Supplement of Atmos. Chem. Phys., 25, 7619–7645, 2025
https://doi.org/10.5194/acp-25-7619-2025-supplement

Atmospheric Chemistry and Physics
Open Access

Supplement of

Machine-learning-assisted chemical characterization and optical properties of atmospheric brown carbon in Nanjing, China

Yu Huang et al.

Correspondence to: Xinlei Ge (caxinra@163.com)

S1

A portion of each quartz filter (10*3.14 cm^2) was cut off for analysis. The piece was cut into small pieces and placed into a 30 ml brown vial. 5 mL methanol (Optima LC/MS grade, Fisher Chemical, USA) was added to the brown vial and sonicated for 30 minutes at ambient temperature, and the procedure was repeated three times. The solution was then filtrated through a 0.22 μm PTFE filter (Fisher Chemical, USA), followed by the blow-drying with nitrogen gas. The blow-dried solid residue was re-dissolved in 300 μL of methanol and transferred to a 1.5 ml brown injection vial for storage. A 30 μL aliquot from each sample was mixed together as a QC (quality check) sample.

S2

The analysis was performed by using an Acquity H Class Ultra Performance Liquid Chromatography system coupled to a Xevo G2-Xs Quadrupole time-of-flight mass spectrometer (UPLC-QTOF-MS, Agilent Technologies Inc. Santa Clara, CA, USA). A C18 column (100 mm × 2.1 mm × 1.6 μm) (Luna Omega, Phenomenex) was used for the chromatographic separation, and temperature of the column was maintained at 40 °C. The sample volume was 1.5 μL for the positive ion (ESI^+) mode and 3 μL for the negative ion (ESI^-) mode. The mobile phase was consisted of solvent A (ultrapure water containing 0.1% v/v formic acid and 5 mM ammonium acetate) and solvent B (acetonitrile containing 0.1% v/v formic acid), and the gradients of eluent were programmed as follows: 2 % B at 0 ~ 1.5 min; linearly from 2 % B to 20 % B at 1.5 ~ 11 min; linearly from 20 % B to 60 % B at 11 ~ 18 min; linearly from 60 % B to 98 % B at 18 ~ 20 min; maintain at 98 % B at 20 ~ 22min, then decrease to 2 % B at 22 ~ 25 min.

The instrument used the electrospray ionization (ESI) technique. A data independent acquisition (DIA) resolution mode was operated with a *m/z* ratio of 50-1200, a scanning interval of 0.1 s, capillary voltages of 0.7 KV for ESI^+ and 2.35 KV for ESI^-, a cone bore voltage of 30 V, an ion source temperature of 120 °C, a cone bore

gas of 50 L min^{-1}, a desolvation gas of 1000 L min^{-1}, and collision energies of 10 ~ 50 eV. The positive or negative ion modes were calibrated with leucine enkephalin and sodium formate polymers, respectively, and data were acquired by using Masslynx 4.1.

S3

The raw UPLC-QTOF-MS data were processed using the Mass Spectrometry-Data Independent Analysis (MS-DIAL, version 4.92) software, involving peak extraction, peak alignment, and deconvolution with a detection probability of 70 %. The summed ions included $[M-H]^-$ (ESI^-) and $[M+H]^+$, $[M+NH_4]^+$, $[M+Na]^+$ (ESI^+). Missing data were replaced by 1/10 of the minimum value (default value: 100). Chromatographic intensities were normalized using the systematic error removal using random forest (a machine learning algorithm) (SERRF) software, which is based on the machine learning (ML) random forest (RF) algorithm. Additionally, corrections were applied for potential intensity drift.

By using MS-DIAL, all deconvoluted MS/MS spectra were exported as individual .mat files. Subsequently, all MS/MS spectra were examined and imported into SIRIUS (version 5.6.2) for the identification of molecular formulas of each *m/z*. In order to obtain a more refined list of molecular formulas, specific constraints below were applied, and those did not comply with these rules were excluded.

