Supplement of Atmos. Chem. Phys., 26, 7539–7554, 2026
https://doi.org/10.5194/acp-26-7539-2026-supplement

Atmospheric Chemistry and Physics
Open Access
EGU

Supplement of

Seasonal variation in aerosol chemistry drives new particle formation and CCN activity in a coastal city, China: insights from year-long online measurements in Fuzhou

Zihan Wang et al.

Correspondence to: Honglei Wang (hongleiwang@nuist.edu.cn)

S1. XGBoost-SHAP analysis for FR attribution

The analysis was performed using the Python XGBoost library (version 1.5.0) and the SHAP library (version 0.41.0).

Feature selection and data preparation. The target variable was the hourly FR (1.5–3 nm particles). Predictor variables included: meteorological parameters (T, RH, wind speed WS, wind direction WD), condensation sink (CS), nucleation-mode particle concentration (N_{nuc}), Aitken-mode particle concentration (N_{ait}), inorganic hygroscopicity parameter (κ_{inorg}), and gaseous precursors (NH_3, SO_2, HNO_2, HNO_3). All variables were hourly averaged and synchronized with FR measurements. No manual feature engineering or dimensionality reduction was applied; the full set of available predictors was used to allow SHAP to identify the most influential drivers.

Model training and validation. The dataset (June 2021 to May 2022) was randomly split into training (69.2 %, n=705), validation (10.8 %, n=111), and testing (20 %, n=205) sets. An XGBoost regressor with fixed hyperparameters (n_estimators = 200, max_depth = 3, learning_rate = 0.03, min_child_weight = 5, gamma = 0.1, subsample = 0.8, colsample_bytree = 0.8, reg_alpha = 1) was trained on the training set. The model achieved good generalization with test $R^2 = 0.976$ and RMSE = 0.782 (training $R^2 = 0.985$, validation $R^2 = 0.961$), indicating that the XGBoost model captured the dominant FR variations without overfitting. No hyperparameter optimization was performed for this attribution analysis, as the goal was to obtain stable SHAP values from a reasonably generalized model.

SHAP analysis for interpretation. After training, SHAP values were computed on the test set using shap.TreeExplainer. For each prediction, SHAP decomposes the model output into additive feature contributions, where positive/negative values indicate directional effects on FR. We derived: (1) global feature importance (mean absolute SHAP values) to rank each variable's overall contribution; (2) SHAP dependence plots for individual features to identify nonlinear thresholds (e.g., the NH_3 concentration at which SHAP turns from negative to positive); and (3) SHAP interaction values to quantify pairwise coupling strengths. The main thresholds and interaction strengths obtained from this analysis are discussed in the main text (Sections 3.2–3.3).

Table S1. Summary of measurement instruments, parameters, time resolution, averaging time used for analysis, and measurement uncertainties.

Station	Instrument	Parameter(s)	Time resolution	Uncertainty
The Fujian Provincial Environmental Monitoring Center Station	WPS-1000	Aerosol size distribution (10–350 nm)	6 min	±10%
	AE-33	BC	1 h	±10% (Drinovec et al., 2015)
	OC/EC Analyzer	OC, EC	1 h	OC: ±3.6%; EC: ±6.8% (Zhang et al., 2021)
	MARGA (ADI 2080)	SO_4^{2-}, NO_3^-, NH_4^+, Na^+, K^+, Ca^{2+}, Cl^-; NH_3, HNO_2, HNO_3, HCl, SO_2	1h	±5–20% (Battelle, 2009)
The Fuzhou Meteorological Bureau Station	Meteorological station	T, RH, WS, WD, precipitation	1 h	T: ±0.2°C; RH: ±4–8%; WS: ±(0.5+0.03v) m/s; WD: ±5°; precip.: ±0.4 mm/±4%
	CCNC-100	CCN number concentration and spectrum distribution	10 min per SS	±10%

Table S2. Comparison of air pollutant concentrations at the two sites during NPF event days (mean ± standard deviation, n = 46 days).

Factors	Riverside site (1282A)	Urban site (1283A)	Difference (%)
$PM_{2.5}$ (μg/m³)	19.6±9.7	17.6±8.7	–2.0
NO_2 (μg/m³)	19.3±6.8	20.2±8.5	+0.9
O_3 (μg/m³)	61.4±19.5	61.8±18.3	+0.4
PM_{10} (μg/m³)	38.3±14.0	37.6±13.8	–0.7
SO_2 (μg/m³)	3.9±1.2	3.9±1.0	0.0
CO (mg/m³)	0.50±0.10	0.52±0.15	+0.02

Table S3. Seasonal mean values (± standard deviation) of key NPF parameters (κ_{inorg}, Q, C, GR, CS, CoagS, and FR) for each trajectory cluster. The cluster proportion (%) is given in parentheses.

Season	Cluster (%)	κ_{inorg}	Q ($\times10^5$ $cm^{-3}\cdot s^{-1}$)	C ($\times10^7$ cm^{-3})	GR ($nm\cdot h^{-1}$)	CS ($\times10^{-2}$ s^{-1})	CoagS ($\times10^{-4}$ s^{-1})	FR ($cm^{-3}\cdot s^{-1}$)
Spring	CL1 (23.11)	0.55±0.03	2.23±17.12	3.30±24.31	2.41±17.75	1.95±0.54	1.37±0.65	0.16±0.27
	CL2 (51.33)	0.52±0.03	0.39±15.65	0.68±12.70	0.50±9.27	4.36±2.14	5.61±4.00	6.65±9.92
	CL3 (25.57)	0.56±0.04	9.64±24.23	4.38±12.86	3.19±9.39	2.57±1.59	3.07±3.16	2.57±5.96
Summer	CL1 (20.00)	0.51±0.01	0.80±14.08	0.40±4.89	0.29±3.57	2.49±0.52	1.13±0.22	0.04±0.11
	CL2	0.55±0.	3.63±34.6	0.98±10	0.71±7.	1.08±0.	1.01±0.	0.06±0.

	(60.00)	03	2	.31	52	31	65	24
	CL3 (20.00)	0.71± 0.02	1.63± 13.23	0.75±9. 11	0.55±6. 65	1.71±0. 33	1.98±0. 69	0.29±0. 50
Fall	CL1 (22.22)	0.53±0. 03	1.11±12.3 3	1.06±6. 83	0.77±4. 99	1.92±0. 66	1.29±0. 55	0.05±0. 41
	CL2 (41.20)	0.58±0. 05	2.45±15.3 4	0.90±7. 37	0.65±5. 38	1.70±0. 71	1.18±0. 75	0.08±0. 25
	CL3 (36.57)	0.52±0. 02	1.35±13.5 8	0.68±5. 27	0.50±3. 85	2.32±0. 55	1.45±0. 53	0.11±0. 21
Winte r	CL1 (43.75)	0.59±0. 03	0.84±7.23	0.49±4. 91	0.35±3. 58	1.50±0. 50	0.98±0. 33	0.08±0. 14
	CL2 (31.25)	0.58±0. 04	0.07±6.75	- 0.06±6. 52	- 0.05±4. 76	2.15±0. 56	1.06±0. 49	0.07±0. 20
	CL3 (25.00)	0.58±0. 05	-1.41±9.38	- 1.11±6. 08	- 0.81±4. 44	2.50±1. 02	1.32±0. 52	0.07±0. 16

Table S4. Seasonal mean values (± standard deviation) of precursor gas concentrations (HNO_3, HNO_2, HCl, NH_3, SO_2) for each trajectory cluster.

