Rebuttal to the concerns raised by the reviewers for the paper titled “A Framework for Dynamic Hyper-local Source Apportionment using Low-cost Sensors for Real-time Policy Action”

(MS No.: egusphere-2025-5677)

Concerns of Reviewer 3:

This manuscript introduces a machine learning framework designed for real-time source apportionment (SA) of urban air pollutants. The authors propose a methodology that utilizes Positive Matrix Factorization (PMF) profiles derived from high-resolution reference instrumentation (HR-ToF-AMS and Xact-625i) as baseline target vectors to train the machine learning based models. While the framework addresses a highly relevant and timely topic for hyper-local air quality monitoring, substantial deficiencies regarding technical reporting, physical plausibility of the underlying datasets, and statistical evaluation preclude its publication in its current form. Major revisions are required to ensure the scientific validity and integrity of the proposed methodology before a final editorial decision can be reached.

Concern #1: Technical Characterization of Reference Instrumentation and Mobile Deployment Setup.

The reliability and accuracy of the reference grade instrument (RGI) measurements represent the core baseline of this study, as these observations constitute both the sensor calibration standards and the machine learning training targets. However, the manuscript fails to supply comprehensive descriptions regarding the technical specifications and operational parameters of the reference equipment hosted within the mobile laboratory. Although the low-cost sensors are described briefly, a rigorous accounting of the RGI mobile deployment configuration is missing. The authors must provide explicit details concerning inlet configurations, sampling line loss considerations, and the exact operation protocols used for field validation and calibration of the mass spectrometers and gas analysers during the campaign period.

Response: We thank the reviewer for highlighting the need for a more complete technical description of the reference-grade instrumentation and mobile deployment configuration. We have substantially revised Section 3.3 and appendix.

The reference-grade instruments were housed inside an air-conditioned mobile laboratory to maintain a stable operating temperature and prevent instrument overheating (Lakra et al., 2026). Aerosol and gas inlets were mounted on the roof of the mobile laboratory at approximately (~1.0 m above roof level). During stationary measurements, the vehicle engine was switched off, and the sampling side of the mobile laboratory was positioned away from nearby idling vehicles wherever possible.

The aerosol instruments used separate size-selective inlets and short, straight sampling lines to minimize residence time and particle losses (Chow, 1995) (Fig. 3.1). Bends and unnecessary connections were avoided, and the tubing was inspected and cleaned before each deployment. We have also clarified that no quantitative particle-transmission correction was applied;

therefore, possible diffusional and inertial losses, particularly for particles near the lower and upper ends of the measured size range, are acknowledged as a limitation. Because the lines were short and operated at their prescribed flow rates, such losses were expected to be small for the dominant accumulation-mode particles, although they cannot be considered zero (Chow, 1995).

1. For the **HR-ToF-AMS,** ambient aerosol was sampled through a roof-mounted inlet extending approximately 0.8 m above the mobile-laboratory roof. Below the roof, the sampling line consisted of approximately 0.2 m of stainless-steel tubing with an outer diameter of about 10 mm and an inner diameter of about 8 mm, followed by approximately 0.3 m of black tubing, which was connected to a silica-gel diffusion dryer upstream of the instrument. The dryer was used to reduce the influence of relative humidity on particle collection efficiency (Tobler et al., 2020).

Operation and calibration: The HR-ToF-AMS was operated in V-mode at a vaporiser temperature of approximately 600 °C and an electron-ionization energy of 70 eV (DeCarlo et al., 2006). Key operational parameters, including filament current, inlet pressure, flow rate, turbo-pump temperature, and voltages, were checked daily. Ionization-efficiency calibrations were performed at each campaign site using 300 nm size-selected ammonium nitrate particles and the CPC mass-based method (Jayne et al., 2000). The average ionization efficiency values were 8.67E-08 ions molecule^{-1} at Cycle-1 and 1.05E-07 ions molecule^{-1} at Cycle-2, 1.25E-07 at Kanpur. The calibrations at each site ranged from 1E-08 to 1E-07, indicating stable instrument sensitivity during the campaigns (http://cires.colorado.edu/jimenez-group/UsrMtgs/). HEPA-filter measurements were conducted regularly for air-background correction, and the data were screened to remove periods affected by calibration, filter sampling, or instrument warnings.

2. For the **Xact 625i,** $PM_{2.5}$ aerosol was sampled through a stainless inlet tube with a total length of approximately 1.5 m and an inner diameter of 22.1 mm. Of this length, about 0.4 m was inside the van from the roof entry to the Xact inlet. The instrument operated at a flow rate of 16.7 L min^{-1}.

Operation and calibration\checks: The sample passed through the instrument dew-point heater and was collected onto Teflon filter tape prior to ED-XRF analysis. Calibration and quality assurance included internal energy alignment checks, X-ray intensity checks, validation using NIST-certified thin-film standards, flow and leak tests within the manufacturer-specified tolerances (±10%), and cyclone cleaning before each campaign (Shukla et al., 2021a). The detailed version is provided in our previous group paper by (Lakra et al., 2026). Element-specific detection limits and concentration-to-MDL ratios are now included in Table R3.1. This ratio indicates the strength of the measured elemental signal relative to the minimum detectable concentration. Ratios greater than 3 indicate that the mean concentrations were sufficiently above the detection limit, whereas ratios close to or below 1 indicate weak signals that were near or below the instrumental detection capability (Kadachi and Al-eshaikh, 2012). During Campaign I, the concentration-to-MDL ratios ranged from 0.02 to 584.48 at CSIR-CIMAP and from 0.01 to 837.94 at BBAU. During Campaign II, the corresponding ratios ranged from 0.01 to 464.09 at CSIR-CIMAP and from 0.02 to 997.82 at BBAU. At Kanpur, the ratios ranged from 0.17 to 1322.55. The wide range reflects substantial differences in ambient concentrations among the measured elements. Elements such as S, K, Ca, Ti, Fe, Cu, Zn, Br, and Pb exhibited

high ratios, indicating strong and consistently detectable signals. In contrast, elements such as Al, V, Co, Mo, Cd, In, Sn, Sb, and Bi generally showed low ratios, indicating that their concentrations were frequently close to or below the MDL. However, some elements with lower ratios, such as V, Se, Ba, Rb, and Cr, were retained in the PMF analysis because of their importance as source markers (Table R1).

3. The **gas analyzers** for CO, NOx, SO_2, and O_3 are sampled via separate roof-mounted gas lines made of transparent Teflon tubing, each about ~0.5 meters long above the roof, connected with silica gel to minimise relative humidity.

Operation and calibration\checks: Zero and span checks were performed at the beginning of each site using zero air and certified calibration gases, and instrument diagnostics, flow, and baseline stability were monitored throughout the campaign (Mukta et al., 2021). For CO gas analyzers calibration was performed after every 3-4 days. Periods affected by calibration, flow instability, or analyzer warnings were excluded from analysis.

