# Supplementary Information of *Mass-constrained source apportionment of nitrate-containing particles in eastern China using a SPAMS–NMF framework*

Wenfei Zhu[1,2], Qinghong Wang[1], Qingsong Wang[3], Jialin Shi[1], Yu Sun[1], Juntao Huo[4], Yi Sun[4], Jia Chen[4], Yue Zhao[4], Mei Li[5], Jun Chen[1], Shengrong Lou[1*], Hui Chen[3]*

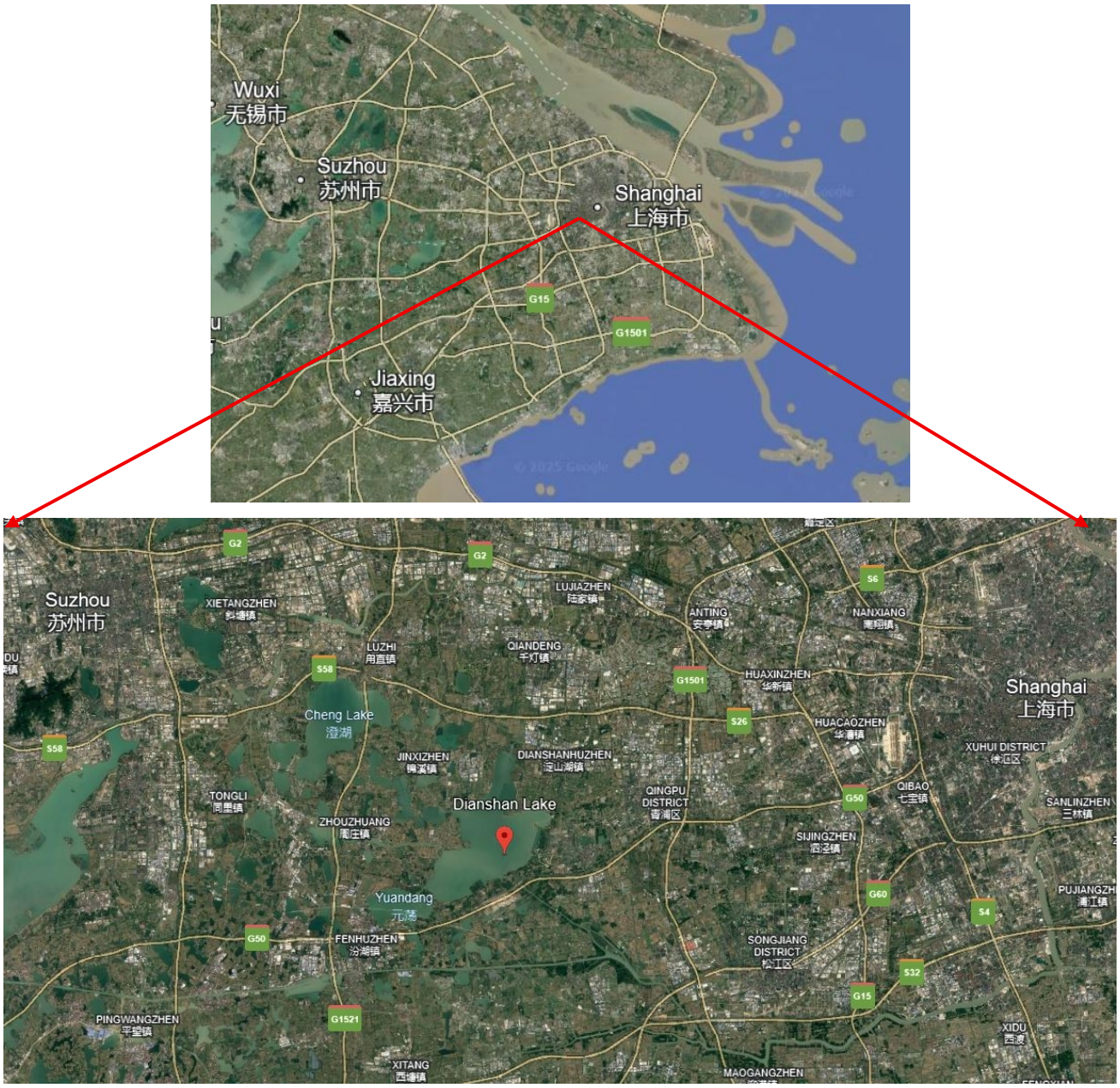
[1]School of Energy and Power Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China P. R.
[2]Institute of Climate and Energy Research, ICE-3: Troposphere, Forschungszentrum, Jülich, 52425 Jülich, Germany
[3]School of Environmental and Chemical Engineering, Shanghai University, Shanghai, 200444, China P. R.
[4]Shanghai Environmental Monitoring Center, Shanghai, 200235, China P. R.
[5]Institute of Mass Spectrometry and Atmospheric Environment, Guangdong Provincial Engineering Research Center for On-line Source Apportionment System of Air Pollution, Jinan University, Guangzhou 510632, China P. R.

*Correspondence to*: Shengrong Lou (lousr@usst.edu.cn); Hui Chen (huichen@shu.edu.cn)

**Tables**



**Figures**




50 **Figure S1. Location of observation sites in Shanghai.** The base map is from © Google Maps 2025.

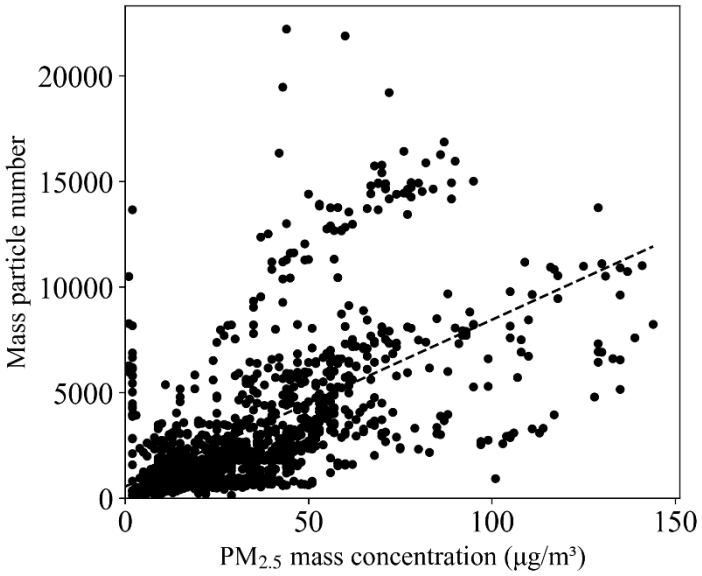



**Figure S2. The correlation between number concentration measured by SPAMS and $PM_{2.5}$ mass loadings**.

**S1. Semi-quantitative calculation of nitrate-containing particles from SPAMS data**

Single-particle aerosol mass spectrometry (SPAMS) provides real-time information on the size, number, and chemical composition of individual particles. However, SPAMS do not directly provide absolute particle mass concentrations because the detected ion signals are affected by particle size, chemical composition, laser desorption/ionization efficiency, and matrix effects. Therefore, a mass-constrained semi-quantitative method was applied in this study to convert SPAMS-derived single-particle information into mass-relevant nitrate concentrations.

**S1.1 Workflow of nitrate particle screening and classification**

The workflow of the semi-quantitative calculation is shown in **Figure S3**. Full-component single-particle data were first imported into the data-processing platform. Nitrate-containing particles were then screened using characteristic nitrate fragment ions at m/z 46 ($NO_2^-$) and m/z 62 ($NO_3^-$). Particles with clear nitrate-related signals were retained for further analysis. After nitrate screening, the selected particles were classified according to their mass spectral characteristics. Hierarchical clustering was conducted based on peak-area information and correlation relationships among the initial clusters. The ART-2a algorithm was then combined with manual consolidation to further group nitrate-containing particles into major classes and subclasses. During manual identification, characteristic ion signal peaks were used to assign individual particles into chemically meaningful particle types. The processed classification results were then exported for subsequent semi-quantitative mass calculation.

**S1.2 Mass-constrained calibration using bulk nitrate measurements**

Because SPAMS provides particle counts and relative peak areas rather than direct mass concentrations, bulk nitrate concentrations measured by MARGA were used as an external mass constraint. Before regression analysis, SPAMS data, MARGA nitrate concentrations, $PM_{2.5}$ mass concentrations, and meteorological parameters were synchronized to the same time resolution.