Season	Cluster (%)	HNO_3 ($\mu g \cdot m^{-3}$)	HNO_2 ($\mu g \cdot m^{-3}$)	HCl ($\mu g \cdot m^{-3}$)	NH_3 ($\mu g \cdot m^{-3}$)	SO_2 ($\mu g \cdot m^{-3}$)
Spring	CL1 (23.11)	0.57±0.03	0.58±0.26	0.06± 0.06	1.26±0.53	1.27±1.10
	CL2 (51.33)	0.95±0.38	2.81±2.16	0.13±0.09	6.76±2.09	0.46±0.37
	CL3 (25.57)	0.66±0.14	1.38±1.39	0.09±0.06	4.00±1.76	0.71±0.48
Summer	CL1 (20.00)	1.22±0.36	1.92±1.26	0.25±0.12	7.44±1.26	0.62±0.33
	CL2 (60.00)	0.95±0.10	0.61±0.22	0.22±0.06	2.85±0.34	0.56±0.17
	CL3 (20.00)	0.93±0.07	0.49±0.08	0.23±0.05	2.98±0.24	0.53±0.07
Fall	CL1 (22.22)	0.68±0.08	0.96±0.60	0.10±0.07	3.13±0.68	0.59±0.34
	CL2 (41.20)	0.48±0.06	0.70±0.54	0.07±0.04	2.09±1.16	0.71±0.51
	CL3 (36.57)	0.50±0.08	1.08±0.87	0.06±0.02	2.59±0.57	0.83±0.32
Winter	CL1 (43.75)	0.49±0.05	0.81±0.37	0.08±0.06	3.24±1.20	1.15±0.69
	CL2 (31.25)	0.57±0.08	1.22±0.68	0.08±0.04	3.59±1.13	1.22±0.88
	CL3 (25.00)	0.58±0.09	2.36±2.83	0.07±0.04	3.83±2.73	0.92±0.64

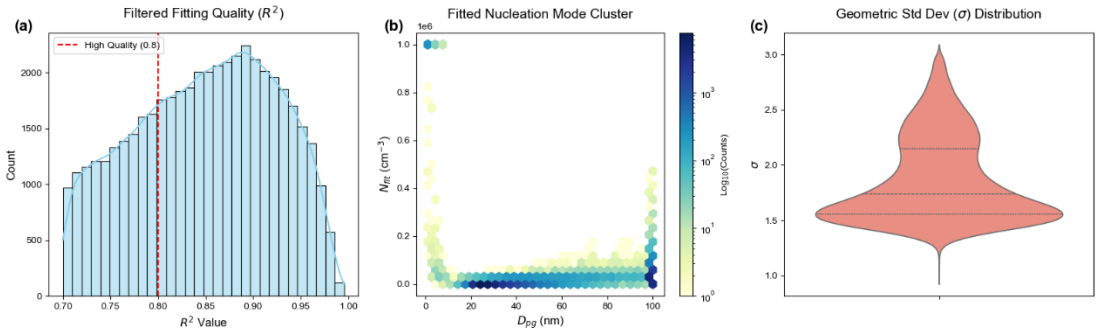

Fig. S1 Evaluation of the single-mode log-normal fitting for nucleation-mode particle size distributions (PSDs). (a) Frequency distribution of the coefficient of determination (R^2) for the retained fits (R^2>0.7). The red dashed line indicates the high-quality threshold of 0.8. (b) Density plot (hexagonal binning) of the fitted nucleation-mode parameter space showing the relationship between geometric mean diameter (D_{pg}) and total number concentration (N). The color scale represents the number of occurrences on a logarithmic scale. (c) Frequency distribution (violin plot) of the retrieved geometric standard deviation (σ). The dashed and dotted lines within the violin plot indicate the median and quartiles, respectively.

To ensure high data fidelity for subsequent calculations of growth rate (GR), formation rate (FR), and condensable vapor concentration, a rigorous quality control (QC) protocol was implemented. Only fits satisfying both of the following criteria were retained: (1) R^2>0.7, and (2) N≤10^6 cm^{-3}. The latter criterion acts as a physically constrained upper limit to exclude spurious mathematical solutions caused by instrumental noise or extreme outliers.

As shown in Fig. S2a, the retained fits account for approximately 58% of the total dataset, a robust proportion considering the intermittent nature of new particle formation (NPF) events and the frequent occurrence of low-signal background conditions. The majority of R^2 values lie between 0.85 and 0.95, with a peak near 0.91, indicating that the log-normal model effectively captures the variance of the observed nucleation-mode particles.

Fig. S3b displays the parameter space of fitted nucleation mode (D_{pg} vs. N). A high-density cluster appears at D_{pg}<10 nm with elevated N, which serves as a definitive signature of NPF events. The observed downward-bending density band reflects the classical aerosol dynamic evolution: a burst of nucleated particles followed by simultaneous diameter growth (increasing D_{pg}) and sink-driven concentration decay (decreasing N).

The frequency distribution of σ is presented in Fig.S1c. The median σ is approximately 1.7, and most values fall within 1.5–2.0. This is highly consistent with the theoretical expectations for a well-defined, unimodal nucleation mode in atmospheric aerosol physics.

In summary, the log-normal distribution provides a statistically robust (high R^2), physically meaningful (reasonable D_{pg}-N evolution, σ within 1.5–2.0), and stable representation of the nucleation-mode PSDs in our dataset. Fits with $R^2 \leq 0.7$ (e.g., during non-NPF or mixed-mode conditions) are excluded from analyses requiring a well-defined nucleation mode.

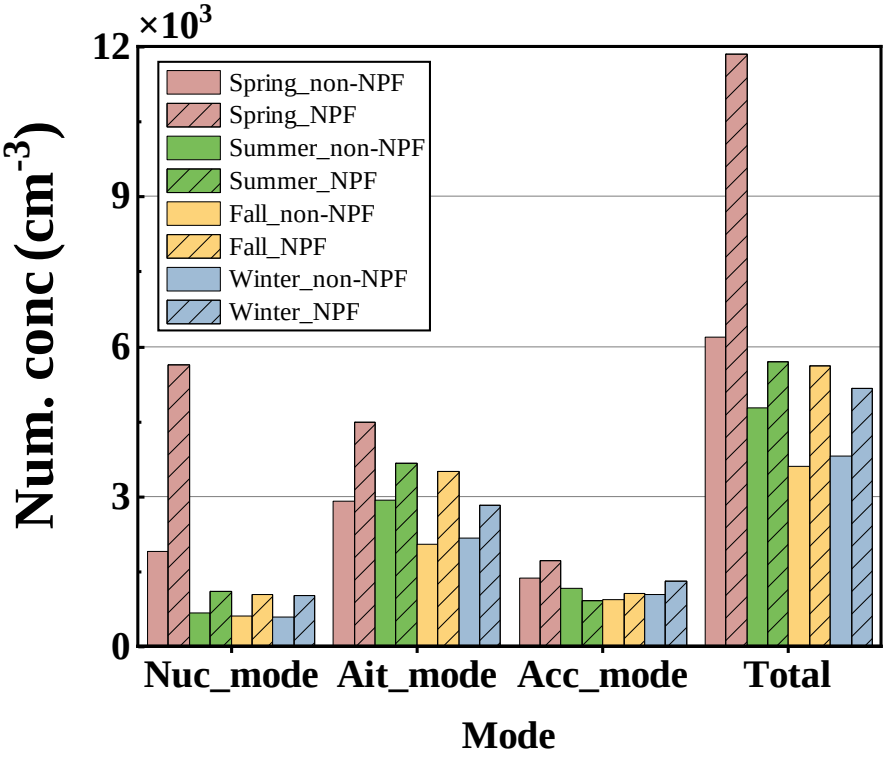

Fig. S2 Particle number concentrations of different modes between NPF days and non-NPF days.

