Supplemental Information for:

Traffic, marine ships and nucleation as the main sources of ultrafine particles in suburban Shanghai, China

Qingsong Wang,‡[a] Juntao Huo,‡[b] Hui Chen,*[ahi] Yusen Duan,*[b] Qingyan Fu,[b] Yi Sun,[b] Kun Zhang,[a] Ling Huang,[a] Yangjun Wang,[a] Jiani Tan,[a] Li Li,*[a] Lina Wang,[c] Dan Li,[c] Christian George,[cd] Abdelwahid Mellouki[efg] and Jianmin Chen[cj]

a Key Laboratory of Organic Compound Pollution Control Engineering, School of Environmental and Chemical Engineering, Shanghai University, Shanghai, 200444, China

b Shanghai Environmental Monitoring Center, Shanghai, 200235, China

c Shanghai Key Laboratory of Atmospheric Particle Pollution and Prevention (LAP3) Department of Environmental Science and Engineering, Fudan University, Shanghai, 200433, China

d Univ Lyon, Université Claude Bernard Lyon 1, CNRS, IRCELYON, F-69626 Villeurbanne, France

e University Mohammed 6 Polytechnic (UM6P), Lot 660, Hay Moulay Rachid Ben Guerir, 43150, Morocco

f Institut de Combustion Aérothermique, Réactivité et Environnement, Centre National de la Recherche Scientifique (ICARE-CNRS), Observatoire des Sciences de 1'Univers en région Centre, 45071, Orleans, France

g Environmental Research Institute, Shandong University, Jinan, 250100, Shandong, China

h State Environmental Protection Key Laboratory of Formation and Prevention of Urban Air Pollution Complex, Shanghai Academy of Environment Sciences, Shanghai, 200233, China

i Key Laboratory of Atmospheric Chemistry, China Meteorological Administration, Beijing, 100081, China j Institute of Eco-Chongming, Shanghai, 200062, China

‡ The authors contributed equally.

* Corresponding authors: Hui Chen, huichen@shu.edu.cn; Yusen Duan, duanys@sheemc.cn; Li Li, lily@shu.edu.cn.

1. Sampling and data

DSL site is located in Qingpu District in a western suburb of Shanghai, and it is surrounded by Dianshan Lake, arable land, residences, rivers, and road network (Fig. 1a). DSL site is 0.4 km from the nearest G318 national highway and 1.5 km from the nearest G50 expressway. There are no traffic restrictions on the vehicle type and traffic time around DSL site. Since it is located at the junction of Shanghai, Zhejiang Province, and Jiangsu Province, all of which are well-developed areas with large populations, this site is often affected by regional pollution transport and suffers from photochemical pollution episodes.[1] DT site is inside the Chongming Dongtan Birds National Nature Reserve, which is on the east corner of Chongming island, as the entrance of Yangtze River to East China Sea. The reserve area is not open to the public. Chongming island has formulated a world-class ecological island construction plan, and the development of highly polluting industries is constrained. The main expressway G40 across Chongming Island is 15 km away from DT site.

SMPS used in this observation campaign has a single-channel uncertainty of ±15%. The sampling protocol for PNC measurement adheres to the EUSAAR-ACTRIS protocol.[2] The sampler inlet is equipped with an air dryer situated 10 m above the ground. From January 1 to March 31, 2021, approximately 5% of the 5-min and 1% of the hourly PNSD data are missing or invalid at both sites (Table S1). The mass concentrations of $PM_{2.5}$ were measured by a tapered element oscillating microbalance with filter dynamics measurement system (1405-F, TEOM-FDMS, Thermo Scientific™, USA).

The gaseous species were continuously measured using a set of gas analyzers including an O_3 analyzer (model 49i), NO/NO_2 analyzer (model 42i), SO_2 analyzer (model 43i), and CO analyzer (model 48i, Thermo Scientific™, USA). The concentrations of OC/EC, water-soluble ions and trace metal elements in $PM_{2.5}$ were measured by a semi-continuous thermo-optical carbon analyzer (Sunset Laboratory, USA), a Monitor for AeRosols and Gases in Ambient air (MARGA, ADI 2080, Metrohm, Netherlands), and a Multi-Metals Monitor System with dispersive X-ray fluorescence analysis (Xact® 625, Cooper Environmental, USA), respectively. More details can be found in the earlier publications.[3]

Meteorological parameters, including wind speed (Ws), wind direction (Wd), surface pressure (Press), relative humidity (RH), and ambient temperature (Temp) were simultaneously measured at respective site. Surface solar radiation (Ssr), total cloud cover (Tcc), and boundary layer height (Blh) are extracted from the gridded data of ERA5 (the fifth generation ECMWF reanalysis, accessible at https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels, last accessed: June 2023).

The median PNC_{Total} at DSL and DT sites are 7,300 # cm^{-3} and 4,500 # cm^{-3}, respectively, which is lower than the reported value of 15,300 # cm^{-3} in urban Shanghai in 2013.[4] A recent land-use regression study in Shanghai reported higher PNC_{UFPs} with a portable UFPs monitor.[5] In urban environments, there are more pollution sources, such as cooking activities,[6] which leads to higher PNC. The measurement duration is probably not the driving factor since studies in cities near Shanghai (i.e., Nanjing and Hangzhou) showed that PNC varied in a small range (<10%) among the seasons.[7] This needs further investigation because Shanghai is a coastal city and can be more easily influenced by ocean air masses. Effective measures against $PM_{2.5}$ in recent years in China can be another reason for the reduction in PNC_{UFPs}. PNC/$PM_{2.5}$ ratio is used as a quantitative measure of the relationship between PNC and $PM_{2.5}$ and to investigate how it varies for different cities and for different durations.[4] The highest PNC/$PM_{2.5}$ (> 1×10^9 # μg^{-1}) has been observed at roadside sites in the cities of very low $PM_{2.5}$.[4] PNC/$PM_{2.5}$ at the two sites in this study is 0.19×10^9 # μg^{-1} (DSL) and 0.22×10^9 # μg^{-1} (DT), respectively (Fig. S7), which are comparable to those Chinese cites of high $PM_{2.5}$.[4] This can be attributed to the fact that $PM_{2.5}$ originating from primary emissions and secondary formations with larger size do not contribute to UFPs, which leads to low PNC/$PM_{2.5}$. In addition, the size range of PNSD is the other important factor for PNC measurement. Our measurements were in a smaller range compared to the other studies and resulted in underestimated PNC (Fig. S7).

2. Implementation details of NMF model

In our study, the NMF analysis is performed using the "NMF" package[8] in R version 4.2.3. We removed invalid values from the original data and filled in approximately 1% missing hourly data by using the "missForest" package.[9] The principle of "missForest" is to use a random forest that has been trained on the observations of the data matrix to predict the missing values.[10]

It is important to determine the optimal number of factors (rank number, r). A common way of deciding on r is to try different values and choose the best value according to this quality criteria.[11] Brunet et al.[12] proposed to take the first value of r for which the cophenetic coefficient starts decreasing, Hutchins et al.[13] suggested to choose the first value where the RSS curve presents an inflection point. The quality criteria is provided as NMF rank survey in "NMF" package.[8] The results of the NMF rank survey are shown in Fig. S9. NMF rank survey and reordered consensus matrices suggest the rank number an objective consideration of the quantitative cophenetic coefficient rather than a subjective evaluation.[12]

DSL site. Great drop can be observed when r increases from 4 to 5 and from 5 to 6. The recorded consensus matrices show a nested structure as k increases from 2 to 5. Clear block diagonal patterns attest to the robustness of models with 2, 3, 4, and 5 clusters, whereas a r-6 factorization shows increased dispersion. Cophenetic correlation and consensus matrices do not suggest a rank higher than 6.

DT site. There are two plate value (r = 2–4 and r = 5–7, respectively) in cophenetic correlation when rank increases from 2 to 10. The recorded consensus matrices show a nested structure as r increases from 2 to 6. The boundaries among the clusters for r = 5, 6, are less distinct than those for r = 2–4. Cophenetic correlation and consensus matrices do not suggest a rank higher than 6.

The boundaries of consensus matrices for r = 5 at both sites are less distinct than those for smaller r. This is considered reasonable since PNSD from the sole emission source are not completely monodisperse (e.g., different types of vehicles and ships) and can contribute to the neighboring size bins. The PNSD reaching the site can also be influenced by the transport time (coagulation and condensation processes). The overlapping of the factors (clusters) cannot be prevented. Hence, we determine r by carefully investigating if the resolved factors provide meaningful information.

The meanings of resolved factors for $r = 5$ at both sites have been introduced in the main body text of this study. The results of r-6 factorization at both sites are illustrated in Fig. S12. The uncentered correlation coefficient (UCC) reported by Ulbrich et al. is also introduced to further validate the r.[14] The UCC is the cosine of the angle between a pair of PNSD or time spectra (TS) as vectors, such that $UC = \cos\theta = \boldsymbol{x} \cdot \boldsymbol{y} / (\| \boldsymbol{x} \| \| \boldsymbol{y} \|)$, where x and y denote a pair of PNSD or TS as vectors. The results of the UCC assessment are shown in Fig. S11.

Comparing Fig. 2 and S12, the largest difference between the results of 5 and 6 factors is that N5 at DSL site and N'3 at DT site are further separated to two factors, which are noted as F5a and F5b, and F'3a and F'3b, respectively (Fig. S13). At DSL site, F5a and F5b both exhibit clear bimodal distribution (Fig. S12a), which do not provide further meaningful interpretation. And the 6-factor solution results in high UCC_{PNSD} (> 0.7) for F4_F5b and F1_F2 and high UCC_{TS} (> 0.8) for F4_F5a. Similarly, 6-factor solution at DT site leads to high UCC_{PNSD} (> 0.8) for F'4_F'5 and high UCC_{TS} (> 0.7) for the neighboring factors among F'2, F'3a, F'3b and F4 (Fig. S11). This suggests that if r increases from 5 to 6, there will be even more overlap in PNSD and time series. 5-factor results show that N'3 is marine ship emissions but occasionally influenced by transport pollution.

The rank number of 5 is ultimately determined. The resolved profiles are positive, sparse, localized, and relatively independent, which makes a natural compact decomposition for interpretation. Moreover, the good fit of the NMF output results with the observation proves the good performance of NMF (Fig. S10).

3. Nonparametric tests on $PM_{2.5}$-bound V at DT site

In order to explore the differences in the distribution of V concentration among the groups in Fig 5b, the data of each group in Fig 5b were analyzed for significance. Firstly, IBM SPSS software was used to test the chi-square of the five groups of data, and the results showed that none of the five groups of data satisfied the chi-square. Therefore, we used the nonparametric independent sample test, again using IBM SPSS software. The results of the nonparametric test are shown in Table S3. The results show that there is no significant difference only between N'1, N'4, and N'5, while all other groups are significantly different from each other. Therefore, it can be stated that at the DT site, the concentration of V corresponding to the data above the 75th-percentile of N'2 and N'3 is significantly higher than the concentration of V corresponding to the other factors. And the concentration of V corresponding to the data above the 75th-percentile of N'3 is significantly higher than the concentration of V corresponding to N'2.

4. Respiratory deposits of particles

ICRP model. Impaction, sedimentation, diffusion are the three primary mechanisms of particle deposition in the airways. Particle transport and deposition in the airways varies depending on physiological factors, such as breathing rates, and on particle characteristics, such as size. Particles can be deposited in three different regions of the respiratory system: Head airway (HA, including nasal, pharyngeal and laryngeal passages), Tracheobronchial (TB), Alveolar (Alve, including pulmonary). International Commission on Radiological Protection (ICRP) obtained mathematical models of these three sedimentation modes through experimental data simulation in healthy adults.[15] The ICRP model defines the formula for calculating the lung deposition efficiency as follows. The deposition fraction for the HA (DF_{HA})is:

$$DF_{HA} = IF\left(\frac{1}{1+exp^{6.84+1.183lnD_p}} + \frac{1}{1+exp^{0.924-1.885lnD_p}}\right),$$

where IF is the inhalable fraction, $IF = 1 - 0.5\left(1 - \frac{1}{1+0.00076{D_p}^{2.8}}\right)$.