A multiple linear regression (MLR) model was used to relate bulk nitrate mass concentration to SPAMS-derived and environmental variables. The fitted equation was:

$$[NO_3^-] = -406.353 - 6.387 \times 10^{-7} \times A\_NO_3 + 0.462 \times [PM_{2.5}] + 7.174 \times 10^{-5} \times N\_PM_{2.5} + 0.356 \times T + 0.532 \times TP + 0.254 \times WS + 0.038 \times RH + 0.386 \times P \quad (R = 0.68)$$

where $[NO_3^-]$ is the nitrate mass concentration measured by MARGA, $A\_NO_3$ is the nitrate-related peak area measured by SPAMS, $[PM_{2.5}]$ is the $PM_{2.5}$ mass concentration, $N\_PM_{2.5}$ is the SPAMS particle number concentration within the $PM_{2.5}$ size range, T is temperature, TP is total precipitation, WS is wind speed, RH is relative humidity, and P is atmospheric pressure. These variables were selected because they are physically related to nitrate loading, aerosol abundance, SPAMS detection response, and meteorological influences on nitrate formation and sampling conditions. In the final regression model, all

retained predictors were statistically significant ($p < 0.001$). The regression model was used as a semi-quantitative mass-constrained correction rather than as a direct mechanistic model.

### S1.3 Allocation of nitrate mass to particle classes and subclasses

After the bulk nitrate mass was constrained by the regression model, nitrate mass was allocated to nitrate-containing particles using SPAMS-derived particle information. First, the fitted nitrate mass was assigned to nitrate-containing particles according to nitrate-related peak-area information. Then, the mass contribution was distributed among the major nitrate-containing particle classes and subclasses based on their number fractions.

The nitrate-containing particles were grouped into four major classes, $NO_3$Lv1–$NO_3$Lv4, and 12 chemically distinct subclasses. These subclasses included elemental carbon particles ($NO_3$_EC), organic carbon particles ($NO_3$_OC), organic/elemental carbon mixed particles ($NO_3$_OCEC), ammonium-containing particles ($NO_3$_Ammonium), potassium–nitrate particles ($NO_3$_K_N), potassium–carbon–nitrogen nitrate particles ($NO_3$_K_CN), potassium–sulfate particles ($NO_3$_K_S), sodium–potassium particles ($NO_3$_NaK), sodium-rich nitrate-containing particles ($NO_3$_Na-rich), dust-related nitrate particles ($NO_3$_Dust), Fe-rich nitrate particles ($NO_3$_Fe), and other nitrate-containing particles ($NO_3$_Other).

This allocation procedure converted SPAMS particle-count information and relative ion signals into semi-quantitative, mass-relevant concentrations for nitrate-containing particle classes and subclasses. These mass-relevant concentrations were then used for temporal variation analysis, size-resolved analysis, and NMF source apportionment.

### S1.4 Assumptions of the semi-quantitative approach

The semi-quantitative calculation was based on several assumptions. First, the SPAMS-detected particles were assumed to represent the major nitrate-containing particle population during the observation period. Since only particles with both positive and negative ion spectra were included in the analysis, the chemical classification was more reliable, but particles with weak ion signals may have been excluded. Second, nitrate-related peak areas were used as proxies for relative nitrate signal intensity. However, ion peak areas in SPAMS are not strictly proportional to particle mass because they may be affected by laser-particle interaction, particle phase state, matrix effects, and coexisting chemical species. Third, bulk nitrate concentrations measured by MARGA were assumed to provide an appropriate external mass constraint for the SPAMS-derived nitrate-containing particle population. The MARGA nitrate concentration was therefore used to constrain the mass scale of the SPAMS-derived nitrate information. Fourth, after mass constraint, the relative contributions of nitrate-containing particle subclasses were estimated using their number fractions. This assumes that subclass number fractions can reasonably represent the relative distribution of nitrate-containing particles after mass correction. Therefore, the calculated values should be interpreted as semi-quantitative, mass-relevant estimates rather than direct absolute mass concentrations of individual particles.

**S1.5 Uncertainties associated with the semi-quantitative method**

Several sources of uncertainty should be considered when interpreting the semi-quantitative results. First, SPAMS has size- and composition-dependent detection efficiency. Particles with different aerodynamic diameters, shapes, optical properties, and chemical compositions may have different transmission, sizing, and laser desorption/ionization efficiencies. Therefore, the detected particle population may not fully represent the ambient particle population. This uncertainty mainly affects the absolute magnitude of the estimated mass concentration, while the relative temporal trends and differences among dominant nitrate-containing particle classes are expected to be more robust. Second, particle-type-dependent ionization efficiency may introduce uncertainty. Even if different particles contain similar nitrate mass, their nitrate-related ion signals may differ because of matrix effects and the presence of coexisting species, such as sulfate, organics, metals, chloride, ammonium, and carbonaceous material. Therefore, the nitrate peak area should not be regarded as a direct mass signal. In this study, MARGA bulk nitrate was used as an external mass constraint to reduce the uncertainty caused by the nonlinearity between ion peak area and nitrate mass. Third, uncertainties may arise from the coupling between SPAMS and bulk measurements. MARGA measures bulk water-soluble nitrate mass concentration, whereas SPAMS detects individual particles within a limited aerodynamic size range and with instrument-specific detection efficiency. Differences in sampling inlet, particle size range, time resolution, and detection principle may lead to imperfect correspondence between bulk nitrate and SPAMS-derived nitrate-containing particles. Therefore, the regression-based calibration was treated as an empirical mass-constrained correction rather than a strict physical conversion factor. Fourth, the MLR model itself introduces uncertainty. The model included SPAMS nitrate peak area, $PM_{2.5}$ mass concentration, SPAMS particle number concentration, and meteorological variables. These variables were selected based on their physical relevance, and all retained predictors were statistically significant. However, nitrate formation is affected by nonlinear atmospheric processes, including gas-particle partitioning, heterogeneous reactions, aerosol liquid water content, aerosol acidity, and atmospheric oxidizing capacity. These processes cannot be fully represented by a linear regression model. The correlation coefficient of the fitted model was $R = 0.68$, indicating that the model captured a meaningful fraction of nitrate variability but still left unexplained variance. Fifth, the allocation of nitrate mass among particle subclasses also carries uncertainty. After mass constraint, nitrate mass was distributed among particle subclasses based on peak-area information and number fractions. Low-abundance subclasses may have larger relative uncertainties because small particle numbers are more sensitive to counting statistics and classification errors. Therefore, subclass-level mass concentrations, especially for minor particle types, should be interpreted with caution. Sixth, particle classification and manual consolidation may introduce uncertainty. ART-2a clustering depends on selected parameters, such as vigilance factor, learning rate, and iteration number. Although manual consolidation was performed based on mass spectral characteristics, temporal behavior, and known marker ions, some particle types may have overlapping chemical signatures. For example, particles with strong $Na^+$ and nitrate-related signals but weak chloride-related peaks were conservatively classified as $NO_3$_Na-rich rather than fresh sea-salt particles to avoid overinterpretation of source identity. Finally, the NMF source apportionment inherits uncertainties from the semi-quantitative input matrix. The resolved factors depend on the

particle-type mass allocation, classification results, and factor interpretation. Therefore, the NMF factors were interpreted jointly using source profiles, temporal variations, diurnal patterns, correlations with external chemical tracers, and backward-trajectory analysis. Stable sources such as dust and industrial emissions are expected to be more robust, whereas dynamic secondary sources and mixed Na-rich nitrate factors may have larger uncertainties.