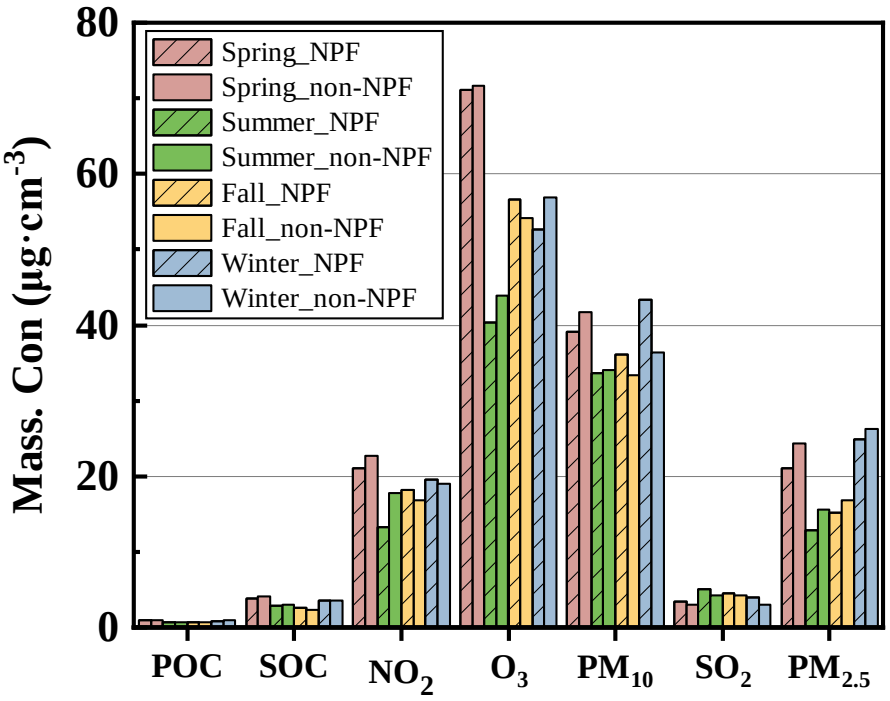

Fig.S3 Mass concentrations of particulate pollutants between NPF days and non-NPF days.

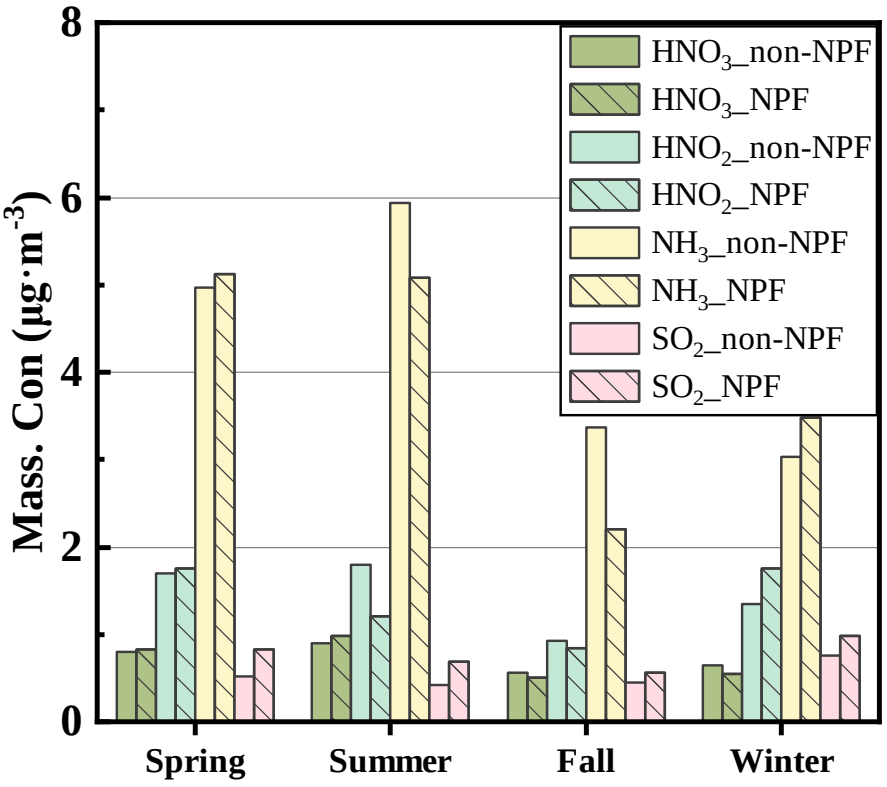

Fig.S4 Mass concentrations of precursor gases (e.g., SO_2, NO_2, O_3, NH_3) between NPF days and non-NPF days.

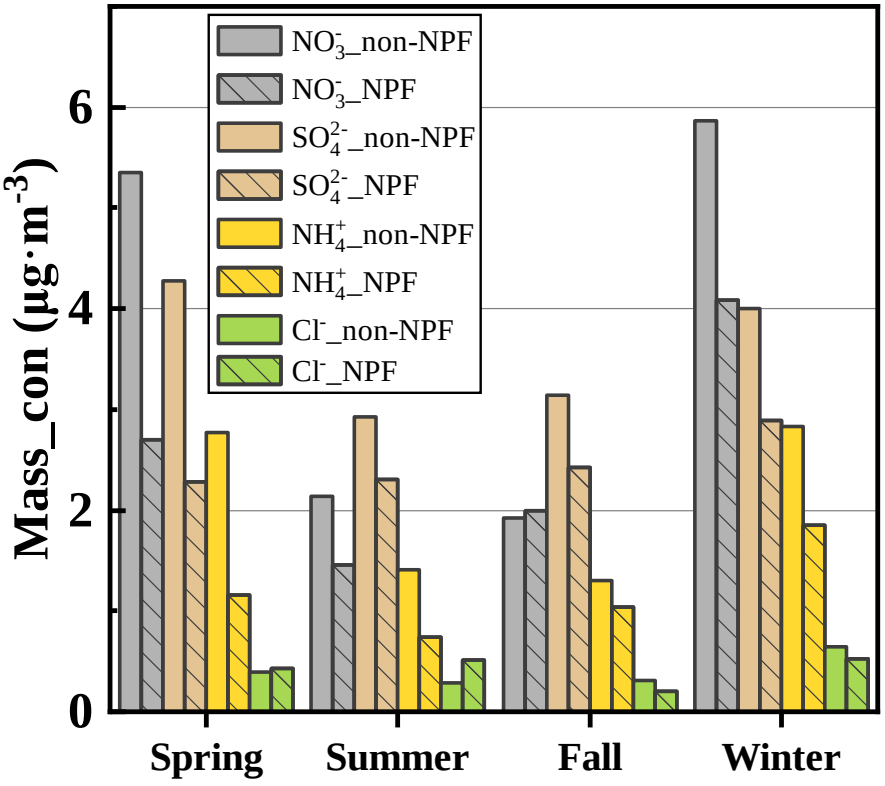

Fig.S5 Mass concentrations of major ions (SO_4^{2-}, NO_3^-, NH_4^+) between NPF days and non-NPF days.

