Supplement of

Long–term Trends in $PM_{2.5}$ Chemical Composition and Its Impact on Aerosol Properties: Field Observations from 2007 to 2020 in Pearl River Delta, South China

Yunfeng He et al.

Correspondence to: Xiang Ding (xiangd@gig.ac.cn), Michael Boy (michael.boy@helsinki.fi)

Text S1. Eliminating anthropogenic impact on Cl^- by XGBoost (eXtreme Gradient Boosting)

XGBoost, a promising tool in machine learning (ML), has recently been used in atmospheric research. The performance of XGBoost surpasses that of traditional analysis methods for both nonlinear and linear questions. (Hou et al., 2022). The introduction and Python package is available online (https://github.com/dmlc/xgboost). In this study, we used this method to decouple the impacts of anthropogenic sources on Cl^- in $PM_{2.5}$ as much as possible. A total of 6 independent variables, including levoglucosan (biomass burning marker), picene (coal combustion marker), meteorological parameters (wind speed, temperature, RH, and solar radiation, representing marine activity) and one dependent variable, Cl^-, were fed to the model. The annual data are listed in Table S1. All these samples were randomly divided into two groups: a training set accounting for 70% and a testing set accounting for 30% ($R^2 = 0.41$). Then, we replaced daily concentrations of levoglucosan and picene with their average concentrations in 2007-2020 to eliminate changes in anthropogenic sources on Cl^-. As shown in Fig S6, it decreased slightly at a rate of –2% yr^{-1}, indicating that the influence of marine emissions on $PM_{2.5}$ had been almost unchanged.

Text S2. Prediction of aerosol acidity and liquid water content (ALWC)

The thermodynamics model ISORROPIA-II has been widely used to calculate aerosol pH (Nenes et al., 1998; Fountoukis and Nenes, 2007; Zhou et al., 2022; Wen et al., 2018). Meteorological parameter (temperature and relative humidity), aerosol phase water-soluble ions (SO_4^{2-}, NO_3^-, Cl^-, NH_4^+, Na^+, K^+, Mg^{2+}, Ca^{2+}) and gaseous precursors (HNO_3, HCl, NH_3) are needed for the model execution. Given high relative humidity (RH) in the PRD region, the model was run by assuming aerosol solutions were metastable (only a liquid phase) in forward mode; previous studies suggested this would produce better performance than the stable state solution (solid + liquid) (Guo et al., 2015; Bougiatioti et al., 2016). A recent study suggested that the model was run in the forward mode but did not include gas-phase data, which may capture the general trend of aerosol acidity but underestimate pH (Fang et al., 2025).

We used the data collected in Guangzhou Institute of Geochemistry (GIG) to run the model, and the result (G1) showed there was a great agreement in gaseous precursors between simulation and observation (Fig. S13). Due to the lack of gas-phase concentrations during our campaign period in WQS, we ran the model by performing iterations to avoid the underestimation of pH: only aerosol phase data were used as input for the first run. Then we added the output gas-phase concentrations from the first run to the initial aerosol-phase chemical concentrations to derive total (gas and aerosol) concentrations, which serve as input for the second run and so on. To determine an optimal number of iterations, we ran GIG data (without gas-phase data) to compare with G1. Our result indicated four iterations generated the closest outcome to G1 when gas-phase data was unavailable (Fig. S14). Thus, only water-soluble ions were fed to the model and performing four iterations to calculate pH and ALWC for WQS samples.

Table S1. Annual data of Cl-, levoglucosan, picene, and meteorological parameters.

40

Year	Cl- (μg m^{-3})	Levoglucosan (ng m^{-3})	Picene (ng m^{-3})	Wind speed (m s^{-1})	Temperature (°C)	RH (%)	SSR (W m^{-2})
2007	1.01 ± 0.54	203.84 ± 76.82	0.13 ± 0.07	1.3 ± 0.2	22.2 ± 2.1	57 ± 11	161.3 ± 41.3
2008	1.61 ± 1.27	347.41 ± 143.95	0.29 ± 0.25	1.3 ± 0.4	17.2 ± 2.9	47 ± 12	156.5± 28.3
2009	1.80 ± 1.03	93.23 ± 58.37	0.19 ± 0.15	1.7 ± 0.5	17.0 ± 3.1	67 ± 13	135.2 ± 36.3
2010	1.49 ± 1.16	253.92 ± 148.86	0.21 ± 0.17	1.7 ± 0.7	19.7 ± 3.2	64 ± 11	141.8 ± 51.4
2011	1.46 ± 0.93	208.33 ± 152.46	0.29 ± 0.20	1.8 ± 0.9	19.9 ± 3.8	57 ± 11	134.7 ± 36.9
2012	1.21 ± 0.74	48.36 ± 42.80	0.19 ± 0.10	1.2 ± 0.6	22.1 ± 1.4	61 ± 7	101.2 ± 40.4
2013	1.44 ± 1.12	224.83 ± 109.90	0.22 ± 0.11	1.3 ± 0.3	20.9 ± 1.1	50 ± 13	125.2 ± 46.2
2014	0.38 ± 0.33	204.60 ± 130.77	0.18 ± 0.11	1.9 ± 0.6	20.2 ± 4.4	57 ± 14	95.3 ± 49.1
2015	0.55 ± 0.37	109.92 ± 65.76	0.17 ± 0.08	1.5 ± 0.3	25.1 ± 2.4	63 ± 8	149.0 ± 36.2
2016	0.52 ± 0.34	91.74 ± 43.83	0.24 ± 0.11	1.6 ± 0.4	23.8 ± 3.8	67 ± 7	122.4 ± 44.0
2017	0.75 ± 0.66	143.07 ± 62.51	0.26 ± 0.14	1.7 ± 0.4	21.3 ± 3.2	56 ± 13	128.9 ± 53.1
2018	0.75 ± 0.72	62.62 ± 32.87	0.12 ± 0.07	1.6 ± 0.5	22.4 ± 2.9	63 ± 12	122.2 ± 39.9
2019	0.36 ± 0.36	93.42 ± 36.27	0.08 ± 0.07	1.6 ± 0.3	21.7 ± 2.7	48 ± 11	165.0 ± 29.4
2020	0.36 ± 0.16	73.40 ± 30.27	0.02 ± 0.01	1.7 ± 0.4	20.2 ± 5.1	55 ± 13	145.5 ± 31.9

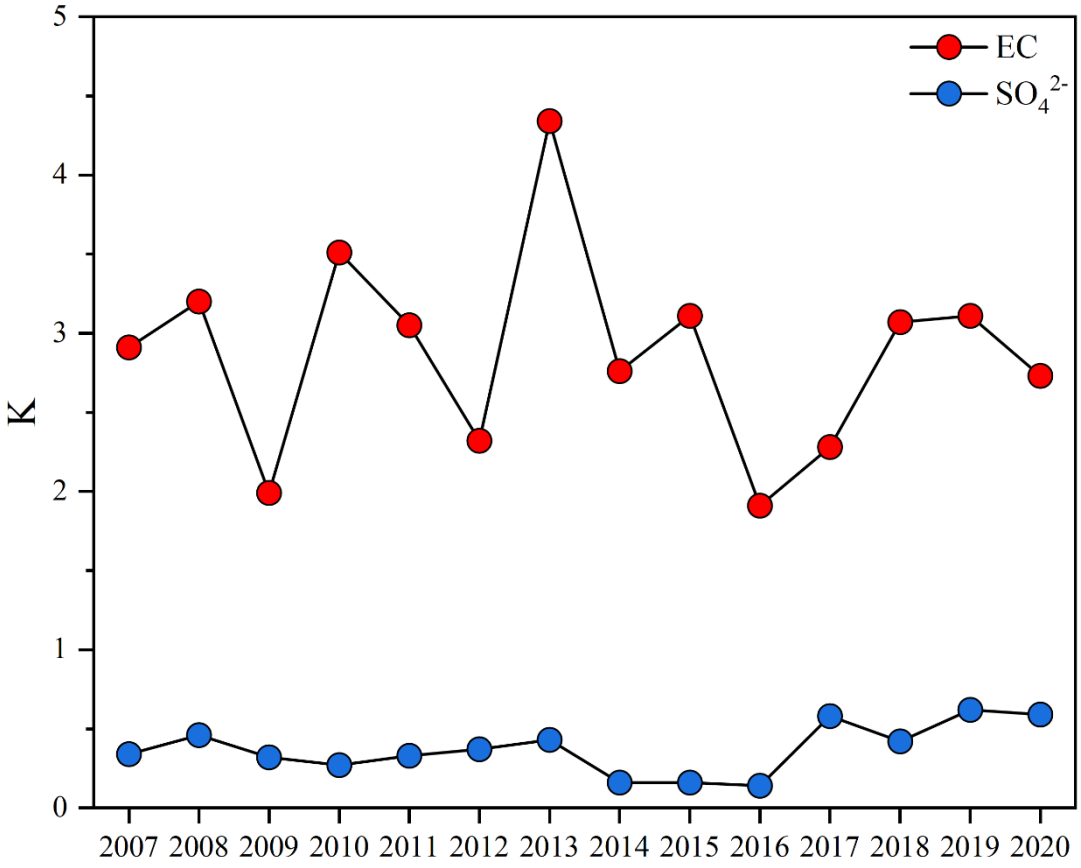


Figure S1. Variations of *K* values for EC and SO_4^{2-}, calculated by Bayesian Inference Approach.