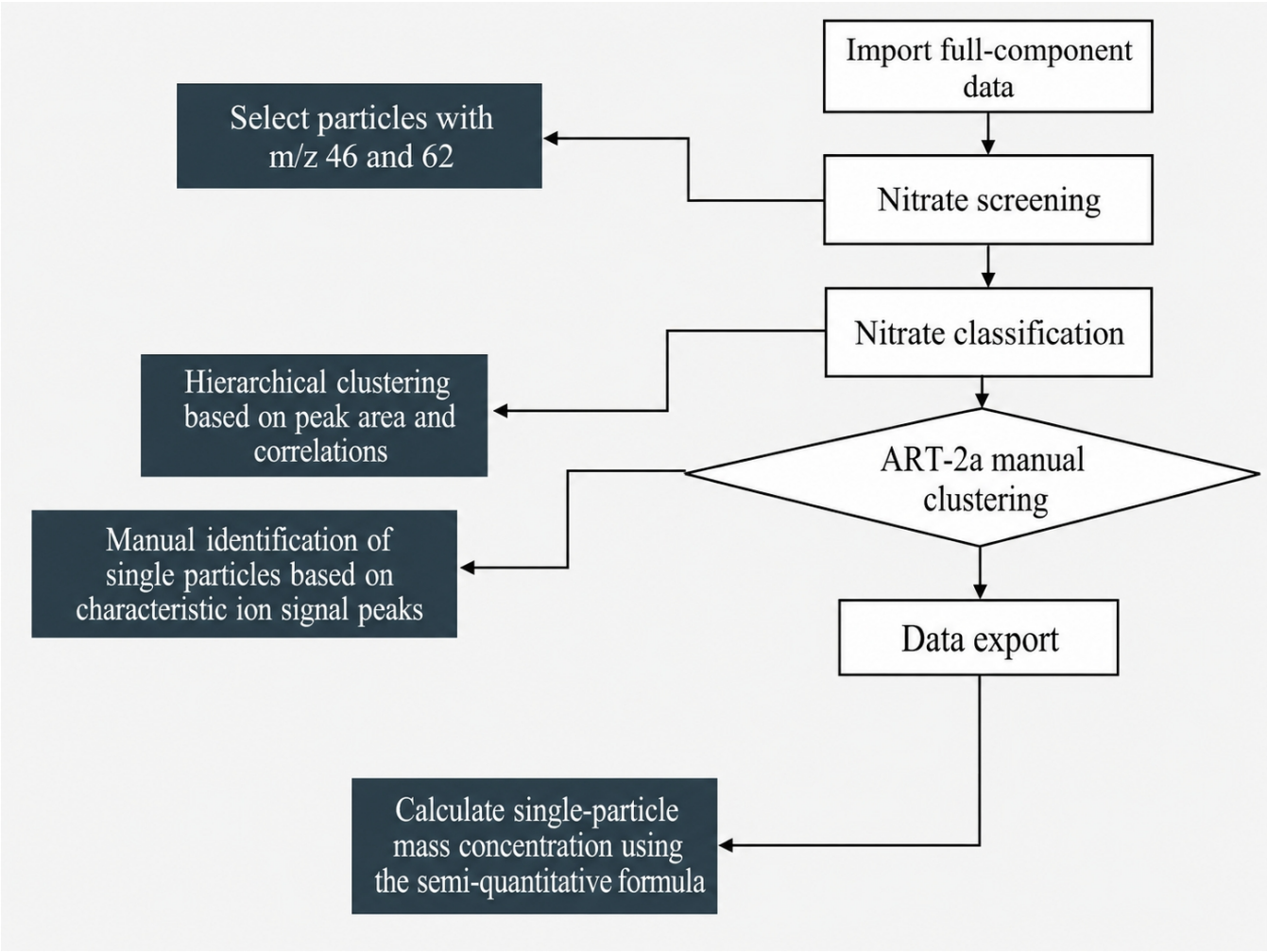



**Figure S3. Workflow of the semi-quantitative calculation for nitrate-containing particles based on SPAMS data.**

Full-component single-particle data were first imported, and nitrate-containing particles were screened using characteristic nitrate fragment ions at m/z 46 and m/z 62. The selected particles were then classified according to their mass spectral characteristics. Hierarchical clustering was performed based on peak-area information and correlation relationships. ART-2a clustering combined with manual identification based on characteristic ion signal peaks was used to consolidate nitrate-containing particle classes and subclasses. The processed data were exported and used to calculate single-particle mass-relevant concentrations using the semi-quantitative formula.

## S2 Classification of nitrate-containing particles

According to the principle of clustering, nitrate particles collected during the autumn–winter period of 2024 were classified into ten categories (C1–C10). The clustering results are presented in **Figure S5**. As shown, the correlations between C1–C4 and C5–C10 are relatively weak. The maximum correlation between C1 and C5–C10 is 0.66, generally below 0.5, while that between C2 and C5–C10 reaches 0.75, mostly below 0.6, indicating that C1 and C2 differ significantly from C5–C10 in their data characteristics ~~(Figure 3-1)~~ **Figure S5**. In contrast, the correlations among C1–C4 are high (mostly >0.7), with the strongest correlation (0.96) observed between C3 and C4, implying strong similarity. Similarly, the correlation between C5 and C6 reaches 0.95, suggesting almost identical characteristics across all variables. Further examination shows that C7–C10 also maintain high correlations ($r > 0.85$), indicating compact clustering in feature space.

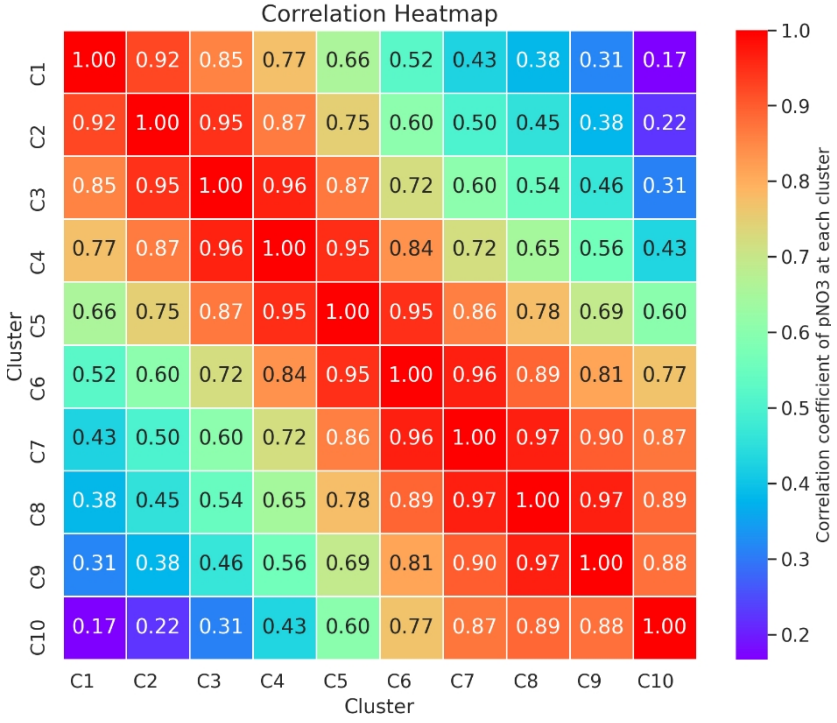



**Figure S5. Classification and correlation heatmap of nitrate particles.**

Based on these correlation characteristics, hierarchical clustering was used to systematically classify the particle groups as follows: C1 and C2 were assigned to NO3Lv1, C3–C4 to NO3Lv2, C5–C6 to NO3Lv3, and C7–C10 to NO3Lv4. The Art-2a core clustering algorithm was used to divide the 4 types of $NO_3$ factors into 12 subcategories.

**Table S1. Summary of clustering of nitrate-containing particles during the observation period.**

| $NO_3$Lv1 | | | $NO_3$Lv2 | | | $NO_3$Lv3 | | | $NO_3$Lv4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Types** | Number | Fraction | **Types** | Number | Fraction | **Types** | Number | Fraction | **Types** | Number | Fraction |
| **Ammonium** | 4846 | 1.3% | **Ammonium** | 6524 | 1.6% | **Ammonium** | 17028 | 3.3% | **Ammonium** | 15457 | 1.8% |
| **Dust** | 1735 | 0.4% | **DUST** | 634 | 0.1% | **DUST** | 1702 | 0.3% | **DUST** | 6654 | 0.5% |
| **EC** | 97507 | 25.7% | **EC** | 94960 | 23.8% | **EC** | 156000 | 30.4% | **EC** | 55860 | 38.8% |
| **Fe** | 6735 | 2.4% | **Fe** | 9919 | 2.6% | **Fe** | 19053 | 4.1% | **Fe** | 96997 | 7.3% |
| **K_CN** | 30732 | 8.1% | **K_CN** | 38270 | 9.6% | **K_CN** | 38665 | 7.6% | **K_CN** | 24459 | 1.7% |
| **K_N** | 6992 | 1.8% | **K_N** | 15320 | 3.8% | **K_N** | 32202 | 6.3% | **K_N** | 264820 | 18.4% |
| **K_S** | 29955 | 9.2% | **K_S** | 40995 | 12.4% | **K_S** | 32713 | 6.4% | **K_S** | 11163 | 1.3% |
| **NaK** | 63343 | 16.7% | **NaK** | 53096 | 13.2% | **NaK** | 89338 | 17.4% | **NaK** | 175660 | 12.2% |
| **OC** | 69275 | 18.3% | **OC** | 72625 | 18.2% | **OC** | 88462 | 17.3% | **OC** | 204360 | 14.2% |
| **OCEC** | 28626 | 7.6% | **OCEC** | 46843 | 11.7% | **OCEC** | 18197 | 3.6% | **OCEC** | 37332 | 2.6% |
| **Na-rich** | 20945 | 5.5% | **Na-rich** | 949 | 0.2% | **Na-rich** | 9978 | 1.9% | **Na-rich** | 7001 | 0.5% |
| **Other** | 10485 | 2.7% | **Other** | 8518 | 2.1% | **Other** | 6440 | 1.3% | **Other** | 10622 | 0.7% |

180

**S3 Source apportionment**

Both peak areas and number concentration exhibited positive correlations with nitrate mass, but the number concentration showed a stronger association ($R^2$ = 0.4068) compared to peak areas ($R^2$ = 0.1916), suggesting that number concentration provides a more robust proxy for nitrate loading (**Figure S6**). The bulk nitrate mass was initially assigned to individual particles based on peak area and subsequently distributed among subclasses by number fraction. 12 subclasses were assigned to mass concentrations.

