Supplement of Atmos. Chem. Phys., 25, 11975–11989, 2025
https://doi.org/10.5194/acp-25-11975-2025-supplement

Supplement of

Effects of anthropogenic pollutants on biogenic secondary organic aerosol formation in the atmosphere of Mt. Hua, China

Can Wu et al.

Correspondence to: Gehui Wang (ghwang@geo.ecnu.edu.cn)

S1. Random forest analysis and mantel's test for BSOA tracer

Randow forest (RF) was initially introduced by Breiman (2001) to serve as a powerful tool for regression and prediction with high accuracy; and it has been widely applied in environmental field (e.g., pollutants prediction), even the data have complex nonlinear relationships and interactions. In this study, a regression model based on RF algorithm was built to reveal the key factors influencing the BSOA formation at MS site. The potential factors, including O_3, NO_2, SO_4^{2-}, NH_4^+, ALWC, H^+, RH and T, were regarded as explanatory variables for the daytime isoprene-derived SOA prediction. For robust model development and validation, 70% of the original input data were randomly allocated to build the RF model (i.e., training dataset), and the remaining 30% served as the testing dataset for evaluating the model performance. The RF model is user-friendly with only two critical parameters constantly being optimized, namely the number of trees grown (n_{tree}) and the number of variables split at each node (m_{try}). By systematically comparing the results of different parameters settings, we identified that setting n_{tree} to 100 and m_{try} to 4 can provide the optimal prediction accuracy, which can be established by the statistical metrics as shown in Table S1. From Table S1, we can note that a well performance of the model in explaining the importance of these factors to daytime $BSOA_I$, with a robust correlation between predicted and observed BSOA (R^2=0.8). Additionally, the mantel test, as a statistical tool initiated by Mantel (1967), is commonly used for investigating relationships among two matrices. To elucidate the relationships between three types of BSOA tracers and potential influencing factors, a Mantel test was conducted using a network platform (https://www.omicstudio.cn/tool, last accessed: Jan. 10th, 2025).

S2. Estimation the IEPOX reactive uptake on particles at MS site

To quantitatively assess the effects of aerosol water on isoprene-derived SOA formation, the pseudo-first-order heterogeneous reaction rate constant for IEPOX uptake to particle (k_{het}, 1/s) is calculated using the method of Gaston et al. (2014), in which the gas phase diffusion is not considered due to its minor effects on k_{het} (eq. (1)). And the possibility of IEPOX reactive uptake (i.e., uptake coefficient, γ_{IEPOX}) was also calculated using equations S2-S3.

$$k_{het} \approx \frac{\gamma_{IEPOX} S_a \omega}{4} \quad \text{(S1)}$$

$$\frac{1}{\gamma_{IEPOX}} = \frac{R_p \omega}{4D_g} + \frac{1}{\alpha} + \frac{1}{\Gamma_{aq}} \quad \text{(S2)}$$

$$\Gamma_{aq} = \frac{4VRTH_{aq}k_{aq}}{S_a \omega} \quad \text{(S3)}$$

$$k_{aq} = \left(k_{H^+}[H^+]\right) + \left(k_{nuc}[nuc]\alpha_{H^+}\right) + k_{ga}[ga] \quad \text{(S4)}$$

Where S_a represents the total surface area concentration of particles (cm^2/cm^3); The S_a, particle radium (R_p, m) and total particle volume (V, cm^3/cm^3) were measured using a scanning mobility particle sizer (SMPS) here, more details can be found elsewhere (Cao, 2018). In equations S2-S3, ω is the mean molecular speed of an epoxide molecule (231 m/s, at 298 K), D_g the gas-phase diffusion coefficient (0.1 cm^2/s); α is the mass accommodation coefficient, with a mean value of 0.1. R is the universal gas constant (L atm/mol K), T represent the temperature (K). The Henry's law coefficient in the aqueous phase, H_{aq}, is estimated to be 1.7×10^8 (M/atm). The pseudo-first-order aqueous-phase rate constant, k_{aq} (1/s), is based on an acid-catalyzed, epoxide ring-open mechanism. k_{H^+} is the reaction rate taken to be 0.036 (M/s); $[H^+]$ and α_{H+} are the concentration (mol/L) and activity of proton, respectively. k_{nue} is the reaction rate due to the presence of specific nucleophiles (e.g., sulfate, nitrate), which is estimated to be 2.0×10^4 (M/s); [nue] is the nucleophiles concentration (mol/L). [ga] represent the concentration of general acids (e.g., bisulfate, mol/L), of which reaction rate is taken to 2.0×10^4 (M/s). All the key parameters applied here are consistent with Gaston et al. (2014).