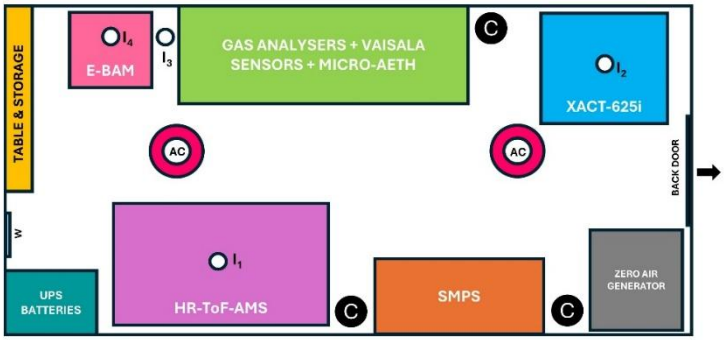


Fig. R3.1. Position of different instruments equipped inside the mobile laboratory. I1 (HR-ToF-AMS and SMPS), I2 (Xact-625i), I3 (gas analyzer, micro-Aeth, and Vaisala sensors), and I4 (E-BAM) represent the sampling inlets from the rooftop. C shows the 3 cylinders (NOx, SOx, CO) for gas analyser calibration. W shows the window, and AC represents air conditioners (Fig. S2 from study Lakra et al. (2026)).

Table R3.1. Average concentration of each trace element (ng/m³) at each site in C-I, C-II, and Kanpur with MDL and S/N ratio.

Trace element	MDL (30 min)	C-I CIMAP Mean (ng/m³)	C-I BBAU Mean (ng/m³)	C-I CIMAP S/N	C-I BBAU S/N	C-II CIMAP Mean	C-II BBAU Mean (ng/m³)	C-II CIMAP S/N	C-II BBAU S/N	Kanpur Mean (ng/m³)	Kanpur S/N
Si	89	649.35	833.61	7.3	9.37	727.65	1293.4	8.18	14.53	1392.19	15.64
S	16	2744.8	2758.6	171.5	172.4	2632.1	3012.7	164.5	188.2	6156.4	384.78
Cl	8.6	169.60	375.09	19.72	43.62	1703.6	1260.1	198.1	146.5	7910.5	919.83
K	5.8	1226.63	1596.41	211.5	275.2	1711.5	1916.8	295.1	330.49	3866.25	666.6
Ca	1.5	320.17	396.5	213.45	264.33	331.89	445.1	221.26	296.73	1139.69	759.79
Ti	0.79	28.64	39.03	36.25	49.41	29.17	40.61	36.92	51.41	64.95	82.22
V	0.6	0.75	0.57	1.25	0.95	0.36	0.56	0.6	0.93	0.8	1.34

Cr	0.58	2.70	5.38	4.66	9.28	1.41	1.61	2.43	2.78	9	15.53
Mn	0.71	8.02	20.52	11.3	28.9	9.28	10.01	13.07	14.1	43.52	61.3
Fe	0.85	329.10	451.44	387.18	531.11	339.43	478.43	399.33	562.86	894.18	1051.97
Ni	0.47	1.39	0.98	2.96	2.09	0.52	0.87	1.11	1.85	2.4	5.1
Cu	0.39	11.85	12.16	30.38	31.18	11.93	10.07	30.59	25.82	44.69	114.59
Zn	0.33	192.88	276.52	584.48	837.94	153.15	329.28	464.09	997.82	436.44	1322.55
As	0.31	2.98	2.31	9.61	7.45	2.01	1.78	6.48	5.74	11.13	35.92
Se	0.4	1.43	1.1	3.58	2.75	1.83	1.93	4.58	4.83	4.02	10.04
Br	0.52	9.88	12.61	19	24.25	24.69	25.8	47.48	49.62	60.71	116.75
Rb	0.95	0.97	1.28	1.02	1.35	1.8	2.41	1.89	2.54	6.89	7.25
Sr	1.1	2.16	2.48	1.96	2.25	2.99	2.63	2.72	2.39	5.52	5.02
Ba	1.9	0.62	2.27	0.33	1.19	6.05	0.7	3.18	0.37	8.66	4.56
Pb	0.63	57.50	76.58	91.27	121.56	47.15	53.23	74.84	84.49	351.65	558.17

Changes to be made in revised manuscript:

- All the details of instruments will be revised in the appendix of the manuscript, including a S/N ratio table.

Concern #2: Methodological Handling and Retention of Unreliable Low-Cost Sensor Signals.

A fundamental conflict exists between the authors' evaluation of sensor data quality and its subsequent inclusion in the machine learning framework. The text explicitly states that ambient SO_2 concentrations continuously fell below the minimum detection limit (MDL) of the sensors, resulting in an unreliable usage in the regression model. Furthermore, the NO_2 data also shows a poor coefficient of determination ($R^2 = 0.22$). Despite acknowledging that these datasets are not reliable in the model, both SO_2 and NO_2 appear to be retained as active predictors within the regression framework. Incorporating input parameters characterized primarily by low signal-to-noise ratios introduces arbitrary noise that can degrade multi-output model performance. The authors must supply a rigorous justification for retaining these specific chemical datasets, perform a focused sensitivity analysis examining model errors with and without these inputs, and clarify the exact processing workflow applied to sub-MDL measurements.

Response: We sincerely thank the reviewer for this important observation regarding the retention of **SO_2** and **NO_2** as predictors despite their comparatively poor calibration performance.

We agree that predictors exhibiting weak agreement with the corresponding reference-grade instruments should be carefully evaluated before being incorporated into the final machine learning framework. In response to this concern, we have substantially expanded the analysis presented in the revised manuscript by performing additional calibration assessment, feature

ablation studies and feature importance analysis using the best-performing **Extra Trees** regression model.

For **elemental source apportionment**, the feature ablation study demonstrated that progressively removing the comparatively unreliable predictors (**SO_2**, **NO_2**, and **VOC**) resulted in comparable or slightly improved prediction performance. Consequently, these variables have been excluded from the final elemental SA framework. The revised elemental model therefore relies only on the more reliable predictors, namely **CO, O_3, $PM_{2.5}$, BC, temperature, relative humidity and cyclical temporal features**.

For **organic source apportionment**, we also revisited the original modelling framework. In addition to the feature ablation study, the original five-factor PMF representation has been replaced by a chemically more meaningful three-factor representation comprising **BBOA, HOA and OOA**. The revised framework is accompanied by a comprehensive feature importance analysis, enabling a quantitative assessment of the contribution of each predictor to the final model. These additional analyses provide considerably greater transparency regarding predictor selection and model behaviour than the originally submitted manuscript.

Regarding the handling of **sub-detection-limit SO_2 measurements**, the original calibration procedure did not apply any special pre-processing or censoring strategy. The calibrated SO_2 concentrations obtained from the calibration model were directly used as input to the machine learning framework. Following the reviewer's suggestion and the additional analyses performed during revision, we concluded that the limited reliability of the SO_2 measurements did not justify their inclusion in the elemental SA model. Accordingly, SO_2 has been removed from the revised elemental source apportionment framework.

Similarly, an additional calibration analysis of the NO_2 sensor before and after the suspected mid-deployment breakpoint revealed a noticeable reduction in calibration performance during the latter part of the deployment. Together with the feature ablation results, this further supported the decision to exclude NO_2 from the final elemental SA model.

Overall, the revised manuscript will now include explicit sensitivity analyses for the predictor set, a revised predictor selection strategy, and substantially expanded discussion regarding sensor reliability and predictor importance. These additions directly address the reviewer's concerns regarding the influence of weakly calibrated sensor channels on the machine learning framework.

We sincerely thank the reviewer for this constructive comment, which has led to a substantially more rigorous evaluation of predictor reliability and a more robust final modelling framework.

