Electronic Supplementary Material to: Understanding distribution patterns of secondary inorganic aerosols in two distinct environments through statistical learning

Zehao Zou[1,2], Wei Du[3], Giancarlo Ciarelli[2,4], Yongchun Liu[1], Chao Yan[4], Jing Cai[5], Qiaozhi Zha[4], Kaspar Rudolf Daellenbach[6], Men Xia[1,2,4], Xinran Zhang[2], Qi Yuan[1,2], Tinghan Zhang[2], Feixue Zheng[1], Yuzheng Wang[1], Chenjie Hua[1], Zongcheng Wang[1], Hongyu Pang[1], Boying Song[1], Li Ma[1], Wenshuo Zhou[1], Junlei Zhan[1], Wei Ma[1], Tuukka Petäjä[2], Nina Sarnela[2], Ilona Jaakkola[2], Lauri Ahonen[2], Veli-Matti Kerminen[2], Markku Kulmala[1,2,4], and Tom Kokkonen[2,4]

[1]Aerosol and Haze Laboratory, Beijing Advanced Innovation Center for Soft Matter Science and Engineering, Beijing University of Chemical Technology, Beijing, 100029, China
[2]Institute for Atmospheric and Earth System Research, University of Helsinki, Helsinki, 00014, Finland
[3]State Key Laboratory of Atmospheric Environment and Extreme Meteorology, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, 100029, China
[4]Nanjing-Helsinki Institute in Atmospheric and Earth System Sciences, Nanjing University, Nanjing, 210023, China
[5]School of Atmospheric Physics, Nanjing University of Information Science and Technology, Nanjing, 210044, China
[6]Paul Scherrer Institute PSI, Forschungsstrasse 111, 5232 Villigen PSI, Switzerland

Correspondence: Wei Du (wei.du@helsinki.fi)

Contents of this file

1. Text S1 to S4
2. Figures S1 to S7
3. Tables S1 to S3

Text S1. Measurement sites, data availability, and instrumentation

We compared high-resolution ambient observations from two contrasting environments: the Aerosol and Haze Laboratory at the west campus of Beijing University of Chemical Technology in Beijing, China (AHL/BUCT; $39.942^\circ\,\mathrm{N},\ 116.298^\circ\,\mathrm{E}$; Liu et al., 2020) and the Station for Measuring Ecosystem-Atmosphere Relations (SMEAR) II, located in the Finnish boreal forest in Hyytiälä, Finland ($61.846^\circ\,\mathrm{N},\ 24.295^\circ\,\mathrm{E}$; Hari et al., 2013). The AHL/BUCT station represents a typical anthropogenically influenced urban environment, including residential emissions, traffic, and cooking emissions, whereas SMEAR II represents a boreal forest background site.

The nominal study period covered 1 January 2018 to 31 December 2022, but actual data availability differed between sites and variables. Measurements at AHL/BUCT started on 6 February 2018. In addition, trace-gas measurements at AHL/BUCT

were temporarily interrupted during the warm season of 2021, from April to August, due to instrument maintenance. At SMEAR II, the datasets used in this study covered 2018–2021 only because of instrument availability. At both sites, meteorological parameters, gaseous species, and particulate inorganic species were measured. To ensure data quality, inter-instrument comparisons and/or instrument calibrations were conducted during the measurement periods. Detailed descriptions of instrument operation and calibration procedures have been reported in previous studies (Neefjes et al., 2022; Dada et al., 2020; Guo et al., 2023).

At the AHL/BUCT site, air temperature, pressure, relative humidity (RH), wind speed, wind direction, and UVB radiation were obtained from sensors connected to the Vaisala AWS310 weather-station data acquisition system. At the SMEAR II site, temperature was measured using a Pt100 thermometer housed in a ventilated radiation shield, while RH was measured using a Rotronic MP102H RH sensor. Wind speed and wind direction were obtained using Gill Solent 1012R ultrasonic anemometers, and radiation was measured with a Solar Light SL501A radiometer.

SO_2 was measured using a Thermo Scientific 43i-TLE analyzer at both stations. NO and NO_x were measured at AHL/BUCT using a Thermo Scientific 42i analyzer and at SMEAR II using a Thermo Scientific TEI 42iTL analyzer. After unit harmonization and quality control, NO_2 concentrations were calculated as the difference between the measured NO_x and NO concentrations. Sulfate (SO_4) and nitrate (NO_3) in non-refractory submicron particulate matter (NR-PM_1) were measured using an Aerosol Chemical Speciation Monitor (ACSM, Aerodyne Research Inc.) at both Beijing and Hyytiälä (Heikkinen et al., 2020).

Text S2. Derived variables

The sulfur oxidation ratio (SOR) and nitrogen oxidation ratio (NOR) were used to characterize the relative abundance of particulate sulfate and nitrate compared with their respective gaseous precursors. Both ratios were calculated using molar concentrations as follows:

$$\mathrm{SOR} = \frac{[\mathrm{SO_4}]}{[\mathrm{SO_2}] + [\mathrm{SO_4}]} \tag{S1}$$

$$\mathrm{NOR} = \frac{[\mathrm{NO_3}]}{[\mathrm{NO_2}] + [\mathrm{NO_3}]} \tag{S2}$$

The combined product and precursor concentrations, S_{tot} and N_{tot}, were defined as follows:

$$S_{\mathrm{tot}} = [\mathrm{SO_2}] + [\mathrm{SO_4}] \tag{S3}$$

$$N_{\mathrm{tot}} = [\mathrm{NO_3}] + [\mathrm{NO_2}] \tag{S4}$$

where the brackets denote molar concentrations, with all species converted to the same units before calculation. For NOR, NO_2 refers to the derived NO_2 concentration calculated from measured NO_x and NO, as described in Text S1.