(1) Atomic numbers: 1~50 12C, 1 ~ 100 1H, 0 ~ 40 16O, 0 ~ 5 14N, 0 ~ 2 32S;

(2) Elemental ratios: In ESI^- mode, 0.3 ~ 3.0 H/C, 0 ~ 3 O/C, 0 ~ 0.5 N/C, 0 ~ 2.0 S/C; in ESI^+ mode, 0.3 ~ 3.0 H/C, 0 ~ 1.2 O/C, 0 ~ 1.0 N/C, 0 ~ 0.8 S/C.

(3) Equivalent double bond (DBE) numbers: 0 ~ 25.

The double-bond equivalent (DBE) value of a molecule is used to indicate the level of unsaturation, which can be calculated by using the following Eq. (S1):

$$DBE = \frac{2 \times C - H + N + 2}{2} \qquad (S1)$$

Here, C, H, N are the number of carbon, hydrogen and nitrogen in the formula of the molecule. Furthermore, the aromaticity equivalent (Xc) has been used to aid the identification of aromatic and condensed aromatic compounds, as described in Yassine

et al. (2014). Compared to the aromaticity index (AI), the advantage of Xc lies in its ability to accurately classify (poly)aromatic compounds with significant alkylations. The Xc value can be calculated with Eq. (S2):

$$X_C = \frac{3 \times \left(DBE - (p \times O + q \times S)\right) - 2}{DBE - (p \times O + q \times S)} \qquad (S2)$$

Where p and q are the fractions of oxygen and sulfur atoms involved in the π-bond structure of the molecule, respectively. In this study, $p = q = 0.5$ was used for the compounds detected in ESI^-, and $p = q = 1$ was selected for ESI^+ (Kourtchev et al., 2016; Tong et al., 2016) because ESI^- is more sensitive to compounds containing carboxylic groups, and compounds with a large diversity of functional groups can possibly be detected in ESI^+. Compounds with Xc < 2.5 were considered to be non-aromatics, with Xc⩾2.5 indicating aromatics and Xc⩾2.71 being considered as condensed aromatics (Yassine et al., 2014).

The O/C, H/C and DBE of a sample was calculated over all identified molecules based on their relative abundances, as follows:

$$\mathrm{O/C} = \sum(\mathrm{Iin} * \mathrm{O/C_i}) \,/\, \sum \mathrm{Iin_i} \qquad (S3)$$

$$\mathrm{H/C} = \sum(\mathrm{Iin} * \mathrm{H/C_i}) \,/\, \sum \mathrm{Iin_i} \qquad (S4)$$

$$\mathrm{DBE} = \sum(\mathrm{Iin} * \mathrm{DBE_i}) \,/\, \sum \mathrm{Iin_i} \qquad (S5)$$

Where Ini represents the relative abundance of molecule i, O/C_i, H/C_i and DBE_i, represent the O/C, H/C and DBE of the molecule i, respectively.

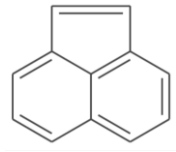
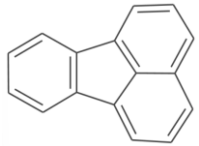
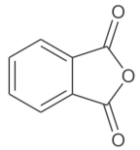
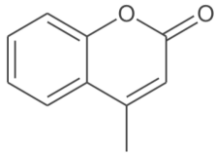
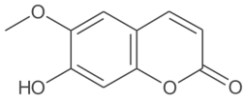
Table S1. Numbers of molecules, number fractions of different types of identified compounds to the total, and their average O/C, H/C and DBE values.