Table S1 RF model performance for testing dataset.

	R^2	MSE	RMSE	MAE
Training dataset	0.91	0.012	0.11	0.07
Testing dataset	0.80	0.008	0.09	0.07

Note: MSE: mean square error; RMSE: root-mean-square error; MAE: mean absolute error.

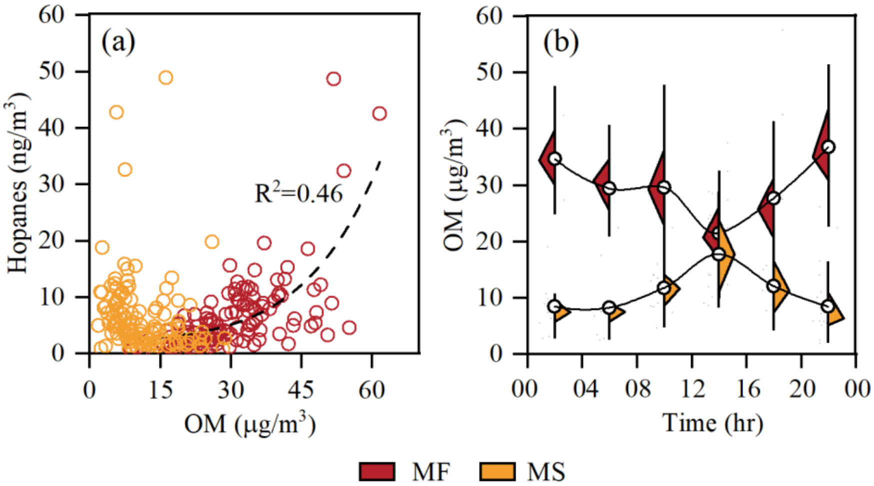

Figure S1 The fit regressions for OM with hopanes **(a)** and diurnal cycle of OM **(b)** at both sampling sites.