Text S3. Air mass analysis and trajectory-based exposure indicators

Air-mass transport histories were analyzed using the HYSPLIT model (Stein et al., 2015). Hourly backward trajectories with a duration of 96 h and an arrival height of 100 m above ground level were calculated for both the Beijing (AHL/BUCT) and Hyytiälä (SMEAR II) sites using meteorological fields from the Global Data Assimilation System (GDAS; $1^\circ \times 1^\circ$ resolution), archived by the NOAA Air Resources Laboratory (ARL) (National Centers for Environmental Prediction, n.d.). To quantify air-mass exposure to contrasting emission environments, two trajectory-based indicators were used: Time over Populated Land (ToPL) in Beijing and Time over Land (ToL) in Hyytiälä. For Beijing, ToPL was defined as the cumulative trajectory residence time within the densely populated East China gigacity region shown in Figure S1a. This indicator was used to represent the duration of air-mass exposure to strongly anthropogenically influenced source areas.

For Hyytiälä, the trajectory classification followed the sector definition shown in Figure S1b and described previously (Petäjä et al., 2022; Tunved et al., 2006). After ToL was calculated for all trajectories, each trajectory was evaluated according to the fraction of its total residence time spent within the predefined sectors shown in Figure S1b. A trajectory was assigned to a given sector only if more than 95% of its total trajectory time fell within that sector; trajectories not meeting this criterion for any sector were excluded. To represent Hyytiälä as a boreal background reference site, only observations assigned to the clean sector were retained for the subsequent Hyytiälä-related analyses.

Together, ToPL and ToL were used to characterize air-mass exposure to populated anthropogenic source regions in Beijing and land-influenced boreal background air masses in Hyytiälä, respectively.

Text S4. Statistical analysis

Long-term concentration trends were assessed by ordinary least-squares regression of monthly median concentrations against elapsed time in years; the reported p values test whether the fitted slopes differ from zero.

To identify concentration regimes of sulfate (SO_4) and nitrate (NO_3) in Beijing, Gaussian mixture models (GMMs) were fitted to $\log_{10}$-transformed concentrations following previous applications in air quality studies (Gómez-Losada et al., 2014; Wang et al., 2024). For each compound and year, the concentration was first log-transformed as:

$$y_t = \log_{10} C_t \tag{S5}$$

where C_t is the SO_4 or NO_3 concentration, expressed in $\mu\mathrm{g\,m^{-3}}$, at the tth observation time. Only positive concentrations were log-transformed. The probability density function of observation y_t was represented by a Gaussian mixture model:

$$p(y_t \mid \lambda) = \sum_{k=1}^{G} \pi_k \mathcal{N}\left(y_t \mid \mu_k, \sigma_k^2\right) \tag{S6}$$

where G is the number of mixture components, π_k is the mixing proportion, and $\mathcal{N}\left(y_t|\mu_k, \sigma_k^2\right)$ is the Gaussian density with mean μ_k and variance σ_k^2. The parameter set is:

$$\lambda = \left\{\pi_k, \mu_k, \sigma_k^2\right\}_{k=1}^{G}, \quad \sum_{k=1}^{G} \pi_k = 1 \tag{S7}$$

Model parameters were estimated using the expectation-maximization (EM) algorithm with the `Mclust()` function in the `mclust` package in R. Models with 1–10 components were evaluated using the Bayesian information criterion (BIC; Figure S2). The posterior probability τ_{tk} that observation t belongs to component k was calculated as follows:

$$\tau_{tk} = \frac{\pi_k \mathcal{N}\left(y_t|\mu_k, \sigma_k^2\right)}{\sum_{j=1}^{G} \pi_j \mathcal{N}\left(y_t|\mu_j, \sigma_j^2\right)} \tag{S8}$$

These GMM-derived regimes differ from predefined pollution level classes. Each observation was assigned to the component with the highest posterior probability. The fitted mixture components define concentration regimes ordered from lower to higher concentration within each annual distribution. The covariance structure allowing variable variance among components (model "V") was adopted. We considered BIC values, how well the fitted curves described the observed distributions, and whether additional components were needed to describe their main features. Four components were used in all years to allow comparisons between annual distributions, although the number of components giving the highest BIC varied among years. This strategy also helped preserve structure at the high concentration end of the distribution. For example, for $[SO_4]$ in 2022 (Figure S3), a solution with fewer regimes collapsed this upper-end structure into a single regime. Based on these criteria, four regimes were selected for Beijing.

To reduce uncertainty due to overlap between classes in the subsequent PLS-DA, the 40% of observations with the highest posterior assignment probabilities within each component were retained for subsequent analysis. To identify environmental variables associated with differences between adjacent GMM-derived concentration regimes, PLS-DA was performed using the `plsda()` function in the `mixOmics` package in R (Haque et al., 2025). The predictor matrix $\boldsymbol{X}$ included the relevant precursor concentration ($[SO_2]$ for sulfate and $[NO_x]$ for nitrate), relative humidity, specific humidity, temperature, pressure, UVB radiation, wind speed, ozone concentration, and ToPL. Predictors were centered and scaled to unit variance, and near-zero-variance predictors were excluded before model fitting. $\boldsymbol{Y}$ denotes the dummy-coded class-response matrix of concentration regimes. In PLS-DA, $\boldsymbol{t}_h$ and $\boldsymbol{u}_h$ denote the latent score vectors for the predictor and class-response matrices on component h, respectively:

$$\begin{aligned} \boldsymbol{t}_h &= \boldsymbol{X}\boldsymbol{w}_h, \\ \boldsymbol{u}_h &= \boldsymbol{Y}\boldsymbol{c}_h. \end{aligned} \tag{S9}$$

where $\boldsymbol{w}_h$ and $\boldsymbol{c}_h$ are the loading weight vectors for $\boldsymbol{X}$ and $\boldsymbol{Y}$ on component h. Here $h = 1$, since only one component was retained. The PLS-DA model seeks latent components that summarize the covariance structure between the predictor variables and the regime classes. One latent component ($H = 1$) was used in each PLS-DA model to summarize the differences between two adjacent regimes. This helps capture the main difference between two neighboring regimes and reduce the risk of overfitting in year-specific and regime-specific subsets.