The deposition fraction for the TB (DF_{TB})is:

$$DF_{TB} = \frac{0.00352}{D_p}\left[exp^{-0.234(lnD_p+3.40)^2} + 63.9exp^{-0.819(lnD_p-1.61)^2}\right],$$

The deposition fraction for the Alve (DF_{Alve}) is:

$$DF_{Alve} = \frac{0.0155}{D_p}\left[exp^{-0.416(lnD_p+2.84)^2} + 19.11 \times exp^{-0.482(lnD_p-1.362)^2}\right],$$

The total deposition fraction (DF_{Total}) is the sum of the regional depositions:

$$DF_{Total} = DF_{HA} + DF_{TB} + DF_{Alve}.$$

Particle number concentrations deposited in respiratory system ($PNC_{Deposits}$). After obtaining the DF as a function of D_p, the PNC deposited in HA ($PNC_{Deposits,\ HA}$), in TB ($PNC_{Deposits,\ TB}$) and in Alve ($PNC_{Deposits,\ Alve}$) can be obtained from the following equations:

$$PNC_{Deposits,HA} = \sum_{D_p} DF_{HA} \times PNC(D_p),$$

$$PNC_{Deposits,TB} = \sum_{D_p} DF_{TB} \times PNC(D_p),$$

$$PNC_{Deposits,Alve} = \sum_{D_p} DF_{Alve} \times PNC(D_p),$$

where $PNC(D_p)$ refers to the PNC of sources at different particle sizes. $PNC_{Deposits}$ is the sum of $PNC_{Deposits,\ HA}$, $PNC_{Deposits,\ TB}$ and $PNC_{Deposits,\ Alve}$:

$$PNC_{Deposits} = \sum_{D_p} DF_{Total} \times PNC(D_p) = PNC_{Deposits,HA} + PNC_{Deposits,TB} + PNC_{Deposits,Alve}.$$

General Description.

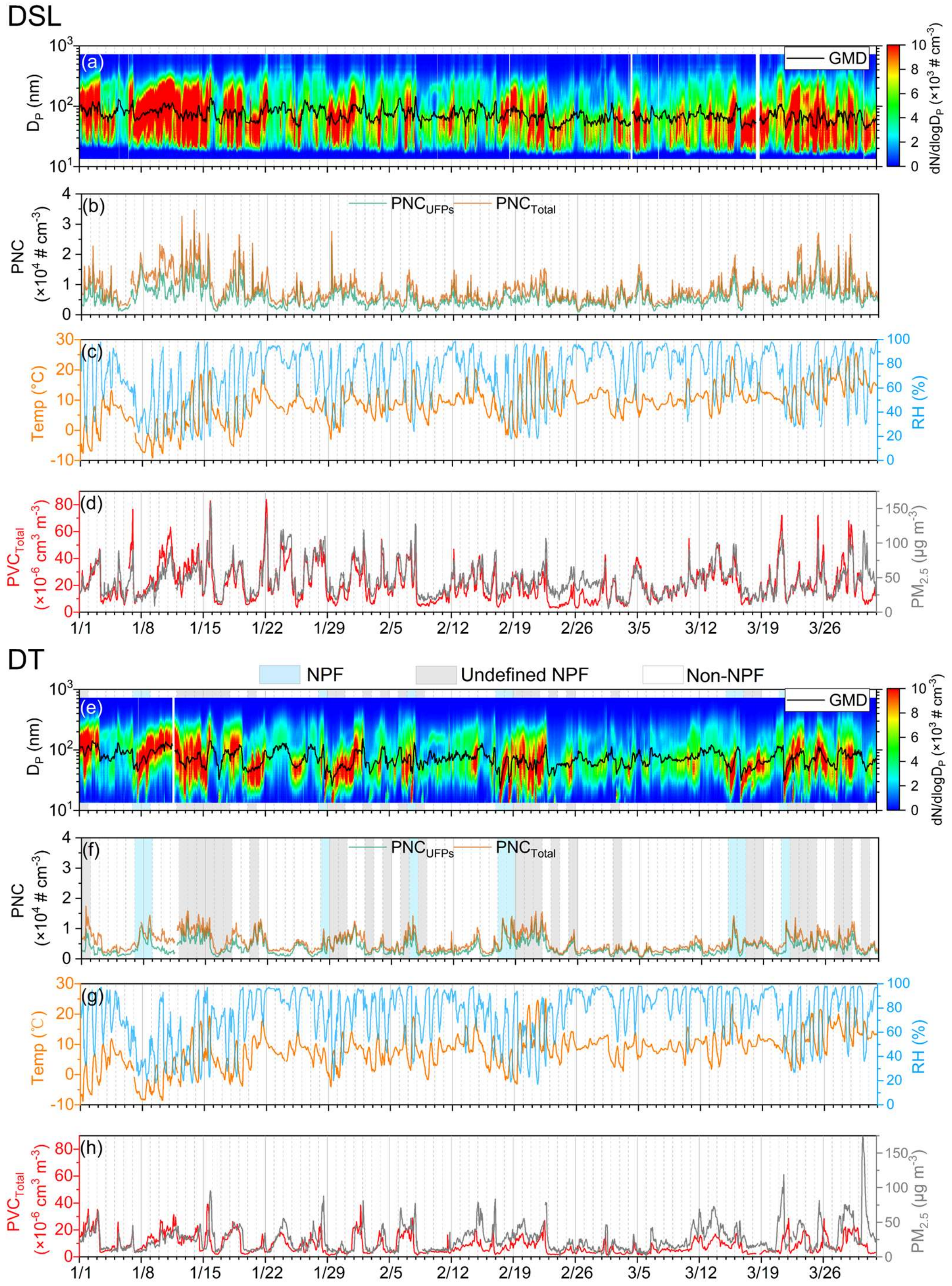

Figure S1. Time series plots of PNSD and GMD (a, e), PNC_{UFPs} and PNC_{Total} (b, f), ambient temperature (Temp) and relative humidity (RH) (c, g), PVC_{Total} and $PM_{2.5}$ (d, h) at DSL (a–d) and DT (e–h) sites. At DT site, NPF, undefined NPF and non-NPF days are marked with different color shading.

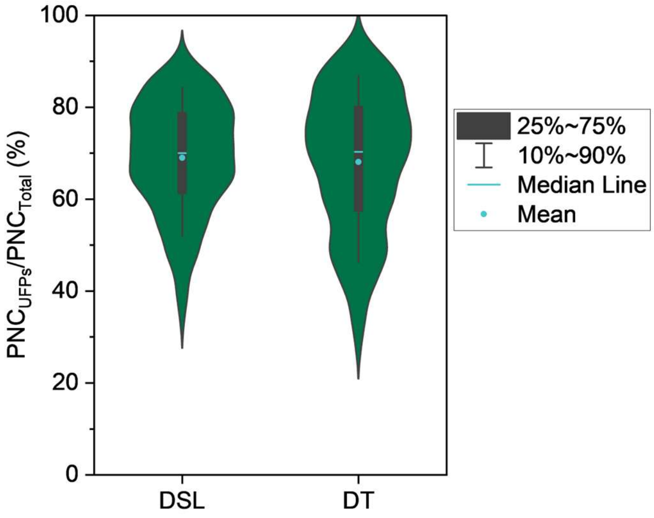

Figure S2. Boxplot of the ratios between PNC_{UFPs} and PNC_{Total} for DSL and DT Sites.

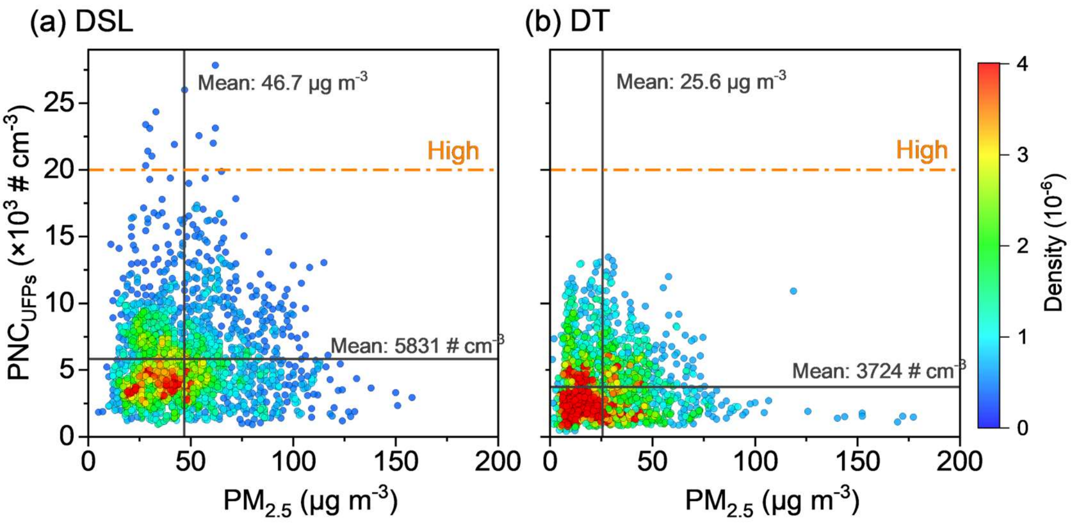

Figure S3. Scatterplots showing the relationship between hourly $PM_{2.5}$ mass concentration and PNC_{UFPs} at DSL (a) and DT (b) sites, the color scales of the dots represent the density distribution of dots. The vertical and horizontal black lines indicate the averages of $PM_{2.5}$ and UFPs, respectively, while the orange dash-dotted line represents the hourly high values (20,000 # cm^{-3}) suggested by WHO AQG2021. [16]

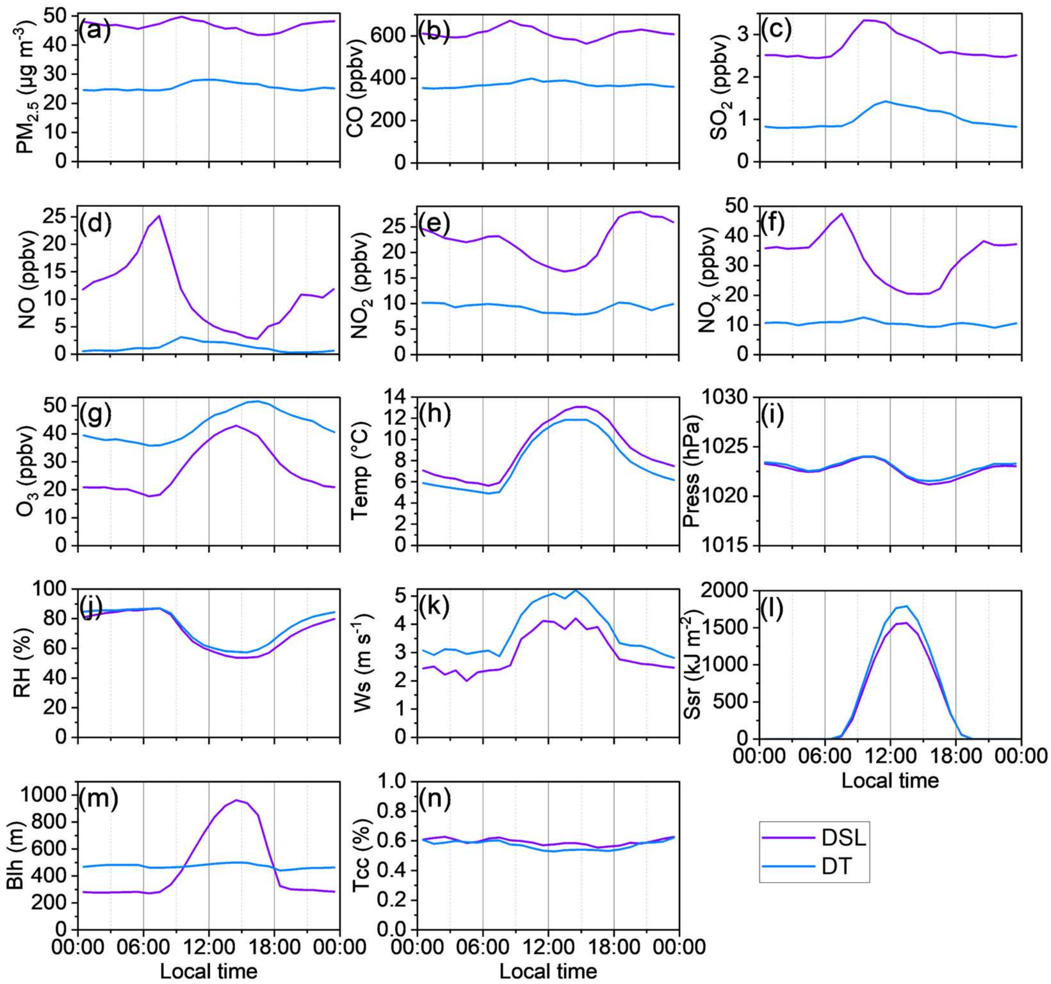

Figure S4. Diurnal variations of the average concentrations of criteria pollutants (a–g) and meteorological parameters at DSL and DT sites (h–n). Temp, Press, RH, Ws, SSr, Blh and Tcc represents ambient temperature, surface pressure, relative humidity, wind speed, surface solar radiation, boundary layer height and total cloud cover, respectively.