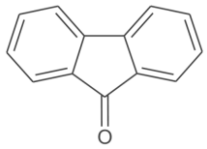
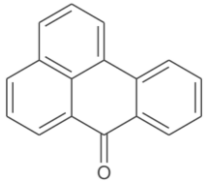
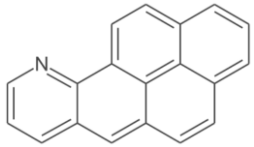
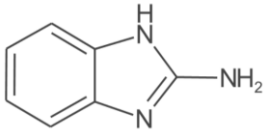
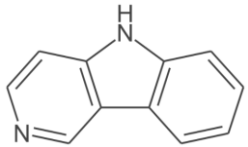
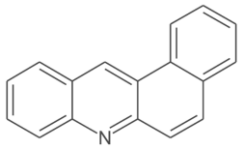
	Ion mode	Molecular types	Number of molecules	Number fractions	O/C	H/C	DBE
Summer Season (SS) Daytime	ESI^-	Total	466	100.00%	0.24	1.65	5.25
		CH	2	0.43%	0.00	1.74	5.99
		CHO	207	44.42%	0.28	1.46	6.61
		CHN	4	0.86%	0.00	1.54	4.75
		CHS	2	0.43%	0.00	0.98	7.06
		CHON	152	32.62%	0.28	1.82	4.33
		CHOS	22	4.72%	0.22	1.78	2.89
		CHNS	2	0.43%	0.00	2.01	1.81
		CHONS	75	16.09%	0.20	1.70	5.99
	ESI^+	Total	644	100.00%	0.17	1.80	4.55
		CH	5	0.78%	0.00	1.48	3.91
		CHO	186	28.88%	0.33	1.55	5.82
		CHN	43	6.68%	0.00	1.84	2.81
		CHS	2	0.31%	0.00	1.55	3.91
		CHON	325	50.47%	0.12	1.88	4.11
		CHOS	3	0.47%	0.21	1.24	5.40
		CHNS	3	0.47%	0.00	1.26	5.60
		CHONS	77	11.96%	0.31	1.73	7.23
	ESI^-	Total	518	100.00%	0.26	1.64	5.41
		CH	2	0.39%	0.00	1.80	5.17
		CHO	197	38.03%	0.24	1.40	6.97
		CHN	2	0.39%	0.00	2.00	1.56
		CHS	3	0.58%	0.00	0.87	7.07
		CHON	180	34.75%	0.30	1.82	4.11

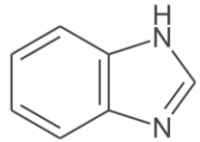
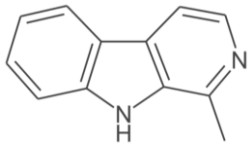
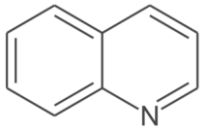
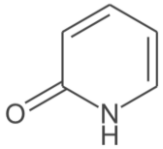
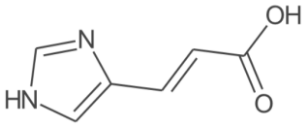
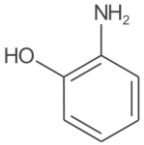
		CHOS	25	4.83%	0.20	1.71	3.21
		CHNS	2	0.39%	0.00	2.00	2.00
Summer Season (SS)		CHONS	107	20.66%	0.28	1.75	6.09
Nighttime		Total	735	100.00%	0.17	1.78	4.54
		CH	9	1.22%	0.00	1.73	2.84
		CHO	225	30.61%	0.30	1.55	5.25
		CHN	63	8.57%	0.00	1.82	2.81
	ESI^+	CHS	2	0.27%	0.00	2.04	0.71
		CHON	345	46.94%	0.12	1.87	4.13
		CHOS	7	0.95%	0.16	1.15	7.24
		CHNS	9	1.22%	0.00	1.66	3.39
		CHONS	75	10.20%	0.47	1.63	7.64
		Total	729	100.00%	0.29	1.55	4.92
		CH	2	0.27%	0.00	1.92	1.60
		CHO	284	38.96%	0.26	1.50	6.43
		CHN	3	0.41%	0.00	1.35	7.51
	ESI^-	CHS	0	0.00%	0.00	0.00	0.00
		CHON	241	33.06%	0.34	1.39	4.99
		CHOS	34	4.66%	0.20	1.81	2.53
		CHNS	2	0.27%	0.01	1.94	2.30
Cold Season (CS)		CHONS	163	22.36%	0.32	1.86	3.69
Daytime		Total	894	100.00%	0.20	1.80	4.23
		CH	14	1.57%	0.00	1.86	2.03
		CHO	217	24.27%	0.27	1.68	4.75
		CHN	63	7.05%	0.00	1.83	2.94
	ESI^+	CHS	3	0.34%	0.00	2.13	0.00
		CHON	484	54.14%	0.13	1.89	3.31
		CHOS	4	0.45%	0.13	0.94	9.61
		CHNS	5	0.56%	0.00	2.24	1.53