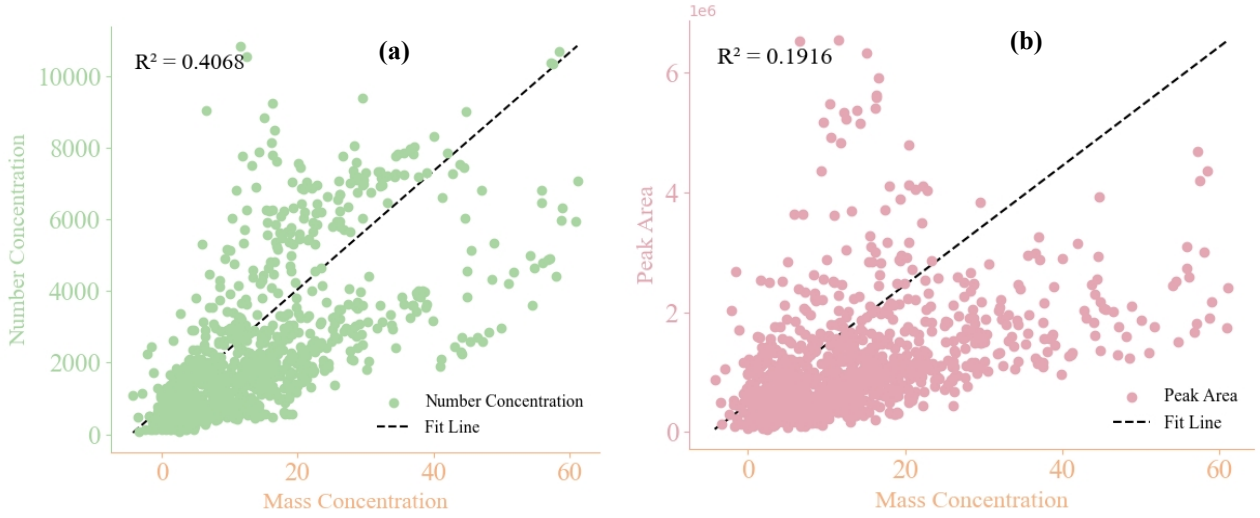



**Figure S6. The correlation between the (a) number concentration, (b) peak area measured by SPAMS and nitrate mass concentration measured by MARGA**.

Based on the "Non-negative Matrix Factorization (NMF) rank evaluation" and hierarchical clustering analyses, the N-NMF algorithm identified a seven-factor solution as the optimal configuration for the single-particle dataset. Ideally, receptor models should distinctly separate various pollution sources; however, in practice, mixed sources often appear. Therefore, this study aimed to identify as many sources as possible while minimizing source mixing to support refined air-quality management. The cophenetic correlation coefficient derived from the consensus matrix reflects factor stability; higher values indicate stronger stability. At seven factors, the cophenetic correlation reaches its maximum, while at eight factors it declines sharply, supporting the selection of seven factors as optimal (**Figure S7**). Evaluation indicators also include explained variance (evar) and residuals: evar quantifies the proportion of data variance explained by the model and should be maximized given satisfactory performance of other indicators. Residuals represent unexplained variability; the RSS serves as a global measure of model fit, indicating the discrepancy between observed and predicted data.

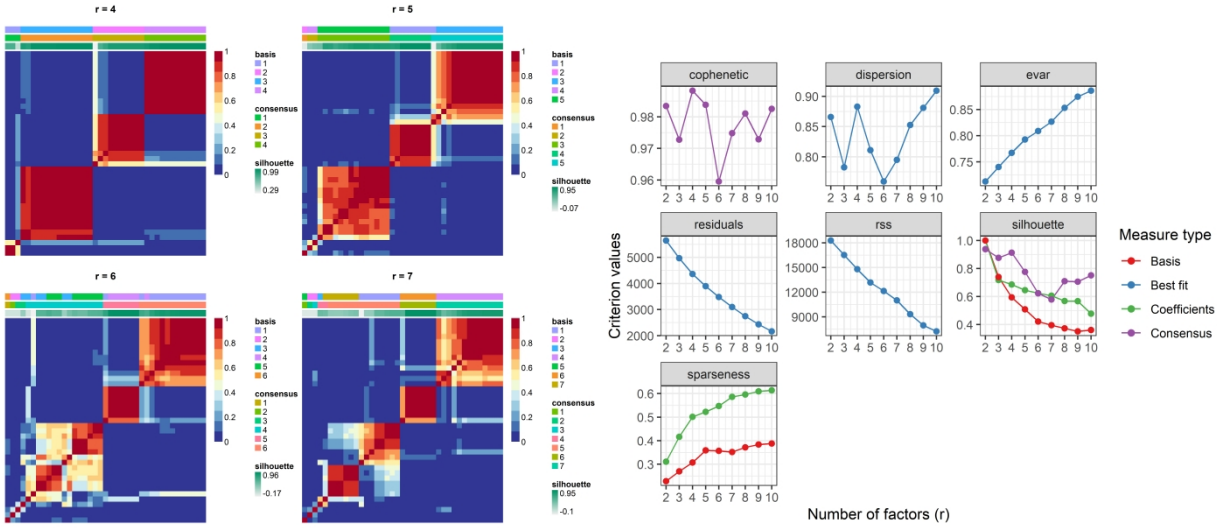



**Figure S7. The NMF rank survey results for single nitrate-containing particles.** It presents the same table correlation coefficient, dispersion score, explained variance, residual sum of squares, profile analysis and sparsity in order; hierarchical clustering heat map.

Nevertheless, increasing factor numbers may artificially improve evar or reduce RSS at the expense of factor stability and cluster consistency. The hierarchical clustering heatmap further visualizes the degree of separation between sub-categories—greater separation implies clearer source differentiation and reduced source mixing. Thus, the seven-factor solution was selected for the single-particle NMF analysis, while a four-factor solution was determined optimal for the bulk ($PM_{2.5}$) dataset.

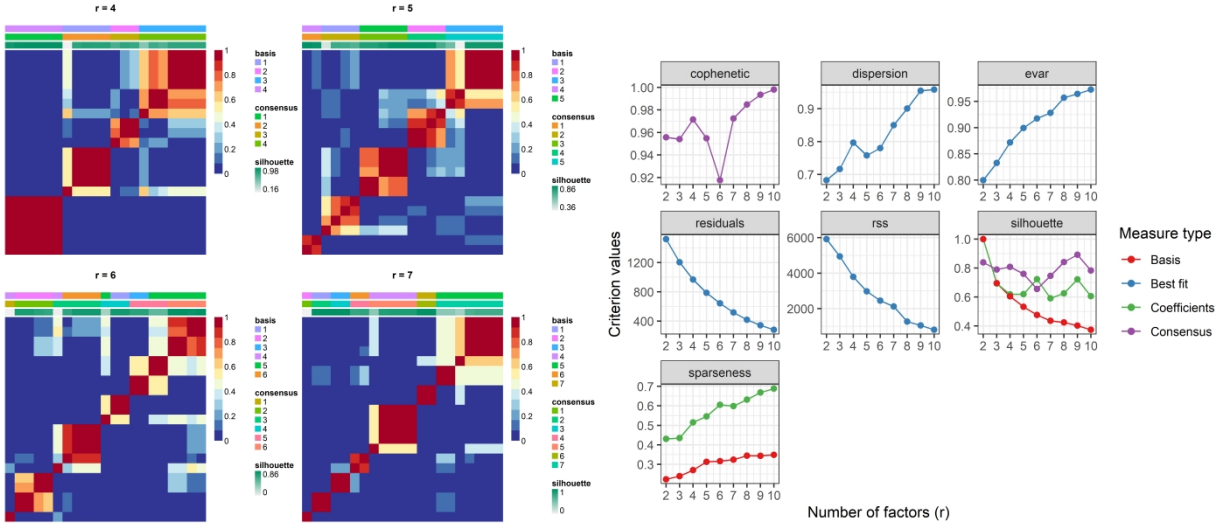



**Figure S8. The NMF rank survey results for bulk.** It presents the same table correlation coefficient, dispersion score, explained variance, residual sum of squares, profile analysis and sparsity in order; hierarchical clustering heat map.

The sources of nitrate-containing particles, with a comprehensive component analysis were resolved by NMF method based on dataset collected at Dianshan Lake site. The NMF analysis identified seven factors (N1–N7) for single-particle data and four factors (N1–N4) for $PM_{2.5}$ mass-based data. Source attributions were determined through joint interpretation of source profiles, temporal evolution, diurnal patterns, and correlations with conventional air pollutants (**Figures S9 to S10**) and $PM_{2.5}$ chemical components (**Figure S11**), summarized in **Table S2**.