Model performance was evaluated by five-fold cross-validation and quantified using the area under the receiver operating characteristic curve (ROC AUC). An AUC value close to 1 indicates strong discrimination between two classes, whereas a value near 0.5 indicates weak discrimination. Variable importance in projection (VIP) scores were used to assess predictor

influence on the PLS-DA model. For predictor j, the VIP score was calculated as follows:

$$\mathrm{VIP}_{j,1} = \sqrt{p\, w_{j,1}^2} = \sqrt{p}\, |w_{j,1}| \tag{S10}$$

where p is the number of predictors included in the model, and $w_{j,1}$ is the weight of predictor j on the first and only retained PLS-DA component. The weight vector $\boldsymbol{w}_1$ is normalized to unit Euclidean norm, such that $\sum_{j=1}^{p} w_{j,1}^2 = 1$. Consequently, the mean squared VIP score across predictors equals 1, and VIP scores greater than 1 indicate above-average relative importance in the pairwise classification. Figure 2 presents VIP scores for selected predictors.

The GMM and PLS-DA workflow was not applied to the Hyytiälä clean-sector data. Instead, the Hyytiälä clean-sector subset was used as a single boreal background reference dataset for comparison with the Beijing regime-based results.

Low- and high-RH conditions were defined as $\mathrm{RH} < 40\%$ and $\mathrm{RH} > 60\%$, respectively. For each RH class, ordinary least-squares regression with an intercept was applied to median concentrations in 10 h ToPL bins, using the bin upper bounds (10, 20, ..., 90 h) as the predictor. Each included bin received equal weight.

References

Carslaw, D. C. and Ropkins, K.: *openair* — an R package for air quality data analysis, Environ. Modell. Softw., 27–28, 52–61, https://doi.org/10.1016/j.envsoft.2011.09.008, 2012.

Dada, L., Ylivinkka, I., Baalbaki, R., Li, C., Guo, Y., Yan, C., Yao, L., Sarnela, N., Jokinen, T., Daellenbach, K. R., Yin, R., Deng, C., Chu, B., Nieminen, T., Wang, Y., Lin, Z., Thakur, R. C., Kontkanen, J., Stolzenburg, D., Sipilä, M., Hussein, T., Paasonen, P., Bianchi, F., Salma, I., Weidinger, T., Pikridas, M., Sciare, J., Jiang, J., Liu, Y., Petäjä, T., Kerminen, V.-M., and Kulmala, M.: Sources and sinks driving sulfuric acid concentrations in contrasting environments: implications on proxy calculations, Atmos. Chem. Phys., 20, 11 747–11 766, https://doi.org/10.5194/acp-20-11747-2020, 2020.

Gómez-Losada, Á., Lozano-García, A., Pino-Mejías, R., and Contreras-González, J.: Finite mixture models to characterize and refine air quality monitoring networks, Sci. Total Environ., 485, 292–299, https://doi.org/10.1016/j.scitotenv.2014.03.091, 2014.

Guo, Y., Deng, C., Ovaska, A., Zheng, F., Hua, C., Zhan, J., Li, Y., Wu, J., Wang, Z., Xie, J., Zhang, Y., Liu, T., Zhang, Y., Song, B., Ma, W., Liu, Y., Yan, C., Jiang, J., Kerminen, V.-M., Xia, M., Nieminen, T., Du, W., Kokkonen, T., and Kulmala, M.: Measurement report: the 4-year variability and influence of the Winter Olympics and other special events on air quality in urban Beijing during wintertime, Atmos. Chem. Phys., 23, 6663–6690, https://doi.org/10.5194/acp-23-6663-2023, 2023.

Haque, S. M., Umar, Y., and Kabir, A.: Advanced chemometric techniques for environmental pollution monitoring and assessment: a review, Chemosensors, 13, 268, https://doi.org/10.3390/chemosensors13070268, 2025.

Hari, P., Nikinmaa, E., Pohja, T., Siivola, E., Bäck, J., Vesala, T., and Kulmala, M.: Station for measuring ecosystem-atmosphere relations: SMEAR, in: Physical and Physiological Forest Ecology, edited by Hari, P., Heliövaara, K., and Kulmala, L., pp. 471–487, Springer Netherlands, Dordrecht, ISBN 978-94-007-5602-1 978-94-007-5603-8, https://doi.org/10.1007/978-94-007-5603-8_9, 2013.

Heikkinen, L., Äijälä, M., Riva, M., Luoma, K., Dällenbach, K., Aalto, J., Aalto, P., Aliaga, D., Aurela, M., Keskinen, H., Makkonen, U., Rantala, P., Kulmala, M., Petäjä, T., Worsnop, D., and Ehn, M.: Long-term sub-micrometer aerosol chemical composition in the boreal forest: inter- and intra-annual variability, Atmos. Chem. Phys., 20, 3151–3180, https://doi.org/10.5194/acp-20-3151-2020, 2020.

Liu, Y., Yan, C., Feng, Z., Zheng, F., Fan, X., Zhang, Y., Li, C., Zhou, Y., Lin, Z., Guo, Y., Zhang, Y., Ma, L., Zhou, W., Liu, Z., Dada, L., Dällenbach, K., Kontkanen, J., Cai, R., Chan, T., Chu, B., Du, W., Yao, L., Wang, Y., Cai, J., Kangasluoma, J., Kokkonen, T., Kujansuu, J., Rusanen, A., Deng, C., Fu, Y., Yin, R., Li, X., Lu, Y., Liu, Y., Lian, C., Yang, D., Wang, W., Ge, M., Wang, Y., Worsnop, D. R., Junninen, H., He, H., Kerminen, V.-M., Zheng, J., Wang, L., Jiang, J., Petäjä, T., Bianchi, F., and Kulmala, M.: Continuous and comprehensive atmospheric observations in Beijing: a station to understand the complex urban atmospheric environment, Big Earth Data, 4, 295–321, https://doi.org/10.1080/20964471.2020.1798707, 2020.