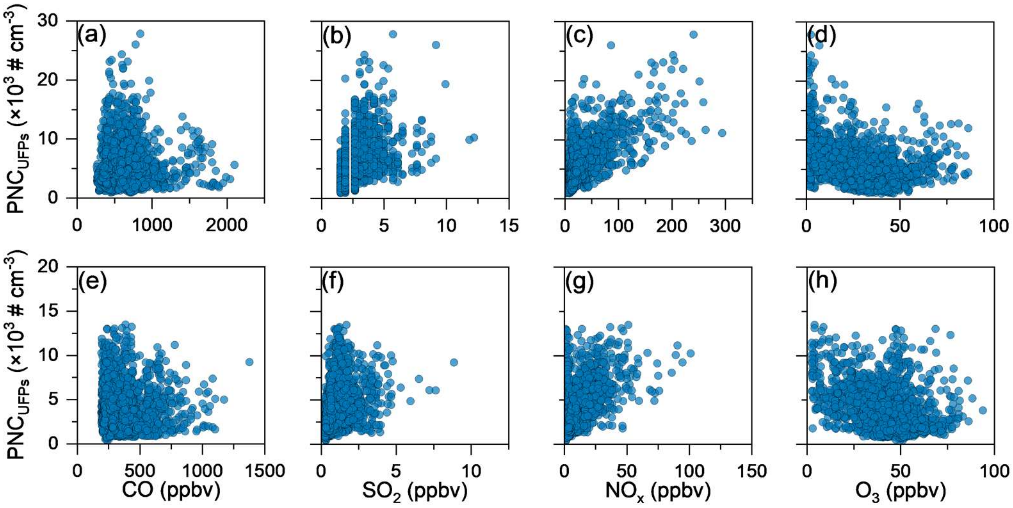

Figure S5. Scatterplots between mixing ratio of CO (a, e), SO_2 (b, f), NO_x (c, g) and O_3 (d, h) and PNC_{UFPs} at DSL (top panel) and DT (bottom panel) sites.

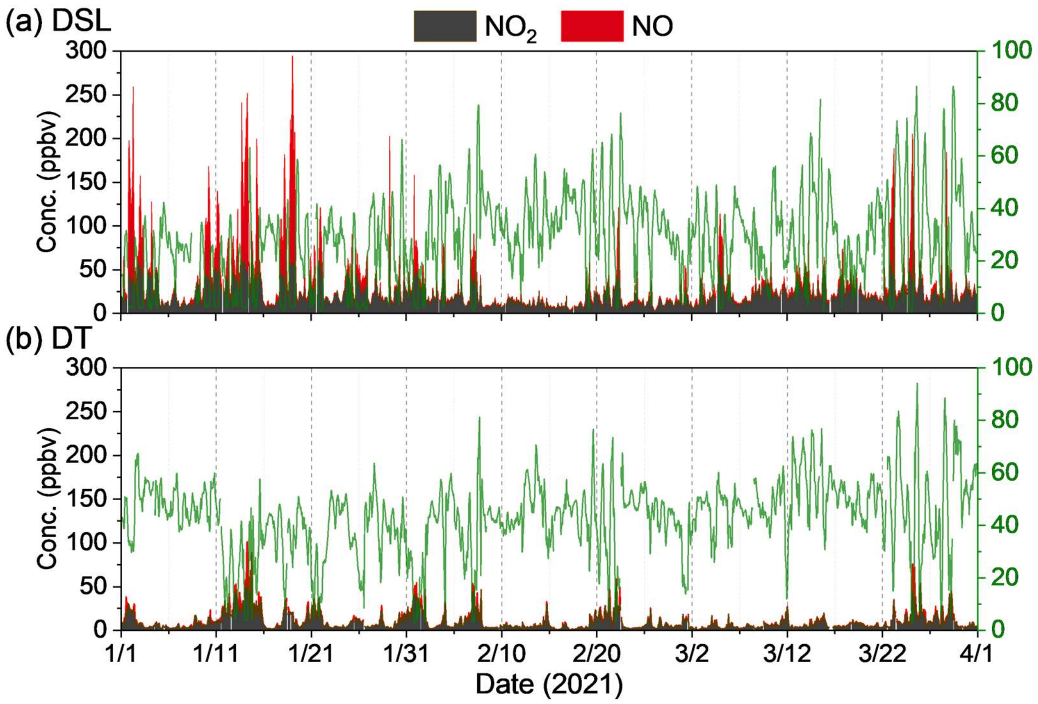

Figure S6. Time series of NO, NO_2 and O_3 at DSL (a) and DT (b) sites.

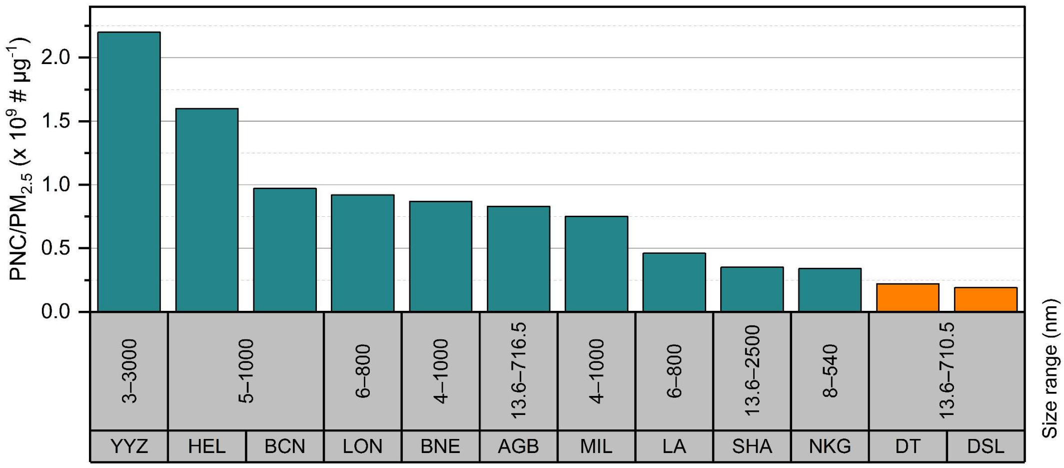

Figure S7. Comparison of the ratios between medians of PNC and $PM_{2.5}$ at cities as well as two suburban background sites (bars marked in orange) in this study. Data for other cities (bars marked in green) are from the investigation by de Jesus et al.[4]. The size ranges of PNSD are summarized in the table under graph.

NMF validation.

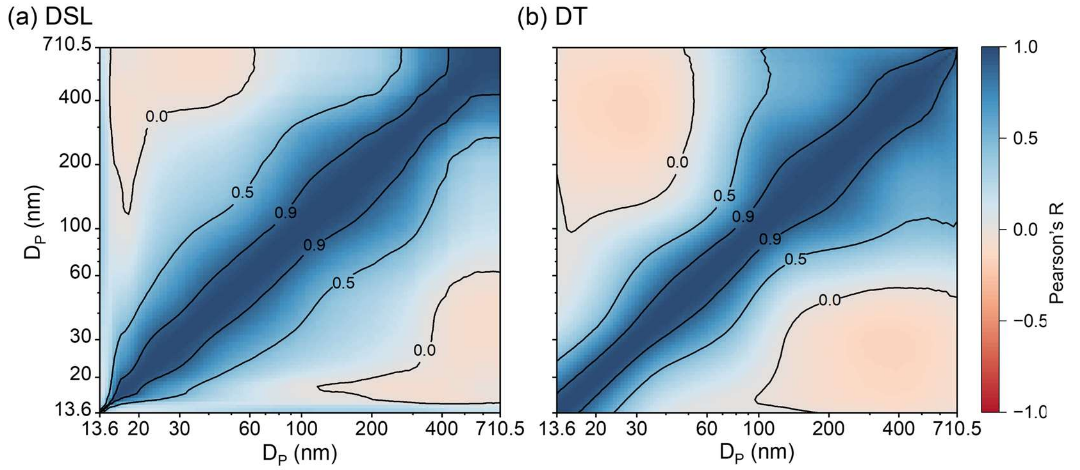

Figure S8. The contour graph of Pearson's R between each particle size bin of PNSD data at DSL (a) and DT (b) sites.

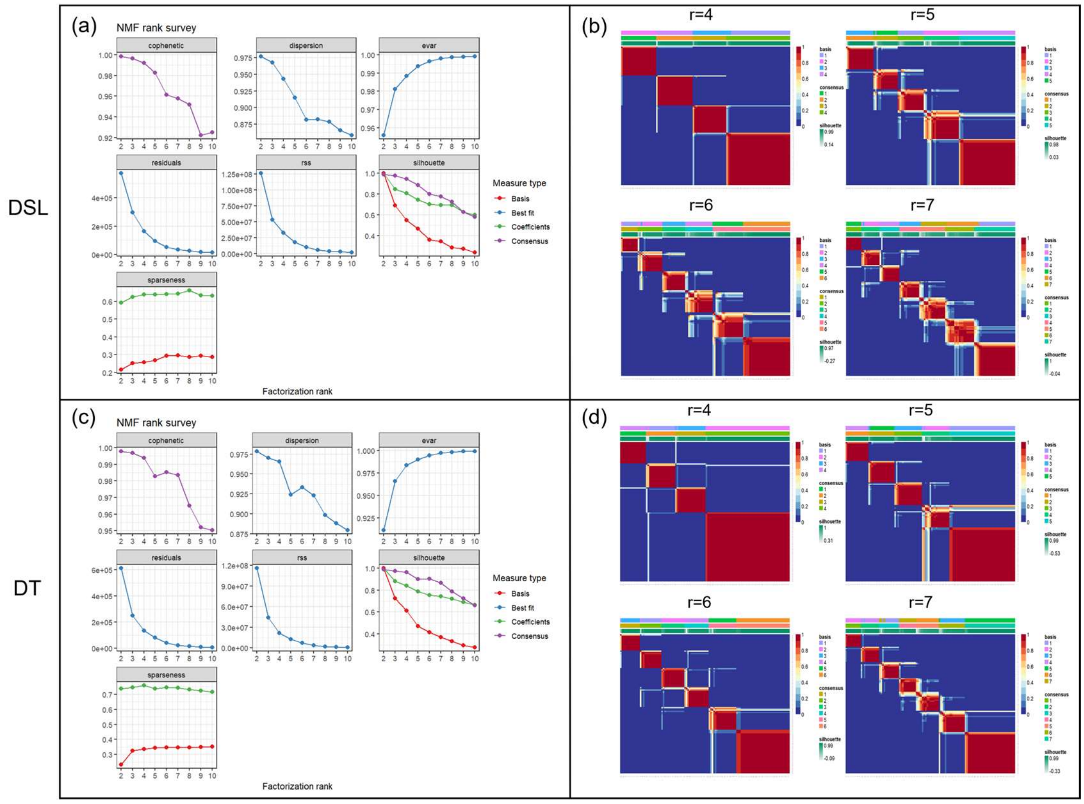

Figure S9. NMF rank survey for determine the optimal number of factors (r) at DSL (a, b) and DT (c, d) sites.