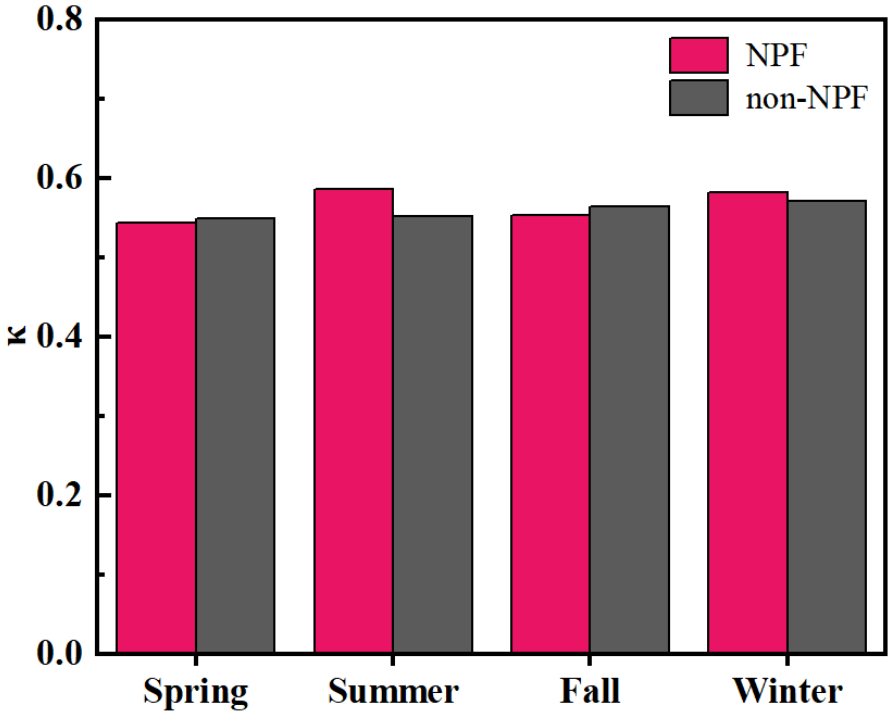

Fig.S6 The mean hygroscopicity parameter (κ) between NPF days and non-NPF days.

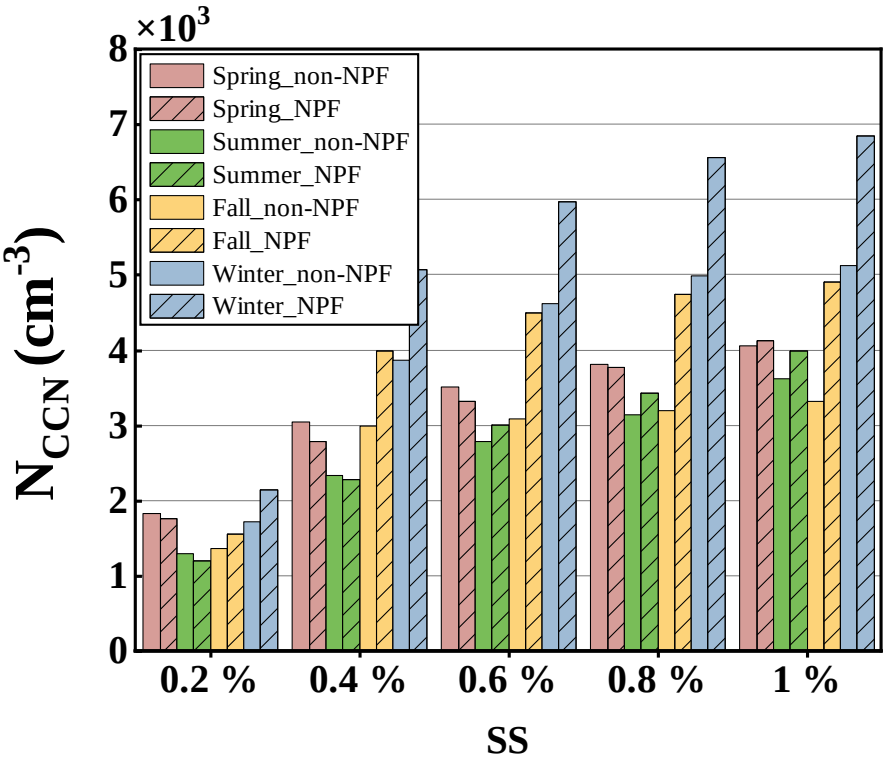

Fig.S7 Cloud condensation nuclei number concentrations (N_{CCN}) at 0.2-1.0% supersaturations between NPF days and non-NPF days.

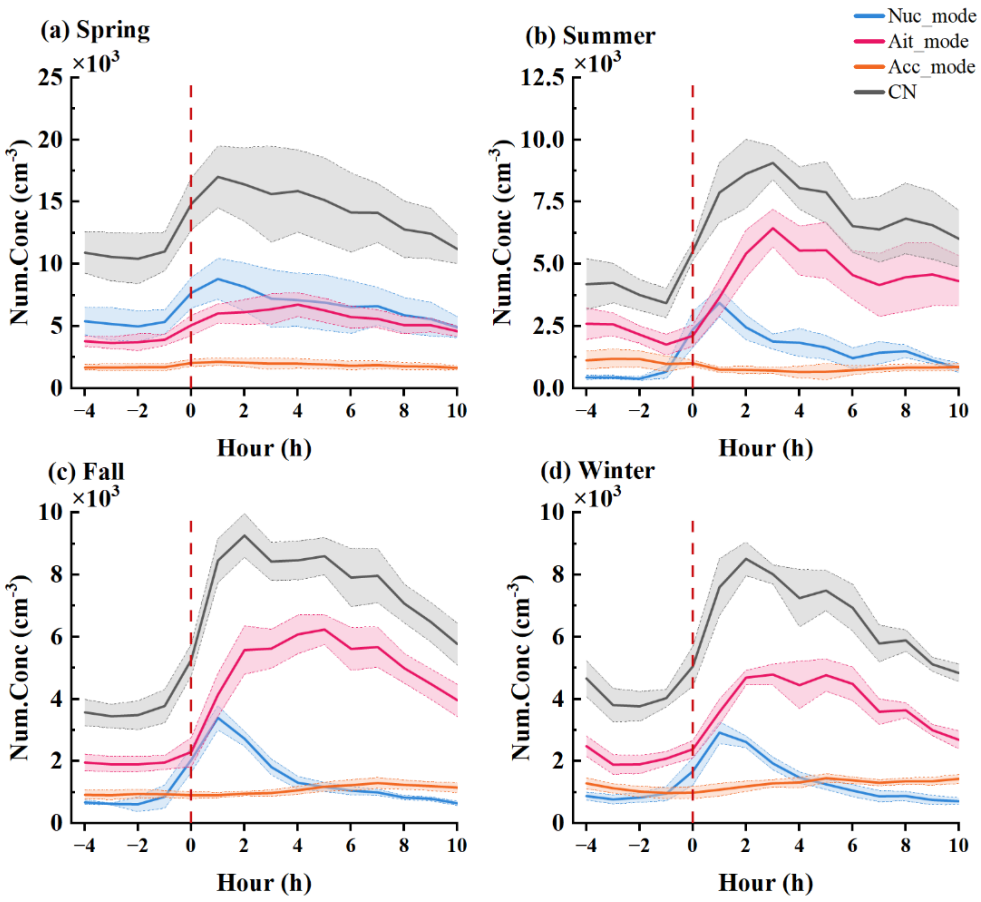

Fig.S8 Temporal evolution of particle number concentrations in the nucleation, Aitken, and accumulation modes during NPF events. Solid lines represent the median values across all NPF events in each season, and shaded bands indicate ±1σ standard deviation.