National Centers for Environmental Prediction: Global Data Assimilation System (GDAS1) one-degree meteorological archive, https://www.ready.noaa.gov/gdas1.php, n.d.

Neefjes, I., Laapas, M., Liu, Y., Médus, E., Miettunen, E., Ahonen, L., Quéléver, L., Aalto, J., Bäck, J., Kerminen, V.-M., Lampilahti, J., Luoma, K., Mäki, M., Mammarella, I., Petäjä, T., Räty, M., Sarnela, N., Ylivinkka, I., Hakala, S., Kulmala, M., Nieminen, T., and Lintunen, A.: 25 years of atmospheric and ecosystem measurements in a boreal forest - seasonal variation and responses to warm and dry years, Boreal Environ. Res., 27, 1–31, 2022.

Petäjä, T., Tabakova, K., Manninen, A., Ezhova, E., O'Connor, E., Moisseev, D., Sinclair, V. A., Backman, J., Levula, J., and Luoma, K.: Influence of biogenic emissions from boreal forests on aerosol–cloud interactions, Nat. Geosci., 15, 42–47, https://doi.org/10.1038/s41561-021-00876-0, 2022.

Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J. B., Cohen, M. D., and Ngan, F.: NOAA's HYSPLIT atmospheric transport and dispersion modeling system, Bull. Am. Meteorol. Soc., 96, 2059–2077, https://doi.org/10.1175/BAMS-D-14-00110.1, 2015.

Stohl, A.: Trajectory statistics—a new method to establish source-receptor relationships of air pollutants and its application to the transport of particulate sulfate in Europe, Atmos. Environ., 30, 579–587, https://doi.org/10.1016/1352-2310(95)00314-2, 1996.

Tunved, P., Hansson, H.-C., Kerminen, V.-M., Ström, J., Maso, M. D., Lihavainen, H., Viisanen, Y., Aalto, P. P., Komppula, M., and Kulmala, M.: High natural aerosol loading over boreal forests, Science, 312, 261–263, https://doi.org/10.1126/science.1123052, 2006.

Wang, J., Han, Y., Yu, X., Zhang, Z., and Song, T.: Improvements on Gaussian mixture model and its application in identifying aerosol types in two major cities in the Yangtze River Delta, China, Sci. Total Environ., 935, 172 743, https://doi.org/10.1016/j.scitotenv.2024.172743, 2024.

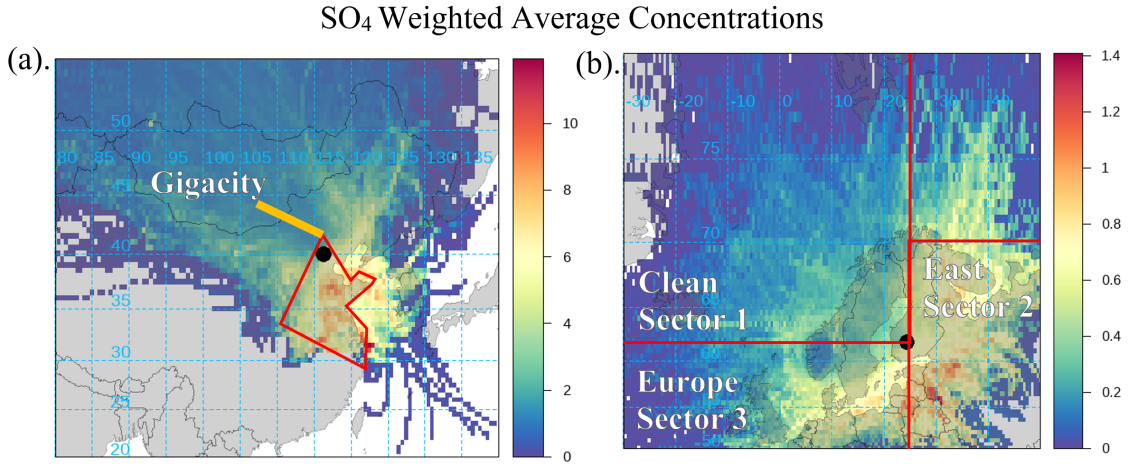


Figure S1. Weighted average sulfate concentrations ($\mu g\,m^{-3}$) of (a) Beijing and (b) Hyytiälä. The maps were produced using concentration-weighted trajectory (CWT) analysis from the `openair` package in R (Carslaw and Ropkins, 2012), based on HYSPLIT backward trajectory output (Stein et al., 2015) for 2019–2021. Black dots indicate the study sites. Red outlines indicate (a) the densely populated gigacity region and (b) the source-sector boundaries (Clean, Europe, and East; see Petäjä et al. (2022)). The CWT method is described by Stohl (1996).