For compositionally stable sources (e.g., dust and industrial emissions), both approaches yielded highly consistent results, confirming the robustness of the analysis. The dust source was identified in both single-particle NMF (N2) and $PM_{2.5}$-NMF (N1), with crustal-element-dominated signatures consistently observed in both number and mass concentrations. Similarly, industrial emissions were resolved independently by both methods (single-particle N4, $PM_{2.5}$ N2), characterized by heavy-metal enrichment and ranking among the top two factors, further supporting their reliability.

In contrast, dynamic secondary sources (e.g., secondary inorganic aerosols) and compositionally complex sources (e.g., sea salt) exhibited notable method-dependent discrepancies. The single-particle NMF demonstrated superior resolution for secondary inorganic aerosols, identifying two distinct factors (N3 and N6), whereas $PM_{2.5}$-NMF extracted only a single composite factor (N4), blurring the distinction between primary emissions and secondary formation. Regarding sea-salt aerosols, the single-particle NMF identified related factors (N5 and N7), whereas $PM_{2.5}$-NMF failed to resolve them, highlighting that low-mass concentration but high-number sources may be obscured in bulk-mass analyses.

For combustion sources, both methods indicated consistent but nuanced differences. Single-particle NMF (N1) showed strong As–Se–OC–EC–$K^+$–$NO_3^-$ signatures characteristic of coal combustion, while $PM_{2.5}$-NMF (N3) exhibited Pb enrichment as independent evidence of combustion emissions. However, in single-particle PMF results, this source (P6) ranked among the lowest contributors. The discrepancy likely arises from algorithmic weighting differences between “toxic low-abundance particles” (As/Se-rich coal combustion aerosols) and “abundant low-toxicity particles” (biomass-derived carbonaceous aerosols). Quantitative assessment of their relative contributions requires further validation using regional emission inventories.

Overall, the cross-validation between the two approaches confirms the robustness of stable sources (dust and industrial) while revealing the sensitivity of dynamic and compositionally complex sources to methodological differences. This comparison offers insight for future development of hybrid or multi-method source apportionment frameworks.

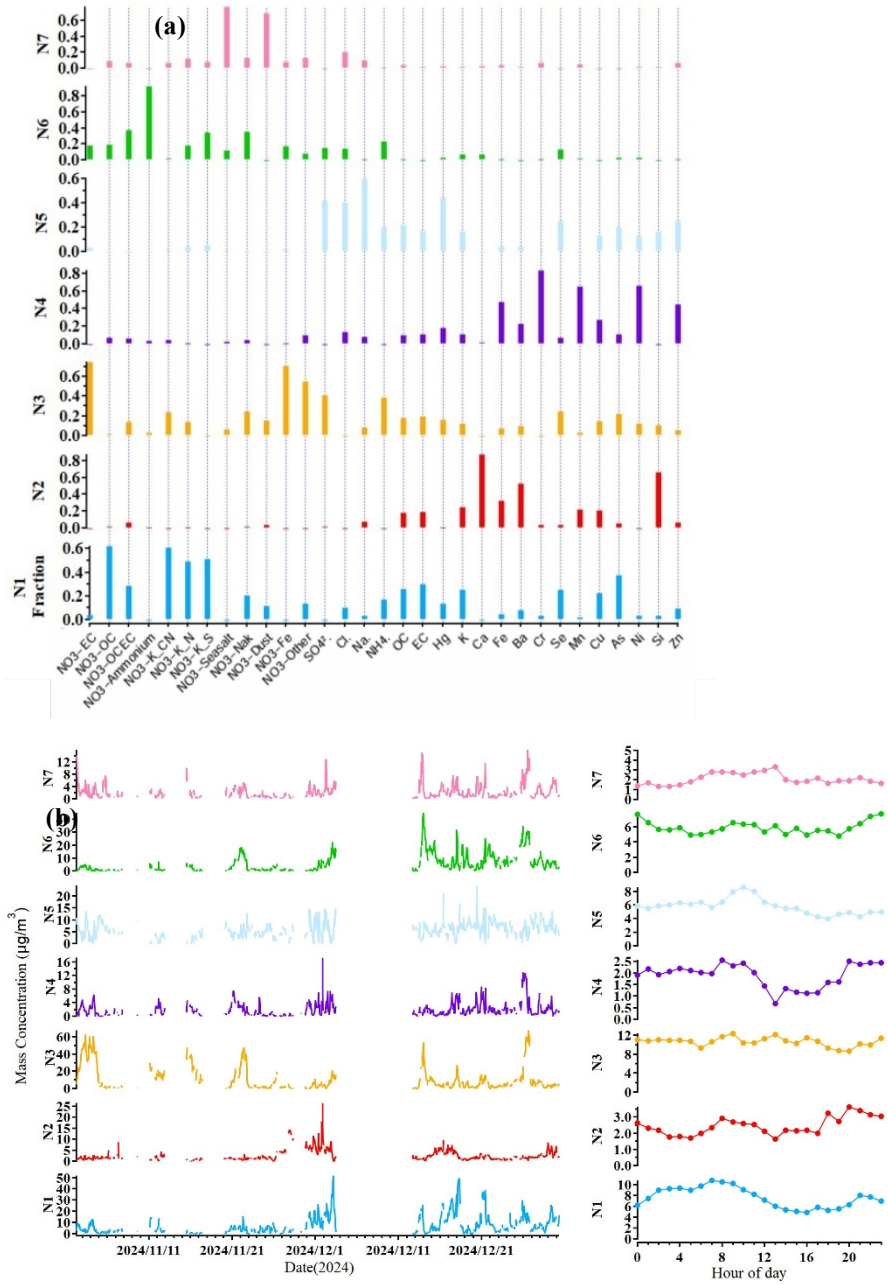
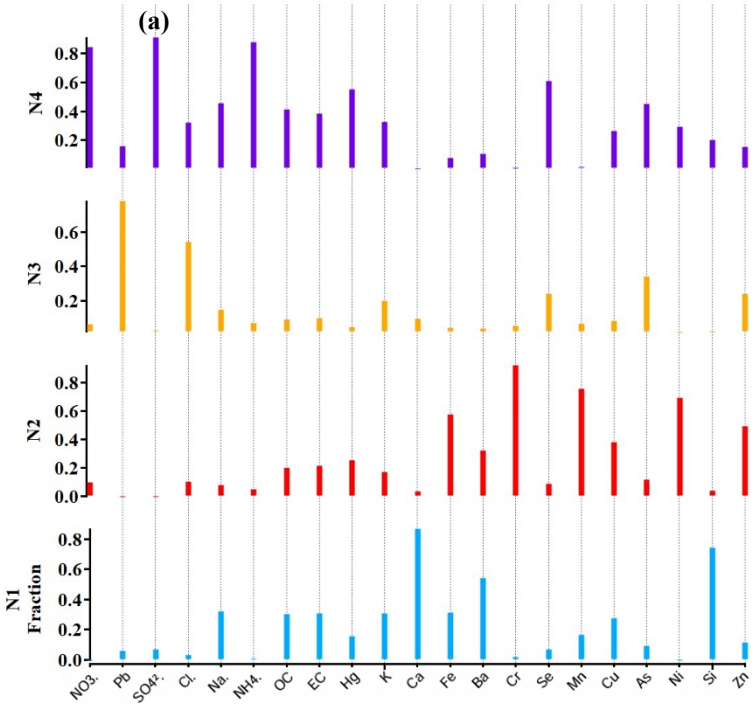



**Figure S9. Source apportionment of single-particle NMF of nitrate-containing particles: (a) Source spectrum and (b) Time series and diurnal variation characteristics.**


(a)
N4
N3
N2
N1
Fraction
NO3-
Pb
SO42-
Cl-
Na+
NH4+
OC
EC
Hg
K
Ca
Fe
Ba
Cr
Se
Mn
Cu
As
Ni
Si
Zn