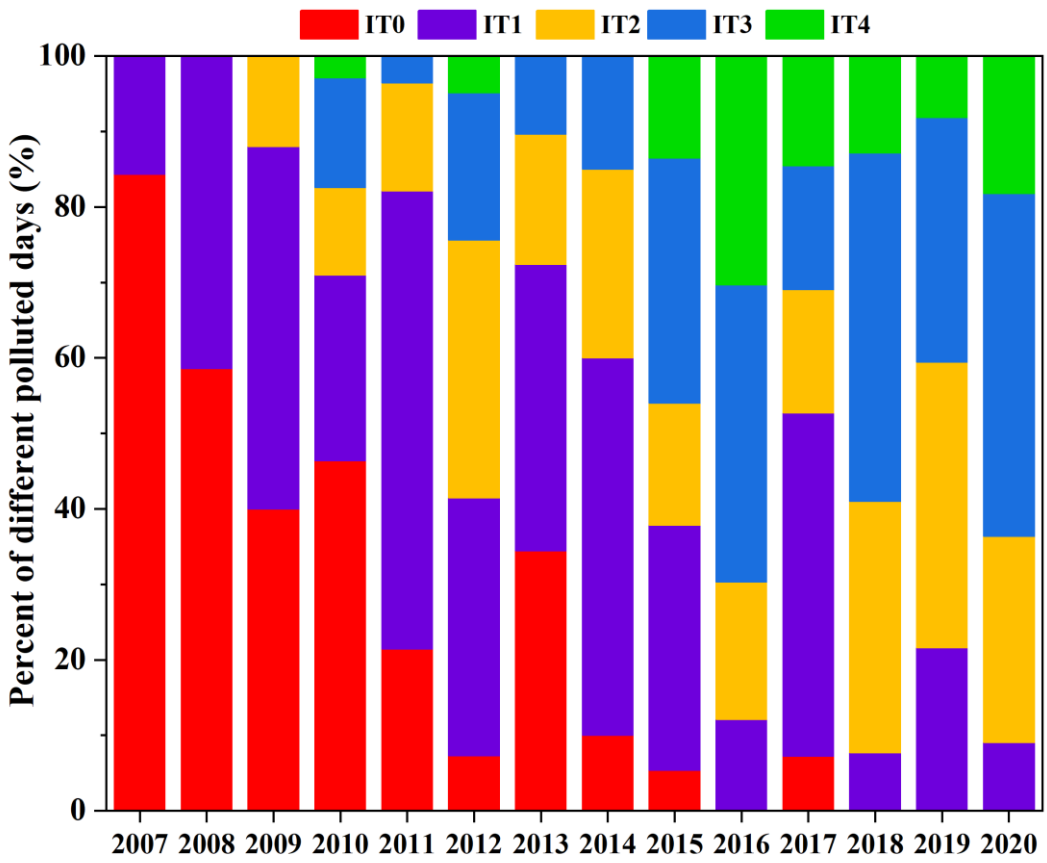


Figure S2. Percent of different polluted days in WQS from 2007 to 2020.

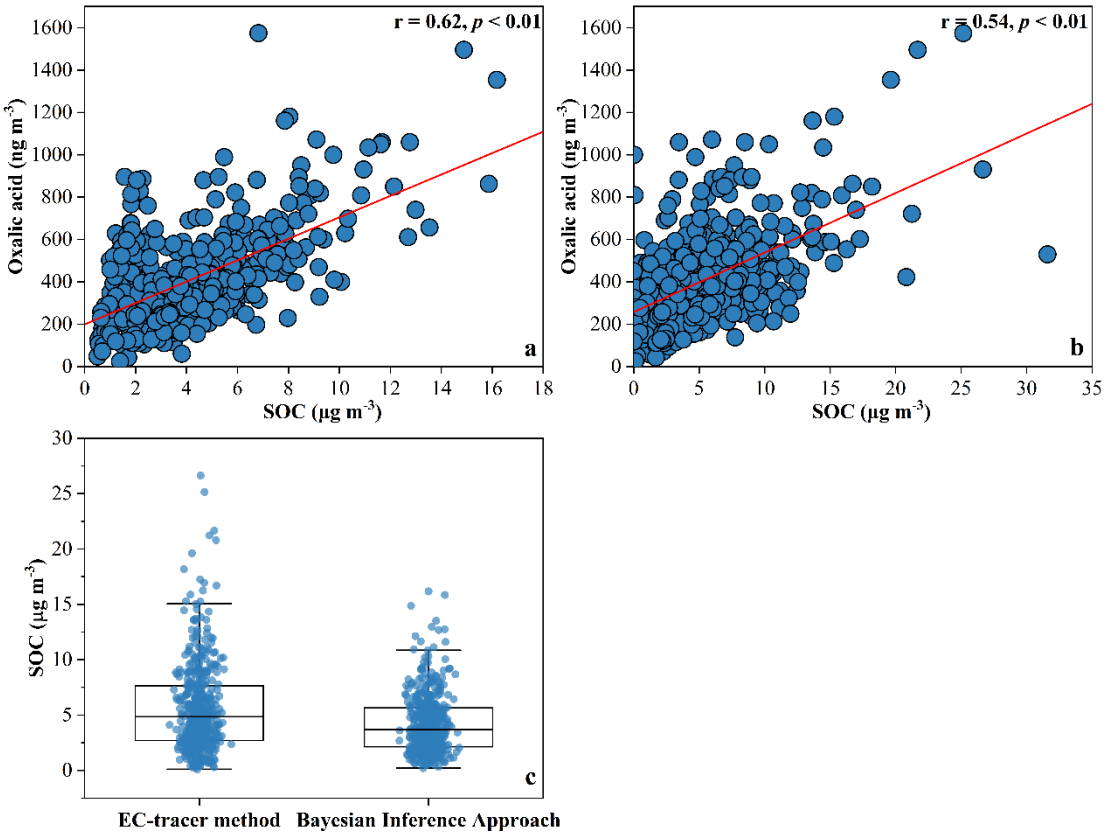


Figure S3. Oxalic acid, a typical secondary organic molecular tracer, exhibited a significant correlation ($p < 0.01$) with SOC estimated using both Bayesian Inference Approach (a) and EC–tracer method (b). The Pearson's correlation coefficient for the Bayesian Inference Approach (0.62) was higher than that for the EC–tracer method (0.54), suggesting that SOC derived from the Bayesian Inference Approach might be more reliable. In addition, the average SOC concentration estimated by the Bayesian Inference Approach (4.1 ± 2.6 μg m^{-3}) was lower and more than that by the EC–tracer method (5.8 ± 4.2 μg m^{-3}) (c).