Changes to be made in the revised manuscript

- Include a **feature ablation study** evaluating the influence of **SO_2**, **NO_2**, and **VOC** on model performance.
- Remove **SO_2**, **NO_2**, and **VOC** from the final **elemental** SA predictor set.
- Replace the original five-factor organic SA framework with the revised **three-factor (BBOA–HOA–OOA)** framework.
- Include **feature importance analysis** for both elemental and organic SA models.
- Expand the discussion on the treatment of **sub-detection-limit SO_2 measurements** and the calibration reliability of **NO_2**.

Concern #3: Mass Balance Discrepancies and Physical Incongruities in Aerosol Data.

A critical review of the baseline chemical concentrations summarized in Figure 5 reveals several profound reporting errors and physical contradictions. The specified concentration unit for Volatile Organic Compounds (VOCs) is entirely omitted from the graphic layout. And, the reported median Black Carbon (BC) mass concentrations (6537.5 µg/m^3 at Site-B and 5512.9 µg/m^3 at Site-C) are multiple orders of magnitude higher than the total $PM_{2.5}$ mass concentrations recorded for the exact same sites. Because BC is inherently a physical constituent of total $PM_{2.5}$ mass, it is physically impossible for the component concentration to exceed the bulk aerosol mass concentration. This implies a fundamental decimal scaling or unit conversion error (such as a systematic confusion between ng/m^3 and µg/m^3) during data analysis.

The authors indicate that Z-score normalization was implemented specifically to mitigate the massive disparities in numerical ranges among variables like CO and BC. If the primary BC dataset contains a systemic mathematical or physical error, the entire subsequent normalization workflow, model training, and regression feature weight allocations are compromised. The authors must comprehensively validate their data units, re-verify statistical summaries, and provide a clear chemical overview justifying why these locations are classified as distinct background or traffic environments based on their ambient profiles.

Response: We sincerely thank the reviewer for carefully examining the reported concentration statistics and for identifying the apparent inconsistency between the reported Black Carbon (BC) concentrations and the corresponding $PM_{2.5}$ mass concentrations.

Upon re-examining the manuscript, we found that the reviewer's observation originated from an error in the **reported concentration units**, rather than from the underlying measurements themselves. Specifically, the BC concentrations reported in Figure 5 were inadvertently labelled as **µg m^{-3}**, whereas the measurements recorded by the **AE-51 MicroAeth®** were expressed in **ng m^{-3}**.

Accordingly, the reported median BC concentrations of **6537.5** and **5512.9** correspond to **6.54 µg m^{-3}** and **5.51 µg m^{-3}**, respectively. These values are physically consistent with the simultaneously measured $PM_{2.5}$ concentrations and therefore do not violate aerosol mass balance.

We thank the reviewer for bringing this important typographical error to our attention. The concentration units for BC have been corrected throughout the revised manuscript, including Figure 5 and all related discussions.

The reviewer also questioned whether the apparent discrepancy in BC concentrations could affect the subsequent Z-score normalization and machine learning analysis. We would like to clarify that the normalization and all machine learning experiments were performed using the **correct underlying BC measurements**. The error was limited solely to the units reported in the manuscript, and therefore **did not affect the numerical values used during**

preprocessing, normalization, model training, or model evaluation. Consequently, none of the reported machine learning results are influenced by this typographical error.

Regarding the **VOC measurements**, we agree that the original manuscript did not clearly specify their physical interpretation. The VOC variable used in this study represents the **ratio of the instantaneous VOC concentration to the average VOC concentration over the preceding 24-hour period**, as provided by the deployed LCAQ sensor unit. Consequently, VOC is a **dimensionless quantity**, rather than a concentration expressed in conventional mass or volume units. This clarification will be incorporated into the revised manuscript to avoid any ambiguity.

We sincerely thank the reviewer for these careful observations, which have improved the clarity and accuracy of the revised manuscript.

Changes to be made in the revised manuscript

- Correct the BC concentration units from **µg m⁻³** to **ng m⁻³** throughout the manuscript.
- Revise Figure 5 and associated discussions accordingly.
- Clarify that the underlying BC data used for normalization and machine learning were always correct, and that only the reported units contained a typographical error.
- Explicitly state that the VOC predictor is a **dimensionless ratio** (current VOC concentration divided by the average VOC concentration during the preceding 24 hours).

Concern #4: Statistical Limitations of the Training Temporal Scale.

The volume of collected data points raises significant concerns regarding the suitability and generalization performance of the chosen data-driven approach. The underlying dataset is highly constrained, encompassing merely 364 and 369 paired observations for Site-B and 229 and 351 observations for Site-C. Because the raw readings were consolidated into 30-minute intervals and partitioned into an 80/20 training-to-testing ratio, the final validation phase evaluates the model over a transient window equivalent to less than two days of active testing data. Such constrained sample sizes increase the vulnerability of regression models to overfitting and fail to capture the atmospheric variance necessary to claim a robust or generalized framework. The authors need to address how a system trained on such restricted, single-month snapshots can provide reliable predictive capabilities across broader, dynamically changing atmospheric conditions.

Response: We sincerely thank the reviewer for highlighting the limitations associated with the temporal extent of the original experimental evaluation.

We agree that the originally submitted manuscript relied on relatively short-duration site-specific experiments, in which the available observations at each deployment site were partitioned into training and testing subsets. Such an evaluation, although suitable for an initial proof-of-concept, was not sufficient to comprehensively assess the robustness and transferability of the proposed framework across different spatial and temporal conditions.

In response to this important observation, we have substantially revised the experimental design presented in the manuscript. Rather than evaluating the proposed framework using only within-site train-test partitions, the revised manuscript now incorporates **five independent transfer-learning experiments** designed to evaluate model generalization under progressively more challenging deployment scenarios.

The principal experiment now trains the machine learning models using the **combined datasets collected during the October 2023 and February–March 2024 monitoring campaigns at two independent sites in Lucknow (CSIR and BBAU)**. The trained models are subsequently evaluated on an **entirely independent dataset collected during December 2024 at Kanpur**, representing a geographically distinct urban environment. Consequently, the external test data are separated from the earliest training observations by approximately **one year**, while also originating from a different city with distinct atmospheric characteristics. This substantially reduces the possibility that the reported performance merely reflects short-term temporal autocorrelation.

To further investigate model robustness, four additional transfer-learning experiments have been incorporated into the revised manuscript:

- Cross-site transfer: **CSIR → BBAU**
- Cross-site transfer: **BBAU → CSIR**
- Cross-season transfer: **October → February–March**
- Cross-season transfer: **February–March → October**

Together, these experiments evaluate the proposed framework across different monitoring locations, different seasonal conditions, and independent deployment campaigns.

In addition, the complete machine learning workflow has been revised to incorporate an explicit **training–validation–test** strategy. For each transfer-learning experiment, the source-domain data are partitioned chronologically into **80% training** and **20% validation** subsets (**shuffle = False**), while the target-domain dataset is reserved exclusively as an independent external test set. Hyperparameter optimisation is performed solely using the validation data, after which the selected model is retrained using the combined training and validation datasets before being evaluated on the external test set. This revised protocol eliminates any possibility of information leakage during model selection and provides a considerably more rigorous assessment of model generalization.

Finally, we agree with the reviewer that evaluation across additional years, cities and atmospheric conditions would provide an even stronger validation of the proposed framework. Such datasets are presently unavailable. Accordingly, the revised manuscript moderates the original claims of robustness and generalizability and explicitly presents the proposed framework as a **proof-of-concept study** demonstrating the feasibility of real-time source apportionment using calibrated low-cost sensor measurements under independent spatial and temporal transfer scenarios.

