Supplementary material

Unravelling the Mechanisms and Chemical Drivers of New Particle Formation in Chennai, a Tropical Coastal Megacity in India influenced by an Urban Forest

Mangattayil Devaprasad[1,2,*,#], Amrutha M. Sathyan[1,2,#], Tanmay Kishor Joshi[1,2], Supriya Dey[1,2], Geethanjali Rajendran Malavika[1,2], Yougal Sapkota[2,3], Rizana Salim[1,2], Shailina Srivastava[1,2], Insha Amin[1,2], Govindan Pandithurai[2], Ravikrishna Raghunathan[2,3], Pengfei Liu[4], and Sachin S. Gunthe[1,2,*]

[1]Environmental Engineering Division, Department of Civil Engineering, Indian Institute of Technology Madras, Chennai, India

[2]Center for Atmospheric and Climate Sciences, Indian Institute of Technology Madras, Chennai, India

[3]Department of Chemical Engineering, Indian Institute of Technology Madras, Chennai, India

[4]School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA, USA

[#]These authors contributed equally to this work.

* Correspondence to: devaprasad@cacs.iitm.ac.in; s.gunthe@iitm.ac.in;

This Supplementary Information contains:

1) Detailed methodology for VOC measurements
2) Derived NPF parameters
3) Auxiliary calculations
4) Supplementary figures
5) Detailed PMF analysis information

1.1 Detailed methodology for VOC measurements

Gas-phase measurements were conducted using an inlet positioned approximately 4 m above from the terrace of the Engineering Design building within the IIT Madras campus. Ambient air was sampled through a multi-stage inlet system designed to minimize particle losses and adsorption of semi-volatile compounds. The inlet section was continuously heated to 80 °C to prevent condensation and wall losses. Inline PTFE particle filters were installed upstream of the instrument to prevent the ingress of coarse particles, insects, and debris, and these filters were replaced on a weekly basis**.**

Volatile organic compounds (VOCs) were quantified using a PTR-TOF-MS 6000 X2 (Ionicon Analytik GmbH, Austria), a soft chemical-ionization technique for real-time trace-VOC detection at ppbv–pptv levels (De Gouw and Warneke, 2007; Graus et al., 2010; Hansel et al., 1995; Lindinger et al., 1998). Ambient air was introduced directly into the drift tube, where VOCs were ionized by proton transfer from hydronium ions (H_3O^+) when their proton affinity exceeded that of water; the major constituents of air, having lower proton affinities, remained unionized, enabling selective detection with minimal fragmentation.

Hydronium reagent ions were generated by directing a continuous flow of 6 sccm of high-performance liquid chromatography (HPLC) grade water into the ion source region. Ion–molecule reactions occurred within the drift tube under controlled electric field and pressure conditions. During the measurements, the drift tube pressure and voltage were maintained at 2.8 mbar and 530 V, respectively, corresponding to an E/N ratio of approximately 130 Townsend (Td). This E/N value represents a compromise between suppressing excessive water-cluster formation and avoiding ion fragmentation, consistent with established PTR-MS operational guidelines. Under these conditions, the relative abundances of NO^+, O_2^+, and $H_3O^+ \cdot (H_2O)_n$ cluster ions were maintained below 1%, 8%, and 12%, respectively, of the primary ion signal, ensuring stable ion chemistry throughout the measurement period.

The sample flow into the PTR-TOF-MS was regulated using the IoniTOF control software, with the flow-control inlet maintained at approximately 300 sccm. Mass spectra were acquired over a range of m/z 21–500, allowing detection of a wide variety of biogenic and anthropogenic VOCs. VOC measurements were performed at a time resolution of 30 seconds, enabling the capture of rapid temporal variability relevant to photochemical processing and new particle formation events.

Mass calibration: The mass axis was calibrated to ensure accurate alignment between the measured and exact ion m/z, using one low-mass and one high-mass reference ion: m/z 21.0221, and m/z 203.9431, a characteristic high-mass ion generated by the PerMaSCal calibration unit (diiodobenzene source). These m/z regions were selected as they are generally free from interfering signals. The mass calibration was subsequently checked and refined during post-processing.

Sensitivity calibration: Instrument sensitivity was determined by multipoint calibration using a certified multicomponent VOC standard (SIAD; ≈1 ppmv per compound in nitrogen) diluted with VOC-free zero air. Zero air was produced with a zero-air generator (Parker Balston, Model 75-83NA): compressed air was passed through a PTFE filter into the generator and further scrubbed through an activated-charcoal trap (Supelpure HC, Supelco) to maintain low VOC backgrounds. The standard gas and zero air were combined at controlled flow rates through PTFE calibration lines (inner diameter 3.175 mm) to generate mixing ratios of approximately 10, 20, 30, and 40 ppbv. The calibration consisted of a zero-air background measurement (~30 min) followed by introduction of the diluted standard (~120 min) and a final zero-air measurement, with the instrument operated at a 1 s acquisition cycle. The standard cylinder contained compounds: methanol, acetonitrile, acetone, isoprene, benzene, toluene, o-/m-/p-xylene, 1,2,4-trimethylbenzene, 1,2-dichlorobenzene, and 1,2,4-trichlorobenzene (with perfluorotributylamine used as a mass reference).

Of the VOC species reported here, methanol, acetone, isoprene, benzene, toluene, xylenes, and acetonitrile were quantified directly against the standard. Mixing ratios of the remaining reported species (monoterpenes, MVK + MACR, and DMS), which were not present in the standard, were estimated using established proton-transfer-reaction rate constants, assuming unit transmission efficiency.

Transmission Factors: Ion transmission through the instrument system is mass dependent, with lighter ions being transmitted more efficiently than heavier ions. Therefore, it is necessary to quantify this transmission efficiency. Transmission factors are determined from a transmission factor curve (as shown in Fig S1) using the same gas standard containing compounds spanning a wide mass range as mentioned above. The calculated concentrations are processed using PTR-MS Viewer 3.4.6, and the corresponding transmission factors are derived and subsequently applied to correct the concentrations of the measured compounds.

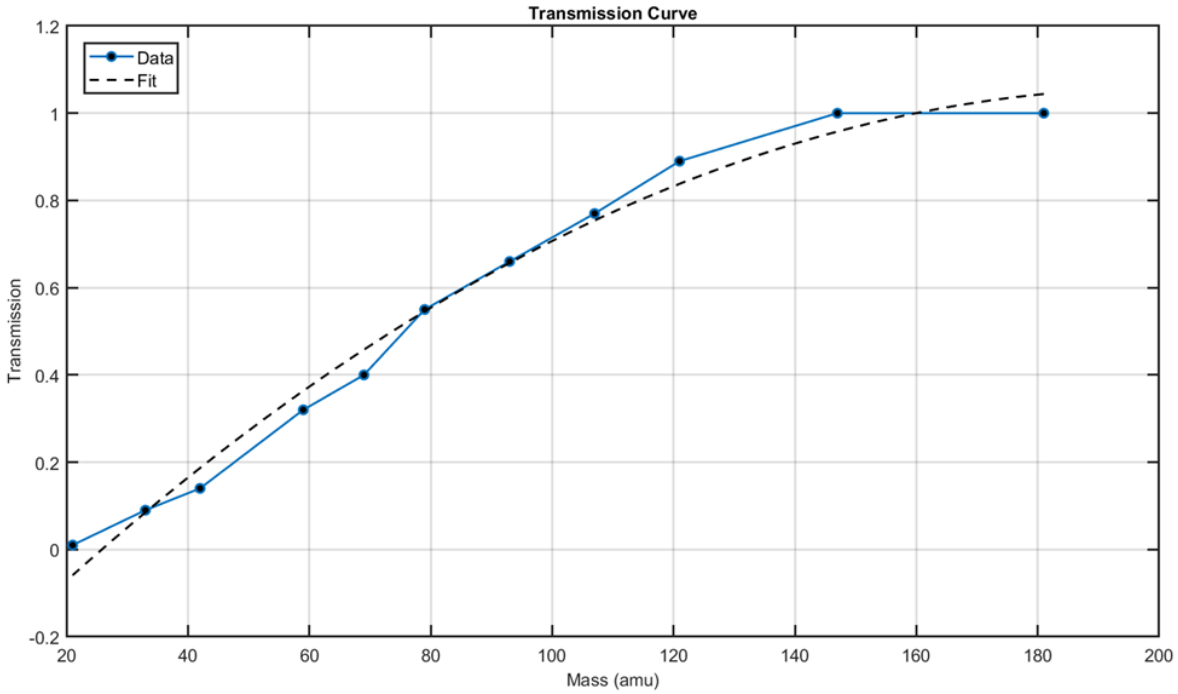

Figure S1: Transmission efficiency curve using hydronium (m/z 21), methanol (m/z 33), Acetonitrile (m/z 42), acetone (m/z 59), isoprene (m/z 69), benzene (m/z 79), toluene (m/z 93), O-P-M-xylene (m/z 107), 1,2,4-trimethylbenzene (m/z 121), 1,2-Dichlorobenzene (m/z 147), 1,2,4-Trichlorobenzene (m/z 181).

1.2 Derived NPF parameters

1.2.1 Particle Growth Rate

The particle growth rate (GR; nm h^{-1}) was determined from the temporal evolution of the geometric mean diameter (GMD) of the growing nucleation/Aitken mode during NPF event days, following Kulmala et al. (2012). Unlike approaches that estimate GR over a fixed particle size range (e.g., 10–30 nm), GR in this study was derived from the continuous temporal evolution of the particle mode, beginning from the lowest detectable GMD and extending until continuous growth was sustained. Such clear and continuous growth was observed only during the two NPF event days (26 May and 14 June 2025). For each event day, a least-squares linear regression was applied to the GMD time series during the sustained linear growth phase identified from the contour plot of the particle number size distribution (Fig. 3); the slope of this fit yields the GR. For non-event days, in which sustained linear growth was not observed, a campaign mean GR of 4.0 nm h^{-1} was used in the $J_{10–30}$ calculation (Sect. 2.4.4).

1.2.2 Condensation Sink

The condensation sink (CS, in s^{-1}) represents the rate at which condensable vapours are scavenged by the pre-existing aerosol population and was calculated following Dal Maso et al. (2002) and Sun et al. (2024) as

	$CS = 2\pi D \sum_i \beta_i \, d_{p,i} \, N_i$	Eq1

where D is the diffusion coefficient of the condensing vapor (here taken as sulfuric acid, H_2SO_4), $d_{p,i}$ (cm) and N_i (cm^{-3}) are the diameter and number concentration of particles in size bin i, and β_i is the Fuchs–Sutugin transition-regime correction factor evaluated as a function of the Knudsen number $Kn_i = 2\lambda/d_{p,i}$, with λ the mean free path of air (Seinfeld, 2016). Air dynamic viscosity was computed from the real-time automated weather station (AWS) temperature using Sutherland's law, and D was obtained via the Stokes–Einstein relation assuming an H_2SO_4 molecular diameter of 0.55 nm. The summation in Eq. (1) was performed over the full reliable SMPS size range (10–280 nm), thereby including the contribution of all detected particles — including those in the nucleation mode — to the total scavenging surface area.

1.2.3 Coagulation Sink

The coagulation sink for a 10 nm particle ($CoagS_{10\ nm}$, in s^{-1}) was calculated following Kerminen and Kulmala, (2002) as

	$\mathrm{CoagS}_{10\,\mathrm{nm}} = \sum_i K(10\,\mathrm{nm}, d_{p,i})\ N_i$	Eq2

where K(10 nm, dp,i) is the Brownian coagulation kernel between a 10 nm test particle and a background particle of diameter dp,i evaluated in the free-molecular regime, and Ni is the number concentration of background particles in bin i (cm^{-3}). Particle thermal velocities were computed from the Maxwell–Boltzmann distribution assuming a particle density of 1500 kg m^{-3}, with real-time AWS temperature used at every time step. The summation was performed over the full reliable SMPS size range (10–280 nm).