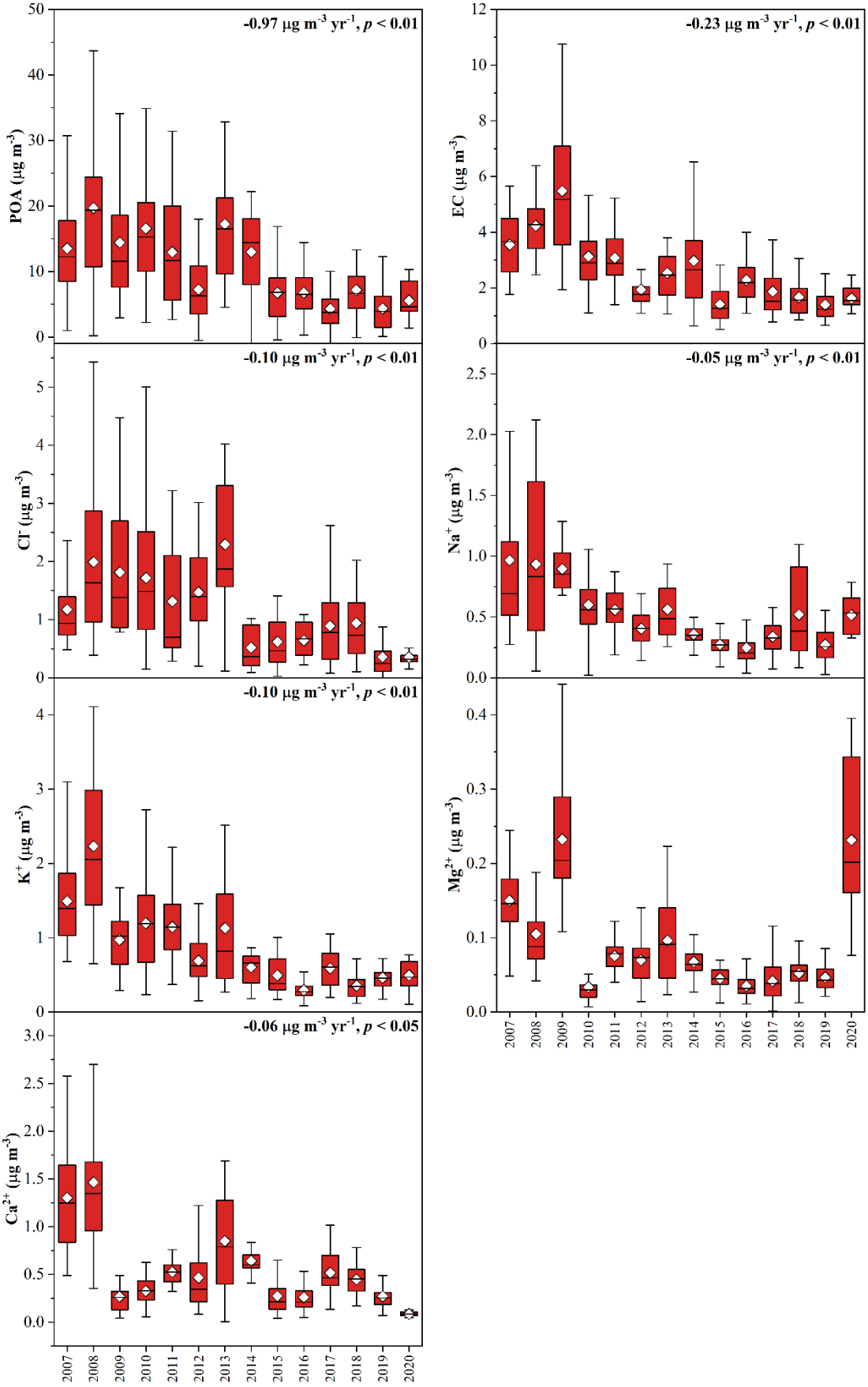
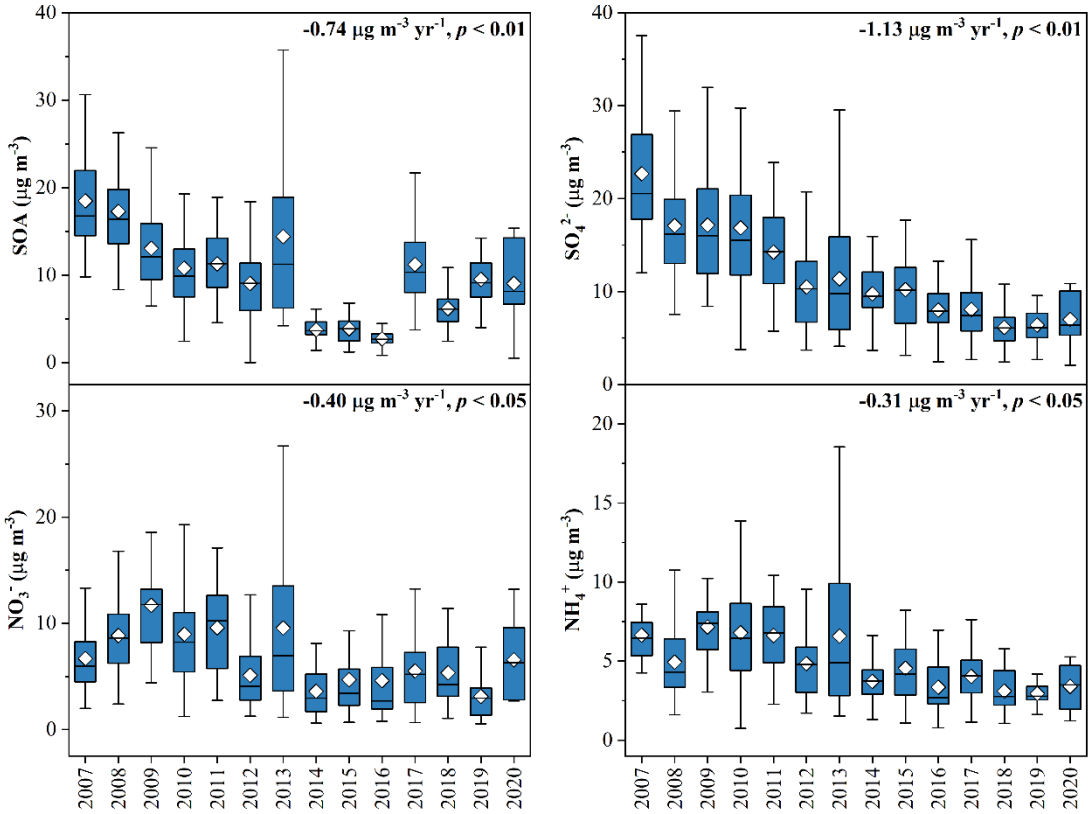


Figure S4. Concentrations of $PM_{2.5}$ primary species during 2007-2020.



55 **Figure S5. Concentrations of $PM_{2.5}$ secondary species during 2007-2020.**

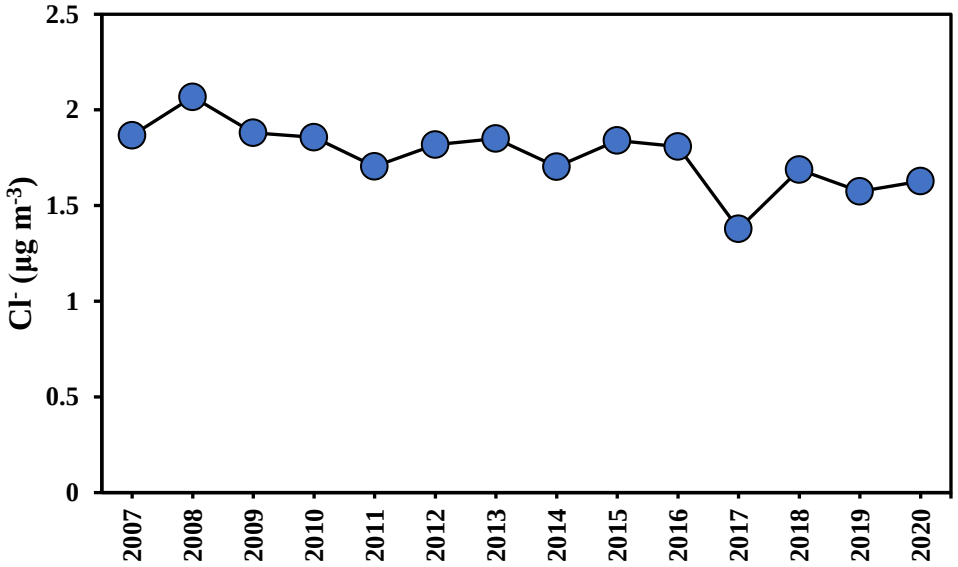


Figure S6. Eliminating changes of anthropogenic sources on Cl^-. It decreased slightly at a rate of –2% yr^{-1} during 2007-2020.