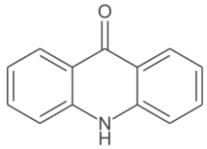
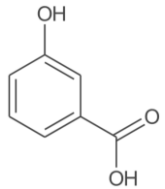
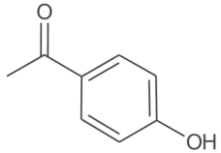
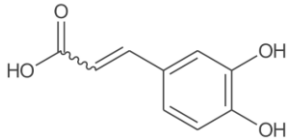
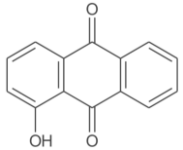
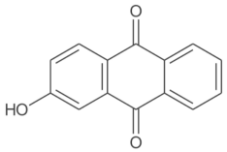
		CHONS	104	11.63%	0.47	1.64	7.91
		Total	865	100.00%	0.32	1.47	5.46
		CH	3	0.35%	0.00	1.91	2.03
		CHO	313	36.18%	0.26	1.49	6.26
		CHN	3	0.35%	0.00	1.03	9.92
	ESI^-	CHS	0	0.00%	0.00	0.00	0.00
		CHON	329	38.03%	0.38	1.37	5.28
		CHOS	34	3.93%	0.25	1.71	3.20
		CHNS	0	0.00%	0.00	0.00	0.00
Cold Season (CS)		CHONS	183	21.16%	0.28	1.72	5.20
Nighttime		Total	1065	100.00%	0.17	1.76	4.72
		CH	13	1.22%	0.00	1.79	2.52
		CHO	245	23.00%	0.26	1.56	5.72
		CHN	86	8.08%	0.00	1.72	3.56
	ESI^+	CHS	3	0.28%	0.00	2.10	0.35
		CHON	587	55.12%	0.14	1.83	4.37
		CHOS	6	0.56%	0.14	1.09	8.14
		CHNS	3	0.28%	0.00	2.44	0.44
		CHONS	122	11.46%	0.44	1.66	7.10

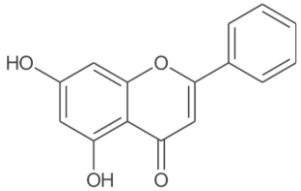
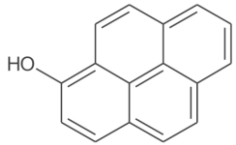
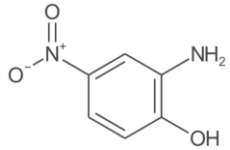
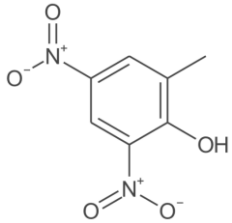
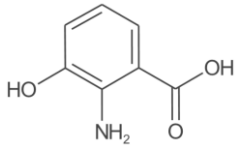
Table S2. Identified key light-absorbing compounds (BrC) assisted by the machine learning (a corresponding reference provides that this compound has been reported as a BrC species)

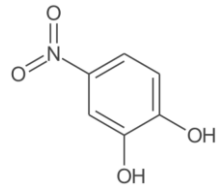
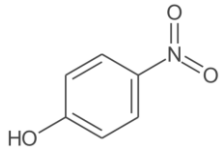
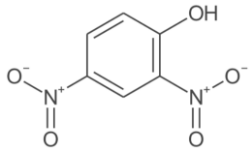
Retention time	***m/z***	**Proposed name**	**Proposed molecular formula**	**Proposed structure**	**Proposed molecular type**	**References**
			ESI$^+$			
14.967	152.0624	Acenapthylene	C12H8		PAH	(Aurell et al., 2015)
18.431	202.0788	Fluoranthene	C16H10		PAH	(Kuang et al., 2021)
19.812	149.0238	Phthalic anhydride	C8H4O3		O-Heterocyclic	(Chen et al., 2022)
20.67	161.0604	4-Methylcoumarin	C10H8O2		O-Heterocyclic	This work
8.908	193.05	Scopoletin	C10H8O4		O-Heterocyclic	(Zhang, 2018)