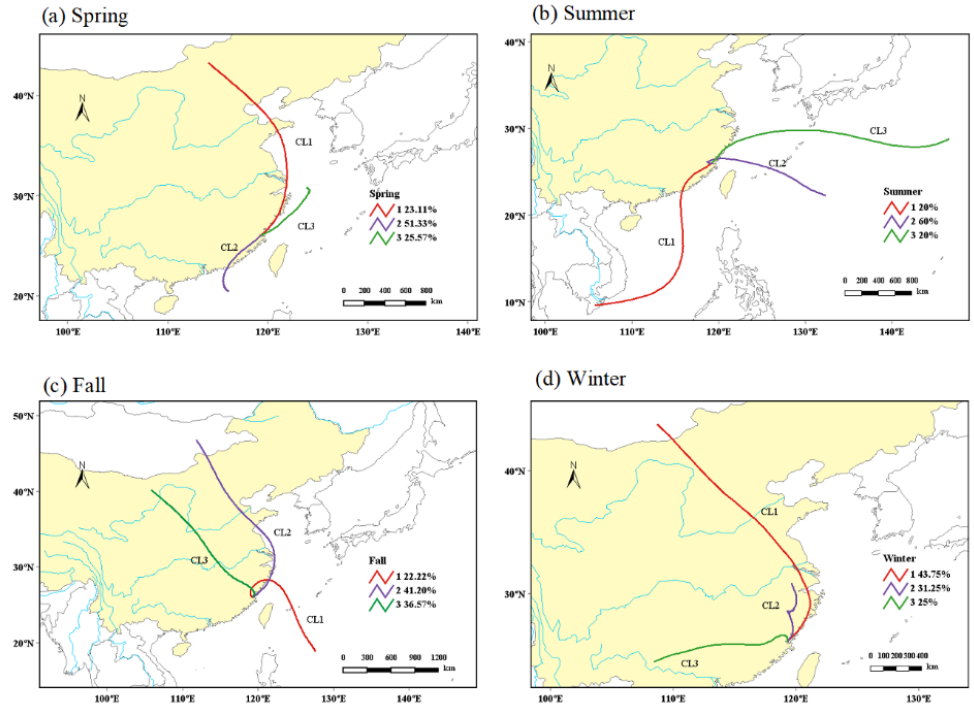

Fig. S9 Seasonal cluster analysis of 72-h back trajectories arriving at the Fuzhou site during NPF event days.

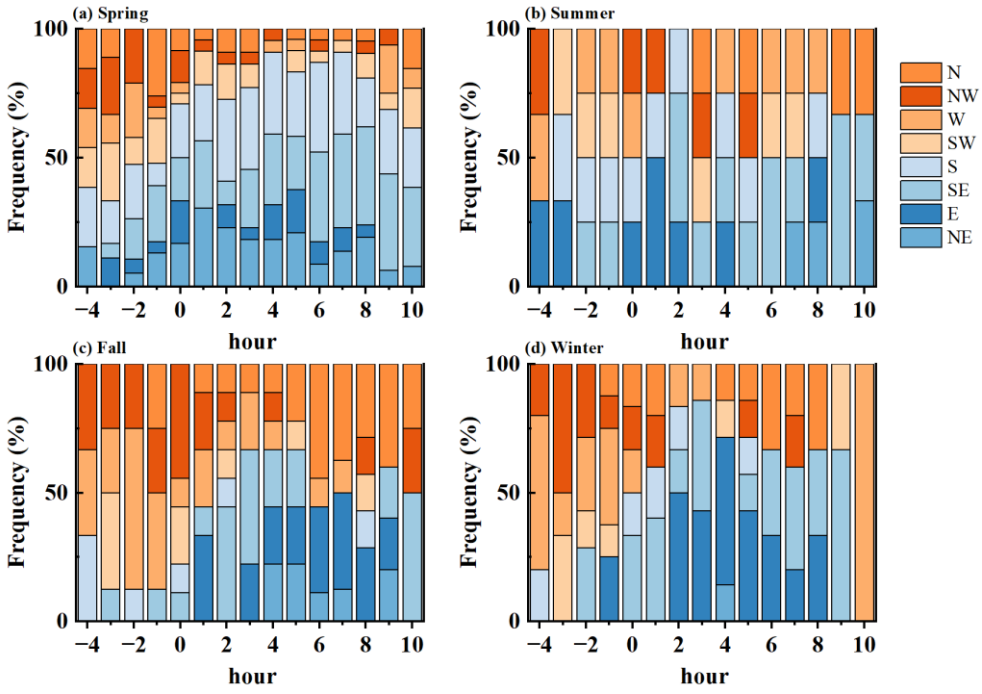

Fig.S10 Normalized diurnal frequency distribution of local wind directions surrounding NPF onset (t=0 h) for different seasons. Time zero (0 h) marks the onset of the NPF event, with negative and positive values indicating hours before and after the onset. The eight wind sectors are color-coded, with warm colors representing continental origins and cool colors representing marine/coastal origins.

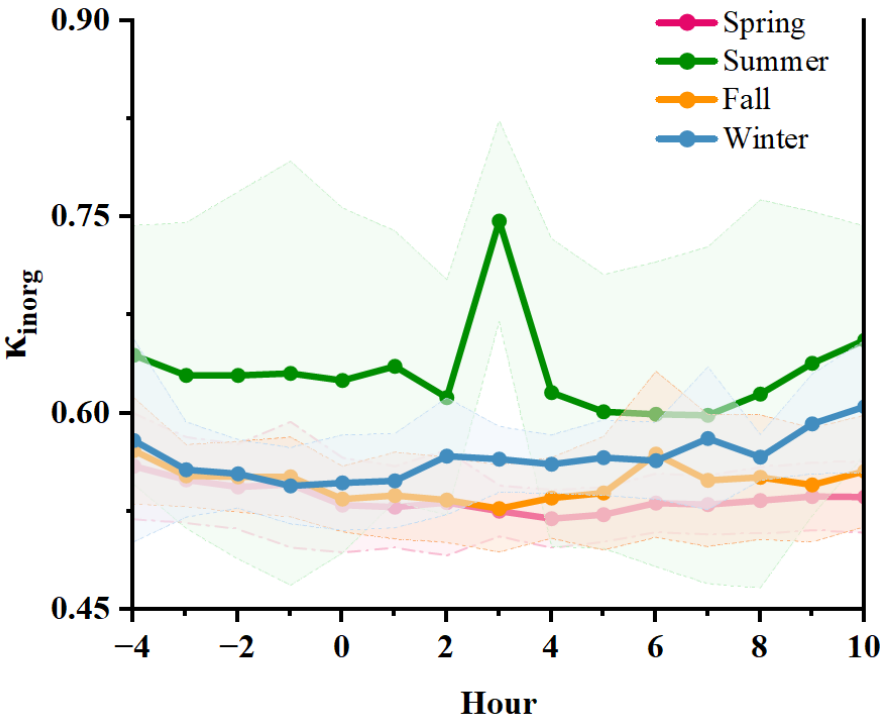

Fig.S11 Temporal evolution of the particle hygroscopicity parameter (κ_{inorg}) during NPF events. The x-axis follows the same normalized time scale as defined in Fig. 4 (t = 0 h represents NPF event start). Shaded bands indicate ±1σ standard deviation.