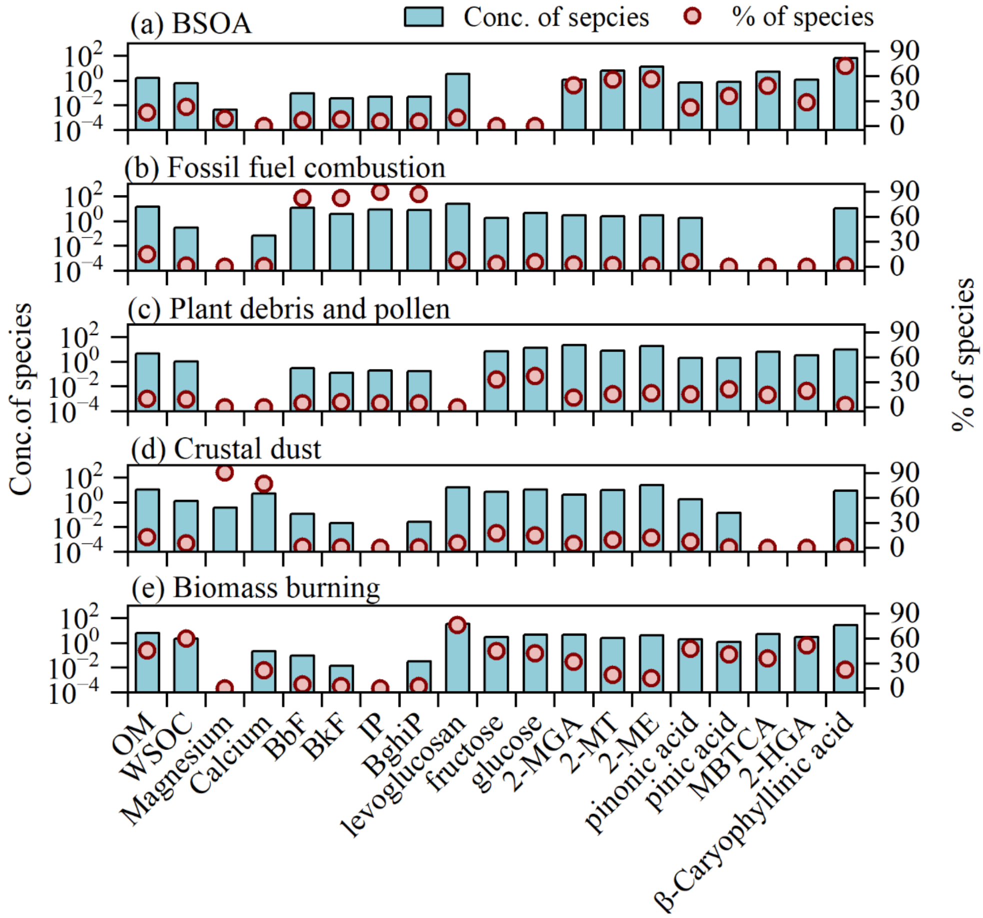

Figure S2 Factor profiles resolved by PMF mode for OM source apportionment.