1.2.4 Particle Formation Rate

The particle formation rate in the 10–30 nm size window ($J_{10–30}$, in cm^{-3} s^{-1}) was calculated following Sun et al. (2024), based on Kulmala et al. (2012) in Eq. 3.

	$J_{10-30} = \frac{dN_{10-30}}{dt} + \mathrm{CoagS}_{10\,\mathrm{nm}}\, N_{10-30} + \frac{\mathrm{GR}}{\Delta D_p}\, N_{10-30}$	Eq3

where $N_{10–30}$ is the integrated nucleation-mode number concentration (cm^{-3}), d$N_{10–30}$/dt is its central-difference time derivative on the time series (calculated over two scan intervals), GR is the measured particle growth rate (in nm s^{-1}), and ΔDp = 20 nm is the width of the nucleation mode. The three terms on the right-hand side represent, respectively, the net rate of change of the nucleation-mode population, the loss of nucleation-mode particles by coagulation onto larger background particles, and the loss by growth out of the upper boundary of the nucleation mode.

1.2.5 Sulfuric Acid Proxy Calculation

Sulfuric acid proxy concentrations were estimated following the parameterization proposed by Mikkonen et al. (2011), which has been widely applied in studies where direct measurements of gaseous sulfuric acid are unavailable. The sulfuric acid proxy was calculated using measured SO_2, condensation sink (CS), relative humidity (RH), and solar radiation, as shown in Eq (4).

	$[H_2SO_4] = 8.21 \times 10^{-3} \times kR \times [SO_2]^{0.62} \times [CS \times RH]^{-0.13}$	Eq4

where [H_2SO_4] is the sulfuric acid proxy concentration (molecules cm^{-3}), R is the solar radiation (W m^{-2}), [SO_2] is the sulfur dioxide concentration (ppb), CS is the condensation sink (s^{-1}), RH is the relative humidity (%), and k is the reaction coefficient for the SO_2 + OH reaction. Solar radiation and RH were obtained from ERA5 reanalysis data and interpolated to the temporal resolution of the gas-phase measurements. The condensation sink was calculated from the measured particle number size distributions. The estimated values represent sulfuric acid proxy concentrations rather than direct measurements of gaseous H_2SO_4; absolute concentrations carry an approximately factor-of-two uncertainty arising from the underlying OH parameterization, although the relative ordering between

day classes is robust to this calibration. The diurnal evolution of the proxy and its event and non-event composites are shown in Fig. S19.

1.3 Auxiliary Calculations

1.3.1 Pollution Wind Rose Analysis

The directional dependence of volatile organic compounds (VOCs), trace gases, aerosol chemical species, and PMF factor contributions on local meteorology was investigated using pollution wind rose analysis at hourly resolution. Concentration data were merged with hourly averaged meteorological parameters obtained from the AWS. Wind speed and wind direction data were quality controlled by removing invalid and negative values, and hourly wind vectors (U and V components) were calculated and averaged prior to reconstructing hourly wind speed and direction. Pollution wind roses were generated in Python using the windrose package, where wind directions were grouped into 22.5° sectors and the radial axis represents the normalized frequency of occurrence within different concentration bins. For PMF factor wind roses, the upper quartile (75th percentile and above) of factor contributions was additionally examined to enhance the identification of dominant source-direction relationships associated with high-concentration episodes.

1.3.2 Aerosol ion balance

Aerosol acidity was assessed via the equivalent ammonium-to-anion ratio (NH_4^+_meas / NH_4^+_pred), where NH_4^+_pred is the molar equivalent ammonium required to fully neutralize the co-measured sulfate, nitrate, and chloride:

	$\mathrm{NH_4^+}_{\mathrm{pred}} = 18\left(2\,\frac{\mathrm{SO_4^{2-}}}{96} + \frac{\mathrm{NO_3^-}}{62} + \frac{\mathrm{Cl^-}}{35.5}\right)$	Eq5

with all species in μg m^{-3}. Ratios below unity indicate insufficient ammonium for full neutralization (i.e., acidic submicron aerosol), while ratios near or above unity indicate near-neutral aerosol.

1.3.3 OH Reactivity

Volatile organic compounds (VOCs) serve as a sink for hydroxyl radicals (OH), which are the primary oxidizing agents in the atmosphere, although a portion of the consumed OH is regenerated during VOC oxidation. Additionally, reactions between VOCs and OH influence the formation of ozone and SOA. OH reactivity is defined as the rate at which OH radicals are removed from the atmosphere and is given in Eq. 6.

	$\mathrm{OH\ Reactivity}\ (s^{-1}) = k_{OH+X_i}[OH][X_i]$	Eq6

where $k_{[Xi + OH]}$ denotes the reaction rate coefficient between the reacting species and OH and [Xi] represents the concentration (molecules cm^{-3}) of that species. The reaction rate coefficients $k_{[Xi + OH]}$

used for calculating total OH reactivity were obtained from the literature and are summarized in Table S1. These coefficients represent the rate constants for the reactions of individual VOC species with OH radicals under atmospheric conditions.

Species	Molecular Formulae	Rate Coefficient with OH	Reference
Carbon-monoxide	CO	1.5×10^{-13}	(Atkinson et al., 2006)
Nitrogen Oxides	NO_x	1×10^{-11}	(Burkholder et al., 2015)
Ozone	O_3	7.8×10^{-14}	(Burkholder et al., 2015)
Methanol	CH_4O	9×10^{-13}	(Atkinson et al., 2006)
Acetonitrile	C_2H_3N	2.2×10^{-14}	(Atkinson et al., 2006)
Acetaldehyde	C_2H_4O	1.5×10^{-11}	(Atkinson et al., 2006)
Acetone	C_3H_6O	1.8×10^{-13}	(Atkinson et al., 2006)
Isoprene	C_5H_8	1×10^{-10}	(Atkinson et al., 2006)
MVK + MACR	C_4H_6O	3.5×10^{-11}	(Burkholder et al., 2015)
Benzene	C_6H_6	1.2×10^{-12}	(Atkinson et al., 2006)
Toluene	C_7H_8	5.6×10^{-12}	(Atkinson et al., 2006)
Xylene	C_8H_{10}	1.3×10^{-11}	(Atkinson et al., 2006)
Monoterpene	$C_{10}H_{16}$	5.3×10^{-11}	(Atkinson et al., 2006)

1.4 Supplementary Figures

Figure S2: Map of the sampling location at the boundary of Guindy National Park, Chennai, India (red triangle). Basemap: Google Maps. Imagery © 2026 Airbus, CNES / Airbus, Maxar Technologies; Map data © 2026 Google.

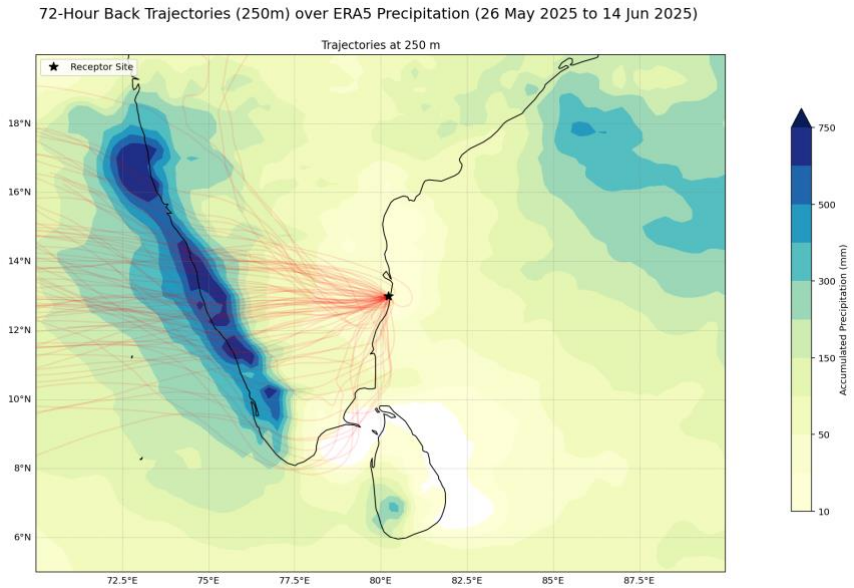

Figure S3: Map showing 72-h HYSPLIT back trajectories arriving at the receptor site in Chennai (black star) during the campaign period (26 May – 14 June 2025). Trajectories were calculated for an arrival height of 250 m above ground level. The background colour shows total accumulated precipitation (mm) over the same period from the ERA5 reanalysis.

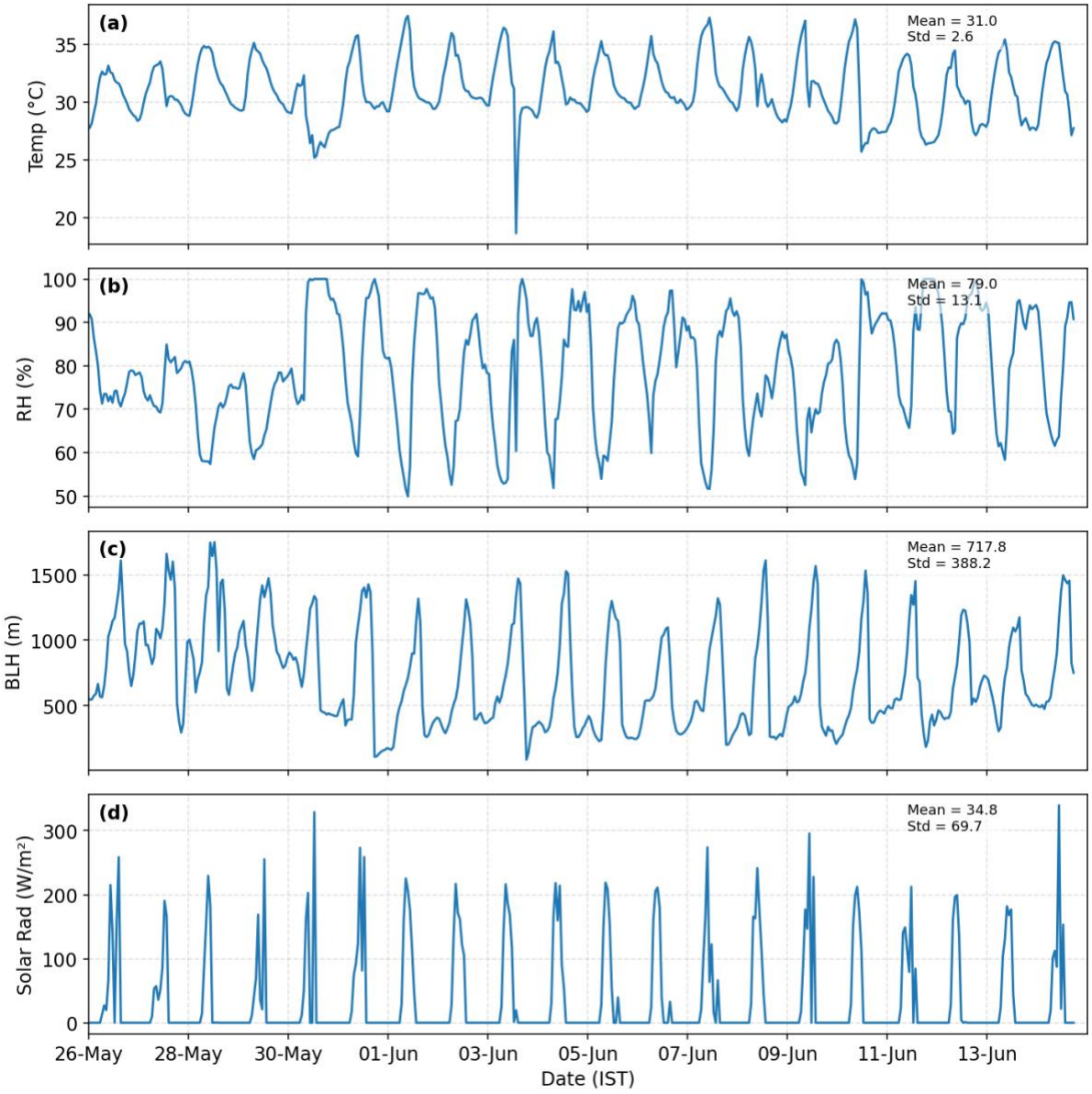

Figure S4: Time series of meteorological parameters during the campaign. Temperature and relative humidity are from the on-site automated weather station; boundary layer height (BLH) and surface solar radiation downward (SSRD) are from the ERA5 reanalysis at the nearest grid point.