We sincerely thank the reviewer for this valuable suggestion, which has significantly strengthened both the experimental design and the evaluation methodology presented in the revised manuscript.

Changes to be made in the revised manuscript

- Moderate the claims regarding robustness and long-term generalizability throughout the manuscript.
- Replace the original short-duration site-specific evaluation with **five transfer-learning experiments** covering cross-city, cross-site and cross-season validation.
- Revise the machine learning workflow to incorporate an explicit **training–validation–external test** protocol.
- Clarify that all hyperparameter optimisation is performed exclusively using the validation set.
- Expand the discussion on the current dataset limitations and identify long-term multi-year validation as an important direction for future work.

Concern #5: Reporting Gaps and Structural Disconnects in Source Apportionment.

The implementation of the PMF source solutions and their integration with the subsequent machine learning steps require significant clarification. While the manuscript notes that factor solutions were selected based on standard statistical criteria (Q/Q_{exp}) and scaled residuals, the manuscript does not present any underlying numerical indices or factor-resolution diagnostic plots for comparison.

Also, the chemical mass spectral markers or particular ion fragments that differentiate the less-oxidized BBOA-1 factor from the more-aged BBOA-2 factor are poorly detailed. The authors should incorporate diurnal profiles or detailed time series profiles to solidify the physical identity of these resolved factors rather than relying strictly on unvalidated mass profiles.

The manuscript suggests that a combined input matrix merging data from both Site-B and Site-C was processed through the initial PMF model, yet this structural step is not explicitly defined. If reference factor profiles were extracted collectively across both locations, the authors must supply a clear scientific rationale for why the subsequent machine learning evaluations and reference-versus-predicted comparisons were divided and executed on a site-segregated basis.

Response: We thank the reviewer for this important comment. We agree that the original manuscript did not provide selection criteria for the PMF solutions or clearly explain how the resolved factor time series were subsequently used in the machine-learning analysis. We have therefore expanded the PMF methodology and added the corresponding diagnostic results to the revised Supplementary Information.

In response to Reviewer 1's response, we substantially expanded the PMF input dataset beyond the original approximately 20-day measurement period by including additional C-II datasets from CSIR-CIMAP and BBAU, conducted from February to March 2024, along with the Kanpur measurements conducted in December 2025. The total number of days has now increased to 63. The PMF analyses were repeated using the expanded dataset, which provided more stable and similar factor solutions. The revised HR-ToF-AMS analysis resolved five factors: HOA, BBOA-I, BBOA-II, SVOOA, and LVOOA, while the Xact 625i analysis resolved seven factors: S-rich, Cl-rich, Dust, Non-Exhaust, SFC-I, SFC-II, and Pb-rich (Figs. 3.3, 3.4) . A detailed procedure for the selection of the optimum solution is provided in our related mobile-lab studies from the same sites by Sethi et al. (2026, under review) and (Lakra et al., 2026). Therefore, only a concise summary of the factor characteristics, Q/Qexp, external

correlation and bootstrap assessment has been included in the revised supplementary information of the manuscript, with appropriate references to these studies.

(a) PMF input matrices and selection of the optimum solutions

For the **HR-ToF-AMS** analysis, the unit-mass-resolution (UMR) organic mass spectra were processed using SQUIRREL after removing spectra affected by instrument malfunction, calibration issues, or other invalid runs. The remaining data were averaged to a 30-minute resolution and used to construct the concentration and corresponding error matrices. The final AMS input matrix contained 2142 observations and 300 (m/z) variables. The large number of valid observations relative to the number of variables provided sufficient temporal variability and degrees of freedom for resolving the five-factor solution.

For the **Xact-625i** analysis, the 30-minute elemental concentration and uncertainty datasets were aligned across all included deployments. Elements with a detection limit below MDL were excluded. The final Xact input matrix contained 2424 observations and 21 elemental variables. The expanded multi-site dataset included substantially more observations than the original analysis and captured variations across sites, seasons, meteorological conditions, and source environments.

Solutions with different numbers of factors were examined for both datasets. The 5-factor AMS and 7-factor Xact solutions were selected based on changes in Q/Qexp, factor-profile interpretability, temporal behaviour, external-tracer relationships, and factor stability (Fig. R3.2). For the AMS analysis, the 5-factor solution was selected because we observed the maximum Q/Qexp relative change from 4 to 5 factors, whereas the 7-factor Xact solution shows the maximum Q/Qexp relative change from 6 to 7 factors. Adding further factors primarily splits physically meaningful factors, whereas solutions with fewer factors showed mixing of chemically distinct source profiles (Fig. R3.2).

Bootstrap assessment

To evaluate the stability of the PMF solutions, 1000 bootstrap resampling runs were performed in SoFi using criterion-based selection. The final five-factor AMS and seven-factor Xact solutions, with all factors unconstrained, were used as the base cases. Bootstrapped factors were mapped using an active Pearson correlation criterion ($R > 0.9$), and only valid, unmixed runs were retained. Overall, 857 runs (85.7%) for AMS and 923 runs (92.3%) for Xact were successfully mapped, indicating good factor stability and reproducibility.

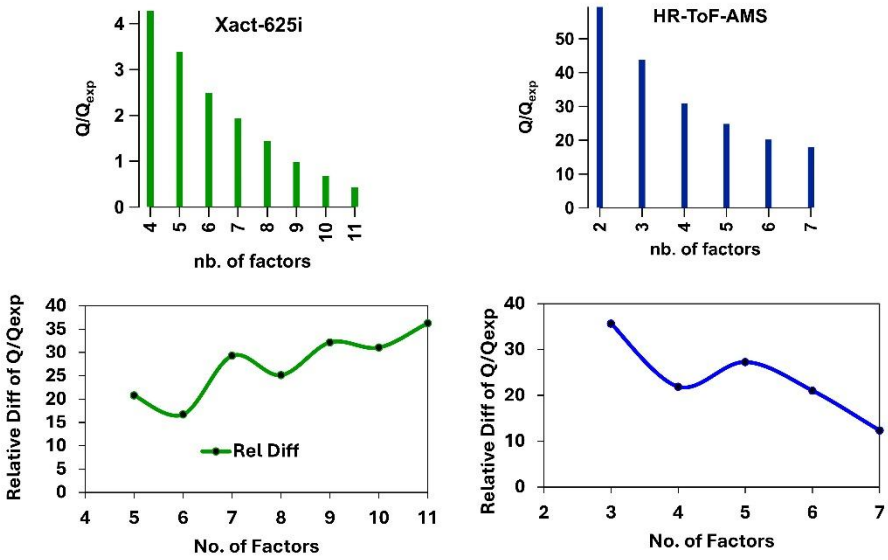


Figure R3.2: Q/Qexp and its relative difference (%) plot with respect to the number of factors for optimum PMF solution selection criteria of elemental SA (Xact-625i) and Organics SA (HR-ToF-AMS)