14.987	181.0648	9-Fluorenone	C13H8O		PAOH	(Kuang et al., 2023)
17.656	231.0816	Benzanthrone	C17H10O		PAOH	(Kuang et al., 2023)
18.533	254.0966	10-Azabenzo[a]pyrene	C19H11N		N-PAH	This work
0.756	134.0712	2-Aminobenzimidazole	C7H7N3		N-Heterocyclic	This work
7.49	169.0761	5-Carboline	C11H8N2		N-Heterocyclic	(Ma and Hays, 2008)
16.29	230.0968	Benz[c]acridine	C17H11N		N-PAH	This work

1.807	119.0604	Benzimidazole	C7H6N2		N-Heterocyclic	This work
9.223	183.0922	Harmane	C12H10N2		N-Heterocyclic	(Ma and Hays, 2008)
1.102	130.0651	Quinoline	C9H7N		Quinoline	This work
1.301	96.0443	2-Hydroxypyridine	C5H5NO		Pyridones	This work
0.863	121.0395	Urocanate	C6H6N2O2		Carboxylic Acids	This work
1.14	110.06	2-Aminophenol	C6H7NO		Aminophenol	(Al-Abadleh et al., 2022)

13.021	196.076	Acridone	C13H9NO		Acridone	(Negron-Encarnacion and Arce, 2007)
ESI$^-$						
4.113	137.0212	3-hydroxybenzoic acid	C7H6O3		Carboxylic acid	This work
7.566	135.0424	4-Hydroxyacetophenone	C8H8O2		Phenol	This work
7.567	179.0325	trans-Caffeic acid	C9H8O4		Carboxylic acid	(Le Person et al., 2013)
15.434	223.0372	1-Hydroxyanthraquinone	C14H8O3		Quinone	(Kuang et al., 2023)
13.817	223.0372	2-Hydroxyanthraquinone	C14H8O3		Quinone	This work

16.344	253.048	Chrysin	C15H10O4		Benzopyrans	This work
17.937	217.0632	1-Hydroxypyrene	C16H10O		Hydroxyl-PAHs	(Huang et al., 2022)
6.45	153.0276	2-Amino-4-nitrophenol	C6H6N2O3		Nitrophenol	This work
14.806	197.0176	2-Methyl-4,6-dinitrophenol	C7H6N2O5		Nitrophenol	(Li et al., 2020)
12.293	152.0322	3-Hydroxyanthranilic acid	C7H7NO3		Aminophenol	This work

6.556	154.0118	4-Nitrocatechol	C6H5NO4		Nitrophenol	(Li et al., 2020)
8.781	138.0173	4-Nitrophenol	C6H5NO3		Nitrophenol	(Li et al., 2020)
10.487	183.0017	2,4-Dinitrophenol	C6H4N2O5		Nitrophenol	(Li et al., 2020)