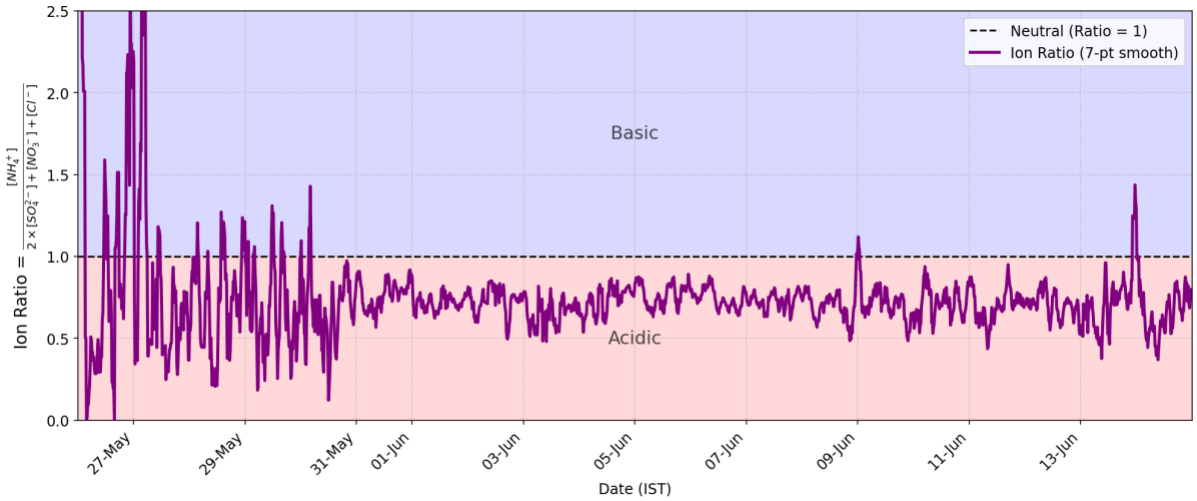

Figure S5. Time series of the aerosol ion balance ratio (NH_4^+_meas / NH_4^+_pred) from 26 May to 14 June 2025. The purple line shows the 7-point moving average. The dashed reference line at ratio = 1 indicates full neutralisation; red and blue shading mark acidic and basic regimes, respectively.

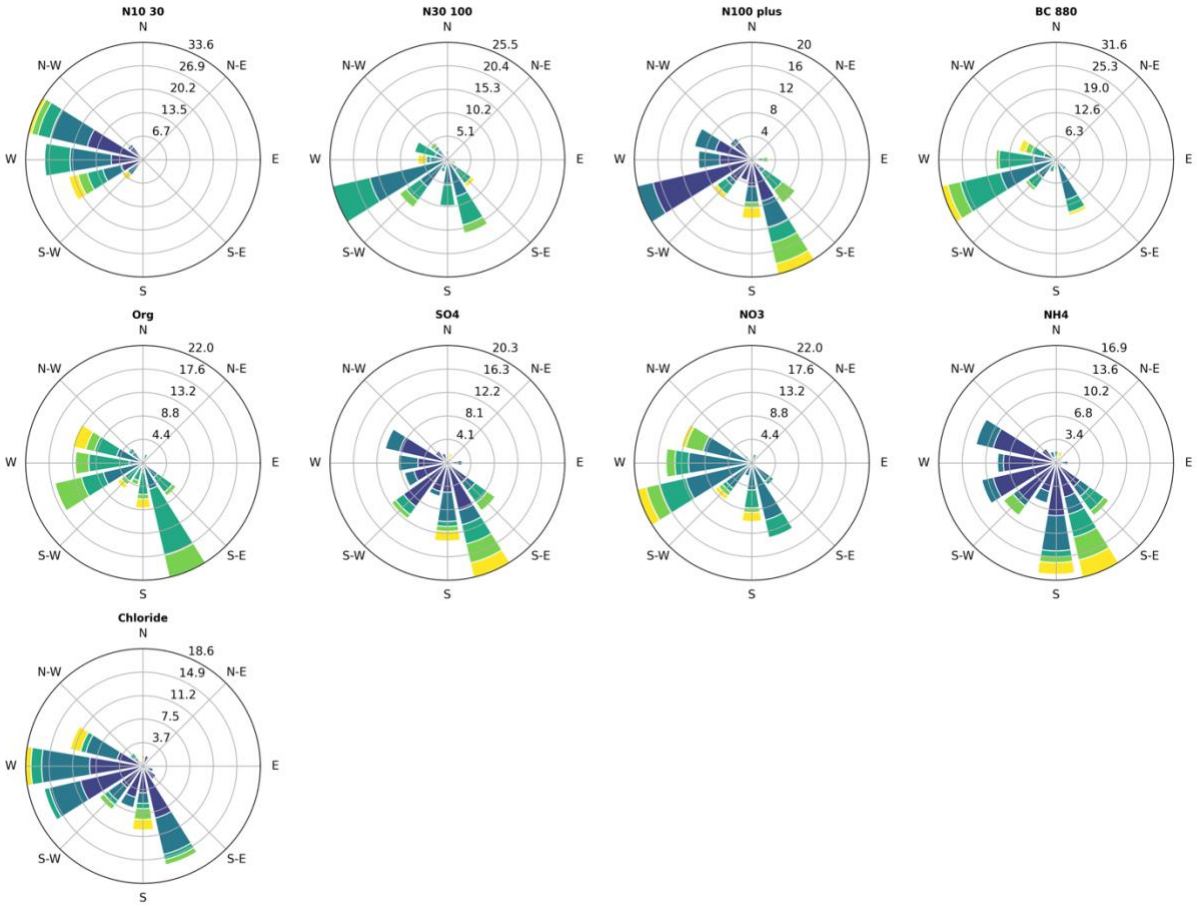

Figure S6. Pollution wind roses for particle number concentrations ($N_{10–30}$, $N_{30–100}$, $N_{100–280}$), equivalent black carbon at 880 nm (BC_{880}), and major non-refractory PM_1 components (organics, SO_4^{2-}, NO_3^-, NH_4^+, chloride) during the campaign. The radial axis shows the

normalised frequency of occurrence (%); colour bins indicate concentration ranges. Wind direction and speed are from the automated weather station.

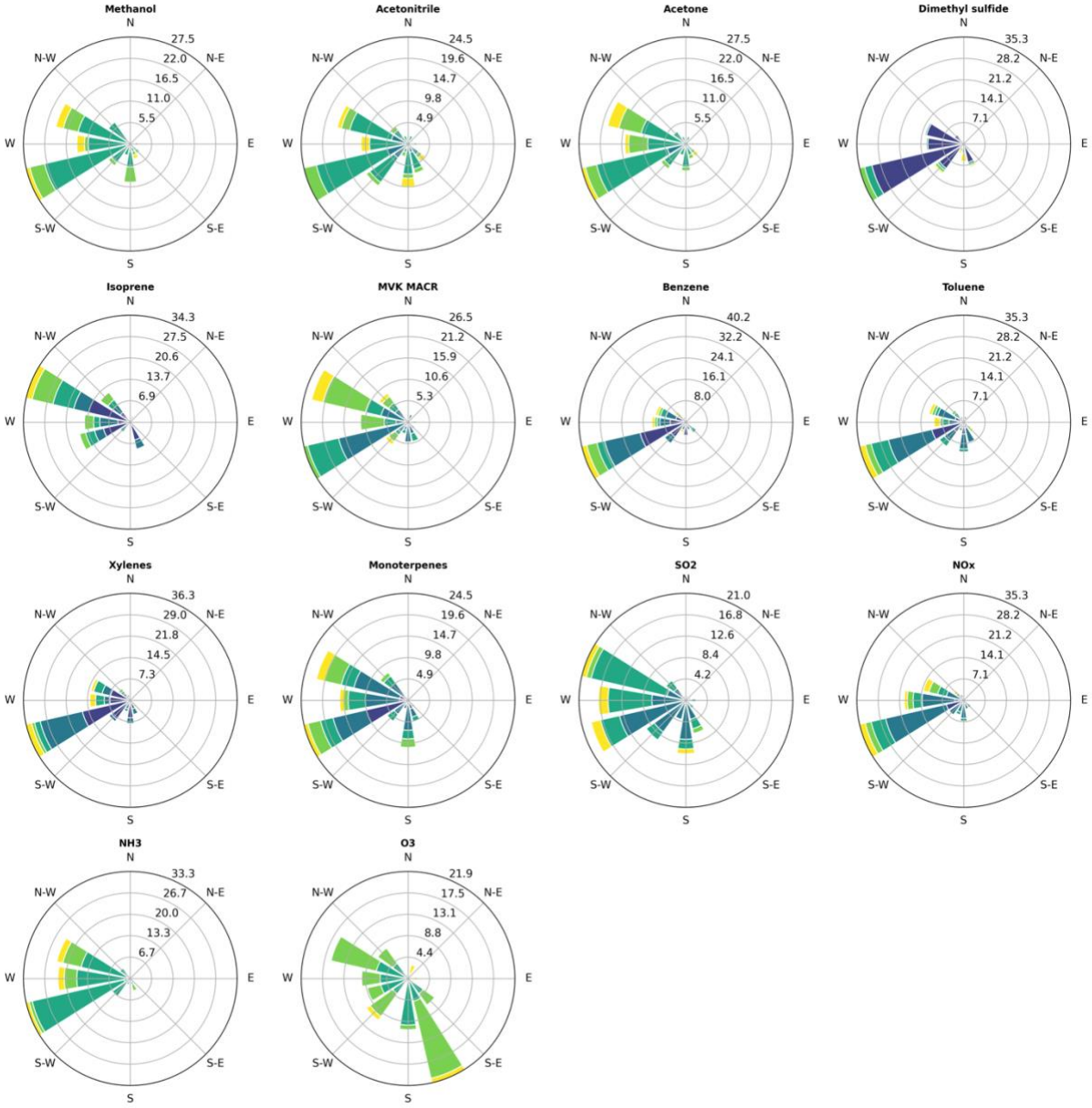

Figure S7. Pollution wind roses for VOCs (methanol, acetonitrile, acetone, DMS, isoprene, MVK + MACR, benzene, toluene, xylenes, monoterpenes) and trace gases (SO_2, NO_x, NH_3, O_3) during the campaign. The radial axis shows the normalised frequency of occurrence (%); colour bins indicate concentration ranges.