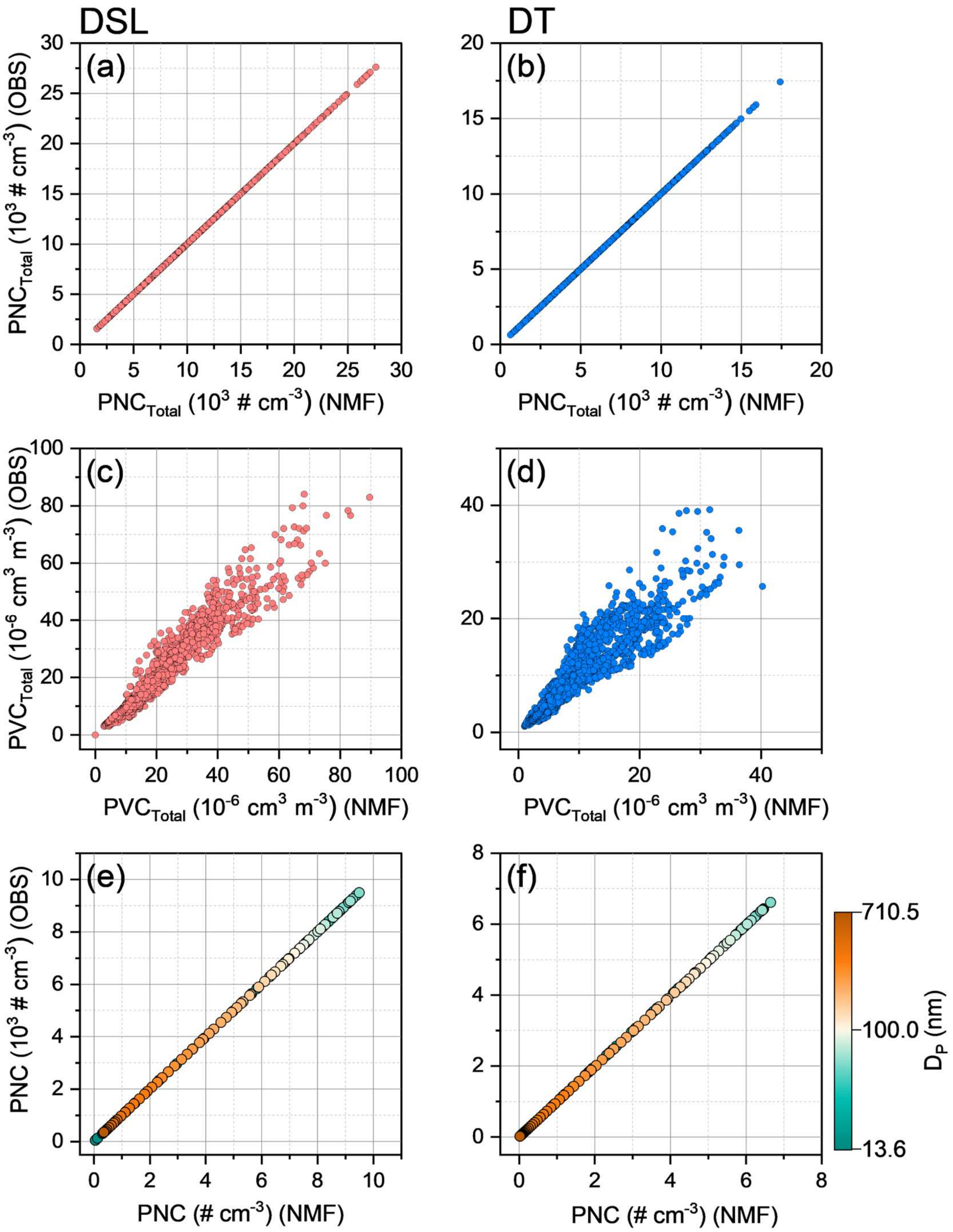

Figure S10. Scatterplots of the PNC (a, b) and PVC (c, d) of observed (OBS) against NMF output (NMF) at the DSL (left panel) and DT (right panel) sites. Scatterplots between the mean concentrations of OBS and NMF at the DSL (left panel) and DT (right panel) sites. The color scales of the dots represent the particle size bin of the SMPS measurement data.

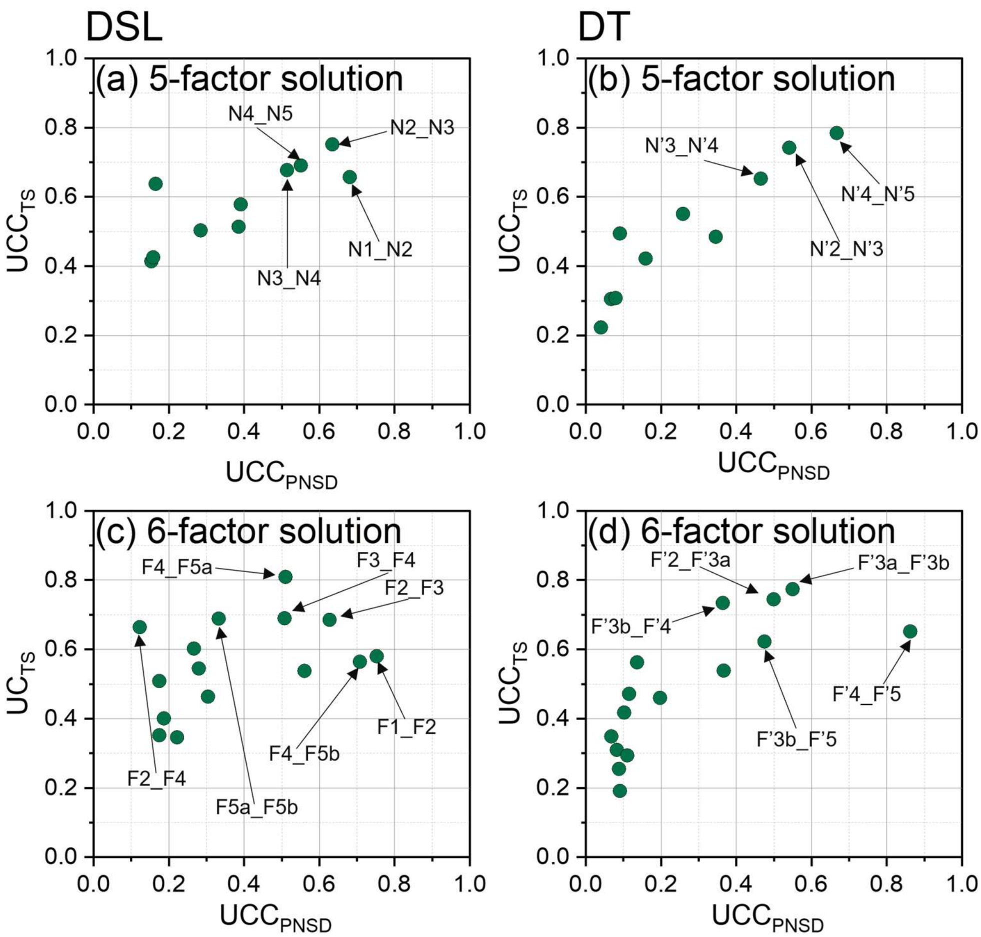

Figure S11. The uncentered correlation coefficient (UCC) of the time spectrum (TS) and PNSD between each of the 5 and 6 factor solutions of DSL (left panel) and DT (right panel) sites. Labels are shown when UCC > 0.6.

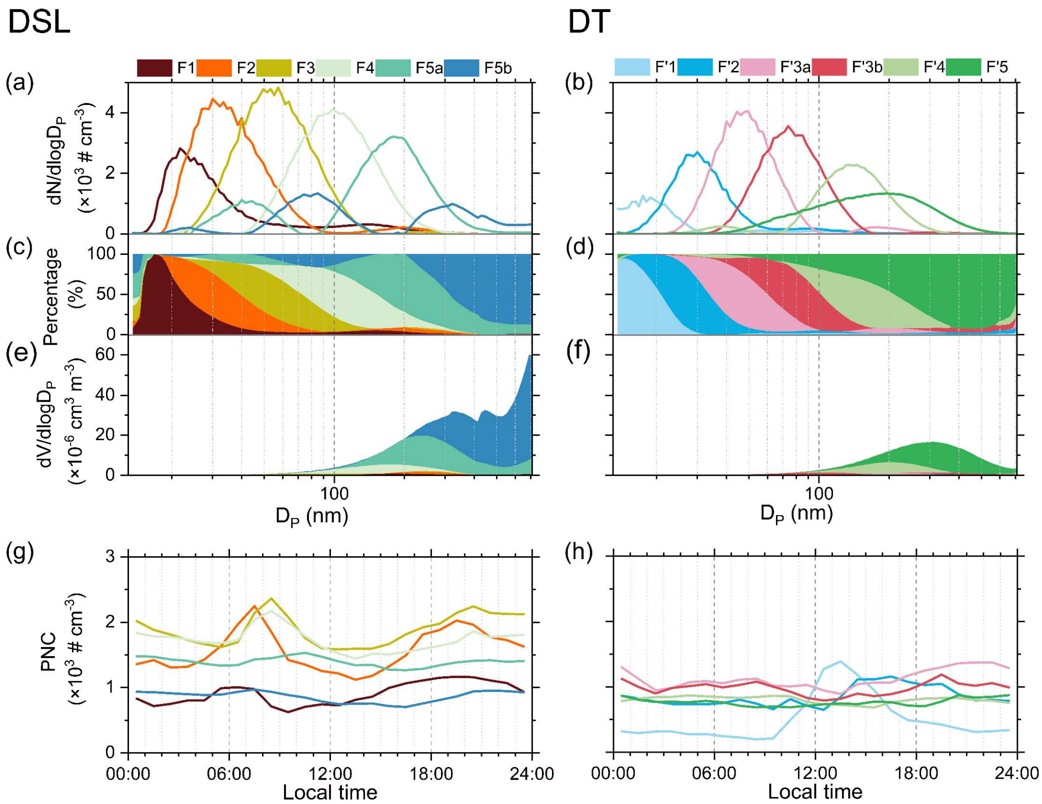

Figure S12. 6-factor NMF solution. PNC distributions (a, b), and corresponding contribution portions to PNC (c, d), PVC distributions (e, f) and diurnal variations (e, f) of the averages resolved factors at DSL (left panel) and DT (right panel) sites, respectively.

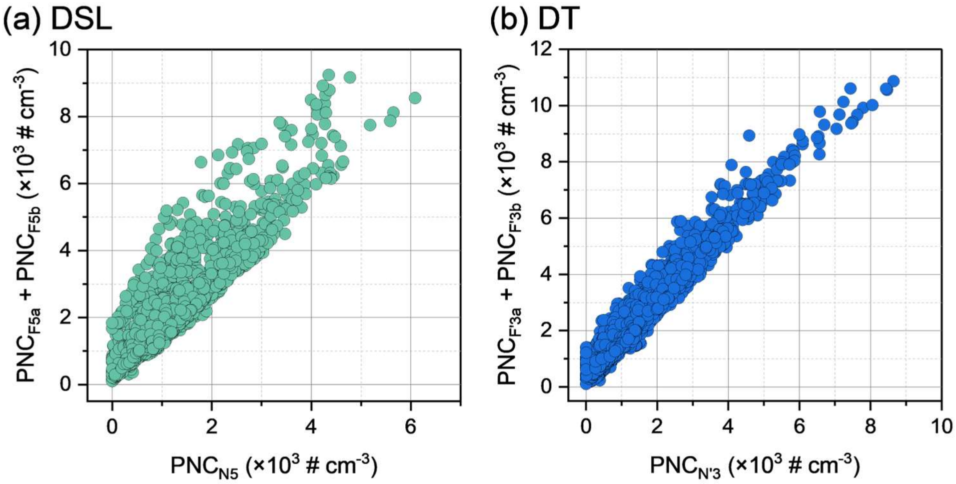

Figure S13. Scatterplots depicting the difference between the 5 factors and 6 factors solutions of DSL (a) and DT (b) sites. "N" and "F" represent the 5- and 6-factor solutions, respectively.

Source identification.