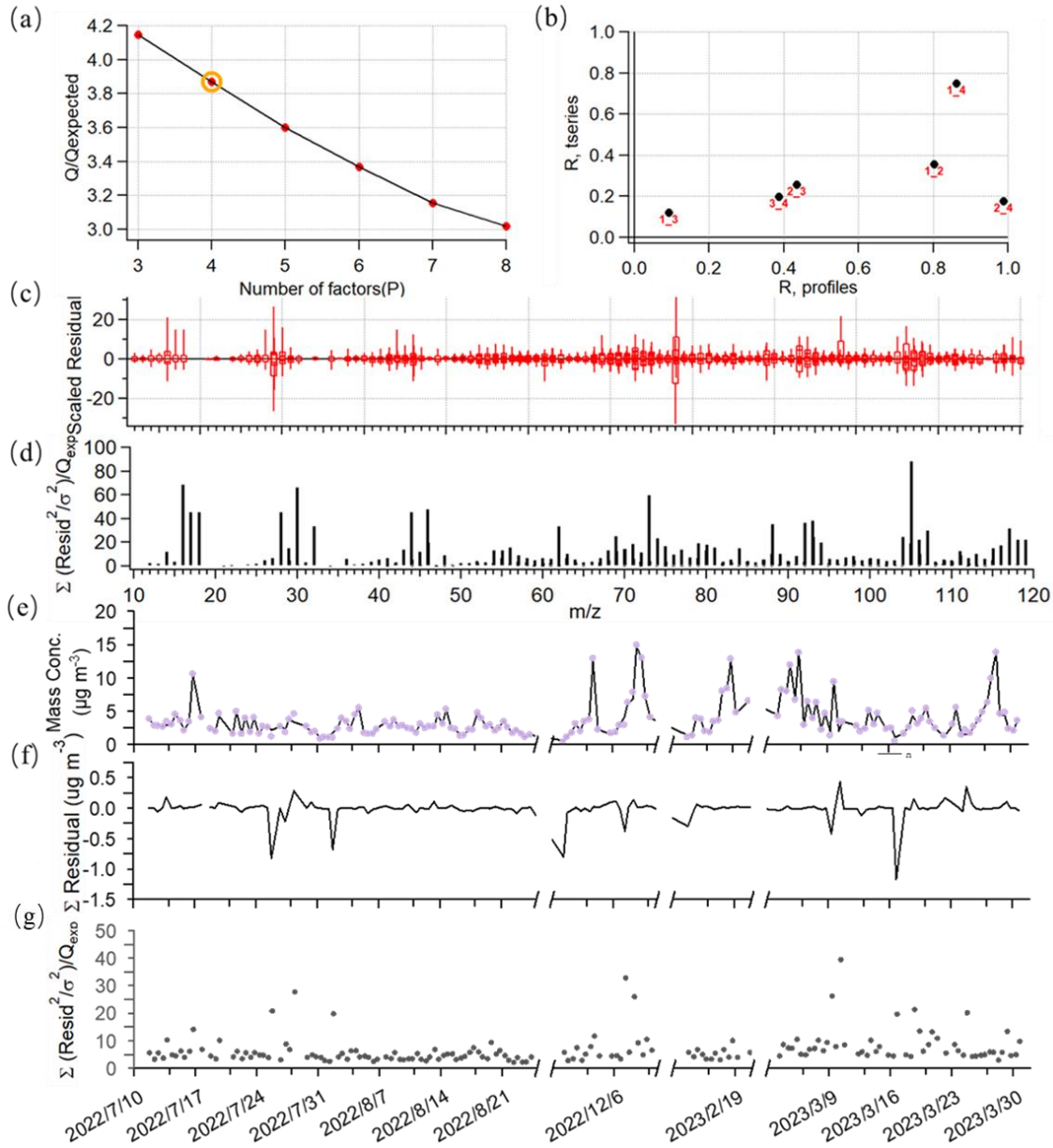


Figure S1. Summary of critical diagnostic plots of the PMF results for the 4-factor solution of WSOA: (a) *Q/Qexp* as a function of the number of factors (P from 3 to 8). For the best solution (4-factor); (b) cross-correlations of the time series and spectral profiles among the four factors; (c) the box and whiskers plot showing the distributions of scaled residuals for each *m/z*; (d) the *Q/Qexp* values for each *m/z*; (e) time series of the measured and the reconstructed WSOA mass loadings; (f) variations of the residuals of the fit; (g) the *Q/Qexp* values for each sample.