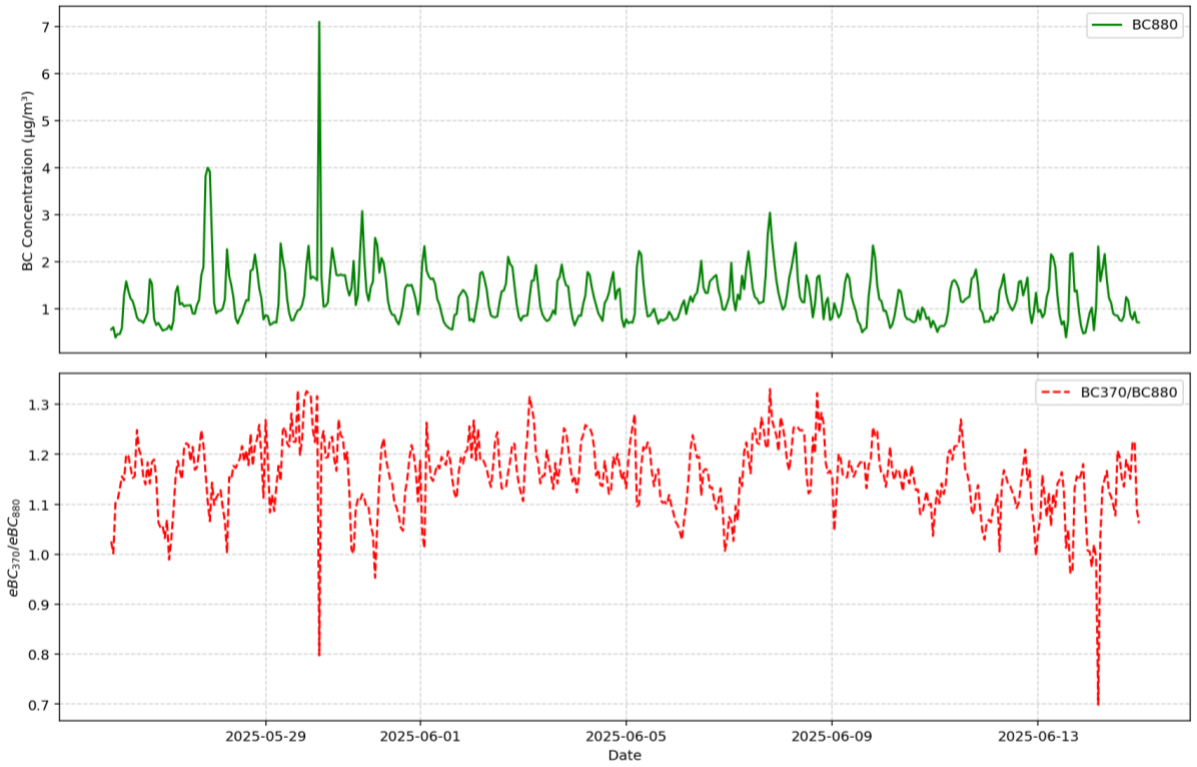

Figure S8. Time series of hourly averaged (a) equivalent black carbon at 880 nm (BC_{880}; µg m^{-3}) and (b) the BC_{370}/BC_{880} ratio during the campaign (26 May – 14 June 2025).

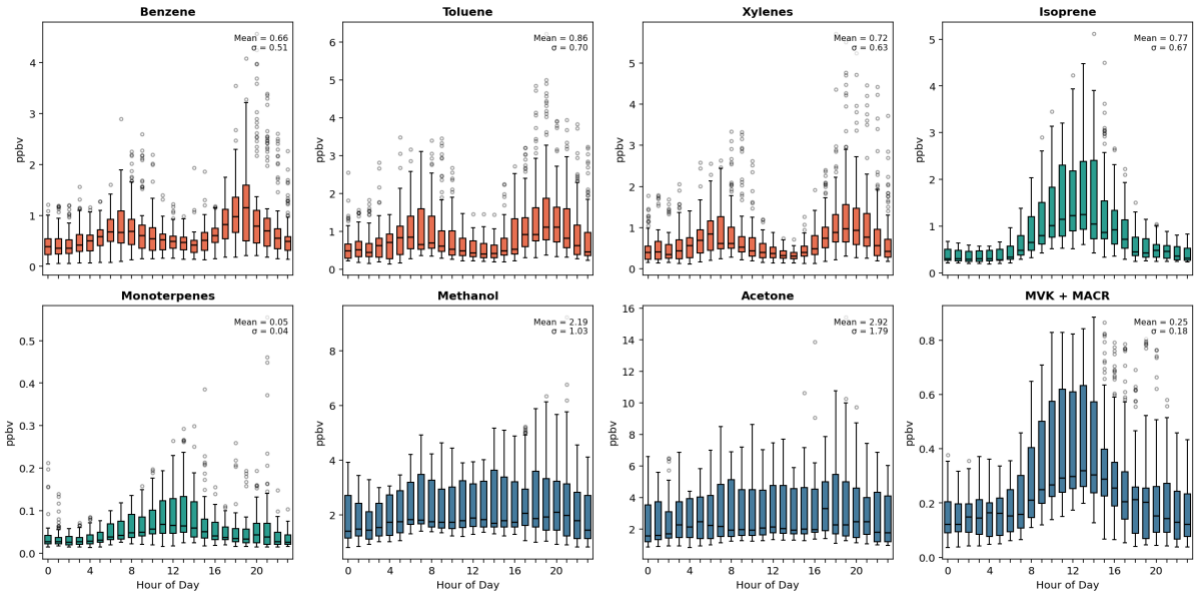

Figure S9. Hourly diurnal distributions of individual VOC species, classified by source: anthropogenic (benzene, toluene, xylenes), biogenic (isoprene, monoterpenes), and oxygenated (methanol, acetone, MVK + MACR). Box-and-whisker plots show the median (centre line), interquartile range (box), and outliers (circles).

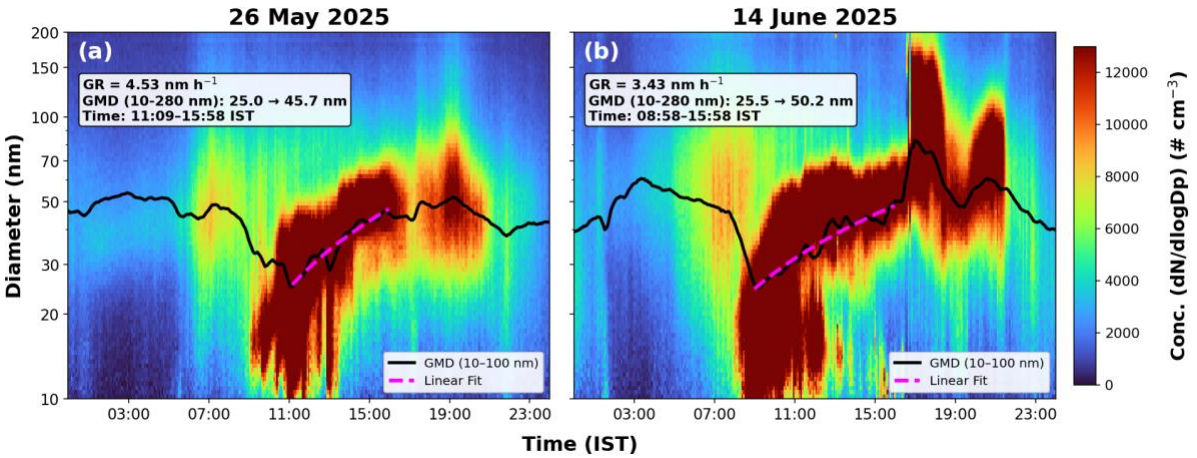

Figure S10. SMPS contour plots showing the temporal evolution of particle number size distributions during NPF events on 26 May 2025 and 14 June 2025. The black line represents the geometric mean diameter (GMD), while the magenta dashed line indicates the linear fit used to estimate particle growth rates (GR).

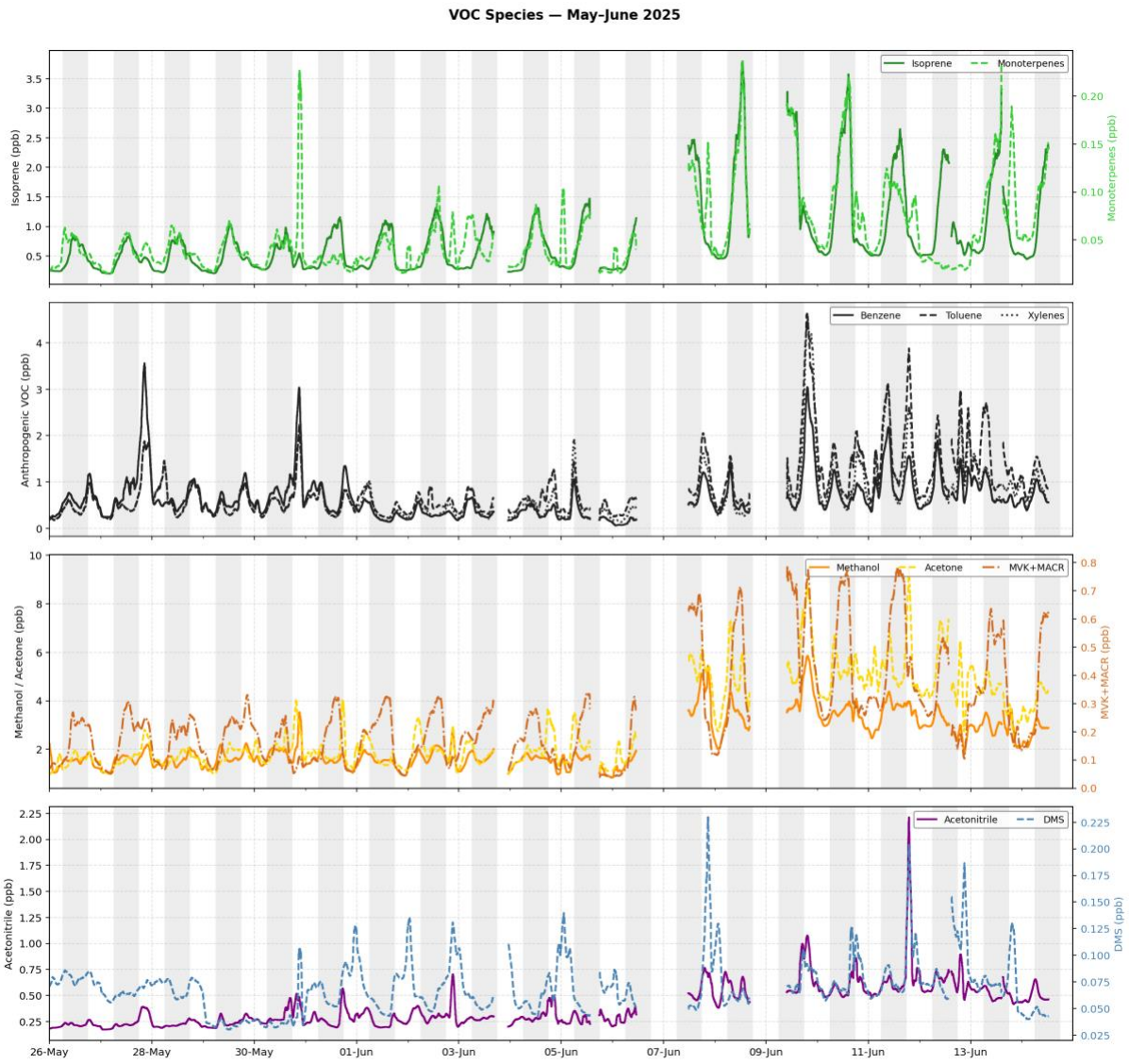

Figure S11. Time series of individual VOC species at 1 min resolution during the campaign (26 May–14 June 2025): (a) biogenic VOCs, isoprene (left axis) and monoterpenes (right axis, dashed line); (b) anthropogenic VOCs, benzene, toluene, and xylenes (BTX), distinguished by line style; (c) oxygenated VOCs, methanol and acetone (left axis) and MVK + MACR (right axis, dash-dot line); and (d) acetonitrile (left axis) and DMS (right axis, dashed line). All concentrations are in ppb. Lines show 105 min centred rolling means. Grey shading indicates daytime hours (06:00–18:00 IST). Gaps correspond to instrument downtime or data-quality screening.