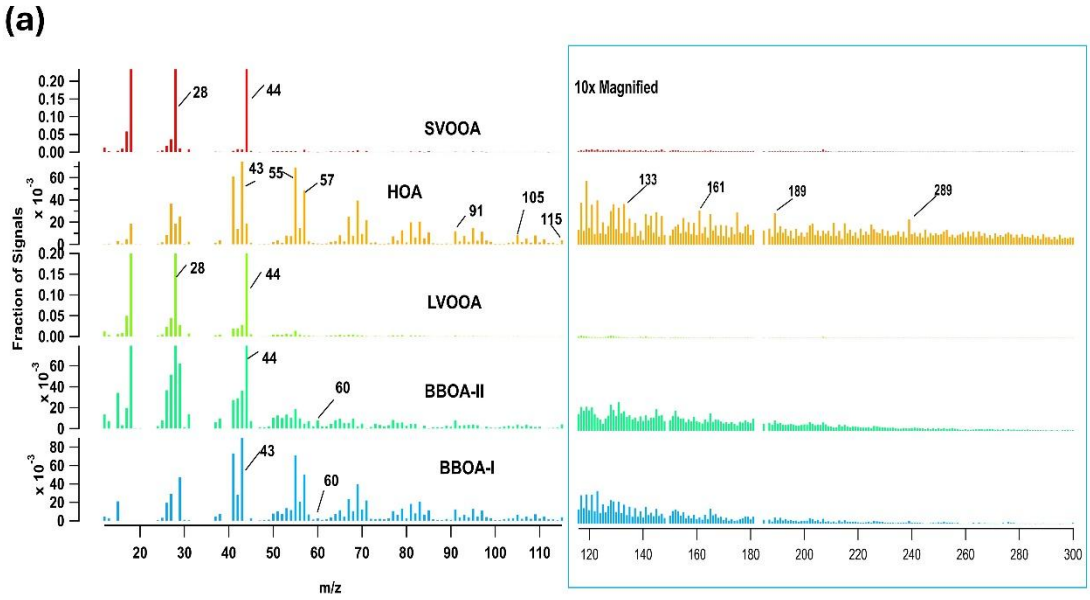
(b) Difference between BBOA-I and BBOA-II

BBOA-I: BBOA-I was also identified as a primary organic aerosol factor with low contribution from m/z44 (oxygenated ion) and contributed 12% of the total organic aerosol. Its mass spectrum was characterized by the biomass-burning marker ions at m/z 60 and 73, together with hydrocarbon and aromatic fragments (m/z 178, 202), indicating relatively fresh biomass and solid fuel combustion emissions (Goetz et al., 2022) (Fig. R3.3). Its diurnal profile showed high concentrations during the late evening, nighttime, and early morning, with a pronounced increase after approximately 19:00 and minimum concentrations during the afternoon. This pattern is consistent with nighttime biomass burning or domestic solid-fuel burning and the accumulation of primary emissions under reduced boundary-layer mixing. BBOA-I was highest at Kanpur, likely due to wintertime biomass and solid-fuel burning for domestic heating, followed by BBAU during Cycle I, where post-monsoon agricultural-residue burning may have contributed (Fig. R3.3). The lowest concentrations were observed at CSIR-CIMAP, reflecting its background nature and limited nearby combustion sources. Its correlations with Chl (R = 0.57) and BC (R = 0.67) further supported its combustion-related origin (Lalchandani et al., 2022) (Fig. R3.5).

BBOA-II: BBOA-II contributed 14% of the total organic aerosol and was identified as a more oxidized biomass-burning factor. Its mass spectrum retained the characteristic biomass-burning fragments at m/z 60 and 73 but showed a stronger contribution at m/z 44, indicating atmospheric ageing and secondary organic aerosol (SOA) formation (Lalchandani et al., 2022). Its diurnal profile exhibited a morning peak around 09:00–10:00, following the earlier BBOA-I increment, suggesting oxidation of freshly emitted biomass-burning aerosol (Fig. R3.3). An

increase during nighttime may be associated with oxidation by NO_3 radicals under stable atmospheric conditions (Lakra et al., 2024). BBOA-II was highest at Kanpur during winter, followed by BBAU during Cycle I and CSIR-CIMAP during Cycle II, reflecting differences in biomass-burning emissions and atmospheric processing across the sites. Its strong correlation with BC ($R = 0.70$) supports a combustion-related origin, while the positive correlation with nitrate ($R = 0.47$) further indicates its oxidised nature, which makes it different from BBOA-I (Fig. R3.5).

Detailed factor profiles with proper differences between BBOA-I and II, diurnal variations, external correlations, and bootstrap results have now been added to the Figures of the revised Supplementary Information.



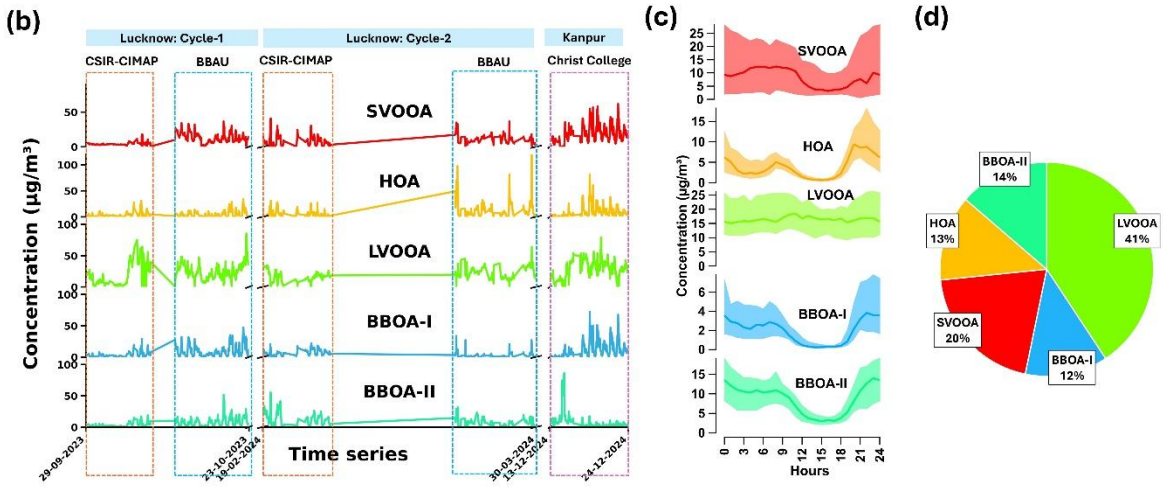


Figure R3.3. (a) m/z spectra for the different organic factors during the measurement period, (b) Time series plots (c) Diurnal (d) Fractional contribution for the different PMF resolved factors of organic aerosol.