240

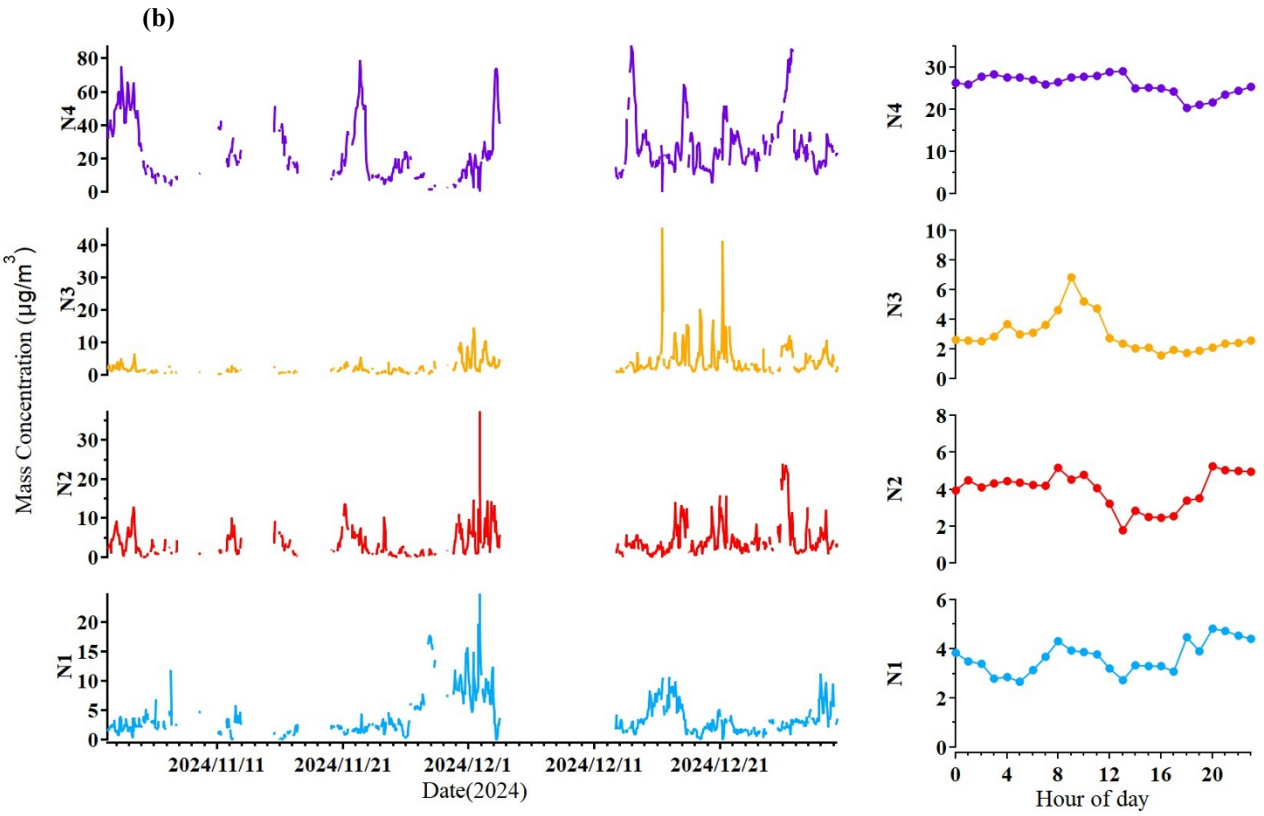



**Figure S10. Source apportionment of bulk $PM_{2.5}$ compositions: (a) Source spectrum and (b) Time series and diurnal variation characteristics.**

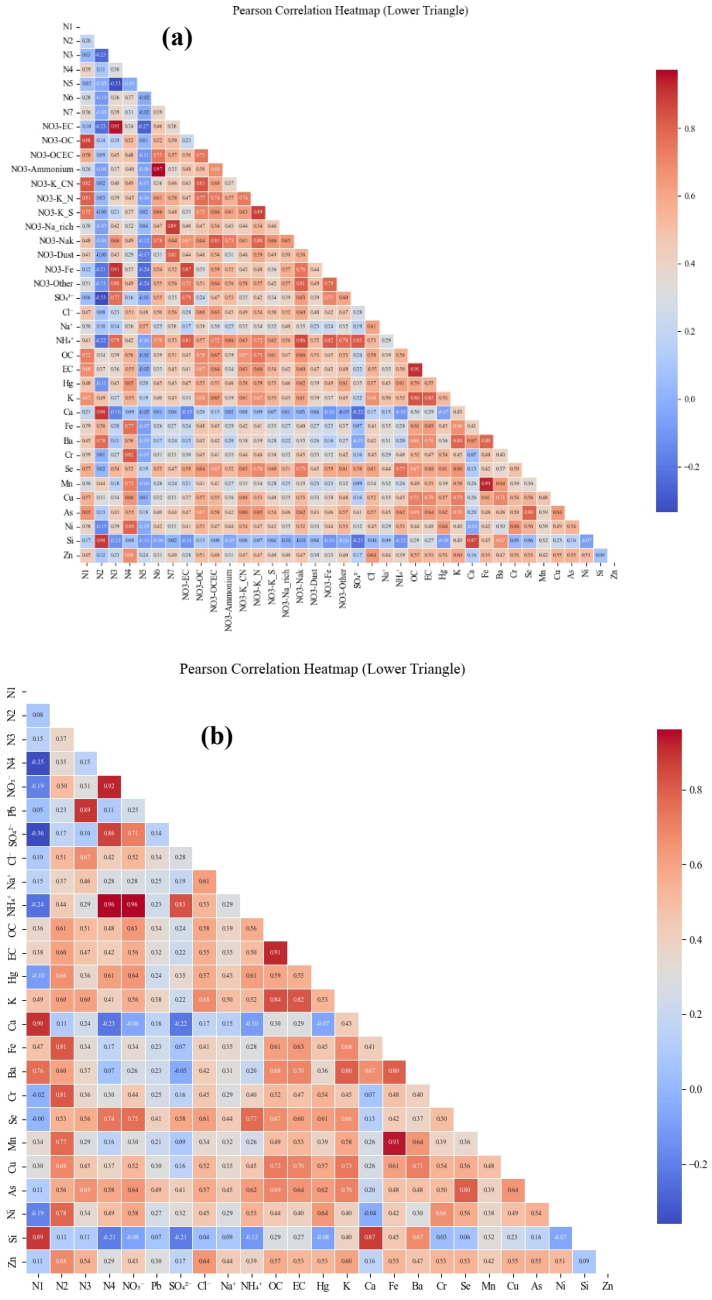



**Figure S11**. Pearson correlation heatmap between NMF source apportionment factors (a) nitrate-containing particles and (b) bulk and other chemical components of $PM_{2.5}$.

**Table S2.** Summary of identified NMF factors and corresponding source types.

| Method | Factor | Source Type |
|---|---|---|
| Single-particle NMF | N1 | Combustion source |
| | N2 | Dust source |
| | N3 | Secondary inorganic aerosol (non-ammonium nitrate) |
| | N4 | Industrial source |
| | N5 | Sodium-rich nitrate factor |
| | N6 | Secondary inorganic aerosol (ammonium nitrate) |
| | N7 | Mixed nitrate–Na-rich factor |
| $PM_{2.5}$-NMF | N1 | Dust source |
| | N2 | Industrial source |
| | N3 | Combustion source |
| | N4 | Composite source |

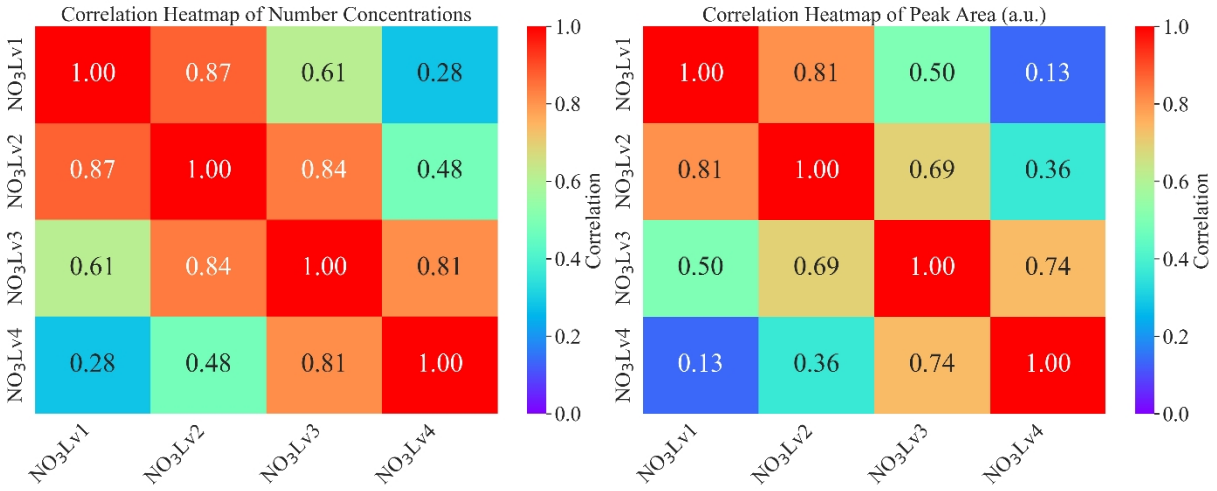



**Figure S12. Heatmap of internal Pearson correlations among four nitrate-containing particles categories for number concentrations and peak area.**