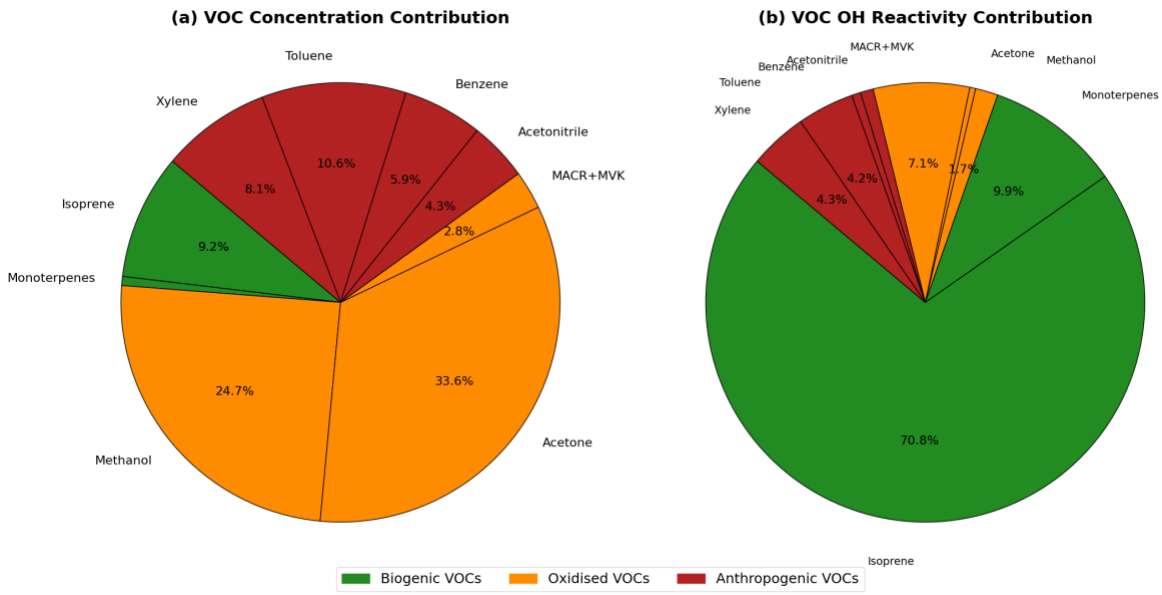

Figure S12. Relative contributions of individual VOC species to (a) total VOC concentrations and (b) total OH reactivity during the study period (26 May–14 June 2025). VOCs are grouped into biogenic, oxidised, and anthropogenic categories.

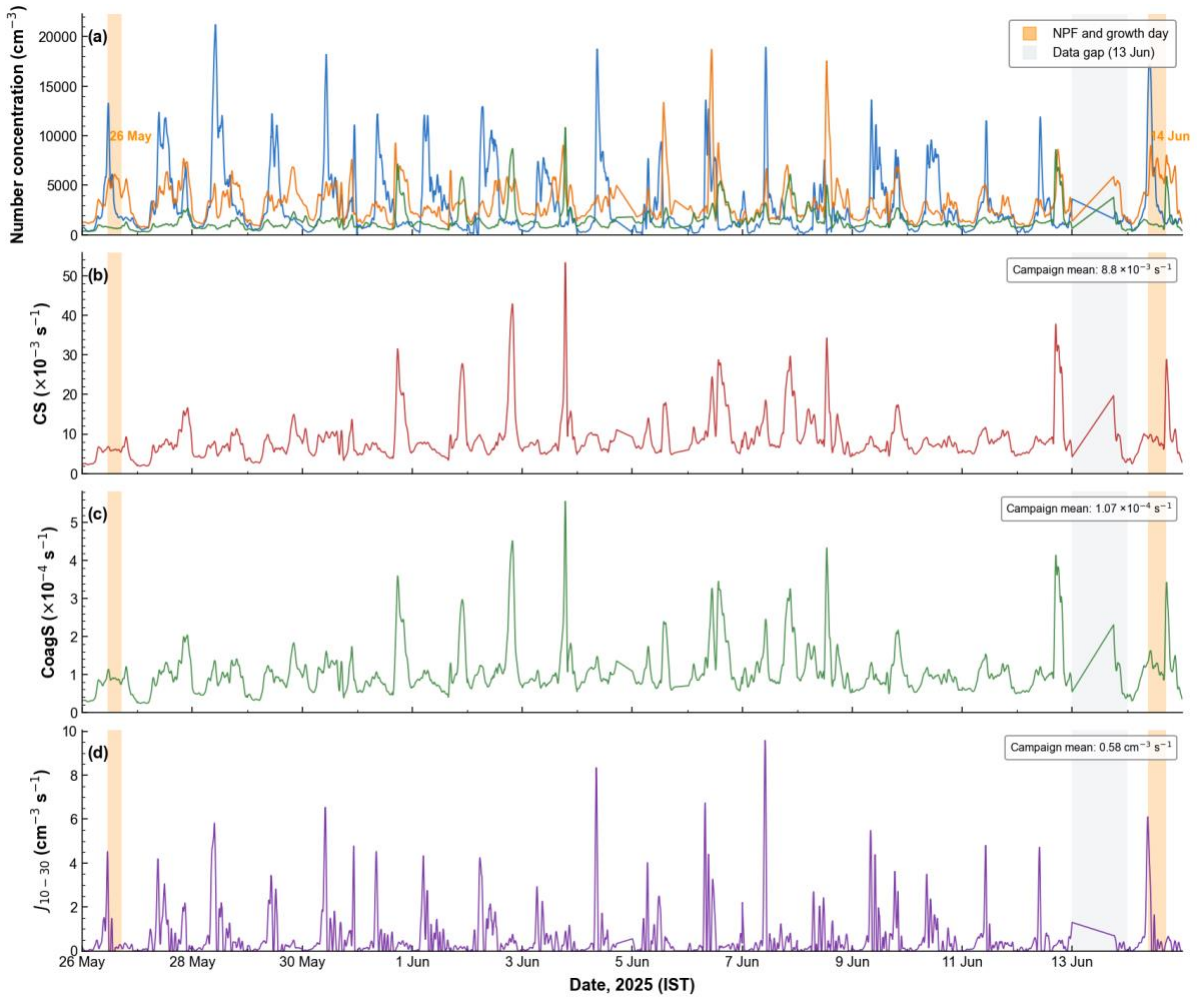

Figure S13. Campaign time series of (a) particle number concentrations in the 10–30 nm ($N_{10–30}$, blue), 30–100 nm ($N_{30–100}$, orange), and 100–280 nm ($N_{100–280}$, green) size ranges; (b) condensation sink (CS; $\times10^{-3}$ s^{-1}); (c) coagulation sink for a 10 nm test particle ($CoagS_{10\ nm}$; $\times10^{-4}$ s^{-1}); and (d) apparent particle formation rate ($J_{10–30}$; cm^{-3} s^{-1}) during 26 May–14 June 2025. CS and $CoagS_{10\ nm}$ were calculated over the reliable SMPS size range (10–280 nm; see Sect. 2.4). Amber shading indicates the confirmed NPF growth-event windows (26 May, 11:00–17:00 IST; 14 June, 09:00–17:00 IST), while grey shading denotes the data gap on 13 June 2025. Boxes in panels (b–d) indicate campaign mean values.

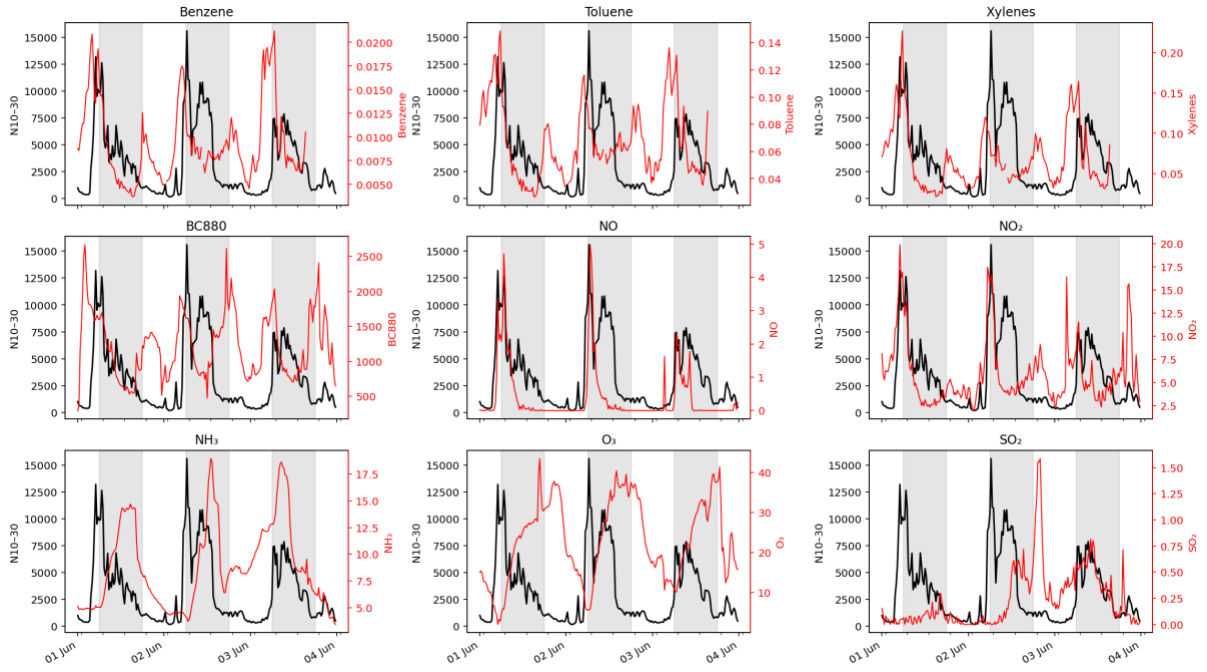

Figure S14. Time series of trace gases, anthropogenic VOCs, and equivalent black carbon during the campaign: benzene, toluene, xylenes, BC_{880}, NO, NO_x, NH_3, O_3, and SO_2. Black lines show the primary species (left axis); red lines show the corresponding secondary axis variable (right axis) where applicable. Grey shading indicates daytime hours.