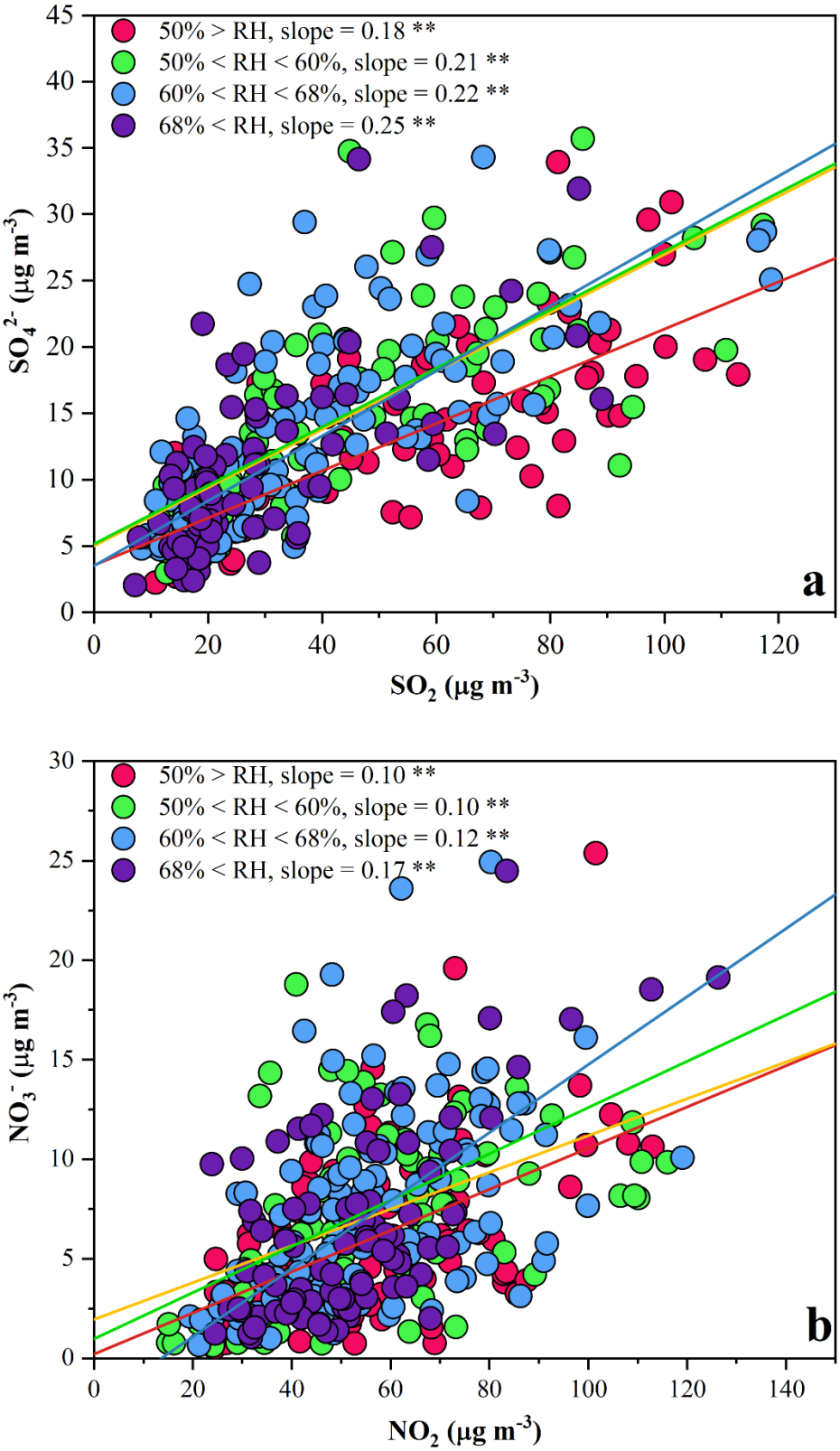


Figure S7. Correlations between SO_4^{2-}/SO_2 (a), as well as NO_3^-/NO_2 (b). Two asterisks denote *p* value less than 0.01. All samples were categorized into four groups according to the quartile ranges of RH. The slope became greater with rising RH, indicating conversion of primary pollutants to secondary species was more efficient.

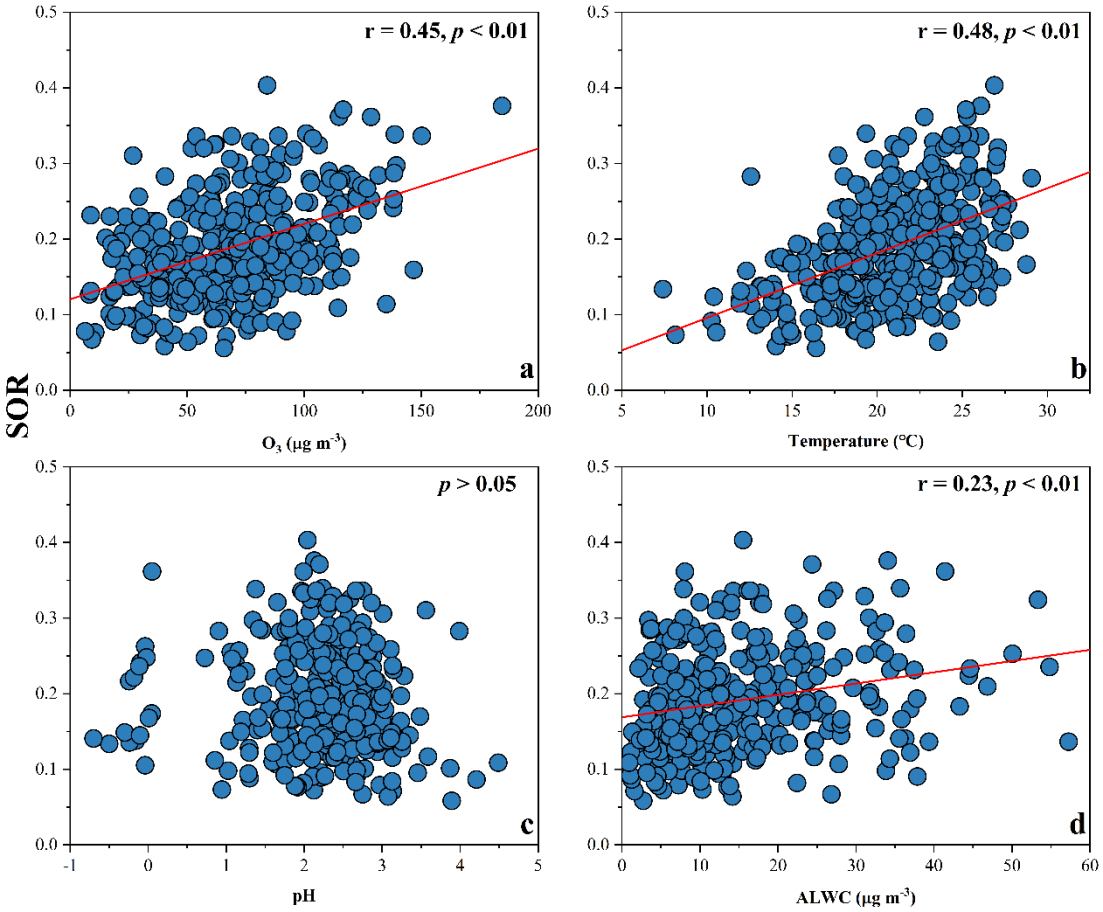


Figure S8. Correlations between SOR and O_3 (a), temperature (b), pH (c), as well as ALWC (d).

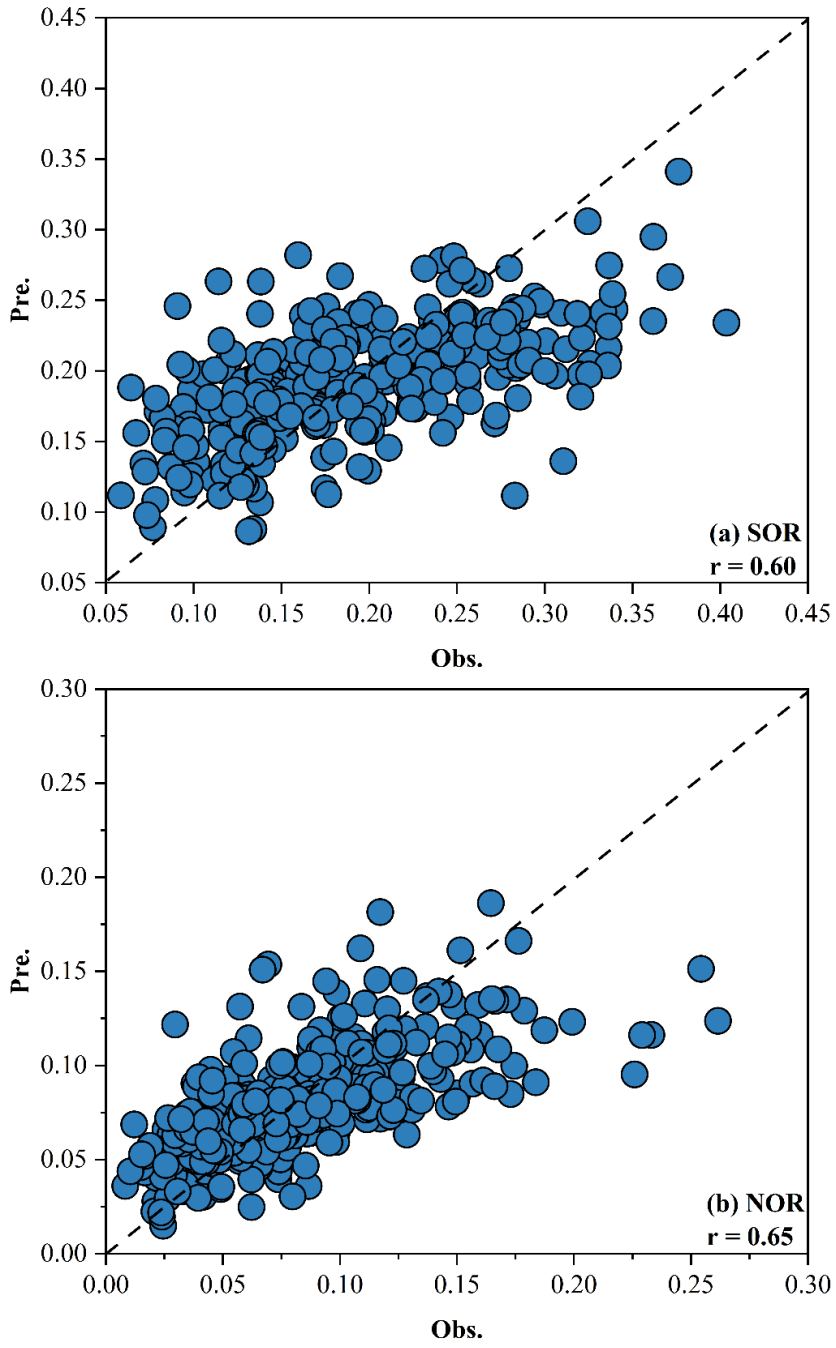


Figure S9. Correlations between predictions and observations of SOR (a) and NOR (b). Solid line is 1:1.