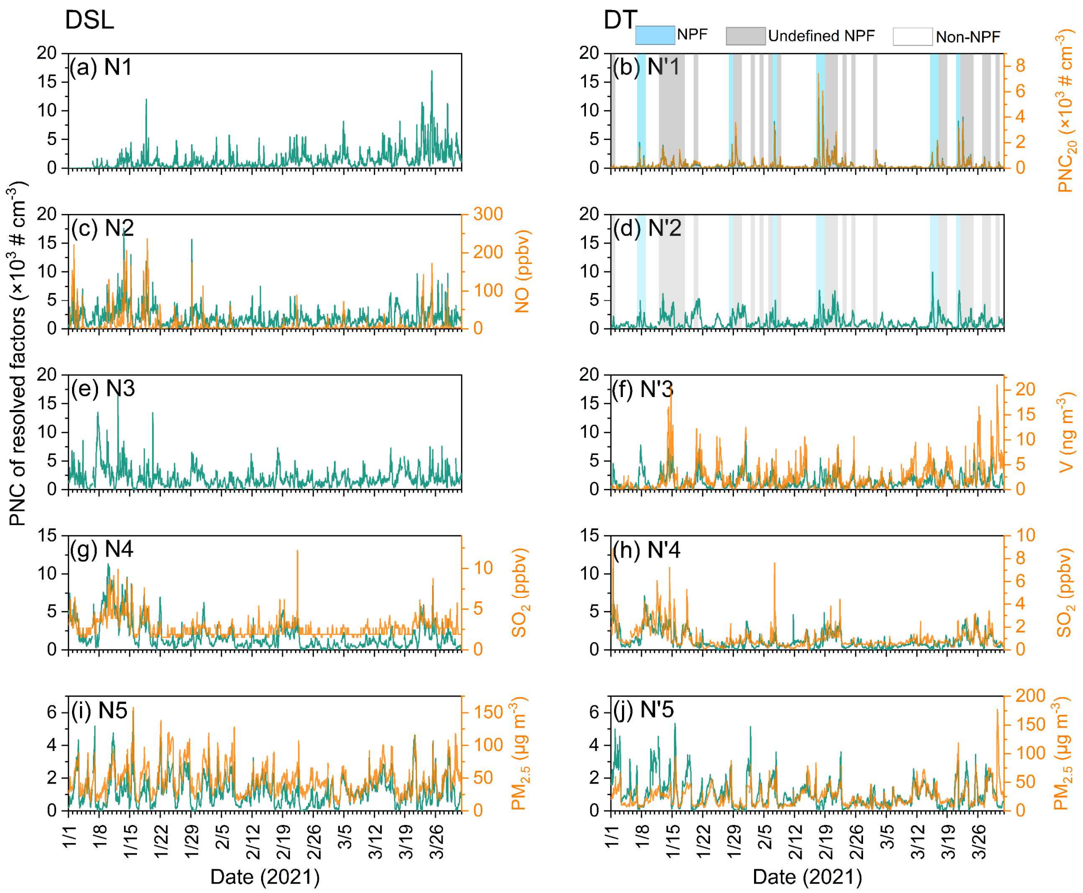

Figure S14. Time series plots of resolved factors at DSL (left panel) and DT (right panel) sites. Some corelated criteria pollutants and $PM_{2.5}$-bound V are plotted correspondingly. At DT site, NPF, undefined NPF and non-NPF days are marked with different color shading.

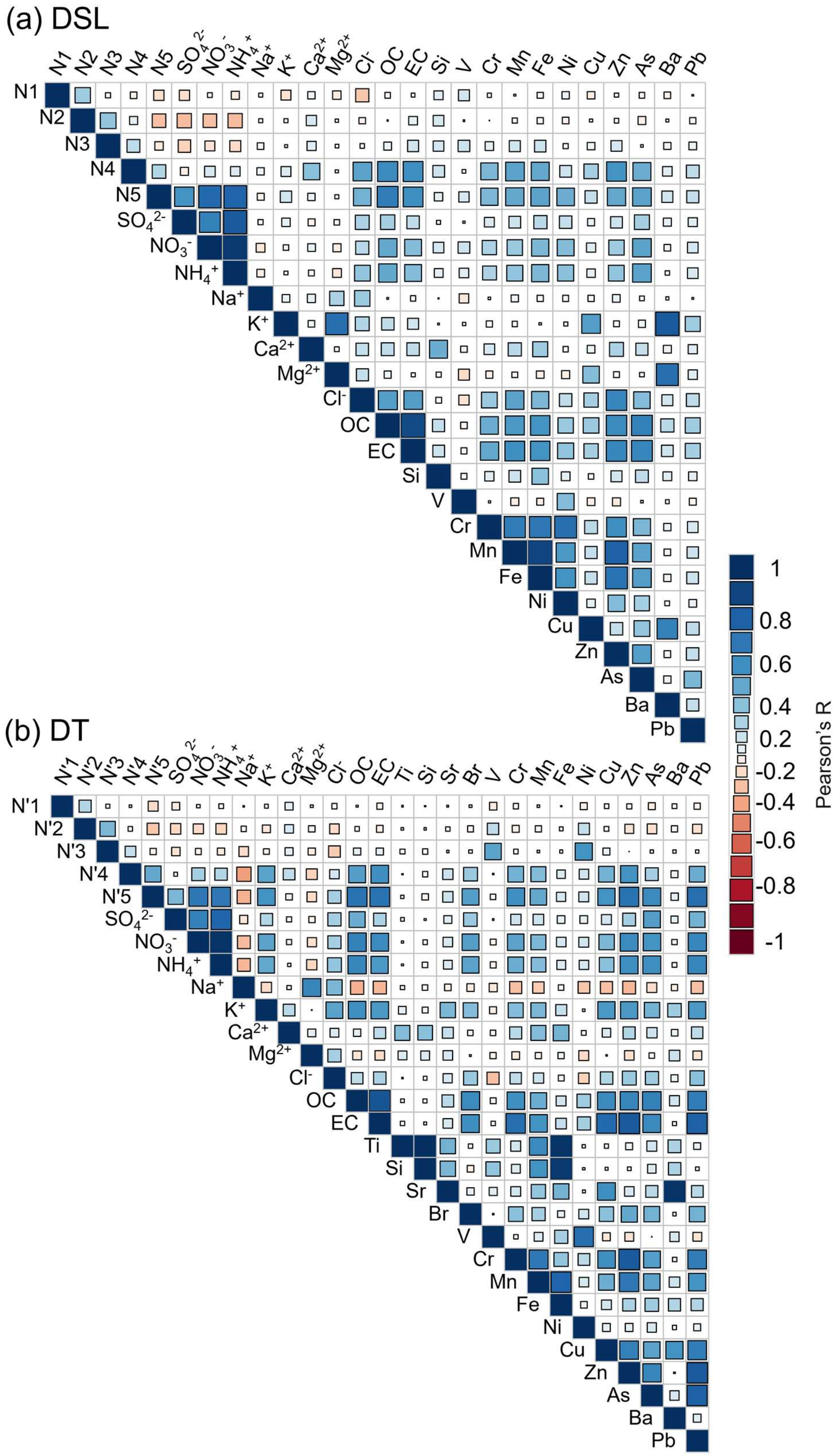

Figure S15. Pearson's correlation heatmaps between the resolved factors and chemical components of $PM_{2.5}$ at DSL (a) and DT (b) sites.

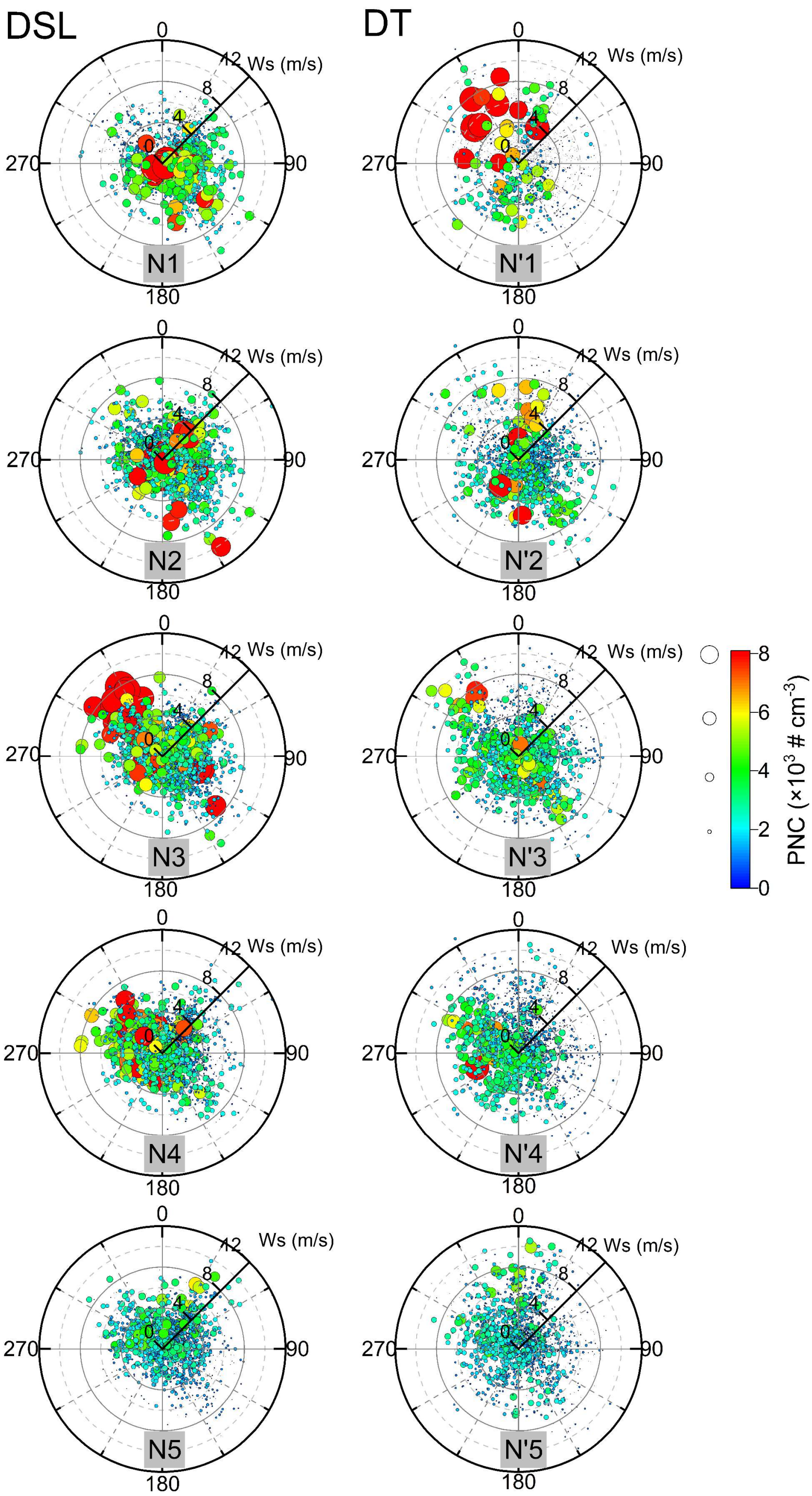

Figure S16. Polar plots of the resolved factors at DSL (left panel) and DT (right panel) sites, the color scales and size of the dots represent PNC of resolved factors.