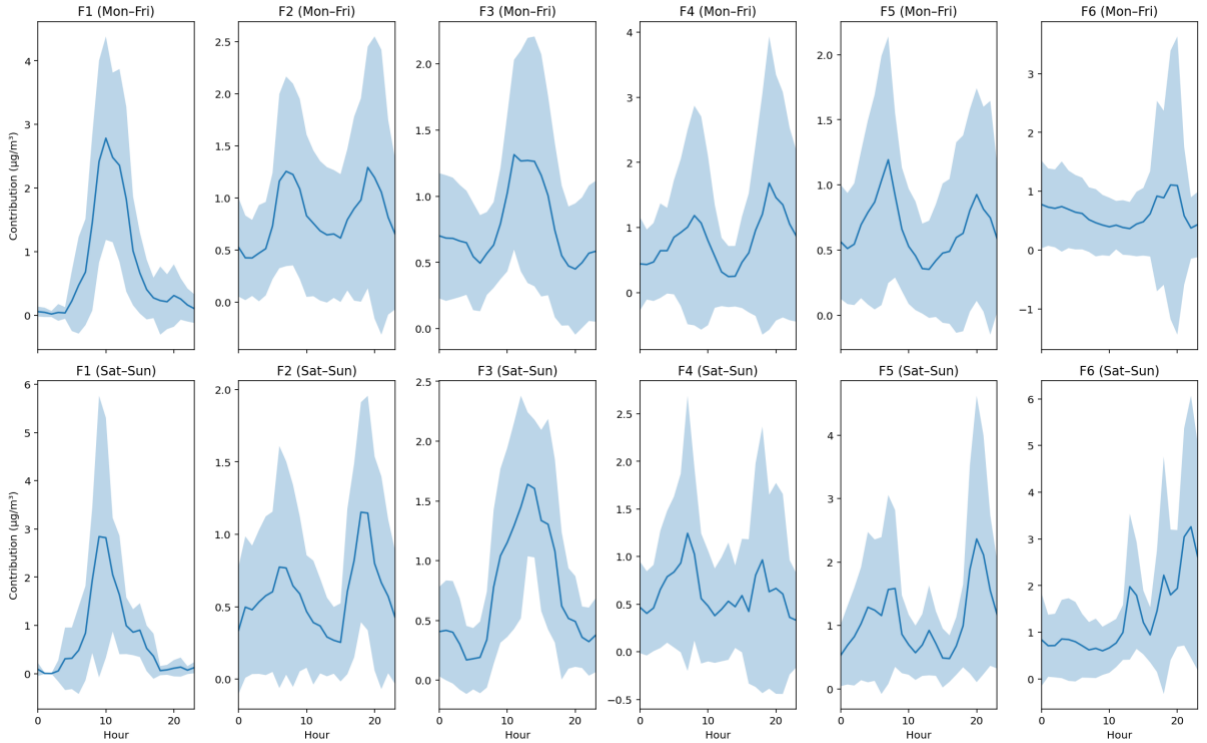

Figure S15. Hourly diurnal profiles of the six PMF factor contributions on weekdays (Mon–Fri; top row) and weekends (Sat–Sun; bottom row). Solid lines show the hourly mean; shading shows ±1 standard deviation.

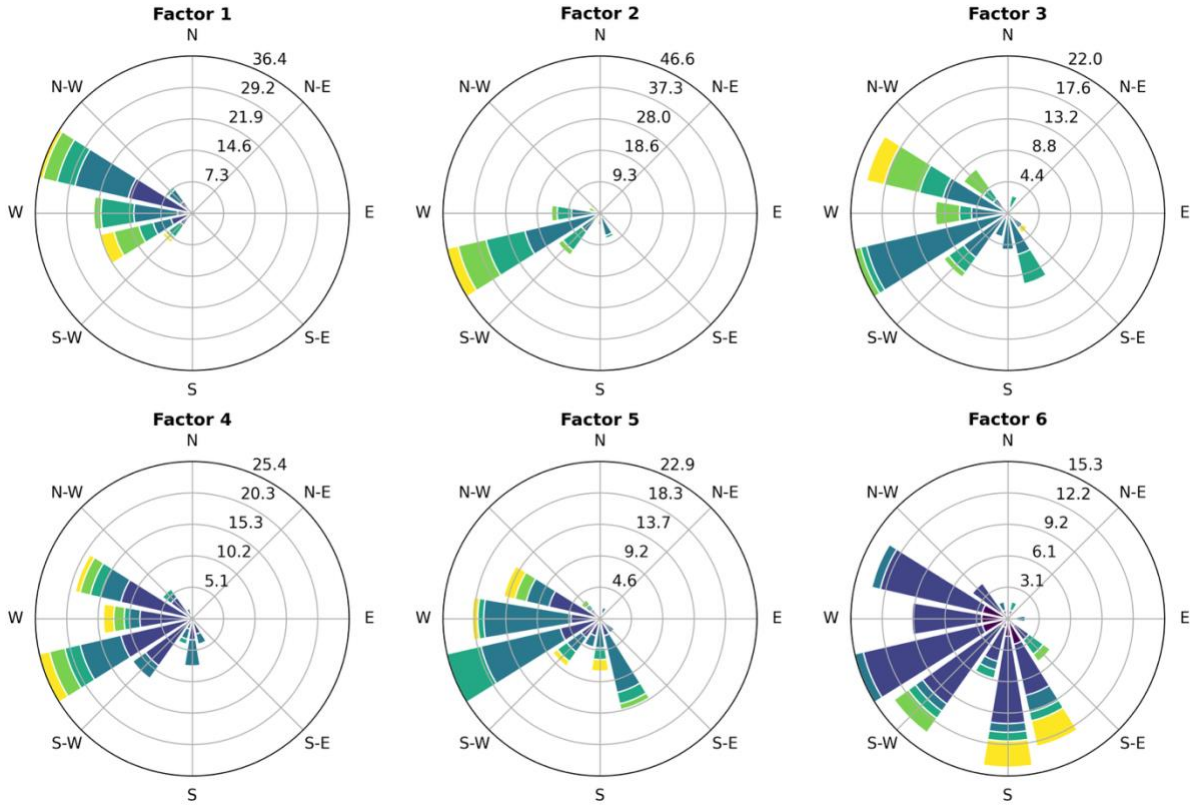

Figure S16. PMF factor wind roses showing the directional distribution of contributions from each of the six PMF factors (Factor 1 — NPF; Factor 2 — vehicular traffic; Factor 3 — biogenic / photochemical precursor reservoir; Factor 4 — aged aromatic VOC plume; Factor 5 — aged combustion-influenced secondary aerosol; Factor 6 — regional secondary sulfate) during the campaign. The radial axis shows the normalised frequency of occurrence (%); colour bins indicate factor-contribution ranges.

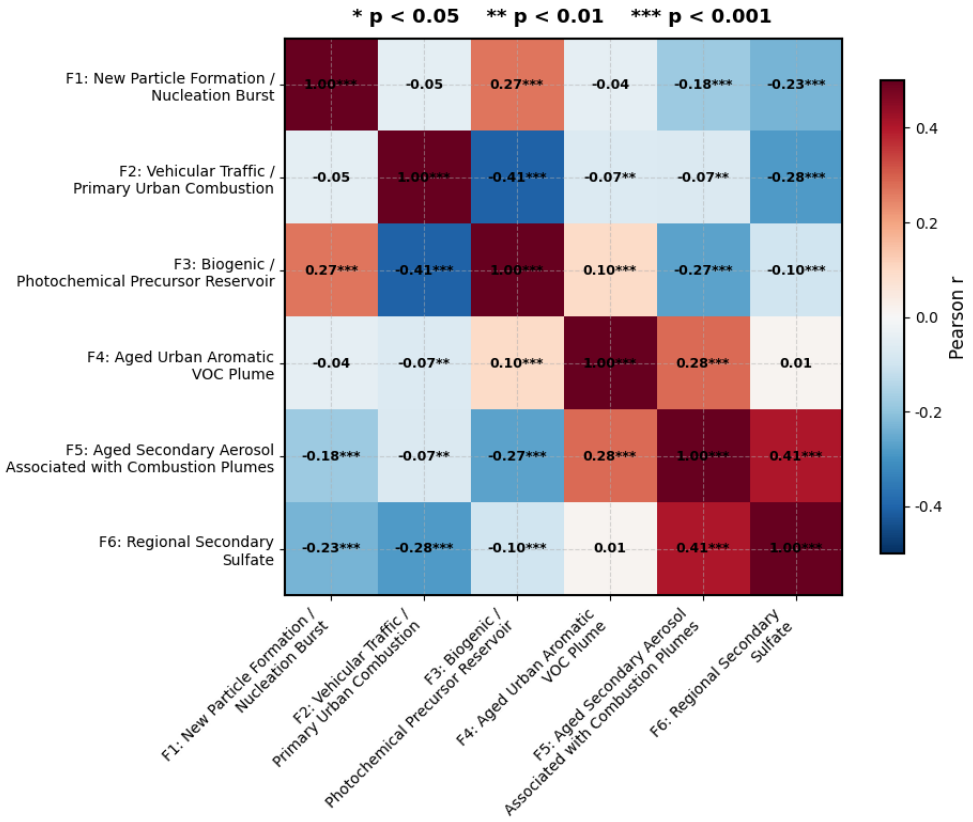

Figure S17. Inter-factor Pearson correlation matrix for the six-factor PMF solution. The colour scale represents the Pearson correlation coefficient (*r*); red — positive; blue — negative. Numerical values are shown within each cell.

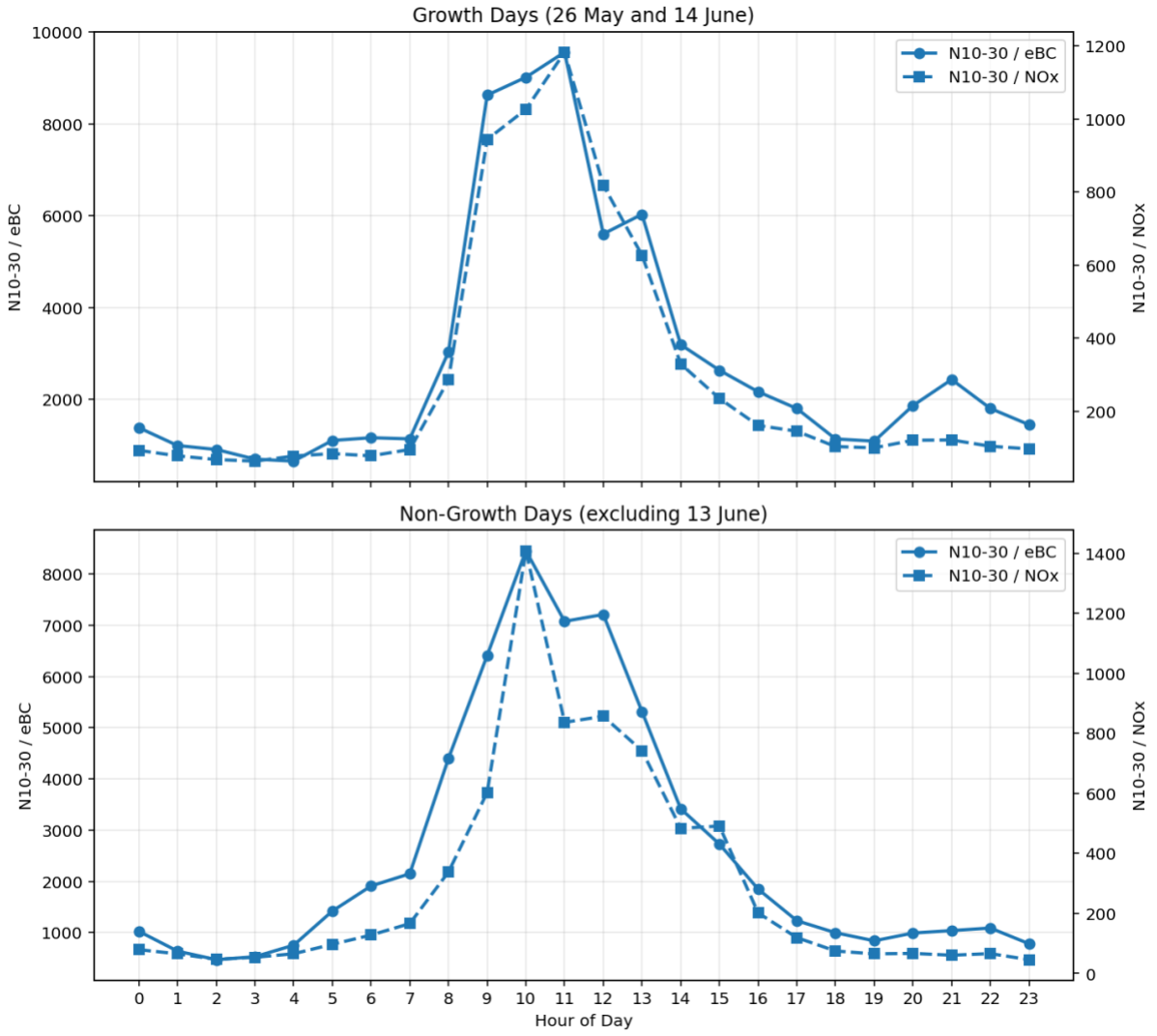

Figure S18. Diurnal variation of N10–30/eBC and N10–30/NOx ratios during (a) growth-event days (26 May and 14 June 2025) and (b) non-growth days (excluding 13 June 2025). The ratios were calculated using measured particle number concentrations (N10–30), equivalent black carbon (eBC), and NOx concentrations to examine differences between growth and non-growth days.