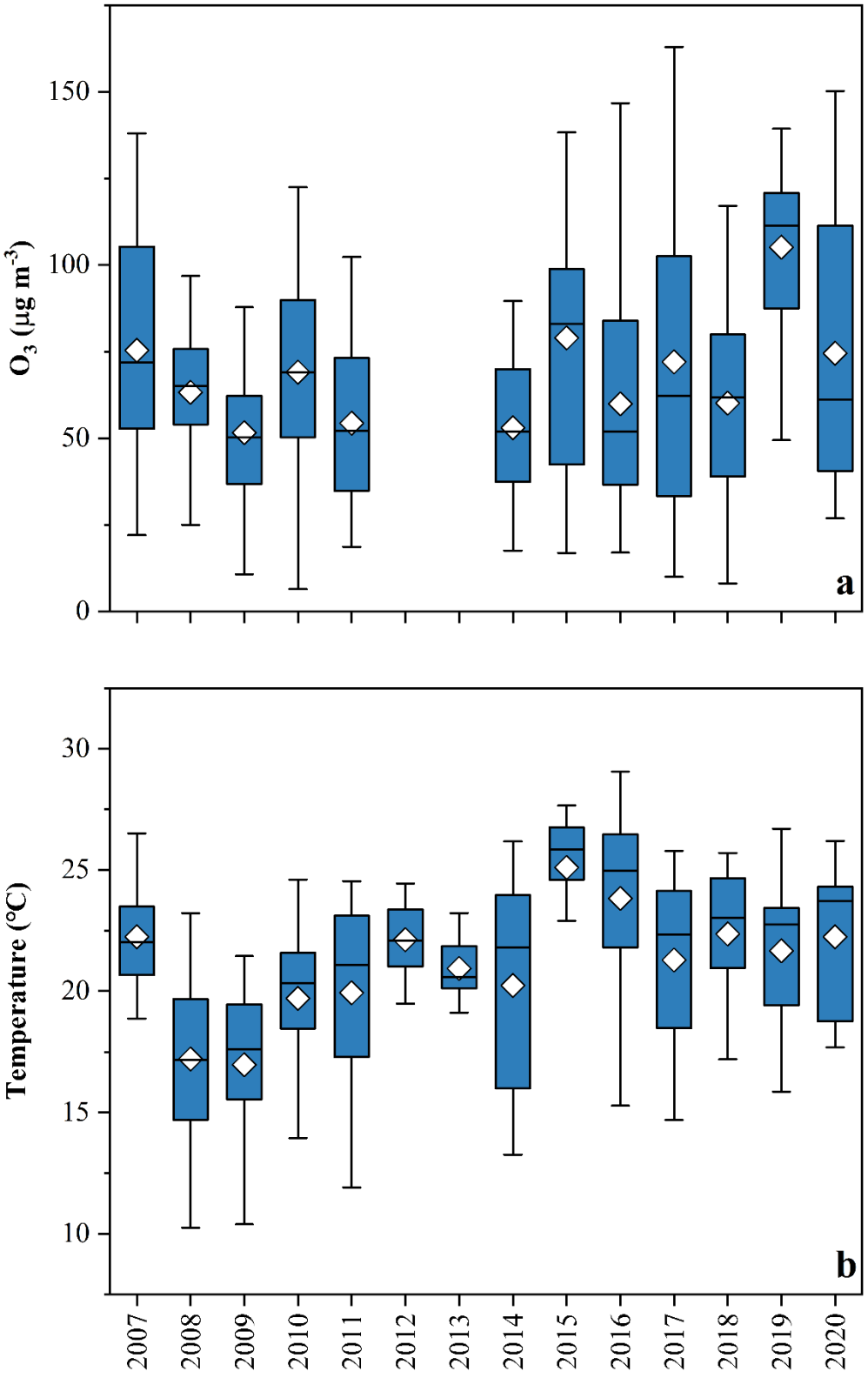


Figure S10. Variations of O_3 (a) and temperature (b) during 2007-2020.

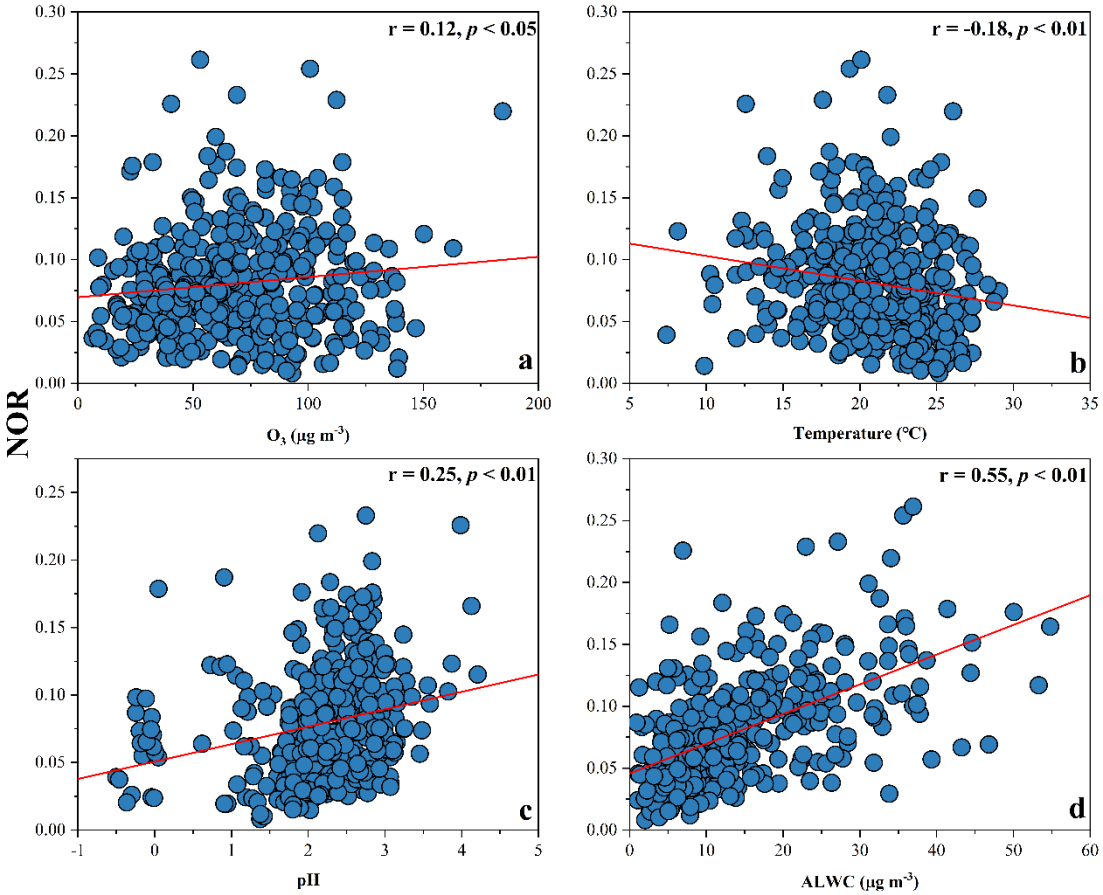


Figure S11. Correlations between NOR and O_3 (a), temperature (b), pH (c), as well as ALWC (d).