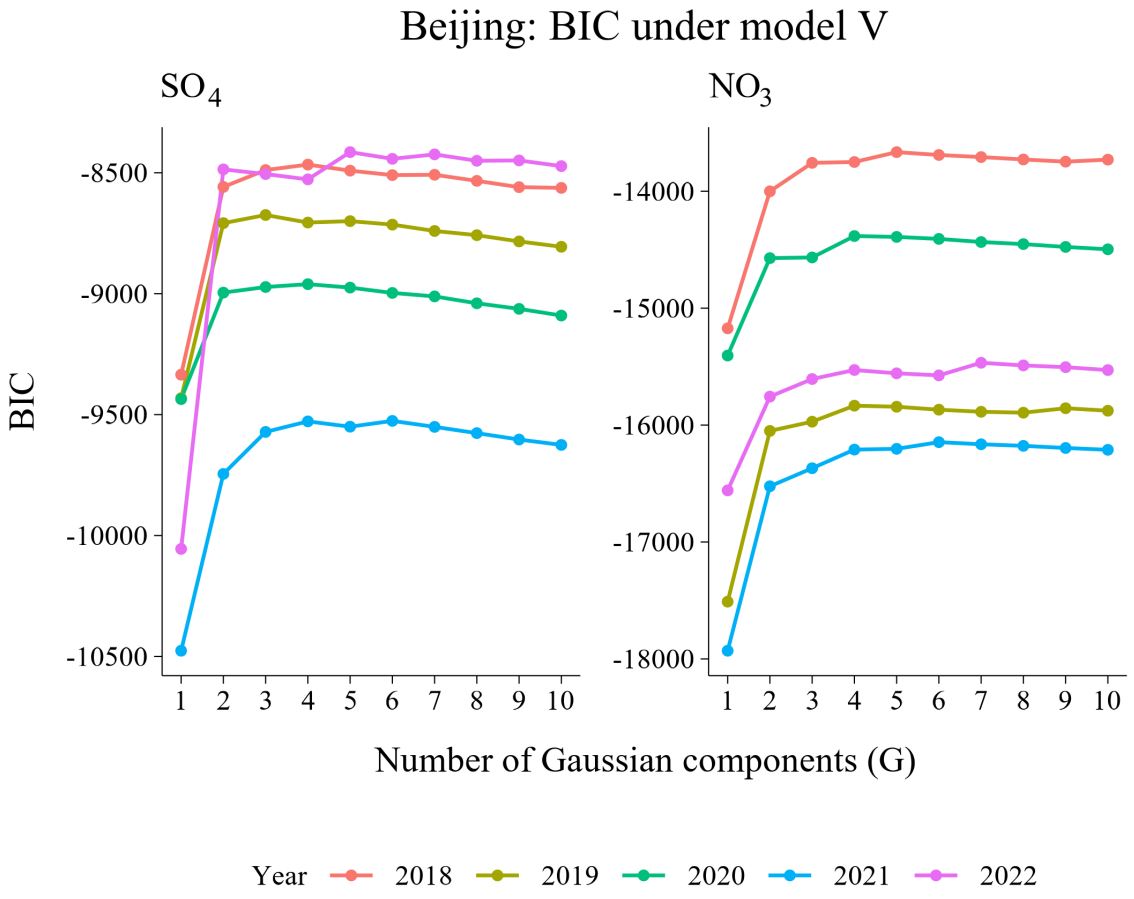


Figure S2. Bayesian information criterion (BIC) values for candidate Gaussian mixture models (model "V") fitted to the annual Beijing sulfate and nitrate distributions. Results are shown for each year and for different numbers of Gaussian components (*G*). Higher BIC values indicate better model support after accounting for model complexity.

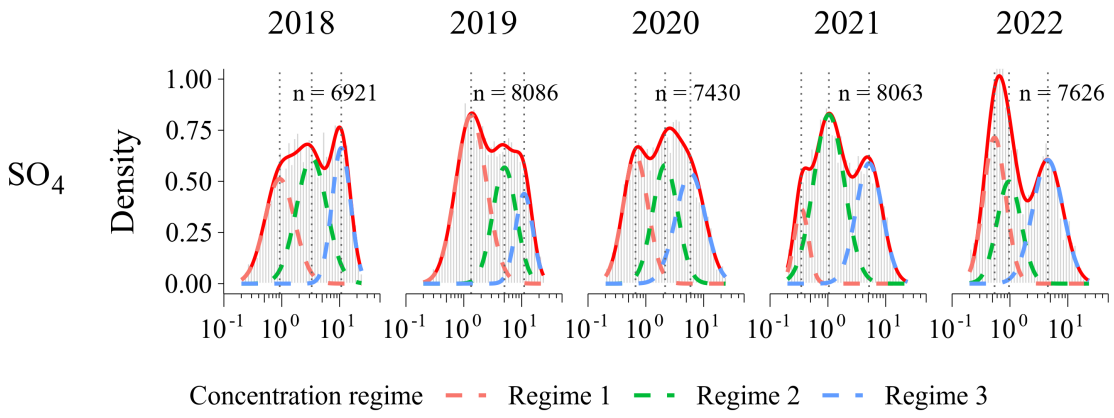


Figure S3. Gaussian mixture model decomposition of Beijing sulfate concentrations using $G = 3$.