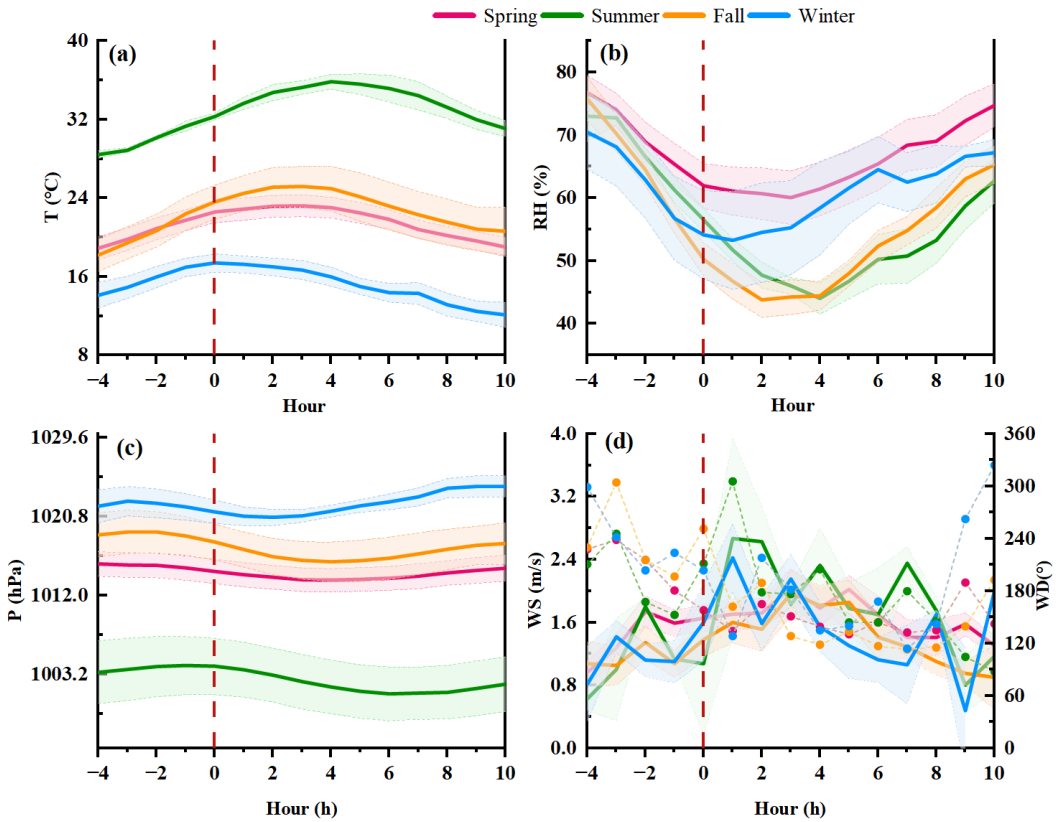

Fig. S12 Temporal evolution of meteorological parameters, including (a) temperature (T), (b) relative humidity (RH), (c) pressure (P), and (d) wind speed (WS, solid lines) and wind direction (WD, dashed lines with dots) during NPF events across four seasons. The x-axis follows the same normalized time scale as defined

in Fig. 4, where t = 0 h represents the onset of the NPF event. In panels (a)-(c), solid lines represent the median values, and shaded bands indicate. In panel (d), the left y-axis corresponds to WS, while the right y-axis corresponds to WD.

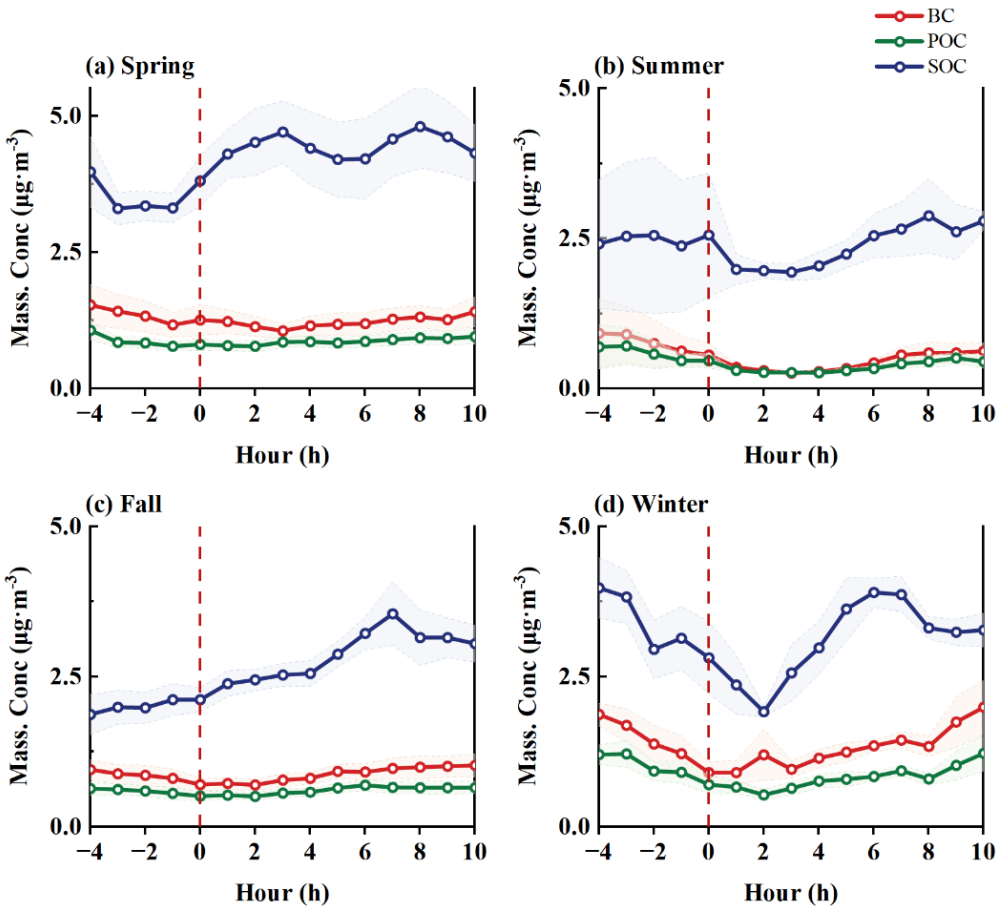

Fig.S13 Temporal variations in the mass concentrations of black carbon (BC), primary organic carbon (POC), and secondary organic carbon (SOC) during NPF events. The x-axis follows the same normalized time scale as defined in Fig. 4 (t = 0 h represents NPF event start). Lines with markers represent the median values across all NPF events in each season; shaded bands indicate ±1σ standard deviation.