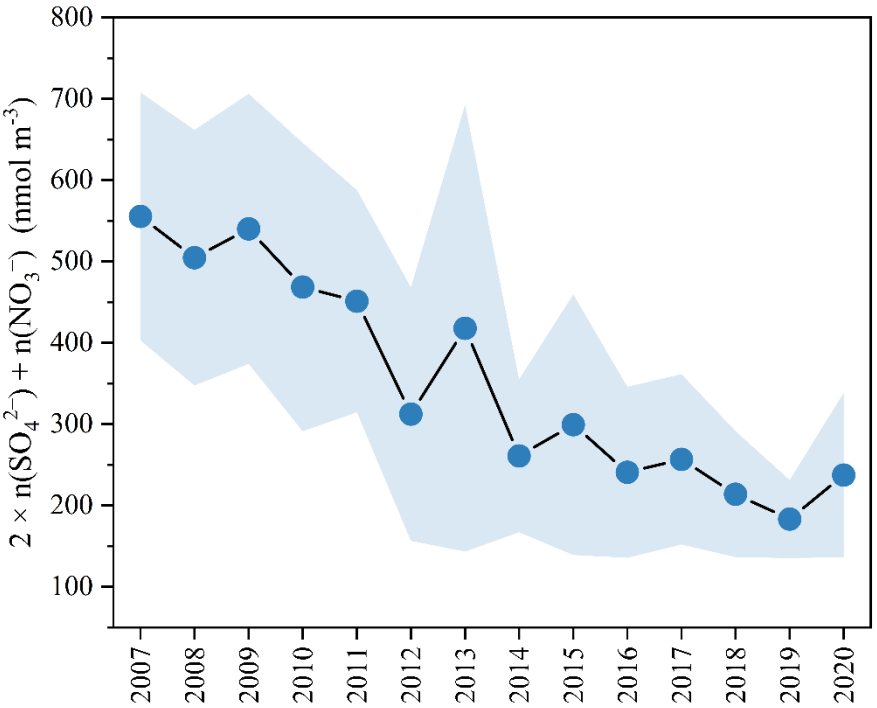
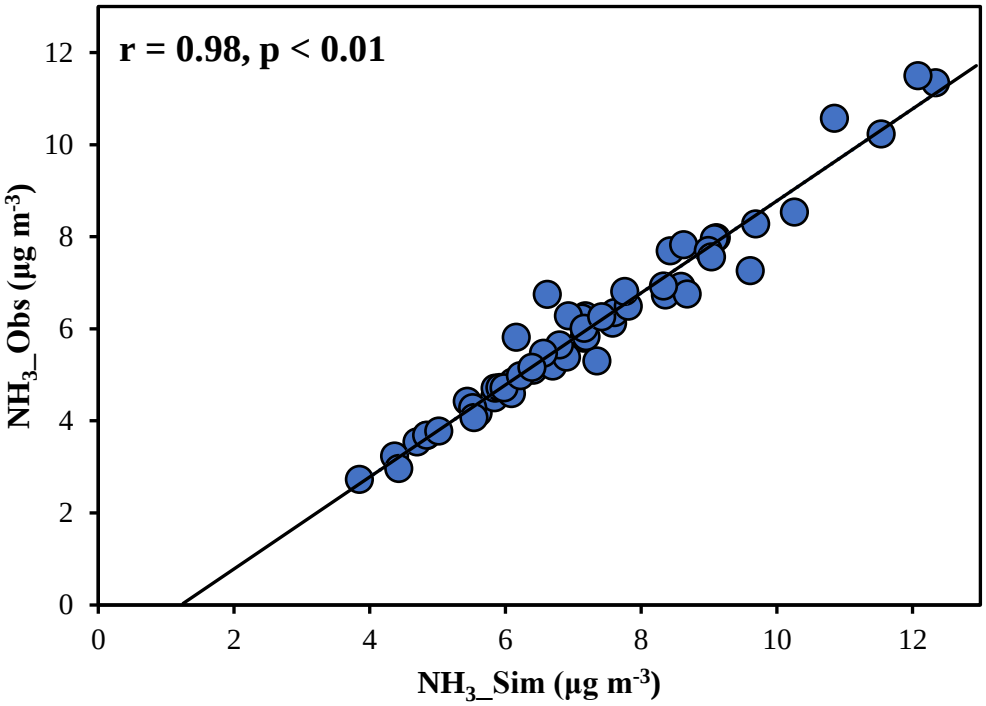


Figure S12. Variations of $2 \times n(SO_4^{2-}) + n(NO_3^-)$.



75

Figure S13. Observations and ISORROPIA-II simulation of gas-phase NH_3 in GIG.

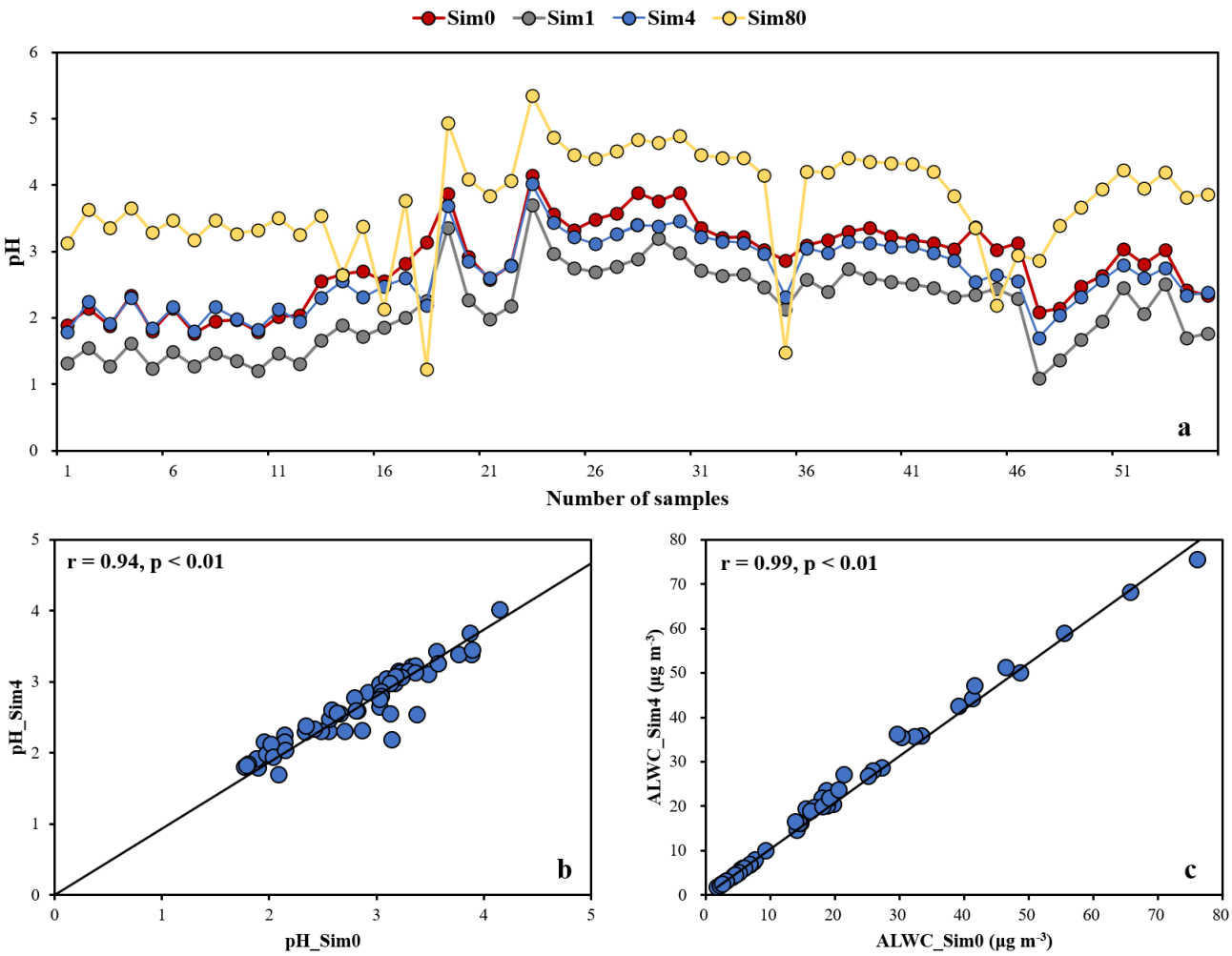
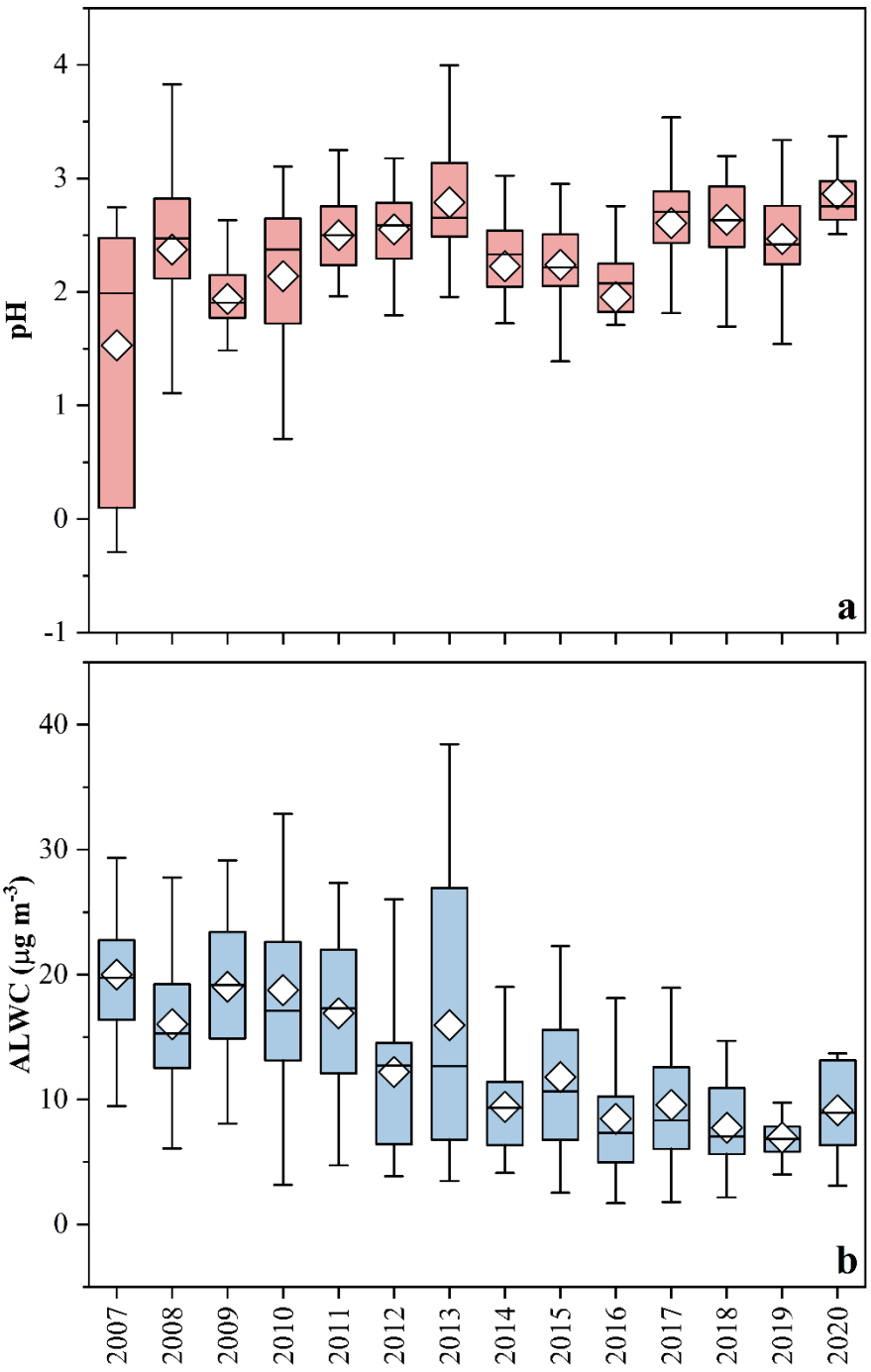