255

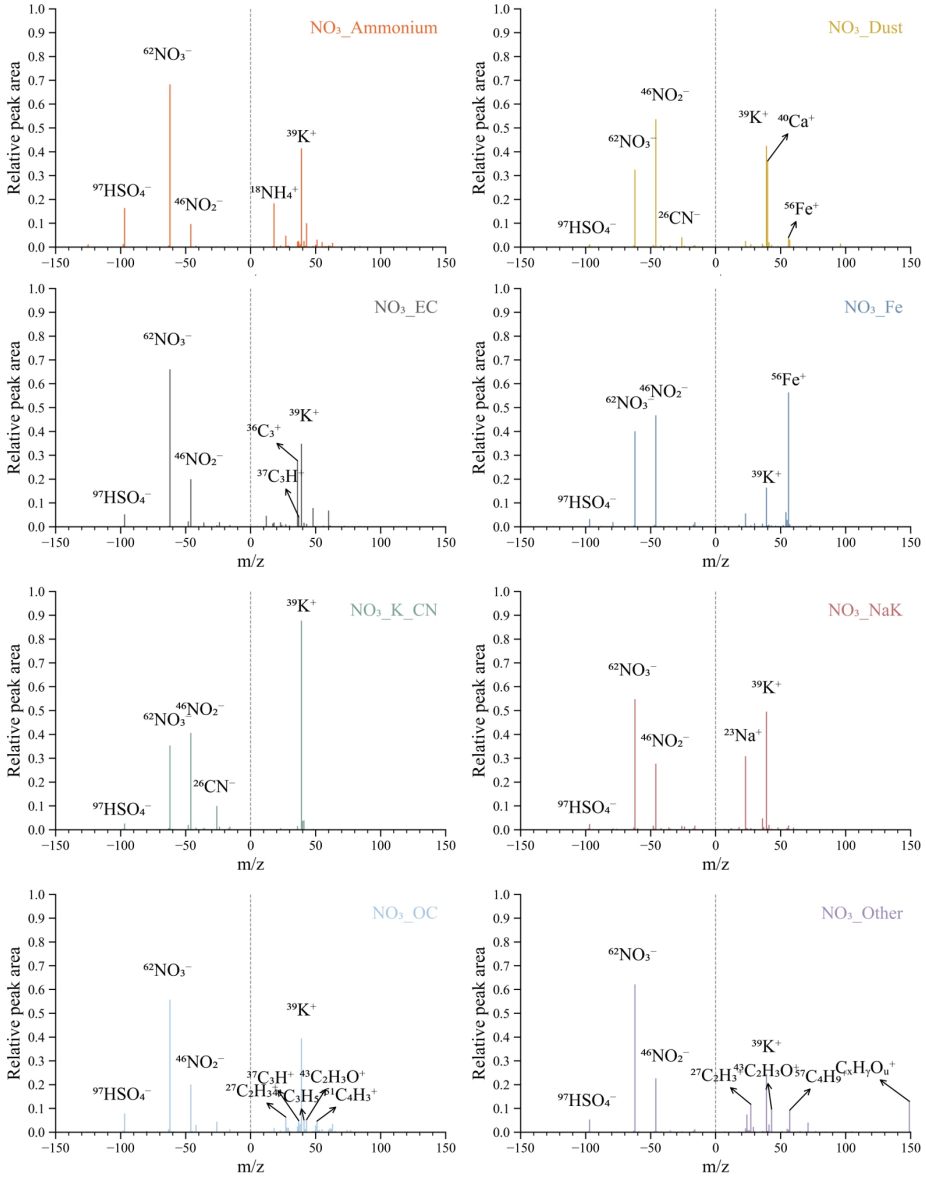
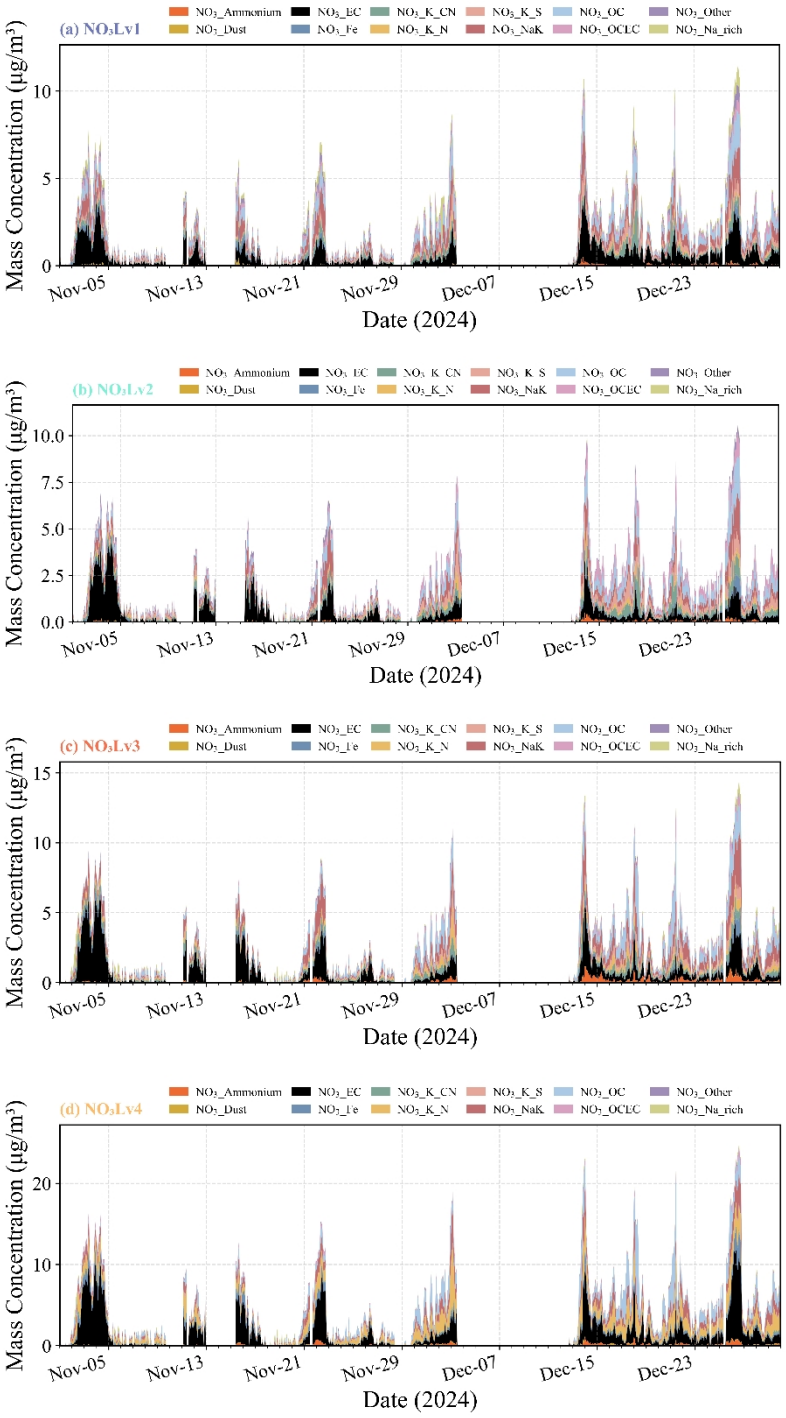



**Figure S13. Representative mass spectra of eight additional nitrate-containing particle types identified by SPAMS, including $NO_3$_Ammonium, $NO_3$_Dust, $NO_3$_EC, $NO_3$_Fe, $NO_3$_K_CN, $NO_3$_NaK, $NO_3$_OC, and $NO_3$_Other.** Positive ion spectra are shown on the right side of the vertical dashed line, and negative ion spectra are shown on the left side. Together with the spectra shown in Figure 1b, these spectra provide representative spectral evidence for all 12 nitrate-containing particle types.