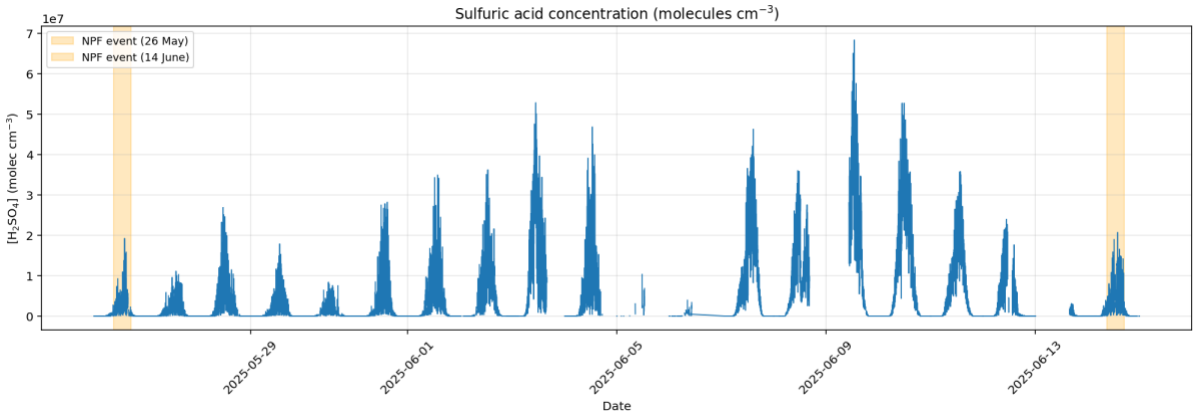

Figure S19: Time series of proxy-derived sulfuric acid concentration (H_2SO_4, molecules cm^{-3}) during the study period. The concentration was estimated using measured SO_2, ERA5 surface solar radiation

downward (SSRD), and condensation sink (CS) following a steady-state proxy approach. Orange shaded regions indicate the particle growth event periods observed on 26 May and 14 June 2025.

1.5 Positive Matrix Factorization

1.5.1 Input generation

Positive Matrix Factorization (PMF) analysis was conducted using EPA PMF 5.0 (Norris et al., 2014). Prior to PMF analysis, all measurements were synchronized to a common 15-min time resolution. The final input matrix included the following parameters:

1. Particle number concentrations in three size bins (10–30 nm, 30–100 nm, 100-280 nm) derived from SMPS measurements
2. Black carbon mass concentration measured by the AE33 Aethalometer
3. Non-refractory submicron aerosol components measured by Q-ACSM (organics, sulfate, nitrate, ammonium and chloride)
4. Trace gases including NOx, NH_3, SO_2 and O_3
5. VOCs measured by PTR-TOF-MS, including methanol, acetone, isoprene, monoterpenes, and methyl vinyl ketone and methacrolein (MVK+MACR), benzene, toluene, xylene, dimethyl sulfide (DMS), and Acetonitrile

Missing values were assigned as a value of -999, Negative NH_4^+ and chloride concentrations were set to zero.

1.5.2 Uncertainty Estimation

Measurement uncertainties were estimated following the guidelines provided in the EPA PMF user manual (Norris et al., 2014). For each observation, the uncertainty was calculated as:

$$\sigma_{ij} = \sqrt{(EF \times C_{ij})^2 + (MDL_j)^2}$$

where C_{ij} represents the measured concentration, EF represents the instrument-specific error fraction, and MDL_j represents the detection limit of species j. Concentrations below the detection limit were replaced with half the detection limit and assigned an uncertainty equal to $5/6$ of the detection limit.

Signal-to-noise ratios were calculated for each species and used to classify variables into three categories:

- Strong species: S/N > 2
- Weak species: 0.2 < S/N ≤ 2 (uncertainties multiplied by three)
- Bad species: S/N ≤ 0.2 (excluded from the analysis)

1.5.3 Model Diagnostics and Selection of Factor Number

PMF solutions with three to eight factors were evaluated to determine the optimal number of sources. The selection criteria included:

- Reduction of the objective function (Q/Qexp)
- Physically meaningful factor profiles
- Residual distributions centered around zero
- Absence of excessive rotational ambiguity

The six-factor solution resulted in a substantial reduction in the Q value relative to the four-factor solution and provided distinct source profiles without introducing nonphysical factors. Therefore, the six-factor solution was selected for further interpretation.

1.5.4 Bootstrap and Rotational Analysis

Model stability was evaluated using bootstrap (BS) analysis with 100 iterations. In this approach, the original dataset was randomly resampled with replacement to generate new datasets that were subsequently analyzed using PMF. Bootstrap factor profiles were mapped to the base solution using a minimum correlation coefficient of 0.6.

All factors were successfully reproduced in more than 85% of bootstrap runs, indicating robust and stable factor solutions.

Rotational ambiguity was assessed using the FPEAK rotational parameter, which was varied between −1.5 and +1.5. The resulting change in the objective function was less than 0.4%, suggesting minimal rotational ambiguity in the selected solution.

1.5.5 Residual Analysis

Scaled residuals were examined to assess model performance. Most residuals fell within the range of −3 to +3, indicating that the PMF model adequately reproduced the measured concentrations without systematic bias.

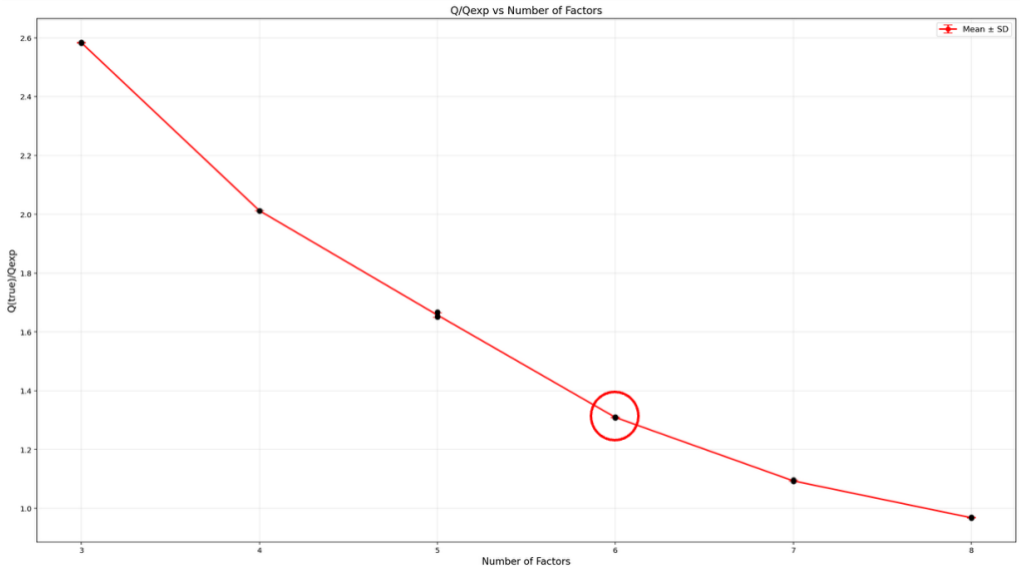

Figure S20. Variation of the PMF objective function (Q/Q_{exp}) as a function of the number of factors evaluated (3–8). The selected six-factor solution is indicated.

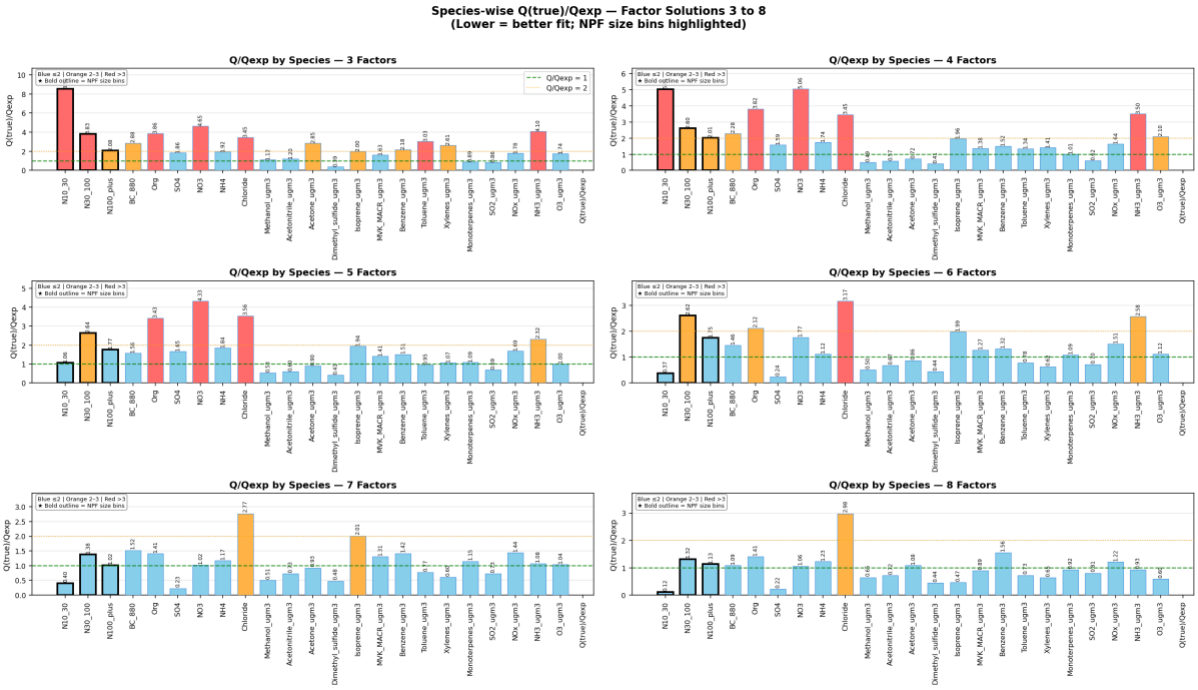

Figure S21. Variation of the PMF objective function (Q/Q_{exp}) as a function of the species included in the PMF input matrix.