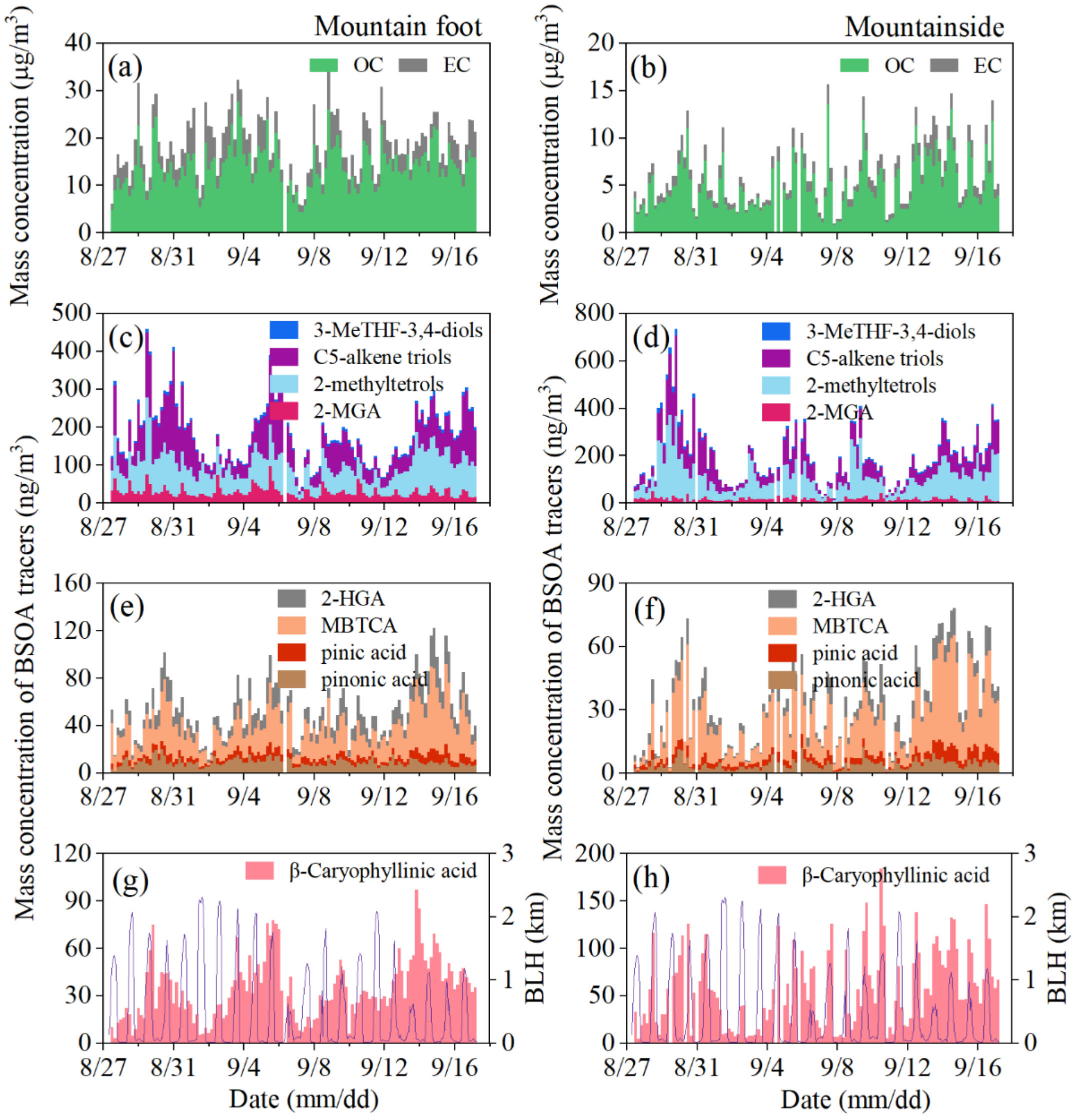

Figure S3 Time series of the mass concentration of BSOA tracers and boundary layer height (BLH) at both sampling sites.

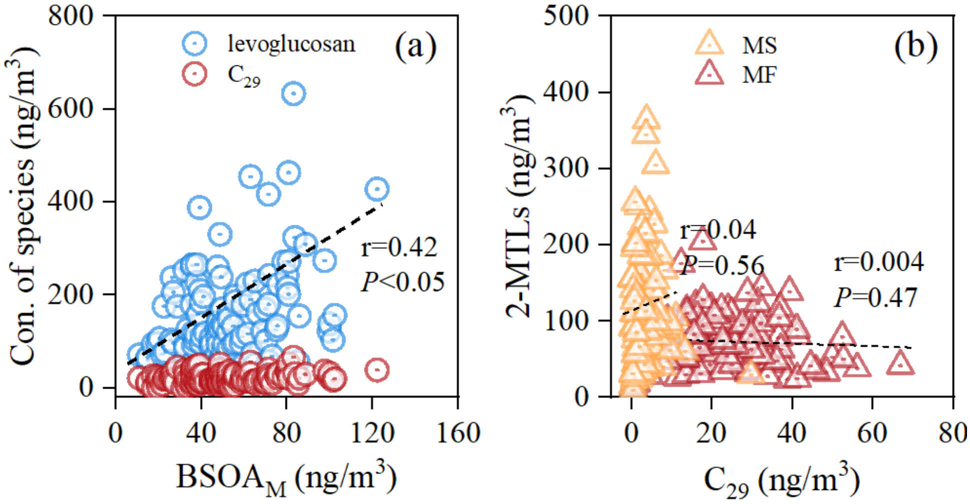

Figure S4 Correlation of $BSOA_M$ with levoglucosan and C_{29}-alkane at MF site

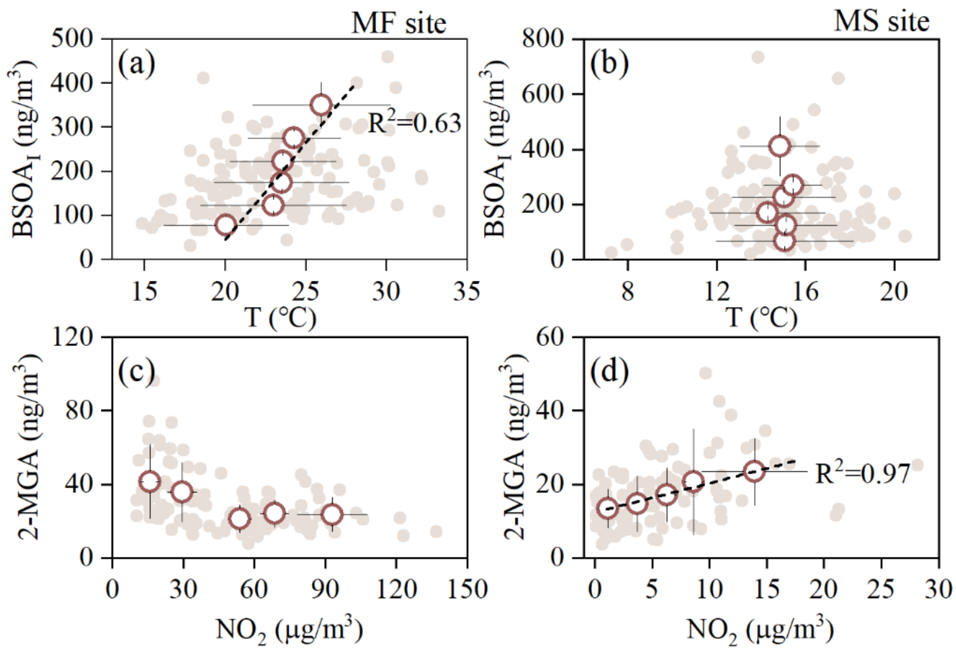

Figure S5 Linear correlations of $BSOA_I$ with temperature (**a** and **b**), and 2-MGA with NO_2 (**c** and **d**) at both sampling sites.