265

**Figure S13. Time series of mass concentrations of chemically distinct types for (a) $NO_3$Lv1, (b) $NO_3$Lv2, (c) $NO_3$Lv3, and (d) $NO_3$Lv4 nitrate-containing particle classes over the sampling period.**

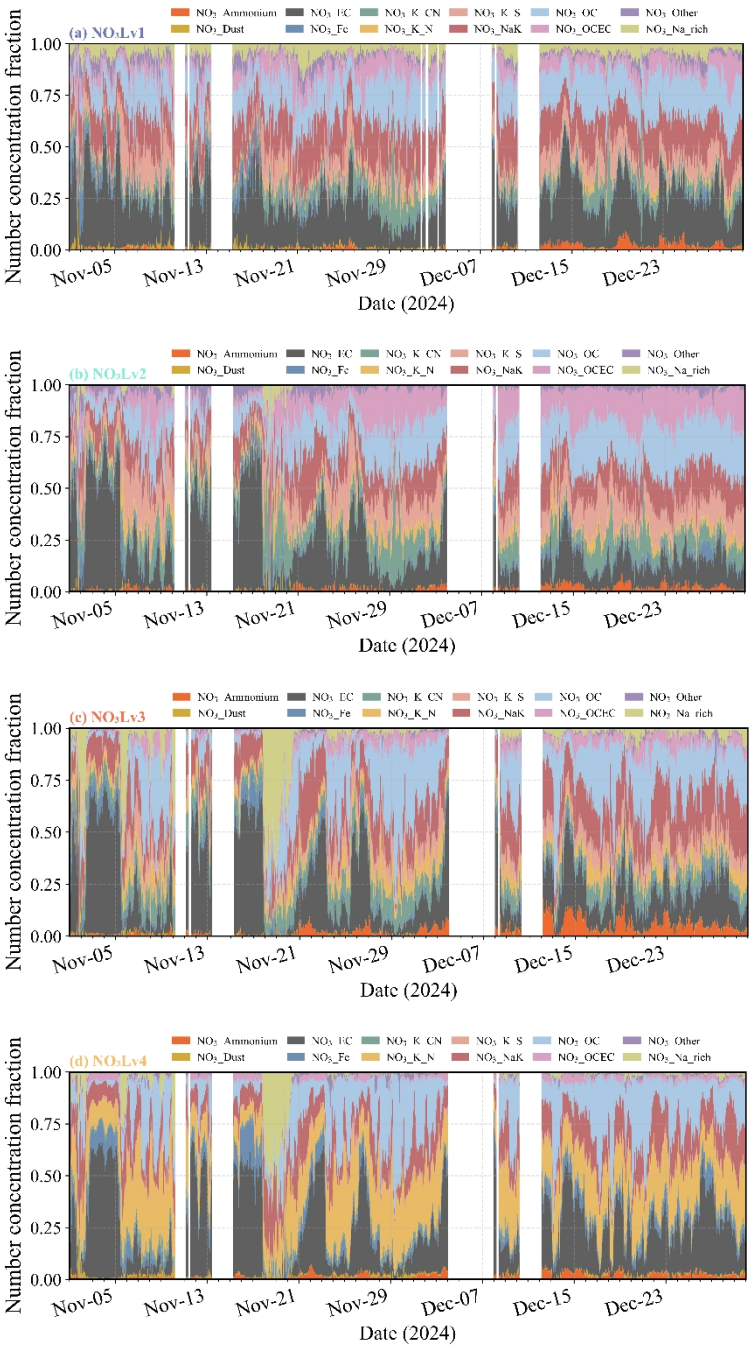



**Figure S14. Time series of number fractions of chemically distinct types for (a) NO3Lv1, (b) NO3Lv2, (c) NO3Lv3, and (d) NO3Lv4 nitrate-containing particle classes over the sampling period.**

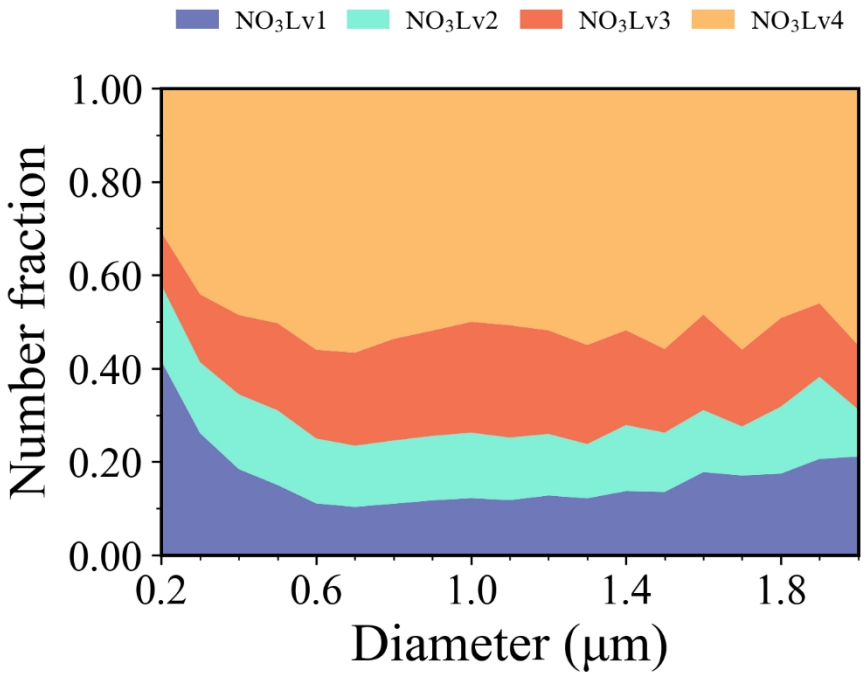



**Figure S15. Size-resolved number fraction of nitrate-containing particle classes during the whole observations.**

280

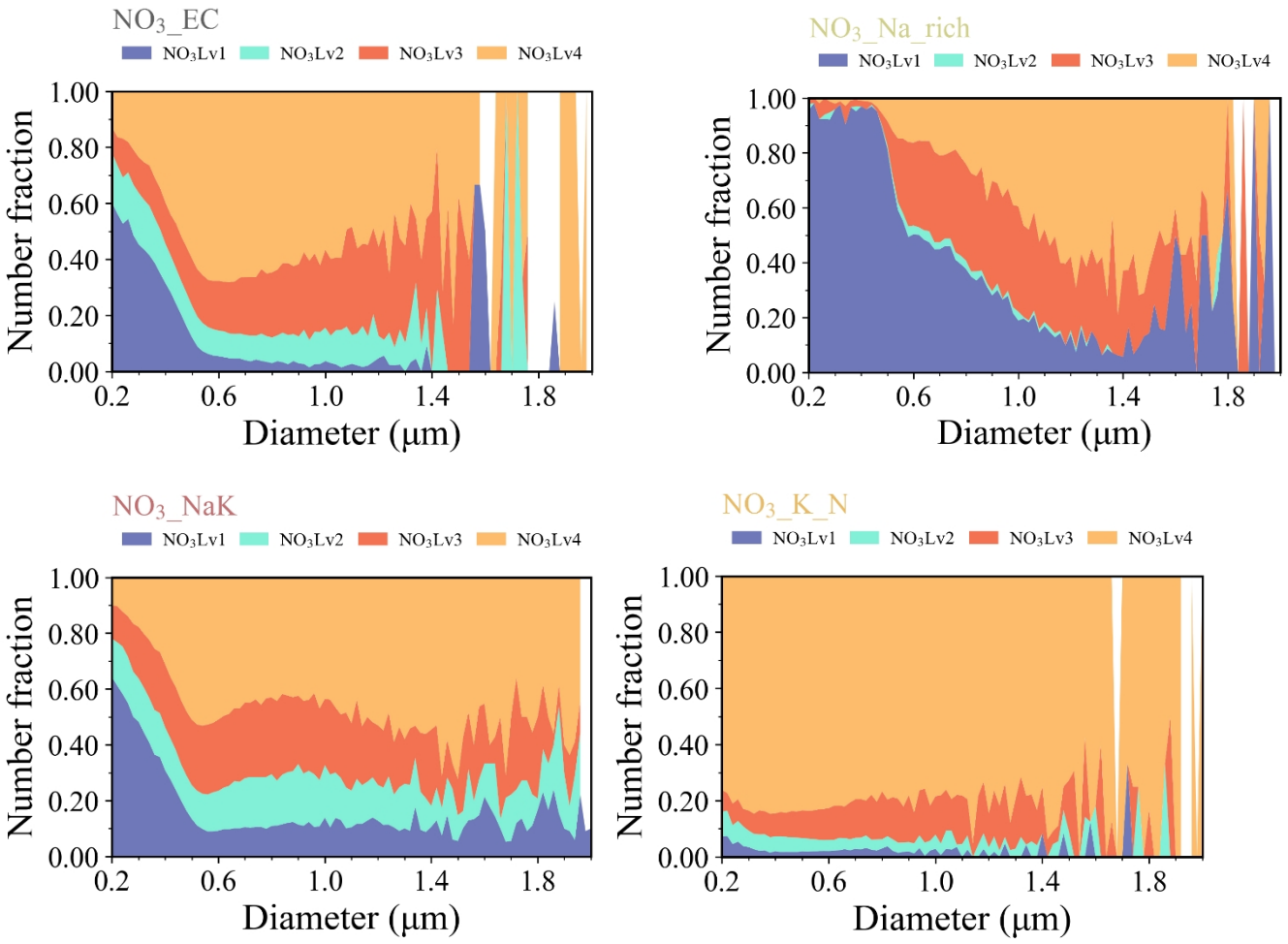



**Figure S16. Size-resolved number fraction of $NO_3$_EC, $NO_3$_Na-rich, $NO_3$_NaK, and $NO_3$_K_N in nitrate-containing particle classes.**

285