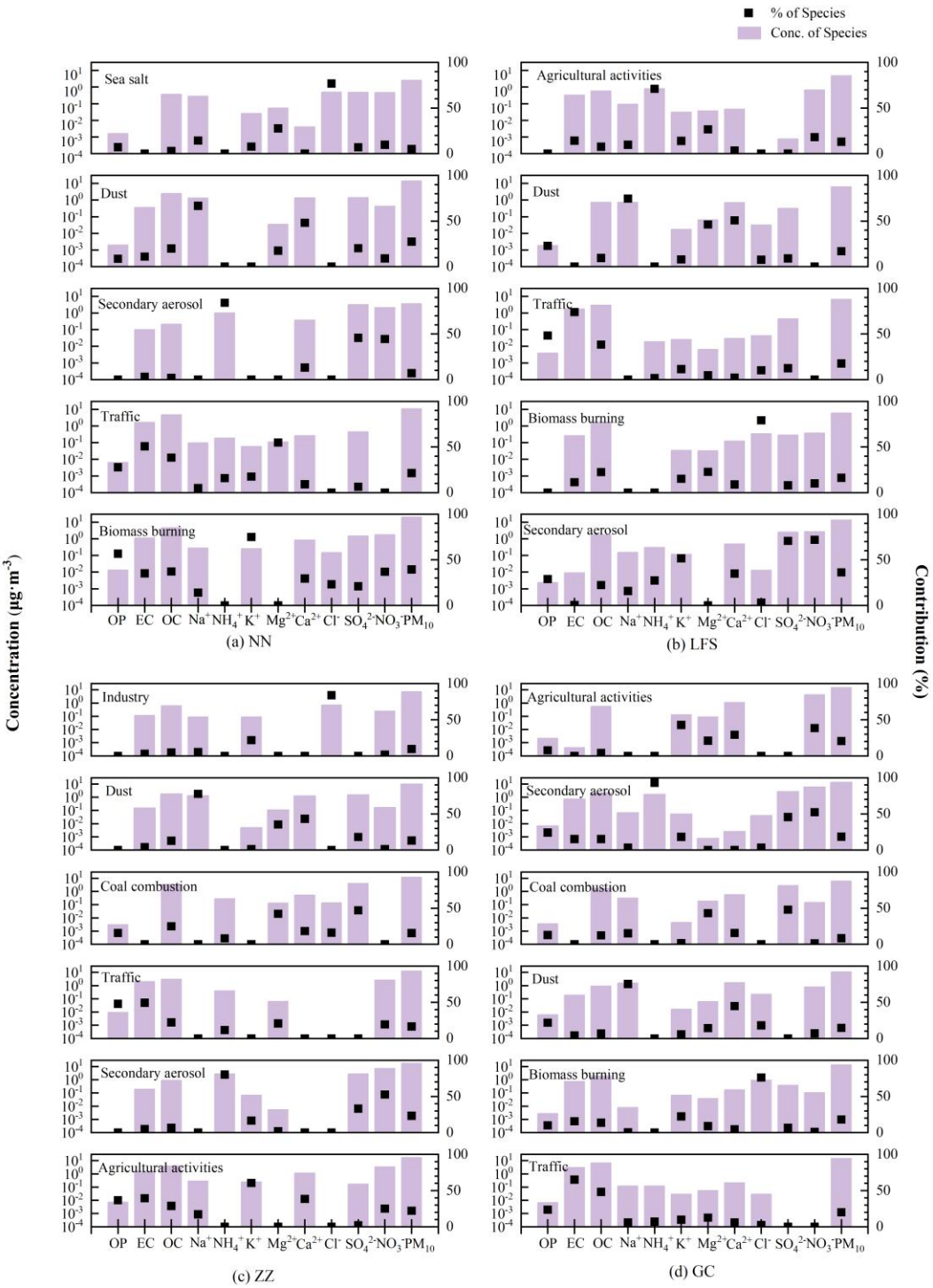

Figure 10. Chemical profiles of the source factors identified at NN, LFS, ZZ and GC. The bars represent the chemical composition profiles (left y-axis) and the dots indicate the contribution values (right y-axis).

32. *Lines 459-460: No contributions from EC and OC were found in the Biomass Burning factor at ZZ site, which is quite unexpected. It should be discussed.*

Response: **Sorry for the confusion.** Upon careful reconsideration of the PMF results and local emission characteristics, we acknowledge that our initial factor identification

may not have been optimal. We have refined our factor identification from biomass burning to industrial emission, which appears to better align with the observed chemical profile lacking significant EC and OC contributions and is more consistent with the local source conditions at the ZZ site. We have corrected in **Lines 509-512**: "The first factor had high contribution of K^+ (21.7%) and Cl^- (83.9%), but low contribution of OC (4.6%) and EC (2.8%), possibly indicating the influence of industrial emissions, such as food manufacturing, cement manufacturing, salt production, or industrial activities involving potassium chloride compounds (Yin et al., 2019; Seo et al., 2019), with a contribution proportion of approximately 9.5%."

33.*Line 464: Please add a reference supporting the coal combustion origin.*

Response: Thanks for your kind reminder. We have added the reference in **Lines 515-518**: "Since SO_4^{2-} primarily originates from fuel combustion (Schwartz, 1993), and Mg is specifically mentioned as an element enriched in the magnetic separation of coal fly ash (Strzałkowska, 2021), this factor is associated with emissions from coal combustion when regional characteristics are considered. Coal combustion accounts for around 15.5% of PM_{10} emissions and is likely to be associated with combined heat and power facilities in the surrounding area."