Figure S14. Comparison of predicted pH, ALWC with and without gas-phase input in GIG. (a) Sim0 represents results with gas-phase input, Sim1, Sim4, Sim80 represent results without gas-phase input and performed one, four, and eighty iterations respectively. It showed predicted aerosol pH rose as number of iterations increased. Four iterations without gas-phase input would generate the optimal result. (b) and (c) showed pH and ALWC calculated by this way had the greatest agreement with the one which has gas-phase input.



85

Figure S15. Recalculation of pH (a) and ALWC(b) using average temperature and RH as model input. The upward trend in pH and downward trend in ALWC still exist.

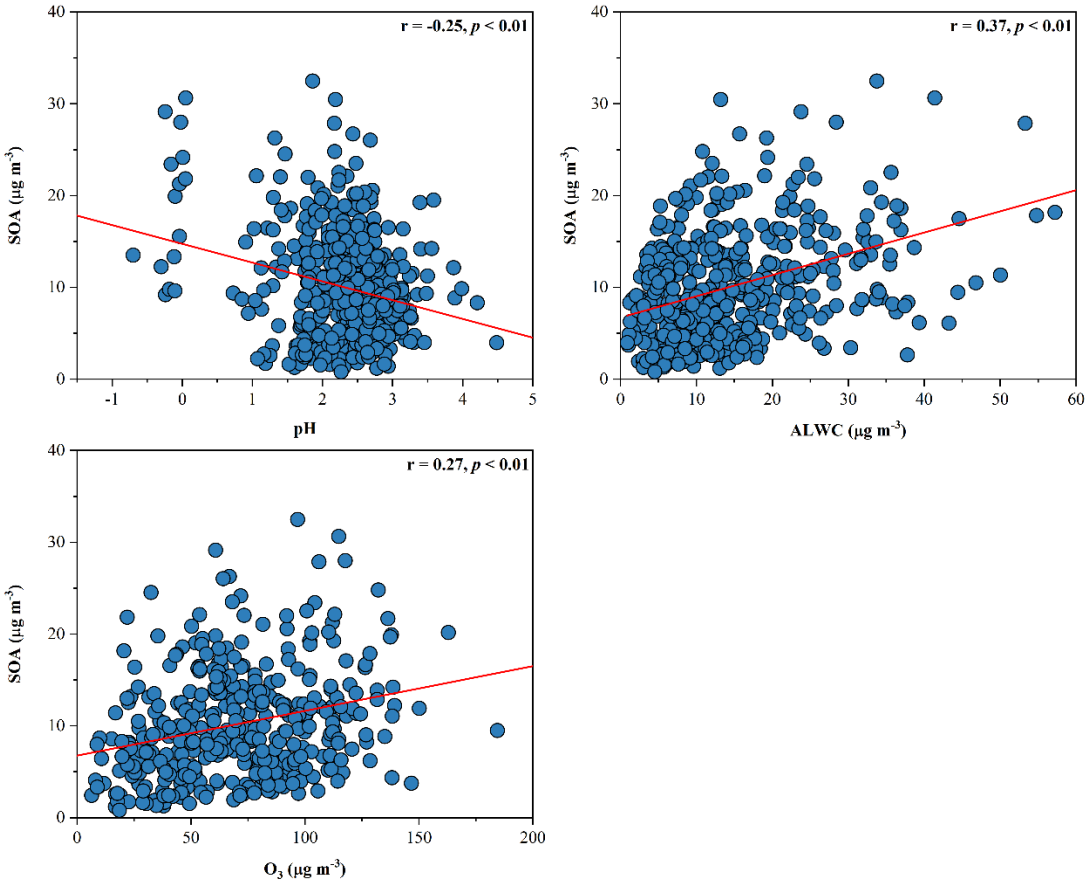


Figure S16. Correlations between SOA and pH, ALWC, as well as O_3.

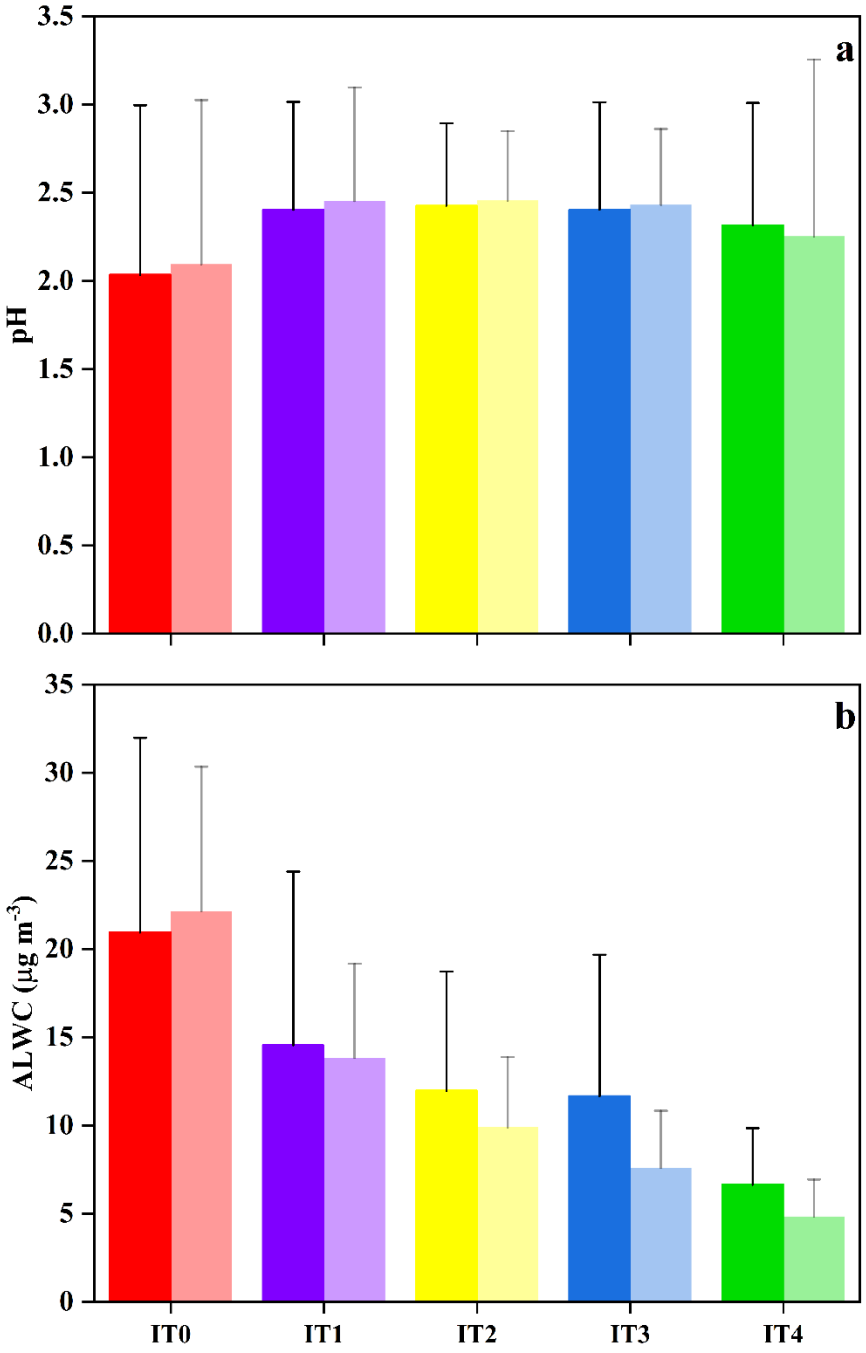


Figure S17. Variations of pH (a) and ALWC (b) under different pollutant levels. The dark color bars represented original model prediction, while the light color bars represented recalculation by average temperature and RH. A low pH occurred under high pollutant level (IT0). With decrease of pollutant levels (IT0-IT4), ALWC exhibited a downward trend.