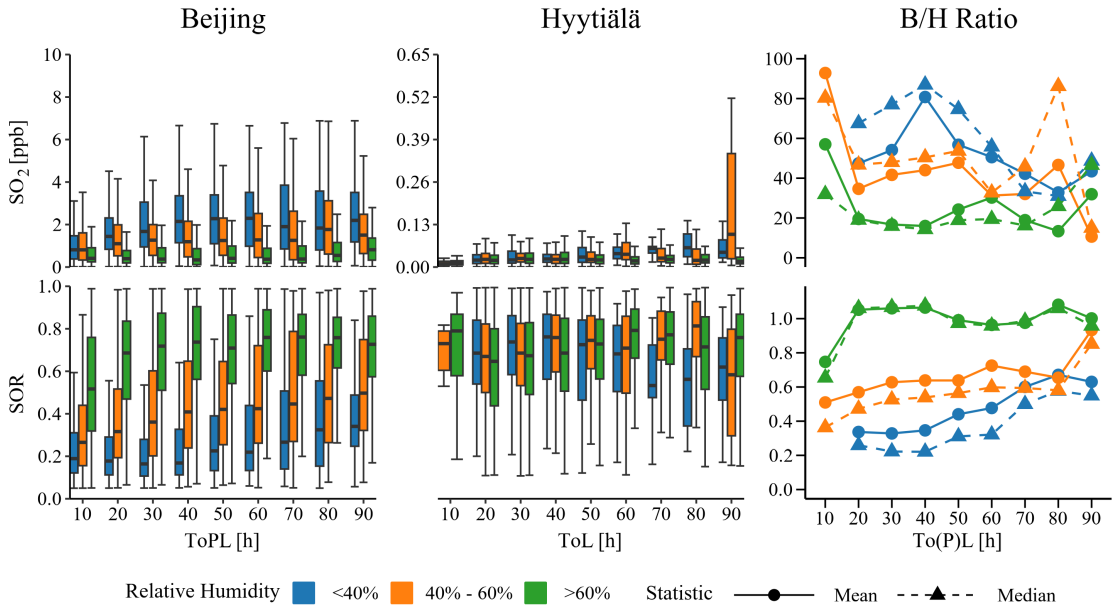


Figure S4. Distributions of SO_2 concentrations and SOR across To(P)L and RH classes in Beijing and Hyytiälä, with corresponding Beijing-to-Hyytiälä ratios. Hyytiälä results shown here are based on the clean-sector subset only. The elevated SO_2 distribution at RH of 40–60% occurs in the 90 h ToL bin, corresponding to $80 < \mathrm{ToL} \leq 90\,\mathrm{h}$.

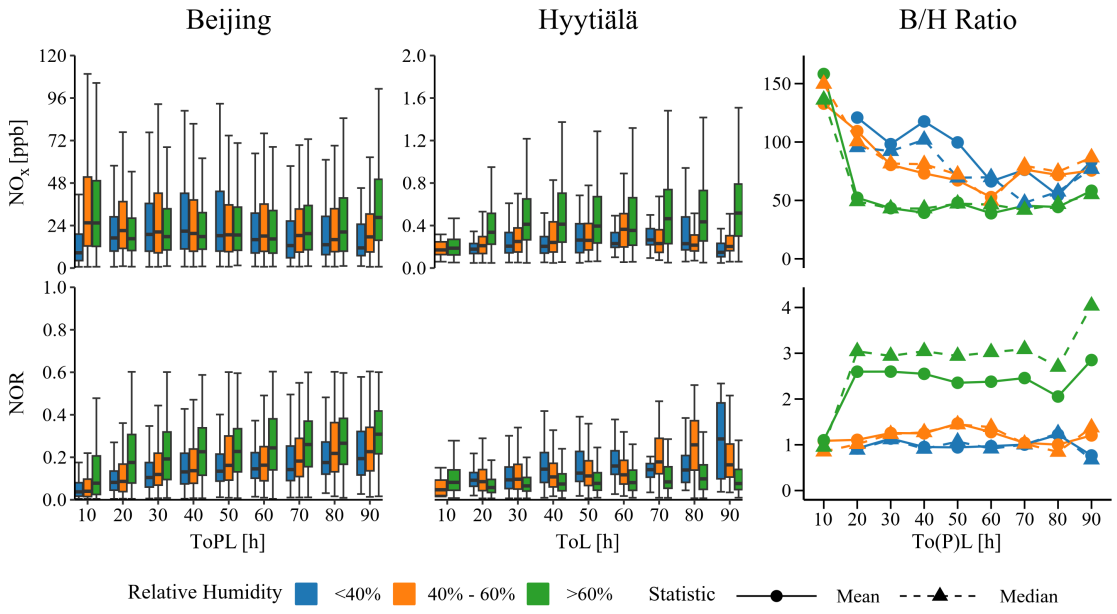


Figure S5. Distributions of NO_x concentrations and NOR across To(P)L and RH classes in Beijing and Hyytiälä, with corresponding Beijing-to-Hyytiälä ratios. Hyytiälä results shown here are based on the clean-sector subset only.

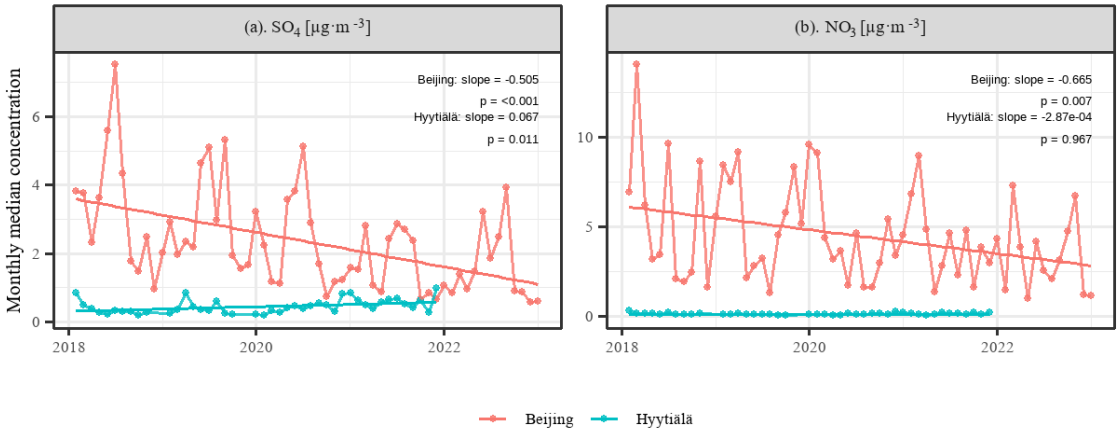


Figure S6. Monthly median concentrations of sulfate (SO_4) and nitrate (NO_3) in Beijing and Hyytiälä during the study period. Points and connecting lines show the monthly median values, and solid lines indicate linear fits to the monthly time series. The slope and corresponding p value shown in each panel are derived from the linear regression of the monthly median concentrations against time. Time was expressed in years, and slopes are reported in $\mu g\,m^{-3}\,yr^{-1}$.