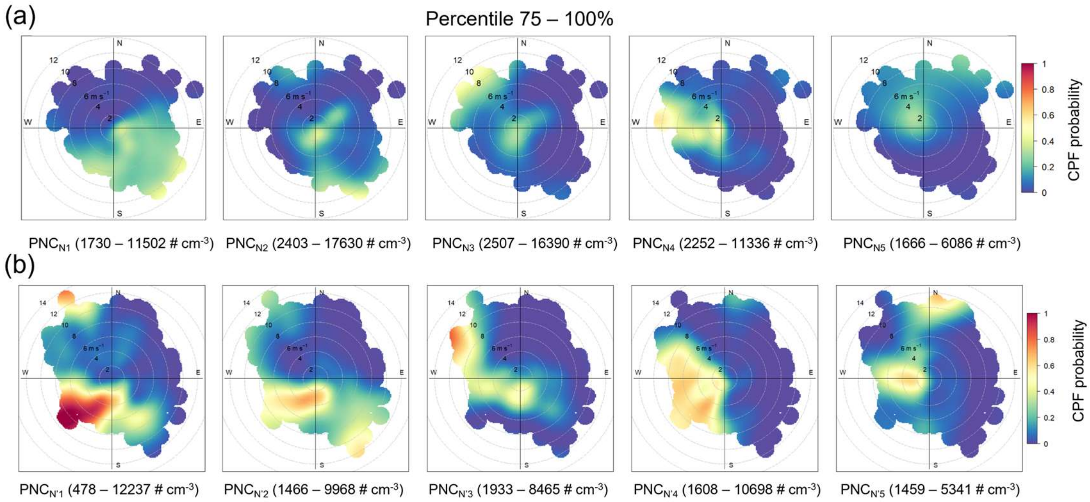

Figure S17. The conditional probability function (CPF) polar plots for the 75th–100th percentiles of resolved factors at DSL (a) and DT (b) sites.

Marine ship emissions.

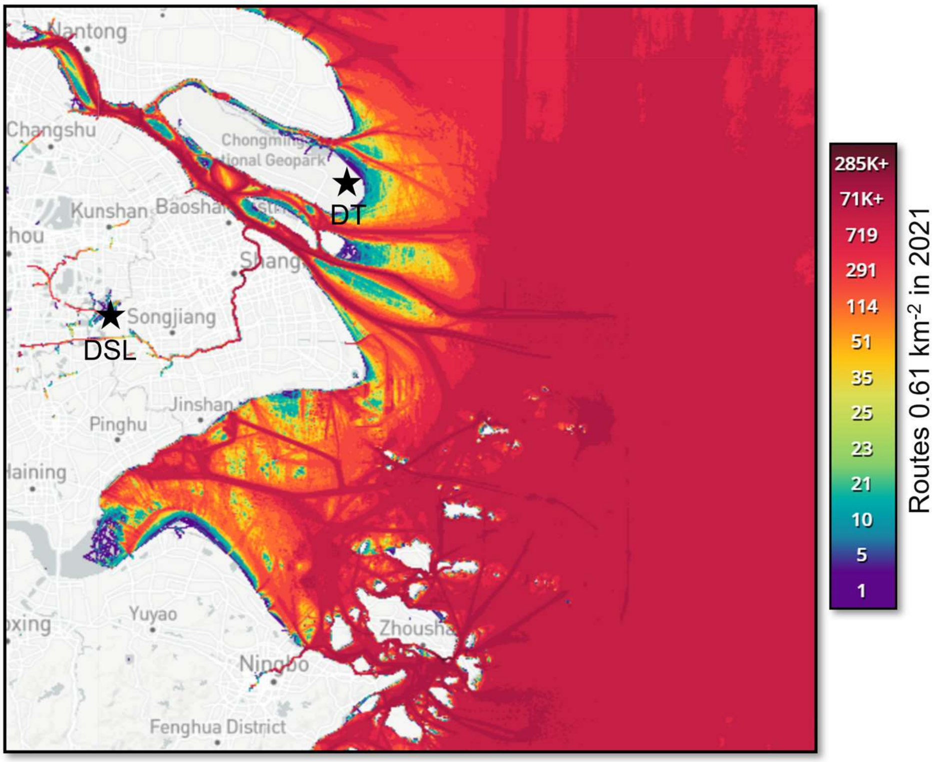

Figure S18. Heatmap of navigation routes of ships close to Shanghai in 2021. (https://www.marinetraffic.com/en/ais/home/centerx, last accesses: May 2023). DSL and DT sites are marked in black stars.

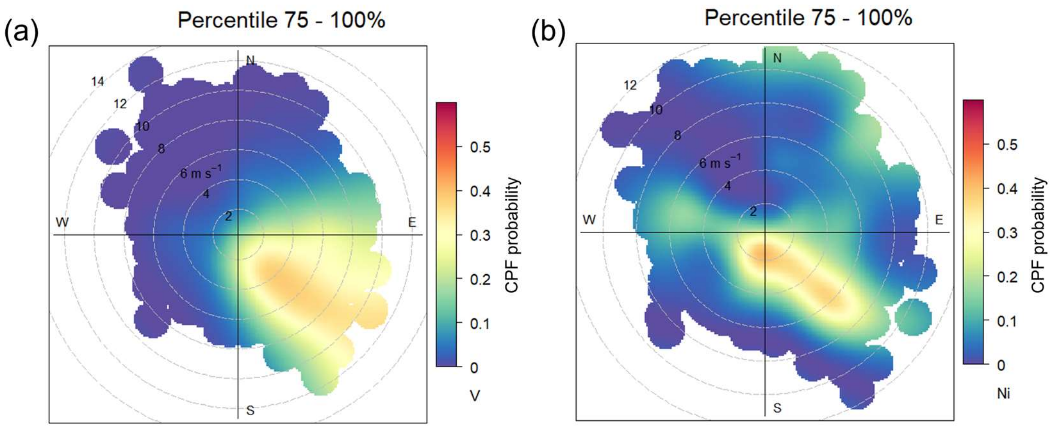

Figure S19. The CPF polar plots for the 75th–100th percentiles of V (a) and Ni (b) at the DT site.

Nucleation.

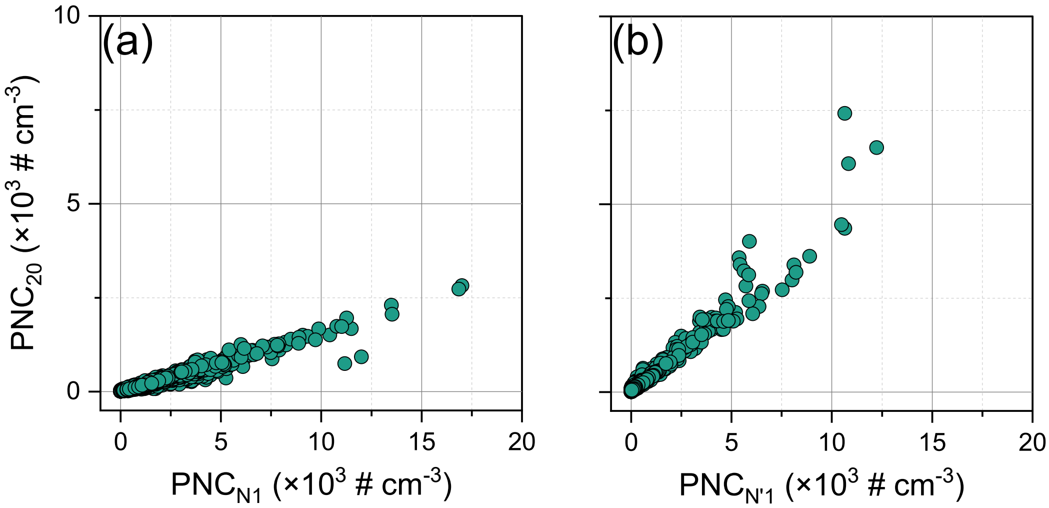

Figure S20. Scatterplots of PNC_{N1} (a) and $PNC_{N'1}$ (b) at DSL and DT sites against the measured PNC below 20 nm (PNC_{20}).

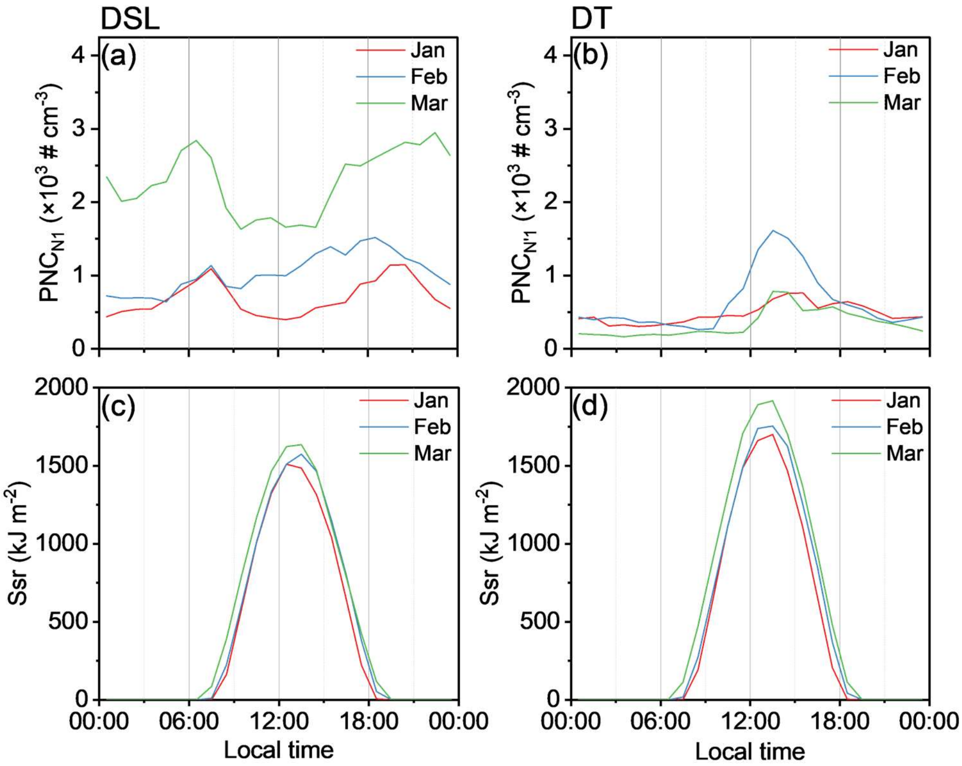

Figure S21. Diurnal variations of the averages of PNC$_{N1}$ (a), PNC$_{N'1}$ (b) and Ssr (c, d) at DSL (left panel) and DT (right panel) sites within each month.