34.*Lines 469-470: What about NO_3^- ? Is there a reason why NH_4^+ is only present in the agricultural activities factor at LFS?*

Response：Thanks for your valuable suggestion. We have added some discussion about NO_3^- in agricultural activities at ZZ, as well as an explanation for why NH_4^+ is only associated with the agricultural activities factor at LFS among the three sites (LFS, ZZ, and GC).

We have added discussion about NO_3^- in **Lines 522-526**: "The sixth factor had high levels of K^+ (60.6%), Ca^{2+} (38.4%), EC (39.3%), OC (28.8%), and NO_3^- (25.0%), based on comprehensive analysis of these characteristic species, this factor may be related to agricultural activity emissions, contributing approximately 22% to PM_{10}. Ca^{2+}, OC and EC may be related to surface soil dust (Yu and Cao, 2023), Jung et al. found elevated

K^+ concentrations at schools near corn farms, supporting the agricultural source attribution (Jung et al., 2024), while NO_3^- would be related to fertilizer application (Cao et al., 2018)."

In addition, the differential NH_4^+ contribution in agricultural emissions among the three sites is primarily attributed to temperature-related atmospheric processes. At the cooler LFS site in northeastern China, lower ambient temperatures favor the stability of particulate NH_4^+ and minimize volatilization, allowing NH_4^+ to remain directly associated with agricultural emission factors. In contrast, at the warmer ZZ and GC sites, higher temperatures promote NH_4^+ volatilization to gaseous NH_3, which subsequently undergoes secondary atmospheric reactions to form ammonium-containing secondary aerosols rather than being retained in primary agricultural emissions. We have added discussions in **Lines 554-564**: "A notable pattern observed among the three sites with agricultural activities (LFS, ZZ, and GC) is the differential contribution of NH_4^+within agricultural emission factors, with NH_4^+ being exclusively associated with agricultural activities at the LFS site. This spatial variation reflects the complex interplay between regional meteorological conditions, agricultural practices, and atmospheric chemistry processes. At the LFS site in northeastern China, cooler climate conditions favor the stability of particulate NH_4^+, allowing its direct retention within agricultural emission factors (Wang et al., 2020). The concentrated fertilizer application during the spring planting season, combined with lower ambient temperatures that minimize NH_4^+ volatilization, preserves the distinct agricultural source signature at this remote location (Huo et al., 2025). Conversely, at the warmer ZZ and GC sites in central and northern China, NH_4^+ undergoes more extensive atmospheric processing due to higher ambient temperatures. These conditions promote the volatilization of NH_4^+ to gaseous NH_3, which subsequently undergoes secondary reactions with acidic species (SO_4^{2-} and NO_3^-) to form ammonium-containing secondary aerosols (Stelson and Seinfeld, 1982; Wang et al., 2015)."

35.*Lines 475-476: Please add a reference supporting this.*

Response: Thanks for your careful review. We have added a reference to support in

Lines 530-534: "This likely reflects the contribution of corn, wheat, and other farming activities around the site to PM_{10}, potentially associated with the agricultural-dominant economic structure of this rural area. Ca^{2+} and NO_3^- may originate from agricultural soil dust during tillage and other agricultural processes, and NO_3^- could be related to fertilizer application (Yu and Cao, 2023; Cao et al., 2018). Similar to the ZZ site, this agricultural source attribution is supported by Jung et al., who found elevated K^+ concentrations at schools near corn farms (Jung et al., 2024)."

Reference

Cao, P., Lu, C., and Yu, Z.: Historical nitrogen fertilizer use in agricultural ecosystems of the contiguous United States during 1850–2015: application rate, timing, and fertilizer types, Earth Syst. Sci. Data, 10, 969-984, https://doi.org/10.5194/essd-10-969-2018, 2018.

Jung, C., Huang, C., Su, H., Chen, N., and Yeh, C.: Impact of agricultural activity on $PM_{2.5}$ and its compositions in elementary schools near corn and rice farms, Sci. Total Environ., 906, 167496, https://doi.org/10.1016/j.scitotenv.2023.167496, 2024.

Yu, Y. and Cao, J.: Chemical Fingerprints and Source Profiles of PM_{10} and $PM_{2.5}$ from Agricultural Soil in a Typical Polluted Region of Northwest China, Aerosol Air Qual. Res., 23, 220419, https://doi.org/10.4209/aaqr.220419, 2023.

36. *Line 495: Transition metals are not measured for this study.*

Response: Sorry for the confusion. We have revised the manuscript, making the following specific modifications in **Lines 572-574**: "The high OP of traffic emissions is mainly attributed to the oxidative components in their particulate matter emissions, including organic carbon as well as potentially present PAHs and transition metals (TMs) (Valavanidis et al., 2008)"

37. *Line 511-512: Several studies highlighted that ammonium nitrate is not redox-active, but can be present with species inducing OP. It should be clarified.*

Response: Thank you for this important clarification. You are correct that ammonium nitrate itself is generally not considered to be directly redox-active. We should clarify that NO_3^- formed during fertilization processes may not directly

contribute to oxidative potential, but can co-exist with redox-active species or influence the chemical environment that affects the oxidative characteristics of particulate matter through ionic strength effects and acidification processes. The presence of NO_3^- may serve as an indicator of agricultural activities that simultaneously release other potentially oxidatively active compounds, rather than being a primary source of oxidative potential itself. We have corrected in **Lines 651-652**: "NO_3^- formed during fertilizer application can influence particle oxidative properties through ionic strength effects and acidification processes (Lodovici and Bigagli, 2011)."