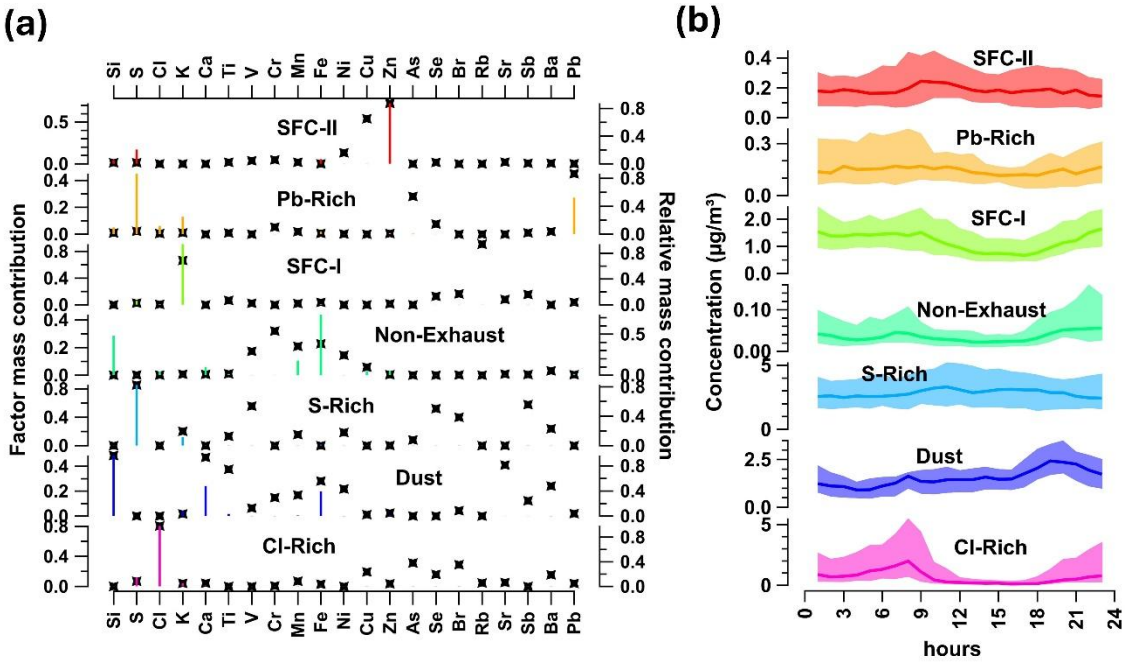


Figure R3.4. (a) Spectra profile for the different elemental factors during the measurement period, (b) Diurnal, (c) Time series plots, (d) Fractional contribution for the different PMF resolved factors of Trace elements.

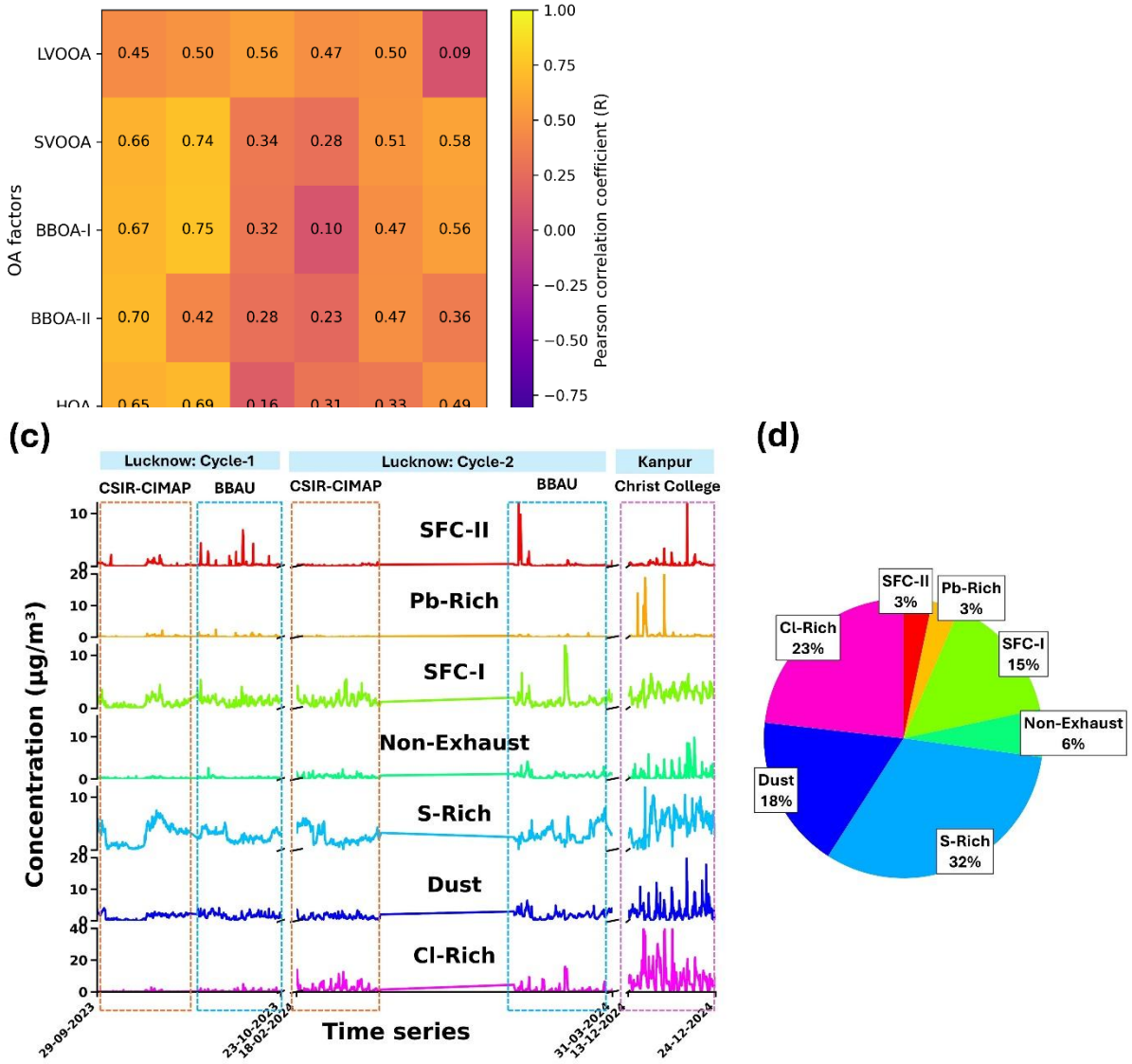


Figure R3.5. Heatmap between PMF resolved OA sources and external parameters during the campaign.

The reviewer is correct that the original manuscript did not clearly explain why the PMF source apportionment was performed using a **combined dataset from both monitoring sites**, whereas the subsequent machine learning models were evaluated separately for each site. We appreciate this opportunity to clarify the rationale.

The PMF analysis was performed on the **combined chemical composition dataset from Site-B (BBAU) and Site-C (CSIR)** to obtain a single, internally consistent set of source profiles representative of the entire Lucknow measurement campaign. Pooling the datasets increases the range of observed chemical variability and provides a more stable PMF solution than analysing each site independently, particularly when the individual deployments are relatively short. The resulting PMF factor profiles therefore represent common source signatures present across the two monitoring locations, while the corresponding time-resolved source contributions remain site-specific.

In the originally submitted manuscript, the subsequent machine learning evaluation was intentionally conducted separately for each site. The objective at that stage was to first establish whether low-cost sensor measurements could reproduce the site-specific PMF source contributions derived from the corresponding reference-grade measurements under controlled conditions, thereby serving as an initial proof-of-concept for the proposed framework.

Following the reviewers' suggestions, we have substantially expanded the scope of the machine learning evaluation in the revised manuscript. Rather than restricting the analysis to site-specific models, the revised framework now explicitly investigates model transferability through five independent transfer-learning experiments, including **cross-site (CSIR → BBAU and BBAU → CSIR)**, **cross-season (October → February–March and February–March → October)**, and **cross-city (Lucknow → Kanpur)** evaluations. In the principal experiment, the machine learning models are trained using the combined datasets from both Lucknow sites collected during the **October 2023** and **February–March 2024** campaigns and are subsequently evaluated on an entirely independent dataset collected at **Kanpur during December 2024**.

Consequently, the revised manuscript now directly investigates the question raised by the reviewer—namely, whether models trained using one spatial and temporal domain can generalize to different monitoring sites and seasons. The revised evaluation therefore extends considerably beyond the original site-specific proof-of-concept and provides a substantially more rigorous assessment of model transferability under realistic deployment conditions.