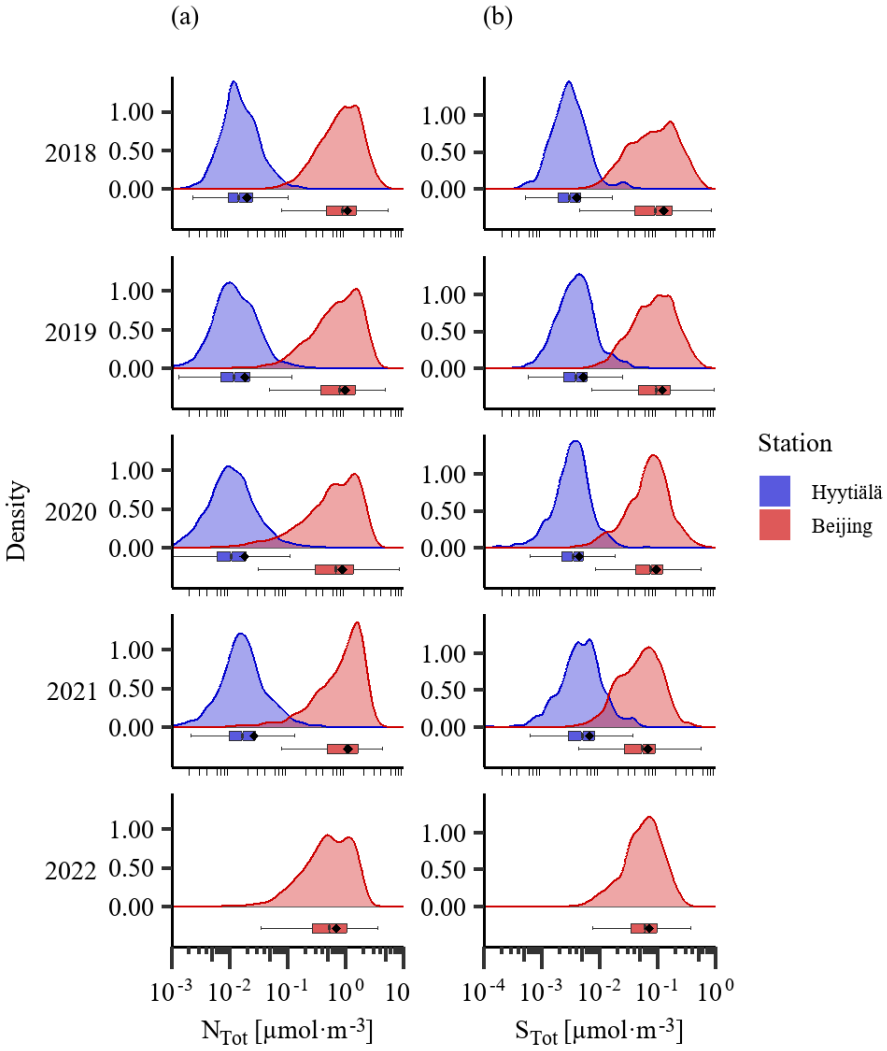


Figure S7. Annual distributions of (a) $N_{\mathrm{tot}} = [\mathrm{NO_2}] + [\mathrm{NO_3}]$ and (b) $S_{\mathrm{tot}} = [\mathrm{SO_2}] + [\mathrm{SO_4}]$, expressed in $\mu\mathrm{mol\,m^{-3}}$, in Beijing (red, 2018–2022) and the Hyytiälä clean-sector subset (blue, 2018–2021). Curves show kernel density estimates of $\log_{10}$-transformed concentrations, using a common bandwidth for each quantity. Each curve integrates to 1 with respect to $\log_{10}$ concentration; the horizontal axis shows the original concentration scale with logarithmic ticks. Box plots below the zero-density baseline show the median and interquartile range, with whiskers extending to the most extreme observations within 1.5 times the interquartile range of the corresponding quartile. Black diamonds indicate means.

Table S1. Annual summary statistics for pollutant concentrations and oxidation ratios in Beijing and Hyytiälä. Q_1 and Q_3 denote the 25th and 75th percentiles, respectively. NA denotes unavailable data.

		Beijing				Hyytiälä			
Variable	Year	Q_1	Median	Q_3	Mean	Q_1	Median	Q_3	Mean
SO_4	2018	1.3	3.0	7.6	5.0	0.2	0.3	0.6	0.5
($\mu g\,m^{-3}$)	2019	1.3	2.7	6.4	4.5	0.2	0.4	0.7	0.6
	2020	0.9	2.1	4.5	3.5	0.2	0.4	0.8	0.6
	2021	0.7	1.4	3.7	2.7	0.3	0.6	1.1	0.8
	2022	0.6	1.2	3.7	2.6	NA	NA	NA	NA
SO_2	2018	0.3	0.8	2.2	1.5	0.0	0.0	0.1	0.1
(ppb)	2019	0.3	0.9	1.9	1.4	0.0	0.0	0.1	0.1
	2020	0.3	0.9	1.8	1.2	0.0	0.0	0.1	0.1
	2021[a]	0.3	0.5	1.3	0.9	0.0	0.0	0.1	0.1
	2022	0.5	0.9	1.4	1.0	NA	NA	NA	NA
SOR	2018	0.43	0.65	0.87	0.64	0.69	0.83	0.91	0.78
	2019	0.42	0.65	0.88	0.64	0.79	0.88	0.94	0.84
	2020	0.38	0.64	0.90	0.62	0.84	0.92	0.97	0.88
	2021[a]	0.36	0.60	0.83	0.59	0.85	0.93	0.97	0.89
	2022	0.30	0.51	0.75	0.53	NA	NA	NA	NA
NO_3	2018	1.0	4.8	15.0	10	0.1	0.2	0.2	0.2
($\mu g\,m^{-3}$)	2019	1.0	5.0	13.8	9.4	0.1	0.1	0.2	0.2
	2020	0.8	3.6	10.5	7.8	0.1	0.1	0.2	0.2
	2021	0.7	3.8	10.3	7.8	0.1	0.1	0.3	0.2
	2022	0.7	2.9	9.4	7.4	NA	NA	NA	NA
NO_x	2018	10.9	19.2	36.4	28.3	0.3	0.6	1.1	0.9
(ppb)	2019	8.0	16.6	35.0	26.5	0.3	0.6	1.1	0.8
	2020	6.4	14.4	30.3	22.4	0.2	0.5	1.2	0.8
	2021[a]	10.2	21.3	41.1	28.8	0.4	0.7	1.3	1.0
	2022	5.1	10.0	21.6	16.0	NA	NA	NA	NA
NOR	2018	0.05	0.17	0.37	0.22	0.11	0.18	0.28	0.21
	2019	0.08	0.19	0.36	0.23	0.09	0.15	0.26	0.19
	2020	0.09	0.19	0.36	0.24	0.08	0.15	0.26	0.19
	2021[a]	0.07	0.17	0.35	0.22	0.09	0.14	0.23	0.17
	2022	0.07	0.17	0.39	0.24	NA	NA	NA	NA