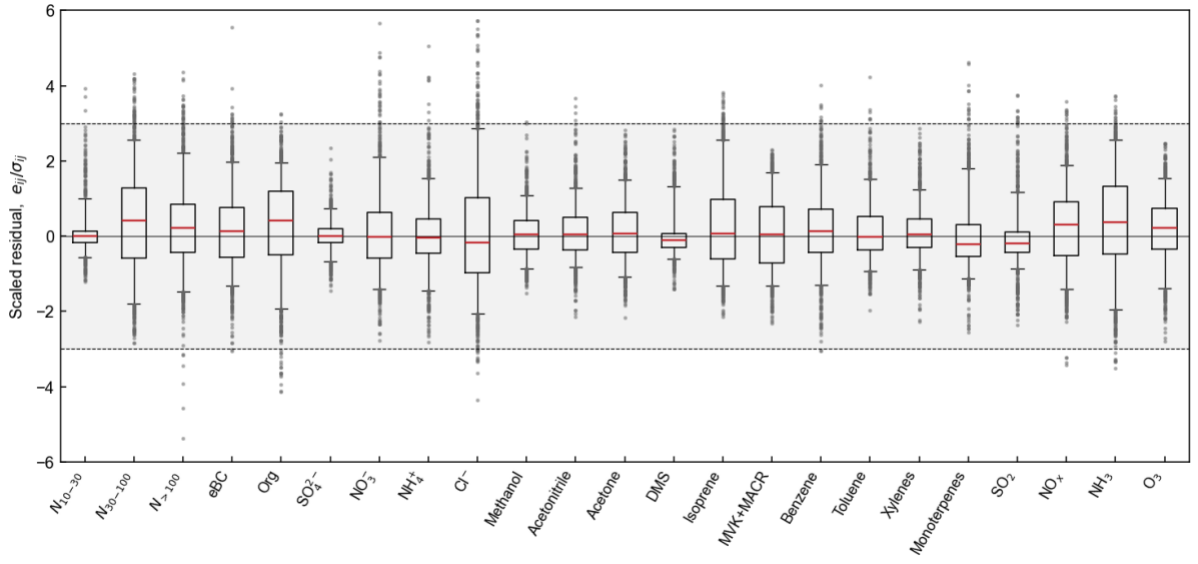

Figure S22. Scaled residuals (e_{ij}/σ_{ij}) for the 23 input species of the six-factor EPA-PMF base solution (n = 1426). Boxes show the interquartile range, red lines the median, whiskers the 5th–95th percentiles, and grey markers individual points beyond. The shaded band marks the $|e_{ij}/\sigma_{ij}| \leq 3$ acceptance range (Norris et al., 2014). Residuals are dimensionless, allowing comparison across the mixed input units (particles cm^{-3} for SMPS bins; $\mu g\ m^{-3}$ for all other species).

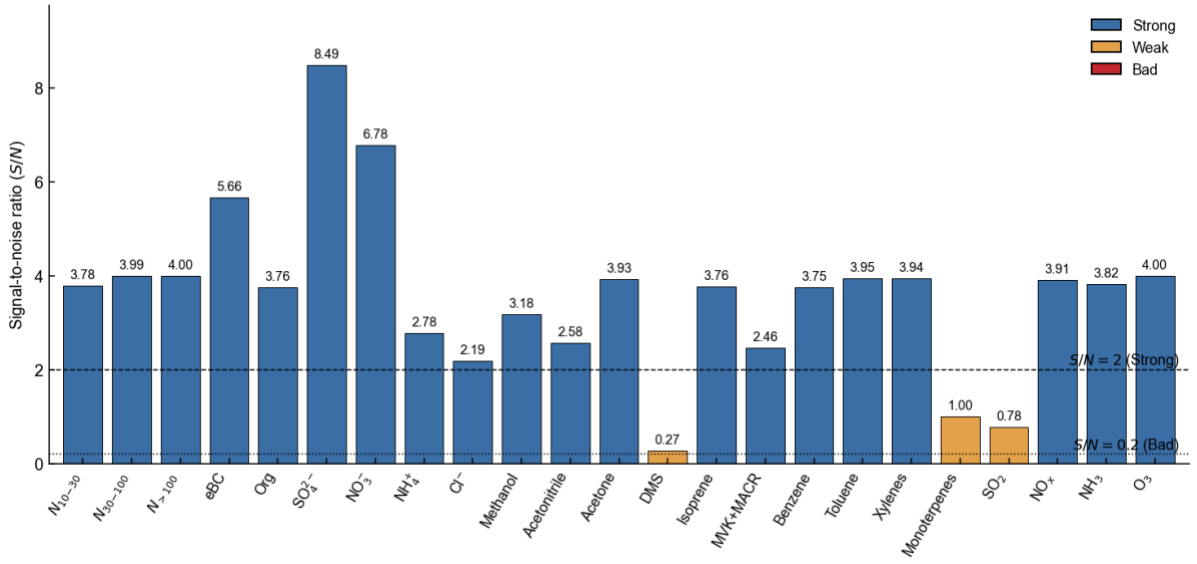

Figure S23. Signal-to-noise ratio (*S/N*) of the 23 input species used in the EPA-PMF analysis. Dashed and dotted lines mark the thresholds for Strong (*S/N* > 2) and Bad (*S/N* < 0.2) categorisation.

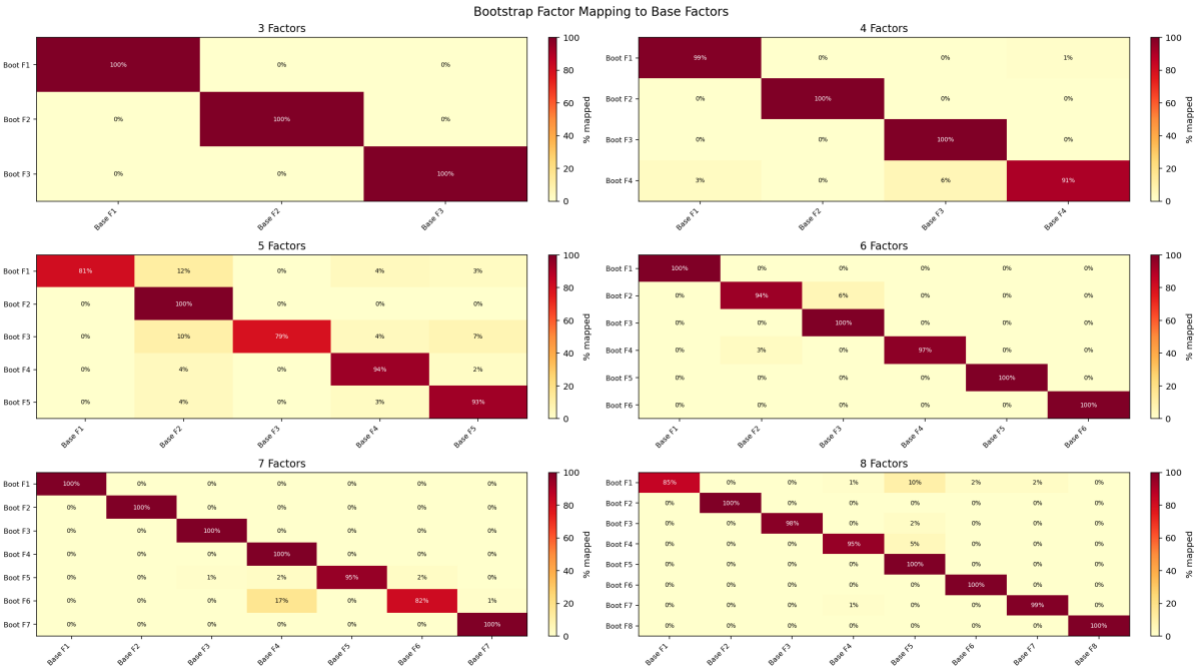

Figure S24. Bootstrap factor mapping for the three- to eight-factor PMF solutions, based on 100 bootstrap runs.

1.6 References

Atkinson, R., Baulch, D. L., Cox, R. A., Crowley, J. N., Hampson, R. F., Hynes, R. G., Jenkin, M. E., Rossi, M. J., Troe, J., and IUPAC Subcommittee: Evaluated kinetic and photochemical data for atmospheric chemistry: Volume II – gas phase reactions of organic species, Atmospheric Chem. Phys., 6, 3625–4055, https://doi.org/10.5194/acp-6-3625-2006, 2006.

Burkholder, J. B., Sander, S. P., Abbatt, J., Barker, J. R., Huie, R. E., Kolb, C. E., Kurylo, M. J., Orkin, V. L., Wilmouth, D. M., and Wine, P. H.: Chemical Kinetics and Photochemical Data for Use in Atmospheric Studies, Evaluation No. 18, Jet Propulsion Laboratory, Pasadena, CA, 2015.

Dal Maso, M., Kulmala, M., Lehtinen, K. E. J., Mäkelä, J. M., Aalto, P., and O'Dowd, C. D.: Condensation and coagulation sinks and formation of nucleation mode particles in coastal and boreal forest boundary layers, J. Geophys. Res. Atmospheres, 107, https://doi.org/10.1029/2001JD001053, 2002.

De Gouw, J. and Warneke, C.: Measurements of volatile organic compounds in the earth's atmosphere using proton-transfer-reaction mass spectrometry, Mass Spectrom. Rev., 26, 223–257, https://doi.org/10.1002/mas.20119, 2007.

Graus, M., Müller, M., and Hansel, A.: High resolution PTR-TOF: Quantification and formula confirmation of VOC in real time, J. Am. Soc. Mass Spectrom., 21, 1037–1044, https://doi.org/10.1016/j.jasms.2010.02.006, 2010.

Hansel, A., Jordan, A., Holzinger, R., Prazeller, P., Vogel, W., and Lindinger, W.: Proton transfer reaction mass spectrometry: on-line trace gas analysis at the ppb level, Int. J. Mass Spectrom. Ion Process., 149–150, 609–619, https://doi.org/10.1016/0168-1176(95)04294-U, 1995.

Kerminen, V.-M. and Kulmala, M.: Analytical formulae connecting the "real" and the "apparent" nucleation rate and the nuclei number concentration for atmospheric nucleation events, J. Aerosol Sci., 33, 609–622, https://doi.org/10.1016/S0021-8502(01)00194-X, 2002.

Kulmala, M., Petäjä, T., Nieminen, T., Sipilä, M., Manninen, H. E., Lehtipalo, K., Dal Maso, M., Aalto, P. P., Junninen, H., Paasonen, P., Riipinen, I., Lehtinen, K. E. J., Laaksonen, A., and Kerminen, V.-M.: Measurement of the nucleation of atmospheric aerosol particles, Nat. Protoc., 7, 1651–1667, https://doi.org/10.1038/nprot.2012.091, 2012.

Lindinger, W., Hansel, A., and Jordan, A.: On-line monitoring of volatile organic compounds at pptv levels by means of proton-transfer-reaction mass spectrometry (PTR-MS) medical applications, food control and environmental research, Int. J. Mass Spectrom. Ion Process., 173, 191–241, https://doi.org/10.1016/S0168-1176(97)00281-4, 1998.

Mikkonen, S., Romakkaniemi, S., Smith, J. N., Korhonen, H., Petäjä, T., Plass-Duelmer, C., Boy, M., McMurry, P. H., Lehtinen, K. E. J., Joutsensaari, J., Hamed, A., Mauldin Iii, R. L., Birmili, W., Spindler, G., Arnold, F., Kulmala, M., and Laaksonen, A.: A statistical proxy for sulphuric acid concentration, Atmospheric Chem. Phys., 11, 11319–11334, https://doi.org/10.5194/acp-11-11319-2011, 2011.

Norris, G., Duvall, R., Brown, S., and Bai, S.: EPA Positive Matrix Factorization (PMF) 5.0 Fundamentals and User Guide, U.S. Environmental Protection Agency, Washington, DC, 2014.

Seinfeld, J. H.: Atmospheric Chemistry and Physics: From Air Pollution to Climate Change, 1st ed., John Wiley & Sons, Incorporated, Newark, 1 pp., 2016.

Sun, J., Hermann, M., Weinhold, K., Merkel, M., Birmili, W., Yang, Y., Tuch, T., Flentje, H., Briel, B., Ries, L., Couret, C., Elsasser, M., Sohmer, R., Wirtz, K., Meinhardt, F., Schütze, M., Bath, O., Hellack, B., Kerminen, V.-M., Kulmala, M., Ma, N., and Wiedensohler, A.: Measurement report: Contribution of atmospheric new particle formation to ultrafine particle concentration, cloud condensation nuclei, and radiative forcing – results from 5-year observations in central Europe, Atmospheric Chem. Phys., 24, 10667–10687, https://doi.org/10.5194/acp-24-10667-2024, 2024.