We thank the reviewer for raising this important point, which has significantly improved both the clarity of the manuscript and the scope of the experimental validation.

Changes to be made in the revised manuscript

- Clarify the rationale for performing PMF on the **combined Lucknow dataset**, emphasizing that pooled analysis yields a single, stable set of source profiles while preserving site-specific source contributions.
- Expand the description of the machine learning framework to explain the transition from the original site-specific proof-of-concept to the revised transfer-learning evaluation.
- Replace the original site-specific machine learning experiments with the new **cross-site, cross-season, and cross-city** transfer-learning experiments.
- Revise the discussion to emphasize that the proposed framework is now evaluated for both **site-specific prediction** and **generalization across independent spatial and temporal domains**.

Concern #6: Representation Biases and Scale Inadequacies in Visualizations.

The graphical methods employed to display and evaluate model output do not provide an objective or statistically comprehensive representation of framework performance. The source contribution comparisons presented in the pie charts (Figures 8, 10, 12, and 14) provide only localized snapshots of model performance. Displaying isolated, single 30-minute data steps selected at disparate times across figures introduces significant selection bias and prevents an objective evaluation of systemic model agreement. These snapshots should be replaced or supplemented with the averages integrated across the entire testing window, or broken down into multi-hour diurnal blocks to demonstrate sustained predictive skill.

In Figure 11, the comparative data resolution appears visibly lower than the defined 30-minute sampling sequence, creating prominent temporal gaps that remain unaddressed in the text. In the time-series trends (such as Figure 13), several trace emission factors—including Pb-rich, Fe-smelting, and Cl-rich sources—exhibit fractional contributions that hover near zero. The absolute scaling of these plots makes it visually impossible to determine whether the model is capturing temporal trends or merely predicting flat baselines. The authors should supplement these figures with explicit cross-correlation scatter plots, residual distribution profiles, or optimized y-axis scale ranges to properly demonstrate real-time tracking fidelity.

Response: We sincerely thank the reviewer for these thoughtful comments regarding the presentation and visualization of the machine learning results.

We agree that the pie-chart comparisons presented in the originally submitted manuscript provide only instantaneous snapshots of the predicted source contributions and therefore do not fully characterize the overall predictive performance of the proposed framework. Although these figures were originally intended as qualitative illustrations of the model output at representative time instants, we agree that quantitative evaluation over the complete testing period provides a more objective assessment.

In response to the reviewer's suggestion, the experimental evaluation has been substantially revised. The original short-duration site-specific experiments have been replaced by a significantly more comprehensive set of **cross-city, cross-site and cross-season transfer-learning experiments**, using considerably larger independent test datasets. The revised manuscript therefore presents model performance over substantially longer testing periods than those considered in the original submission.

Furthermore, rather than relying primarily on qualitative visual comparisons, the revised manuscript emphasizes quantitative performance evaluation using multiple complementary metrics, including **Mean Absolute Error (MAE)**, **Spearman Rank Order Correlation Coefficient (SROCC)** and **Normalized Discounted Cumulative Gain (NDCG)**, computed over the entire independent test dataset. These metrics provide a considerably more comprehensive and objective assessment of prediction accuracy than isolated visual snapshots.

The reviewer also noted the apparent temporal gaps observed in Figure 11. As discussed in our response to Reviewer 2 (Concern 2), these discontinuities resulted from the limited availability of valid reference PMF source apportionment estimates during portions of the original short-duration testing period. In the revised manuscript, these original experiments have been superseded by the new transfer-learning evaluation framework based on substantially larger datasets, and the corresponding time-series figures no longer exhibit these visual discontinuities.

We also appreciate the reviewer's observation regarding the visualization of relatively small source contributions. In the revised manuscript, the original **five-factor organic source apportionment** has been replaced by a chemically more interpretable **three-factor representation (BBOA, HOA and OOA)**. This revision not only improves the machine learning performance but also provides considerably clearer visualization of the predicted source contributions. In addition, the revised manuscript incorporates **feature importance analysis** and **feature ablation studies**, which provide complementary quantitative insight into model behaviour beyond the time-series visualizations.

Finally, we agree that scatter plots provide an effective means of illustrating the agreement between predicted and reference source contributions. Accordingly, the revised manuscript will include representative scatter plots for the principal **cross-city transfer-learning experiment (Lucknow → Kanpur)**, together with the corresponding quantitative performance metrics. These additional visualizations complement the existing time-series comparisons while maintaining a reasonable manuscript length.

We sincerely thank the reviewer for these constructive suggestions, which have led to a clearer, more rigorous and more objective presentation of the machine learning results.

Changes to be made in the revised manuscript:

- Replace the original short-duration site-specific visualizations with results from the revised **cross-city, cross-site and cross-season** experiments.
- Present quantitative performance primarily using **MAE**, **SROCC** and **NDCG** computed over the complete independent test datasets.
- Replace the original **five-factor organic SA** with the revised **three-factor (BBOA–HOA–OOA)** framework.
- Include representative **predicted-versus-reference scatter plots** for the principal cross-city experiment.
- Add **feature importance** and **feature ablation** analyses to provide additional quantitative interpretation of model behaviour.

References:

Chow, J.C., 1995. Measurement methods to determine compliance with ambient air quality standards for suspended particles. J. Air Waste Manag. Assoc. 45, 320–382. https://doi.org/10.1080/10473289.1995.10467369

Davison, A.C., Hinkley, D. V., 1997. Bootstrap Methods and Their Application. Biometrics 54, 795. https://doi.org/10.2307/3109789

Decarlo, P.F., Kimmel, J.R., Trimborn, A., Northway, M.J., Jayne, J.T., Aiken, A.C., Gonin, M., Fuhrer, K., Horvath, T., Docherty, K.S., Worsnop, D.R., Jimenez, J.L., 2006. Field-deployable, high-resolution, time-of-flight aerosol mass spectrometer. Anal. Chem. 78, 8281–8289. https://doi.org/10.1021/ac061249n

DeCarlo, P.F., Kimmel, J.R., Trimborn, A., Northway, M.J., Jayne, J.T., Aiken, A.C., Gonin, M., Fuhrer, K., Horvath, T., Docherty, K.S., Worsnop, D.R., Jimenez, J.L., 2006. Field-deployable, high-resolution, time-of-flight aerosol mass spectrometer. Anal. Chem. 78, 8281–8289. https://doi.org/10.1021/ac061249n

Goetz, J.D., Giordano, M.R., Stockwell, C.E., Bhave, P. V, Puppala, P.S., Panday, A.K., Jayarathne, T., Stone, E.A., Yokelson, R.J., Decarlo, P.F., 2022. Aerosol Mass Spectral Profiles from NAMaSTE Field-Sampled South Asian Combustion Sources. Environ. Sci. Technol. https://doi.org/10.1021/acsearthspacechem.2c00173

Guo, J., Zhou, S., Cai, M., Zhao, J., Song, W., Zhao, W., Hu, W., Sun, Y., He, Y., Yang, C., Xu, X., Zhang, Z., Cheng, P., Fan, Q., Hang, J., Fan, S., Wang, Xinming, Wang, Xuemei, 2020. Characterization of submicron particles by time-of-flight aerosol chemical speciation monitor (ToF-ACSM) during wintertime: Aerosol composition, sources, and chemical processes in Guangzhou, China. Atmos. Chem. Phys. 20, 7595–7615. https://doi.org/10.5194/acp-20-7595-2020