[a] Trace-gas data availability for 2021 in Beijing was limited (missing April–August); trace-gas statistics and the derived oxidation ratios may therefore be biased toward the cold season.

Table S2. Annual GMM-derived median concentrations of sulfate and nitrate regimes in Beijing. Medians were calculated as 10^{μ_k}, where μ_k is the mean of component k on the $\log_{10}$ concentration scale. Regimes are ordered by increasing median concentration, and values are reported in $\mu\mathrm{g\,m^{-3}}$.

	SO_4 Regime				NO_3 Regime			
Year	1	2	3	4	1	2	3	4
2018	0.8	2.3	6.8	13.4	0.2	1.7	9.6	24.1
2019	1.0	1.5	4.1	9.7	0.2	1.4	10.4	28.5
2020	0.6	1.4	3.5	8.7	0.2	1.1	9.4	36.8
2021	0.3	0.8	2.3	6.6	0.2	1.1	7.6	27.8
2022	0.5	0.8	1.9	5.4	0.2	1.0	6.4	23.4

Table S3. Annual median values of selected environmental variables for GMM-derived sulfate and nitrate regimes in Beijing. Within each regime and year, the 40% of observations with the highest posterior assignment probabilities were selected. Medians were calculated after applying the PLS-DA data-screening procedure.

		SO_4 Regime				NO_3 Regime			
Variable	Year	1	2	3	4	1	2	3	4
Precursors	2018	0.4	1.2	1.2	1.9	7.5	19.9	23.3	36.7
(SO_2 or NO_x, ppb)	2019	0.4	1.3	1	0.7	5.3	11.2	27.1	42.4
	2020	0.8	1.0	0.8	0.7	2.8	8.5	23.0	25.6
	2021	0.5	0.9	1.0	0.8	3.6	13.6	25.9	40.9
	2022	0.6	1.1	1.2	1.1	3.4	8.8	14.5	29.2
Oxidation Ratios	2018	0.2	0.4	0.6	0.7	0.0	0.0	0.1	0.3
(SOR or NOR)	2019	0.2	0.2	0.5	0.8	0.0	0.1	0.2	0.3
	2020	0.1	0.2	0.5	0.8	0.0	0.0	0.2	0.4
	2021	0.1	0.2	0.4	0.8	0.0	0.0	0.1	0.3
	2022	0.1	0.2	0.3	0.7	0.0	0.0	0.2	0.4
Air-mass exposure time	2018	1	17	35	50	1	13	27	53
(ToPL, h)	2019	2	5	28	58	1	10	24	50
	2020	2	12	32	48	1	8	37	53
	2021	2	2	28	62	1	3	20	57
	2022	2	3	22	51	1	4	22	61
Relative Humidity	2018	25	48	73	70	24	49	63	74
(RH, %)	2019	25	30	46	64	22	33	42	67
	2020	26	37	59	76	22	47	55	69
	2021	30	31	59	70	24	34	56	72
	2022	25	31	49	67	21	35	45	73
Temperature	2018	14.8	16.4	22.9	25.5	19.0	22.3	20.3	17.3
(T, $^\circ$C)	2019	15.1	8.0	22.0	21.9	11.8	23.9	11.9	10.4
	2020	13.8	8.4	20.1	3.8	12.1	22.4	11.8	6.2
	2021	9.0	7.2	9.6	8.1	7.7	6.7	5.5	9.5
	2022	11.5	3.7	13.7	17.0	14.0	12.8	12.1	9.5
Wind Speed	2018	1.3	0.8	0.7	0.7	1.9	0.8	0.7	0.7
(ws, $\mathrm{m\,s^{-1}}$)	2019	0.9	0.9	0.8	0.8	1.9	1.0	0.8	0.7
	2020	1.0	1.0	0.8	0.7	1.4	1.0	0.8	0.7
	2021	0.8	1.1	0.7	0.8	2.0	0.8	0.8	0.7
	2022	1.2	1.1	0.8	0.7	2.0	1.0	0.9	0.6
Ozone Concentration	2018	24.9	14.7	30.1	53.0	28.9	26.8	21.5	17.0
(O_3, ppb)	2019	25.9	18.0	29.6	28.2	27.7	32.4	17.2	8.7
	2020	22.8	24.8	26.2	26.9	29.1	31.4	20.2	17.3
	2021	22.0	17.6	15.3	9.8	27.5	15.6	11.2	8.5
	2022	26.1	22.2	23.9	27.3	30.4	25.7	24.3	8.5