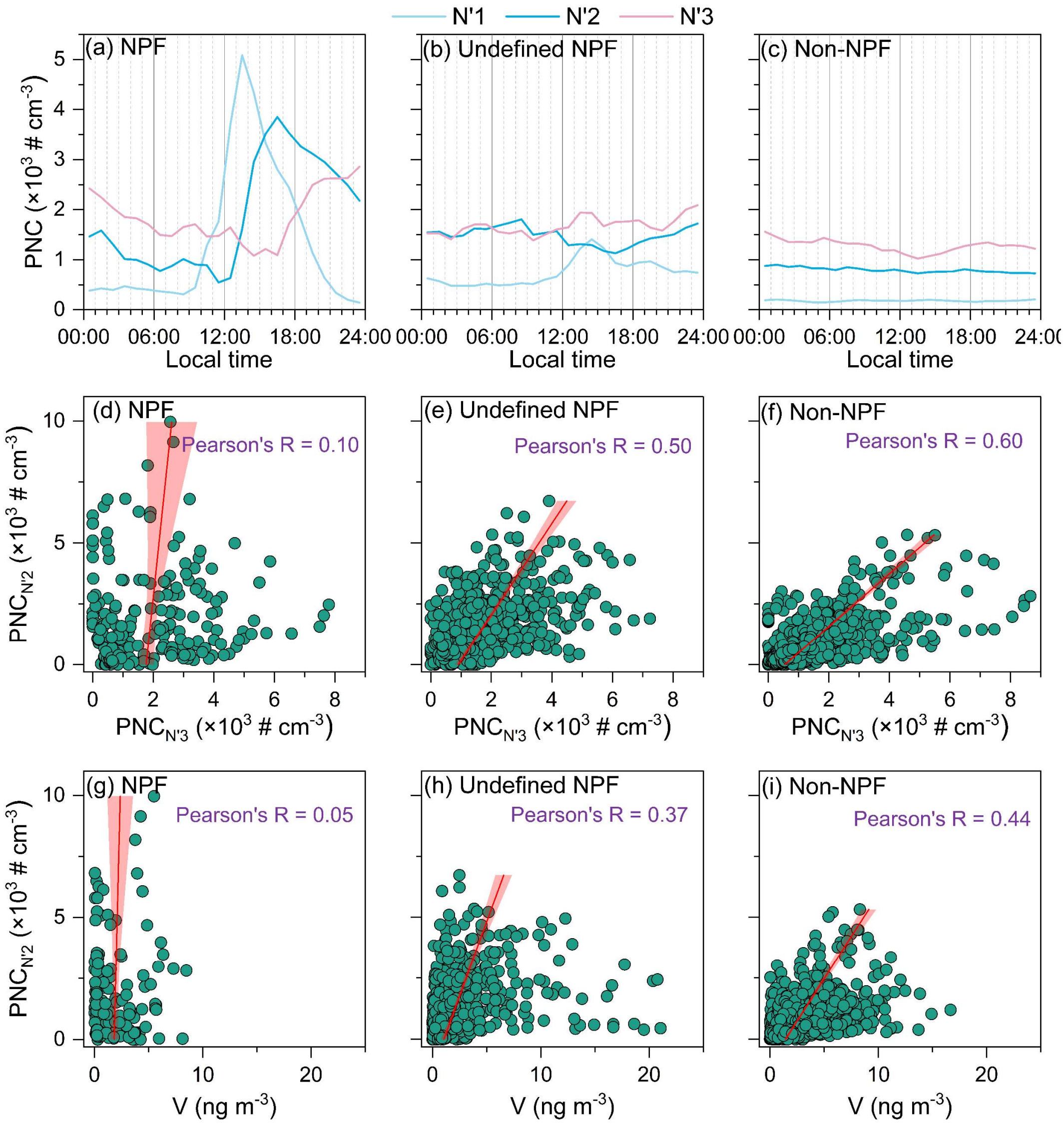

Figure S22. Diurnal variations of average $PNC_{N'1}$, $PNC_{N'2}$ and $PNC_{N'3}$ for NPF (a), undefined NPF (b) and non-NPF days (c). Scatterplots between $PNC_{N'2}$ and $PNC_{N'3}$ (c, d, e), and between $PNC_{N'2}$ and V (g, h, i) are plotted correspondingly.

Industrial emissions and regional background.

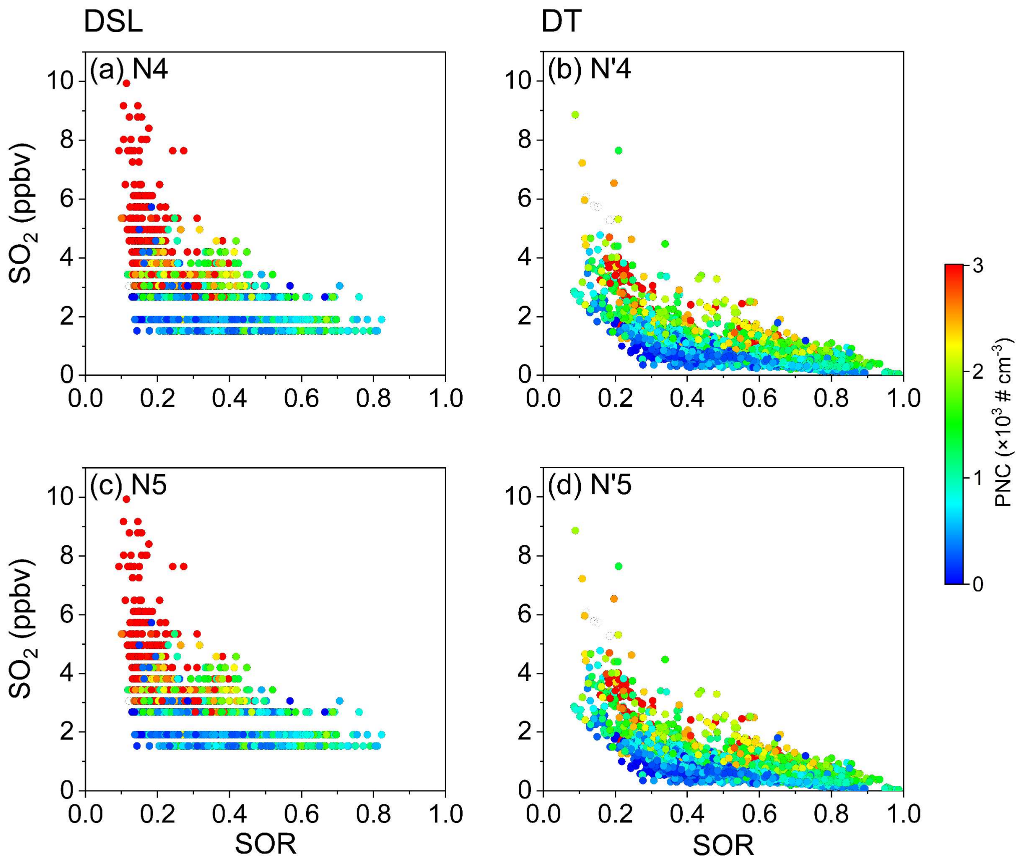

Figure S23. Scatterplots between SO_2 and sulfur oxidation ratio (SOR) for DSL (left panel) and DT (right panel) sites, the color scales of the dots represent the PNC_{N4} (a), $PNC_{N'4}$ (b), PNC_{N5} (c) and $PNC_{N'5}$ (d), respectively.

Specific Events.

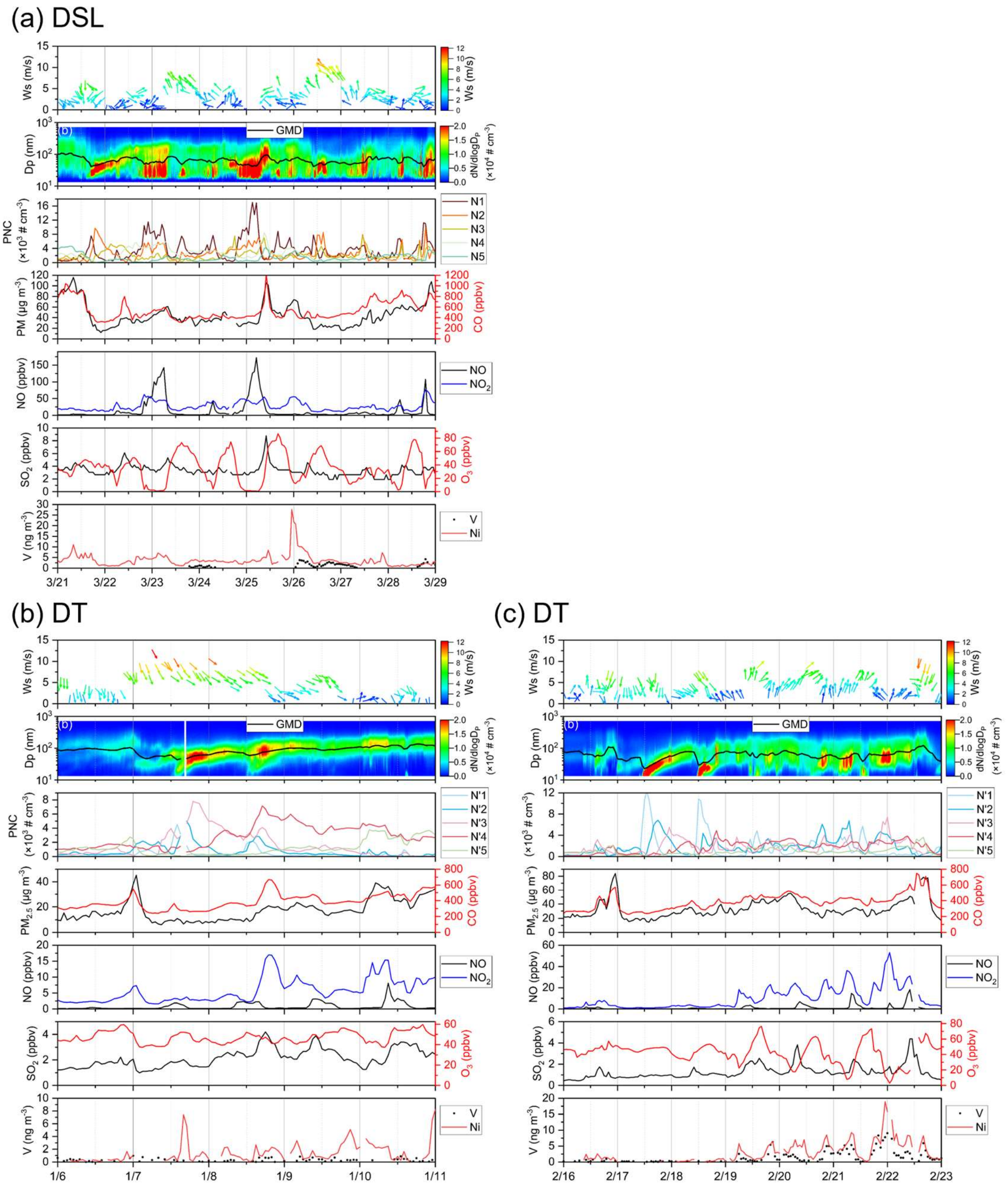

Figure S24. Time series of the resolved PNC factors and $PM_{2.5}$, CO, NO, NO_2, O_3, and V and Ni in $PM_{2.5}$ in specific events at DSL (a) and DT (b and c) sites.

Respiratory deposition.

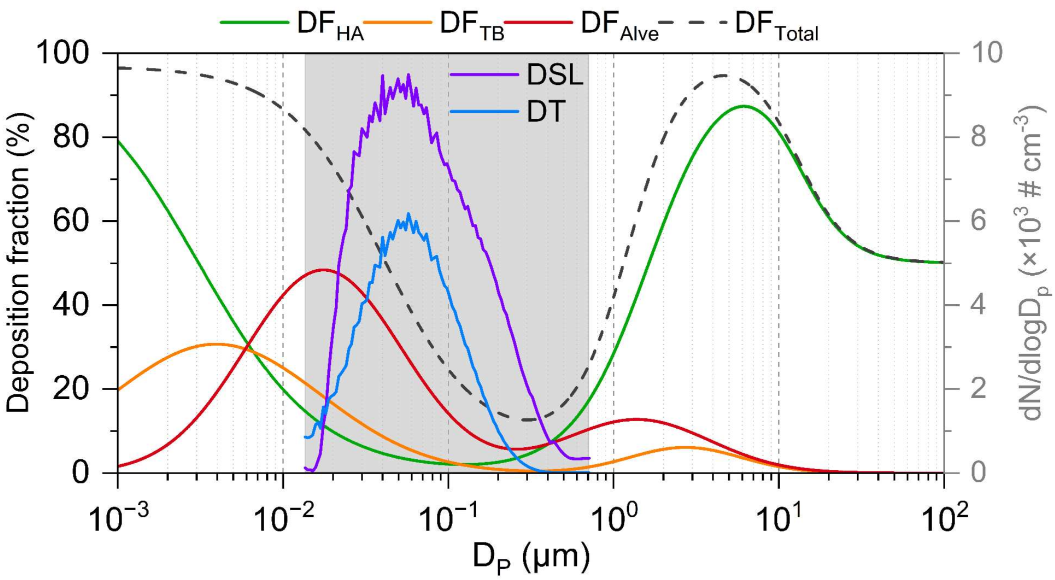

Figure S25. Curve of respiration deposition efficiency defined by the ICRP model. The green, orange and red lines represent the deposition efficiency of particles in head airways (HA), tracheobronchial region (TB) and alveolar region (Alve), respectively, and the black dashed line represents the total deposition efficiency. The blue shaded range indicates the range of PNSD in this study. The blue and purple lines represent the average PNSD observed at DSL and DT sites, respectively.