Jayne, J.T., Leard, D.C., Zhang, X., Davidovits, P., Smith, K.A., Kolb, C.E., Worsnop, D.R., 2000. Development of an aerosol mass spectrometer for size and composition analysis of submicron particles. Aerosol Sci. Technol. 33, 49–70. https://doi.org/10.1080/027868200410840

Kadachi, A.N., Al-eshaikh, M.A., 2012. Limits of detection in XRF spectroscopy. X-Ray Spectrom. 350–354. https://doi.org/10.1002/xrs.2412

Lakra, A., Kumar, A., Sethi, D., Sekhar, H., Jain, V., Tripathi, S.N., 2026. High-resolution source apportionment and health risks of PM 2 . 5 -bound trace elements across a major Indian city : Seasonal and diurnal insights from a multi-site campaign using a mobile laboratory platform. Atmos. Environ. 365, 121675. https://doi.org/10.1016/j.atmosenv.2025.121675

Lakra, A., Shukla, A.K., Bhowmik, H.S., Yadav, A.K., Jain, V., Murari, V., Gaddamidi, S., Lalchandani, V., Tripathi, S.N., 2024. Comparative analysis of winter composite-PM2.5 in Central Indo Gangetic Plain cities: Combined organic and inorganic source apportionment and characterization, with a focus on the photochemical age effect on secondary organic aerosol formation. Atmos. Environ. 338, 120827. https://doi.org/10.1016/j.atmosenv.2024.120827

Lalchandani, V., Srivastava, D., Dave, J., Mishra, S., Tripathi, N., Shukla, A.K., Sahu, R., Thamban, N.M., Gaddamidi, S., Dixit, K., Ganguly, D., Tiwari, S., Srivastava, A.K., Sahu, L., Rastogi, N., Gargava, P., Tripathi, S.N., 2022. Effect of Biomass Burning on PM2.5 Composition and Secondary Aerosol Formation During Post-Monsoon and Winter Haze Episodes in Delhi. J. Geophys. Res. Atmos. 127. https://doi.org/10.1029/2021JD035232

Li, Q., Cheng, H., Zhou, T., Lin, C., Guo, S., 2012. The estimated atmospheric lead emissions in China, 1990-2009. Atmos. Environ. 60, 1–8. https://doi.org/10.1016/j.atmosenv.2012.06.025

Li, Y.J., Lee, B.Y.L., Yu, J.Z., Ng, N.L., Chan, C.K., 2013. Evaluating the degree of oxygenation of organic aerosol during foggy and hazy days in Hong Kong using high-resolution time-of-flight aerosol mass spectrometry (HR-ToF-AMS). Atmos. Chem. Phys. 13, 8739–8753. https://doi.org/10.5194/acp-13-8739-2013

Manchanda, C., Kumar, M., Singh, V., Faisal, M., Hazarika, N., Shukla, A., Lalchandani, V., Goel, V., Thamban, N., Ganguly, D., Tripathi, S.N., 2021. Variation in chemical composition and sources of PM2.5 during the COVID-19 lockdown in Delhi. Environ. Int. 153, 106541. https://doi.org/10.1016/j.envint.2021.106541

Mukta, T.A., Hoque, M.M., Sarker, E., Hossain, N., 2021. Seasonal variations of gaseous air pollutants (SO 2 , NO 2 , O 3 , CO) and particulates (PM 2 . 5 , PM 10) in Gazipur : an industrial city in Bangladesh. Adv. Environ. Technol. 4, 195–209. https://doi.org/10.22104/AET.2021.4890.1320

Rai, P., Furger, M., El Haddad, I., Kumar, V., Wang, L., Singh, A., Dixit, K., Bhattu, D., Petit, J.E., Ganguly, D., Rastogi, N., Baltensperger, U., Tripathi, S.N., Slowik, J.G., Prévôt, A.S.H., 2020. Real-time measurement and source apportionment of elements in Delhi's atmosphere. Sci. Total Environ. 742. https://doi.org/10.1016/j.scitotenv.2020.140332

Rai, P., Slowik, J.G., Furger, M., El Haddad, I., Visser, S., Tong, Y., Singh, A., Wehrle, G., Kumar, V., Tobler, A.K., Bhattu, D., Wang, L., Ganguly, D., Rastogi, N., Huang, R.J., Necki, J., Cao, J., Tripathi, S.N., Baltensperger, U., Prevot, A.S.H., 2021. Highly time-resolved measurements of element concentrations in PM10 and PM2.5: Comparison of Delhi, Beijing, London, and Krakow. Atmos. Chem. Phys. 21, 717–730. https://doi.org/10.5194/acp-21-717-2021

Sethi, D., Lakra, A., Kumar, A., Chakraborty, S., Sekhar, H., Jain, V., Tripathi, S.N., 2026. Hyperlocal air quality monitoring and source apportionment of non- refractory $PM_{2.5}$ at three urban sites using stationary van - based measurements : A Lucknow case study. Atmos. Chem. Phys. 1–41. https://doi.org/10.5194/egusphere-2026-595

Sharma, S.K., Mandal, T.K., 2023. Elemental Composition and Sources of Fine Particulate Matter (PM2.5) in Delhi, India. Bull. Environ. Contam. Toxicol. 110, 1–8. https://doi.org/10.1007/s00128-023-03707-7

Shukla, A.K., Lalchandani, V., Bhattu, D., Dave, J.S., Rai, P., Thamban, N.M., Mishra, S., Gaddamidi, S., Tripathi, N., Vats, P., Rastogi, N., Sahu, L., Ganguly, D., Kumar, M., Singh, V., Gargava, P., Tripathi, S.N., 2021a. Real-time quantification and source apportionment of fine particulate matter including organics and elements in Delhi during summertime. Atmos. Environ. 261, 118598. https://doi.org/10.1016/j.atmosenv.2021.118598

Shukla, A.K., Lalchandani, V., Bhattu, D., Dave, J.S., Rai, P., Thamban, N.M., Mishra, S., Gaddamidi, S., Tripathi, N., Vats, P., Rastogi, N., Sahu, L., Ganguly, D., Kumar, M., Singh, V., Gargava, P., Tripathi, S.N., 2021b. Real-time quantification and source apportionment of fine particulate matter including organics and elements in Delhi during summertime. Atmos. Environ. 261. https://doi.org/10.1016/j.atmosenv.2021.118598

Shukla, A.K., Tripathi, S.N., Canonaco, F., Lalchandani, V., Sahu, R., Srivastava, D., Dave, J., Thamban, N.M., Gaddamidi, S., Sahu, L., Kumar, M., Singh, V., Rastogi, N., 2023. Spatio-temporal variation of C-PM2.5 (composition based PM2.5) sources using PMF*PMF (double-PMF) and single-combined PMF technique on real-time non-refractory, BC and elemental measurements during post-monsoon and winter at two sites in Delhi, India. Atmos. Environ. 293, 119456. https://doi.org/10.1016/j.atmosenv.2022.119456

Tobler, A., Bhattu, D., Canonaco, F., Lalchandani, V., Shukla, A., Thamban, N.M., Mishra, S., Srivastava, A.K., Bisht, D.S., Tiwari, S., Singh, S., Močnik, G., Baltensperger, U., Tripathi, S.N., Slowik, J.G., Prévôt, A.S.H., 2020. Chemical characterization of PM2.5 and source apportionment of organic aerosol in New Delhi, India. Sci. Total Environ. 745. https://doi.org/10.1016/j.scitotenv.2020.140924