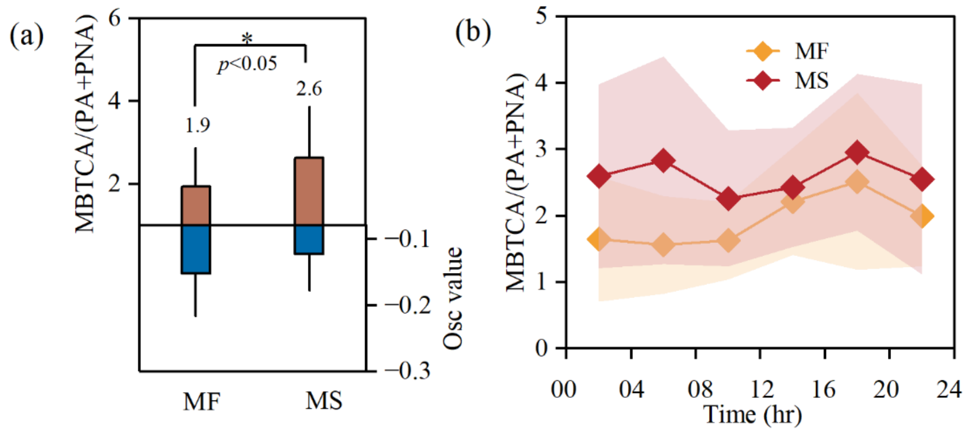

Figure S6 A comparison of $BSOA_M$ tracer among two sampling sites; (a) aging state, (b) diurnal variation of MBTCA/(PA+PNA) ratio. (Error bar of Osc value has been reduced by half)

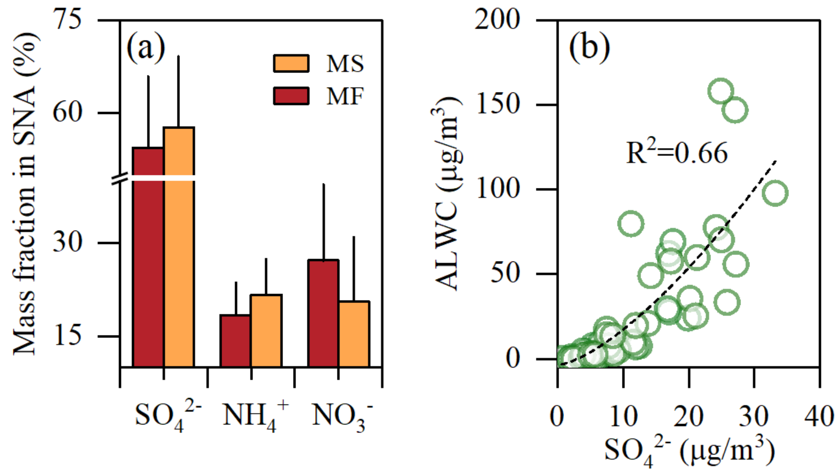

Figure S7 The comparison of proportions in daytime SNA between the two Sampling Sites (a); ALWC as a function of daytime sulfate at MS site (b).

References

Breiman, L.: Random forests, Machine Learning, 45, 5-32, 10.1023/a:1010933404324, 2001.

Cao, C.: Vertical distribution of physical and chemical characteristics of PM2.5 at Mt. Hua and the new particle formation, 2018.

Gaston, C. J., Riedel, T. P., Zhang, Z., Gold, A., Surratt, J. D., and Thornton, J. A.: Reactive Uptake of an Isoprene-Derived Epoxydiol to Submicron Aerosol Particles, Environmental Science & Technology, 48, 11178-11186, 10.1021/es5034266, 2014.

Mantel, N.: The detection of disease clustering and a generalized regression approach, Cancer research, 27, 209-220, 1967.