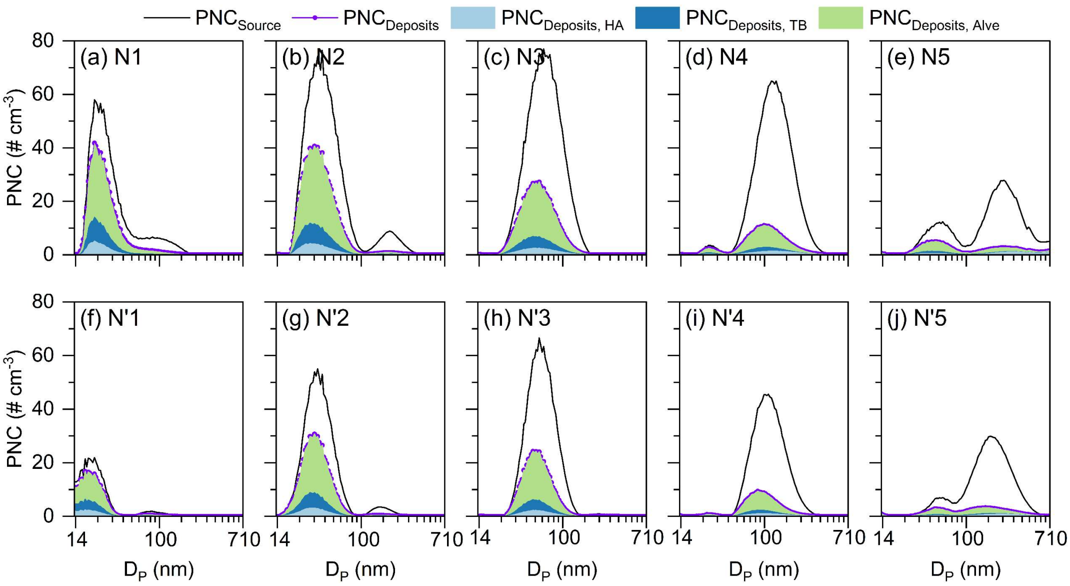

Figure S26. The PNSD and deposition profiles in various regions of the respiratory system of each source for DSL (top panel, N1–N5) and DT (bottom panel, N'1–N'5) sites.

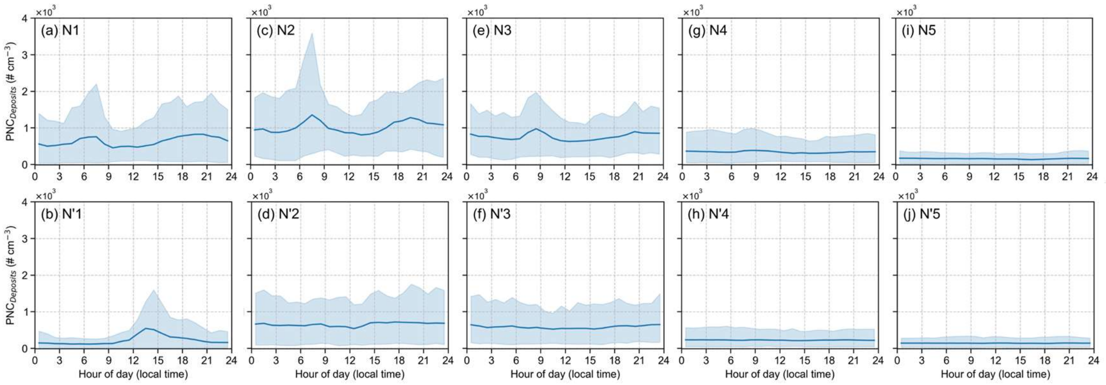

Figure S27. Diurnal variations of the averages PNC$_{\text{Deposits}}$ at the DSL (top panel) and DT (bottom panel) sites, the lower and upper boundaries of the shade represent 10$^{\text{th}}$–90$^{\text{th}}$-percentile of PNC$_{\text{Deposits}}$, respectively.

Table S1. Locations of sites and the number of PNSD during the measurement.

Sites	Location	No. of hourly observations	No. (Proportion) of missing data
Dianshan Lake (DSL)	31.10° N, 120.98° E	237,984	2438 (1.0%)
Dongtan (DT)	31.52° N, 121.97° E	239,095	422 (0.2%)

Table S2. Pearson's R of the hourly PNC and criteria pollutants at DSL and DT sites either in raw or in lognormal form.

Sites	Form	PNC	$PM_{2.5}$	CO	NO_x	O_3	SO_2
DSL	Linear	PNC_{UFPs}	-0.02	0.01	0.67	-0.37	0.47
		PNC_{Total}	0.18	0.21	0.74	-0.42	0.62
	Lognormal	PNC_{UFPs}	-0.01	0.03	0.67	-0.43	0.55
		PNC_{Total}	0.23	0.25	0.75	-0.5	0.67
DT	Linear	PNC_{UFPs}	-0.01	0.01	0.57	-0.36	0.37
		PNC_{Total}	0.18	0.26	0.67	-0.37	0.59
	Lognormal	PNC_{UFPs}	0.07	-0.01	0.61	-0.38	0.49
		PNC_{Total}	0.31	0.27	0.72	-0.39	0.66

Person's R	1	0.5	0	-0.5	-1

Table S3. Nonparametric tests were performed on the concentrations of V corresponding to data above the 90th percentile for the five factors at the DT site.

Sources	N'1	N'2	N'3	N'4	N'5
N'1		<0.001 ***	0 ***	0.847	0.936
N'2	0 ***		<0.001 ***	<0.001 ***	<0.001 ***
N'3	0 ***	<0.001 ***		0 ***	0 ***
N'4	0.847	<0.001 ***	0 ***		0.787
N'5	0.936	<0.001 ***	0 ***	0.787	

Values in the table represent the level of significance and is denoted by P. $P \leq 0.001$ are denoted by ***.

Table S4. Summary of characteristics at DT site related to NPF events. The number of days, $PNC_{N'1}$ and $PNC_{N'2}$, and their corresponding time-weighted contribution to $PNC_{N'2}$ during NPF, undefined NPF and non-NPF days, respectively.

	Number of days	**$PNC_{N'1}$ (# cm^{-3})**	**$PNC_{N'2}$ (# cm^{-3})**	**Time-weighted contribution to $PNC_{N'2}$**
NPF	9	1368	1869	17%
Undefined NPF	28	751	1484	41%
Non-NPF	53	179	794	42%

Table S5. Abbreviations in this study (in the order of initial letter).

Abbreviation	Full name
Alve	Alveolar
CMB	chemical mass balance
CMD	count median diameters
CPF	conditional probability functions
DECA	Domestic Emission Control Area
EURO 7	European emission standard for vehicular exhausts
GMD	geometric mean diameter
HA	Head airway
ICRP	International Commission on Radiological Protection
IMO	International Maritime Organization
$J(NO_2)$	photolysis rate of NO_2
LDSA	lung deposited surface area
Mode	the peak size in the distribution curve that represent the most frequent particle size
MOUDI	multiple-stage inertial impactors
MPPD	Multiple-Path Particle Dosimetry Model
Ni	Nickle
NMF	non-negative matrix factorization
PCA	principal components analysis
$PM_{2.5}$	particulate matter with diameters ≤ 2.5 μm
PNC	particle number concentrations
$PNC_{Deposits}$	PNC deposited in the respiratory system
PNC_{Total}	PNC with size of 13.6–710.5 nm
PNC_{UFPs}	PNC with size of 13.6–100 nm
PNSD	particle number size distributions
PVC	particle volume concentration
SNA	sulphate (SO_4^{2-}), nitrate (NO_3^-) and ammonium (NH_4^+)

SOR	sulfur oxidation ratio
TB	Tracheobronchial
UCC	uncentered correlation coefficient
UFPs	ultrafine particles
V	Vanadium
WHO AQG2021	World Health Organization's air quality guidelines (latest release in 2021)
WD	wind direction

346

References

1. G. Yang, J. Huo, L. Wang, Y. Wang, S. Wu, L. Yao, Q. Fu and L. Wang, Total OH Reactivity Measurements in a Suburban Site of Shanghai, *J. Geophys. Res.: Atmos.*, 2022, **127**, e2021JD035981.
2. A. Wiedensohler, W. Birmili, A. Nowak, A. Sonntag, K. Weinhold, M. Merkel, B. Wehner, T. Tuch, S. Pfeifer, M. Fiebig, A. M. Fjäraa, E. Asmi, K. Sellegri, R. Depuy, H. Venzac, P. Villani, P. Laj, P. Aalto, J. A. Ogren, E. Swietlicki, P. Williams, P. Roldin, P. Quincey, C. Hüglin, R. Fierz-Schmidhauser, M. Gysel, E. Weingartner, F. Riccobono, S. Santos, C. Grüning, K. Faloon, D. Beddows, R. Harrison, C. Monahan, S. G. Jennings, C. D. O'Dowd, A. Marinoni, H. G. Horn, L. Keck, J. Jiang, J. Scheckman, P. H. McMurry, Z. Deng, C. S. Zhao, M. Moerman, B. Henzing, G. de Leeuw, G. Löschau and S. Bastian, Mobility particle size spectrometers: harmonization of technical standards and data structure to facilitate high quality long-term observations of atmospheric particle number size distributions, *Atmos. Meas. Tech.*, 2012, **5**, 657-685.
3. W. Sun, D. Wang, L. Yao, H. Fu, Q. Fu, H. Wang, Q. Li, L. Wang, X. Yang, A. Xian, G. Wang, H. Xiao and J. Chen, Chemistry-triggered events of PM2.5 explosive growth during late autumn and winter in Shanghai, China, *Environmental Pollution*, 2019, **254**, 112864.
4. A. L. de Jesus, M. M. Rahman, M. Mazaheri, H. Thompson, L. D. Knibbs, C. Jeong, G. Evans, W. Nei, A. Ding, L. Qiao, L. Li, H. Portin, J. V. Niemi, H. Timonen, K. Luoma, T. Petäjä, M. Kulmala, M. Kowalski, A. Peters, J. Cyrys, L. Ferrero, M. Manigrasso, P. Avino, G. Buonano, C. Reche, X. Querol, D. Beddows, R. M. Harrison, M. H. Sowlat, C. Sioutas and L. Morawska, Ultrafine particles and PM2.5 in the air of cities around the world: Are they representative of each other?, *Environment International*, 2019, **129**, 118-135.
5. Y. Ge, Q. Fu, M. Yi, Y. Chao, X. Lei, X. Xu, Z. Yang, J. Hu, H. Kan and J. Cai, High spatial resolution land-use regression model for urban ultrafine particle exposure assessment in Shanghai, China, *Science of The Total Environment*, 2022, **816**, 151633.
6. Z. Liu, B. Hu, J. Zhang, J. Xin, F. Wu, W. Gao, M. Wang and Y. Wang, Characterization of fine particles during the 2014 Asia-Pacific economic cooperation summit: Number concentration, size distribution and sources, *Tellus B: Chemical and Physical Meteorology*, 2017, **69**, 1303228.
7. Y. Zhu, I. D. Sulaymon, X. Xie, J. Mao, S. Guo, M. Hu and J. Hu, Airborne particle number concentrations in China: A critical review, *Environmental Pollution*, 2022, **307**, 119470.
8. NMF: Algorithms and Framework for Nonnegative Matrix Factorization (NMF), https://cran.r-project.org/package=NMF, (accessed January 2023).
9. missForest: Nonparametric Missing Value Imputation using Random Forest, https://cran.r-project.org/package=missForest, (accessed January 2023).
10. D. J. Stekhoven and P. Bühlmann, MissForest—non-parametric missing value imputation for mixed-type data, *Bioinformatics*, 2012, **28**, 112-118.
11. R. Gaujoux and C. Seoighe, A flexible R package for nonnegative matrix factorization, *BMC Bioinformatics*, 2010, **11**, 367.
12. J.-P. Brunet, P. Tamayo, T. R. Golub and J. P. Mesirov, Metagenes and molecular pattern discovery using matrix factorization, *Proceedings of the National Academy of Sciences*, 2004, **101**, 4164-4169.

13. L. N. Hutchins, S. M. Murphy, P. Singh and J. H. Graber, Position-dependent motif characterization using non-negative matrix factorization, *Bioinformatics*, 2008, **24**, 2684-2690.
14. I. M. Ulbrich, M. R. Canagaratna, Q. Zhang, D. R. Worsnop and J. L. Jimenez, Interpretation of organic components from Positive Matrix Factorization of aerosol mass spectrometric data, *Atmos. Chem. Phys.*, 2009, **9**, 2891-2918.
15. ICRP, The New ICRP Model for the Respiratory Tract, *Radiation Protection Dosimetry*, 1994, **53**, 107-114.
16. WHO, *WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*, Report 9789240034228, World Health Organization, Geneva, 2021.