95

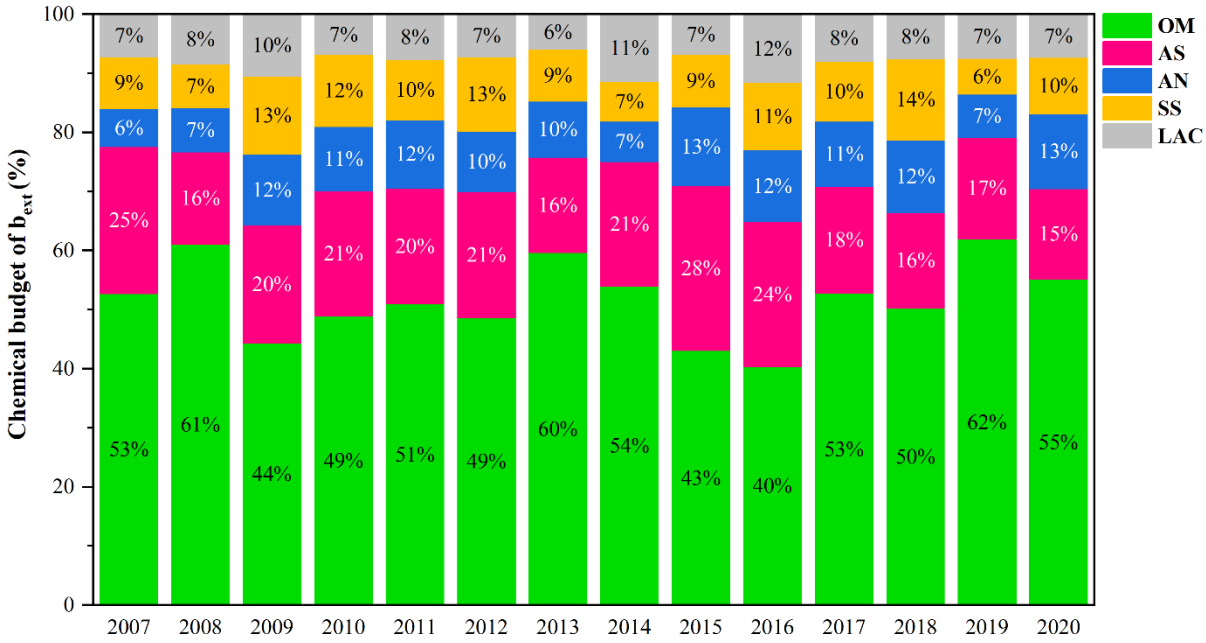


Figure S18. Chemical budget of b_{ext} from different components in $PM_{2.5}$.

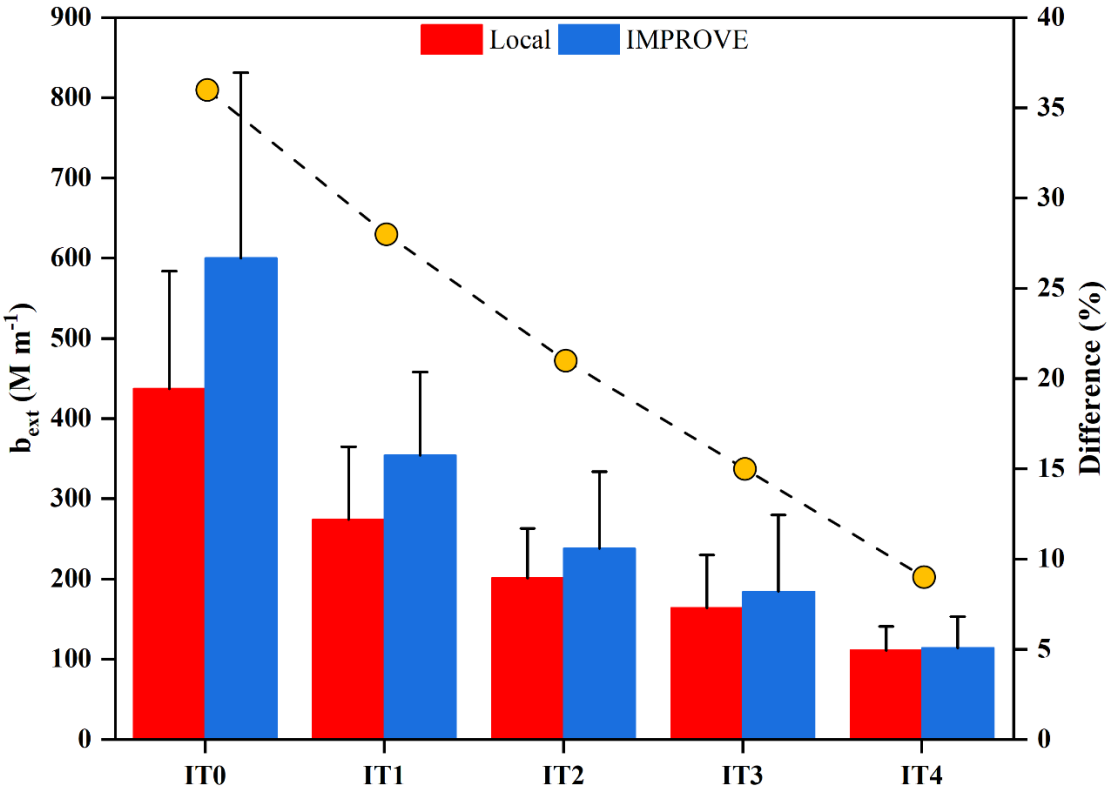


Figure S19. The difference of b_{ext} under different pollutant level. Bars represent bext and circles represent difference between Local scheme and IMPROVE scheme. When pollutant levels were high (IT0), the difference was above 30%, while the difference was under 10% when pollutant levels were low (IT4).

Bougiatioti, A., Nikolaou, P., Stavroulas, I., Kouvarakis, G., Weber, R., Nenes, A., Kanakidou, M., and Mihalopoulos, N.: Particle water and pH in the eastern Mediterranean: source variability and implications for nutrient availability, Atmos. Chem. Phys., 16, 4579-4591, https://doi.org/10.5194/acp-16-4579-2016, 2016.

Fang, Z. Y., Dong, S. W., Huang, C. P., Jia, S. G., Wang, F., Liu, H. M., Meng, H., Luo, L., Chen, Y. Z., Zhang, H. H., Li, R., Zhu, Y. J., and Tang, M. J.: On using an aerosol thermodynamic model to calculate aerosol acidity of coarse particles, J. Environ. Sci., 148, 46-56, https://doi.org/10.1016/j.jes.2023.07.0011, 2025.

Fountoukis, C. and Nenes, A.: ISORROPIA II: a computationally efficient thermodynamic equilibrium model for K+–Ca2+–Mg2+–NH4+–Na+–SO42−–NO3−–Cl−–H2O aerosols, Atmos. Chem. Phys., 7, 4639-4659, https://doi.org/10.5194/acp-7-4639-2007, 2007.

Guo, H., Xu, L., Bougiatioti, A., Cerully, K. M., Capps, S. L., Hite Jr, J. R., Carlton, A. G., Lee, S. H., Bergin, M. H., Ng, N. L., Nenes, A., and Weber, R. J.: Fine-particle water and pH in the southeastern United States, Atmos. Chem. Phys., 15, 5211-5228, https://doi.org/10.5194/acp-15-5211-2015, 2015.

Hou, L. L., Dai, Q. L., Song, C. B., Liu, B. W., Guo, F. Z., Dai, T. J., Li, L. X., Liu, B. S., Bi, X. H., Zhang, Y. F., and Feng, Y. C.: Revealing Drivers of Haze Pollution by Explainable Machine Learning, Environ. Sci. Technol. Lett., 9, 112-119, https://doi.org/10.1021/acs.estlett.1c00865, 2022.

Nenes, A., Pandis, S. N., and Pilinis, C.: ISORROPIA: A New Thermodynamic Equilibrium Model for Multiphase Multicomponent Inorganic Aerosols, Aquat. Geochem., 4, 123-152, https://doi.org/10.1023/A:1009604003981, 1998.

Wen, L., Xue, L. K., Wang, X. F., Xu, C. H., Chen, T. S., Yang, L. X., Wang, T., Zhang, Q. Z., and Wang, W. X.: Summertime fine particulate nitrate pollution in the North China Plain: increasing trends, formation mechanisms and implications for control policy, Atmos. Chem. Phys., 18, 11261-11275, https://doi.org/10.5194/acp-18-11261-2018, 2018.

Zhou, M., Zheng, G. J., Wang, H. L., Qiao, L. P., Zhu, S. H., Huang, D. D., An, J. Y., Lou, S. R., Tao, S. K., Wang, Q., Yan, R. S., Ma, Y. G., Chen, C. H., Cheng, Y. F., Su, H., and Huang, C.: Long-term trends and drivers of aerosol pH in eastern China, Atmos. Chem. Phys., 22, 13833-13844, https://doi.org/10.5194/acp-22-13833-2022, 2022.