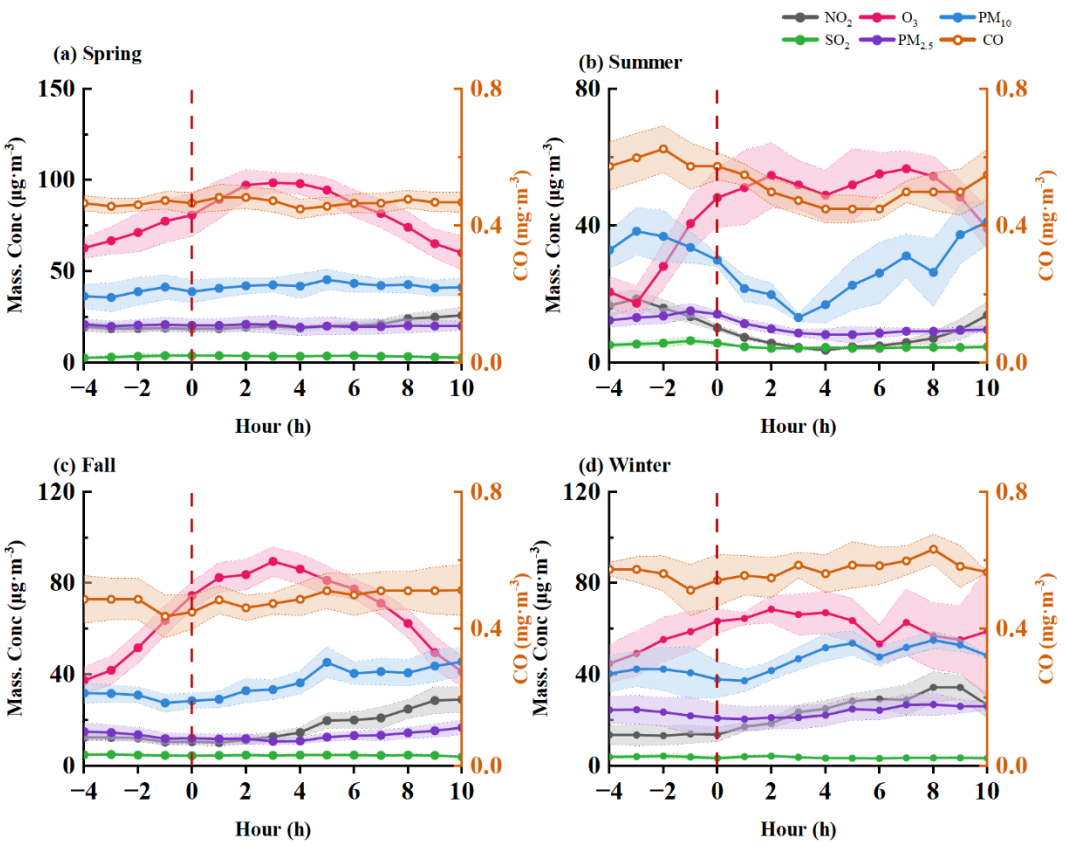

Fig.S14 Evolution of key atmospheric environmental factors during NPF events. The x-axis follows the same normalized time scale as defined in Fig. 4 (t = 0 h represents NPF event start). Lines with markers represent the median values across all NPF events in each season; shaded bands indicate ±1σ standard deviation.

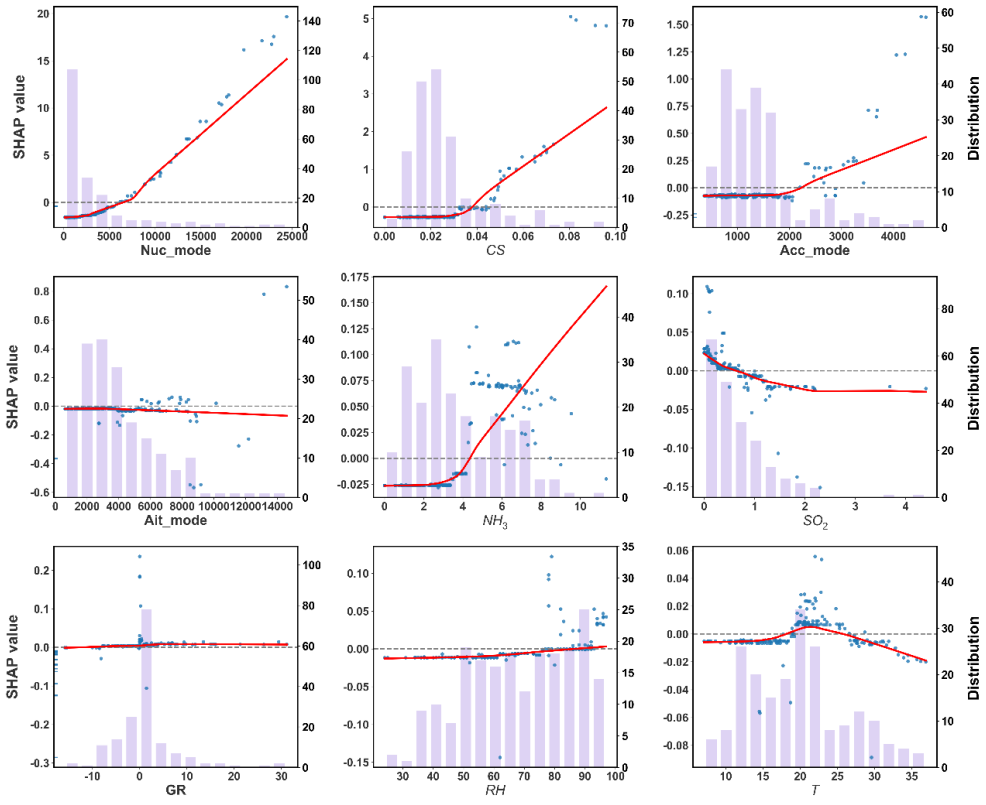

Fig.S15 Global importance analysis of influencing factors for particle formation rate (FR) based on SHAP summary plot.

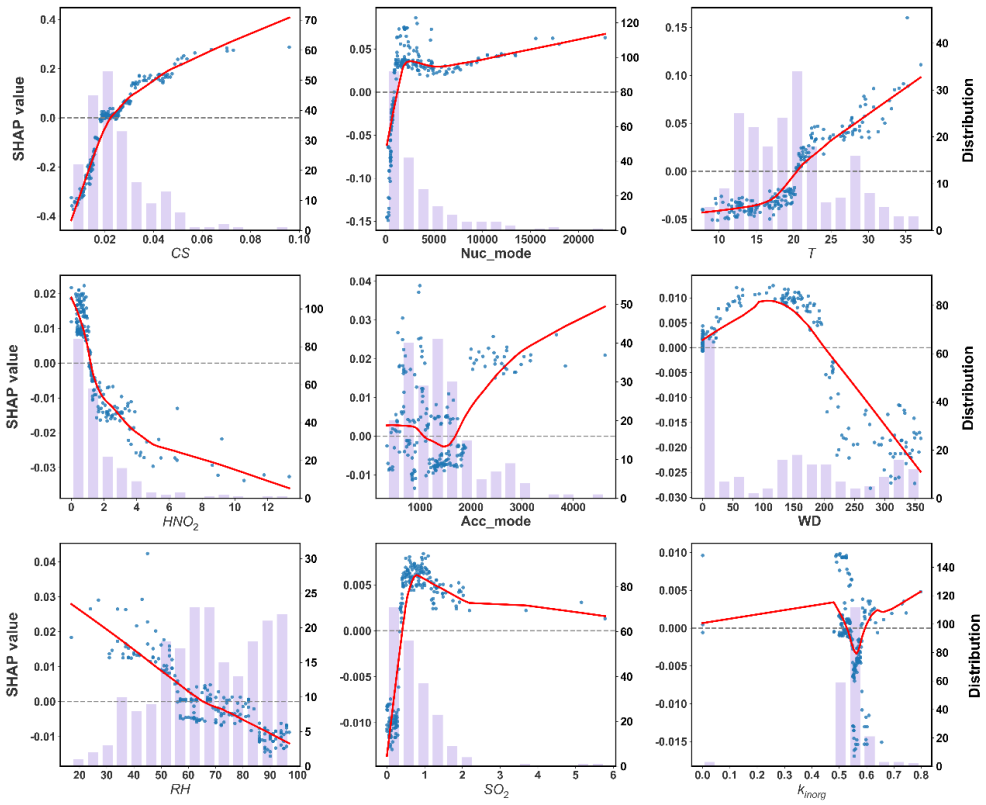

Fig.S16 Global importance analysis of influencing factors for Ait mode based on SHAP summary plot.

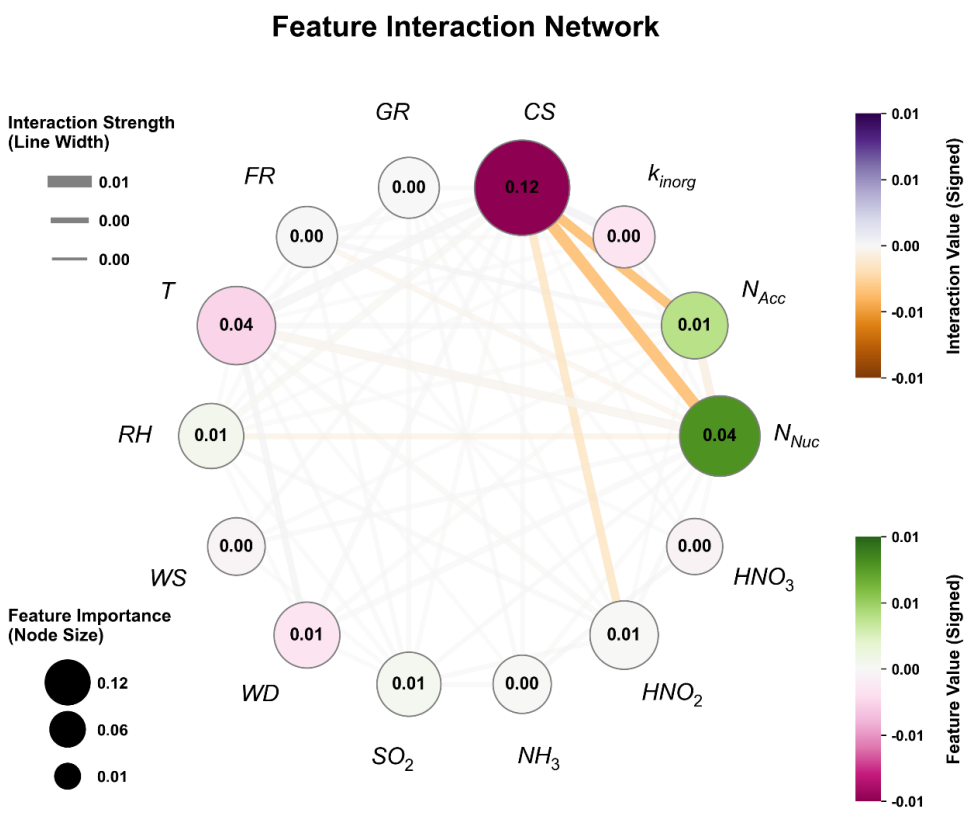

Fig.S17 Non-linear response relationships and interaction analysis between core environmental factors and N_{Ait}. Node size and inner numbers denote feature importance (mean |SHAP|). Node color indicates contribution direction (green: positive; pink: negative). Edge width reflects interaction strength; edge color indicates interaction direction (purple: synergistic; orange: antagonistic).

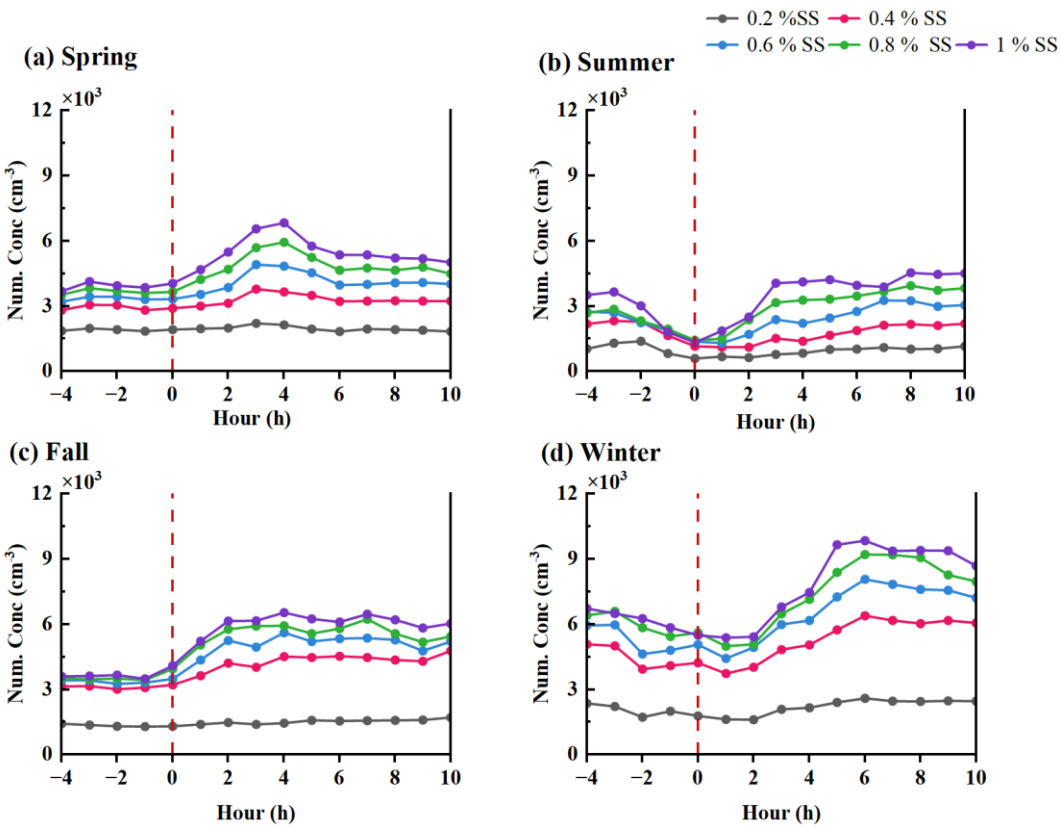

Fig.S18 Temporal evolution of cloud condensation nuclei (CCN) number concentrations at different supersaturations during NPF events. The x-axis follows the same normalized time scale as defined in Fig. 4, where t = 0 h represents the onset of the NPF event.

Reference:

Battelle: Environmental Technology Verification Report: Applikon MARGA Semi-Continuous Ambient Air Monitoring System, U.S. Environmental Protection Agency, available at: https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P100FZOD.pdf (last access: 24 April 2026), 2009.

Drinovec, L., Močnik, G., Zotter, P., Prévôt, A. S. H., Ruckstuhl, C., Coz, E., Rupakheti, M., Sciare, J., Müller, T., Wiedensohler, A., and Hansen, A. D. A.: The "dual-spot" aethalometer: an improved measurement of aerosol black carbon with real-time loading compensation, Atmos. Meas. Tech., 8, 1965–1979, https://doi.org/10.5194/amt-8-1965-2015, 2015.

Zhang, X., Trzepla, K., White, W., Raffuse, S., and Hyslop, N. P.: Intercomparison of thermal–optical carbon measurements by Sunset and Desert Research Institute (DRI) analyzers using the IMPROVE_A protocol, Atmos. Meas. Tech., 14, 3217–3231, https://doi.org/10.5194/amt-14-3217-2021, 2021.