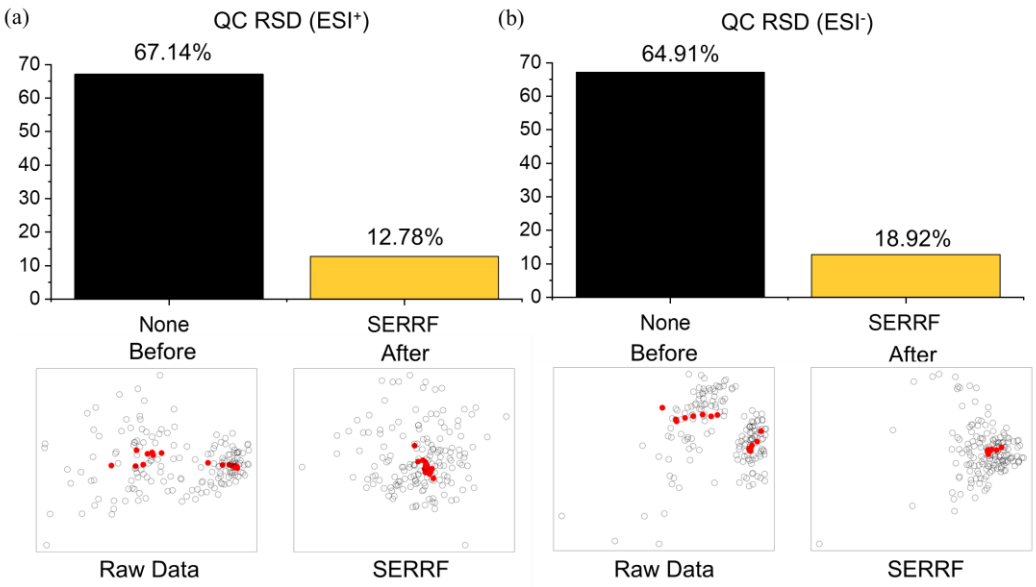


Figure S2. Relative standard deviations before and after the SERRF normalization (The gray and red markers represent real samples and QC samples, respectively). (a) ESI^+ mode, (b) ESI^- mode.

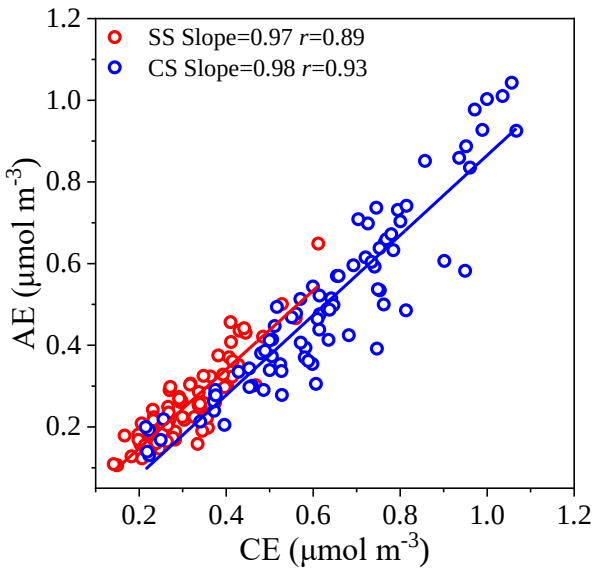


Figure S3. Scatter plot of the molar concentrations of cations versus anions (SS: Summer season; CS: Cold season).

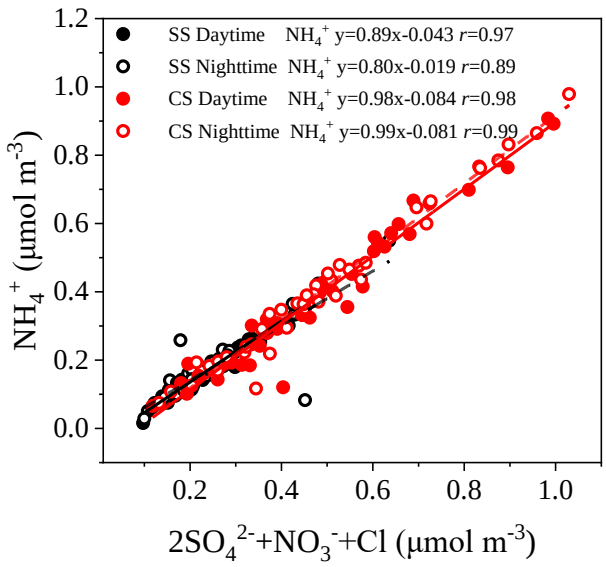


Figure S4. Scatter plot of the molar concentrations of ammonium versus sum of sulfate, nitrate and chloride (SS: Summer season; CS: Cold season).

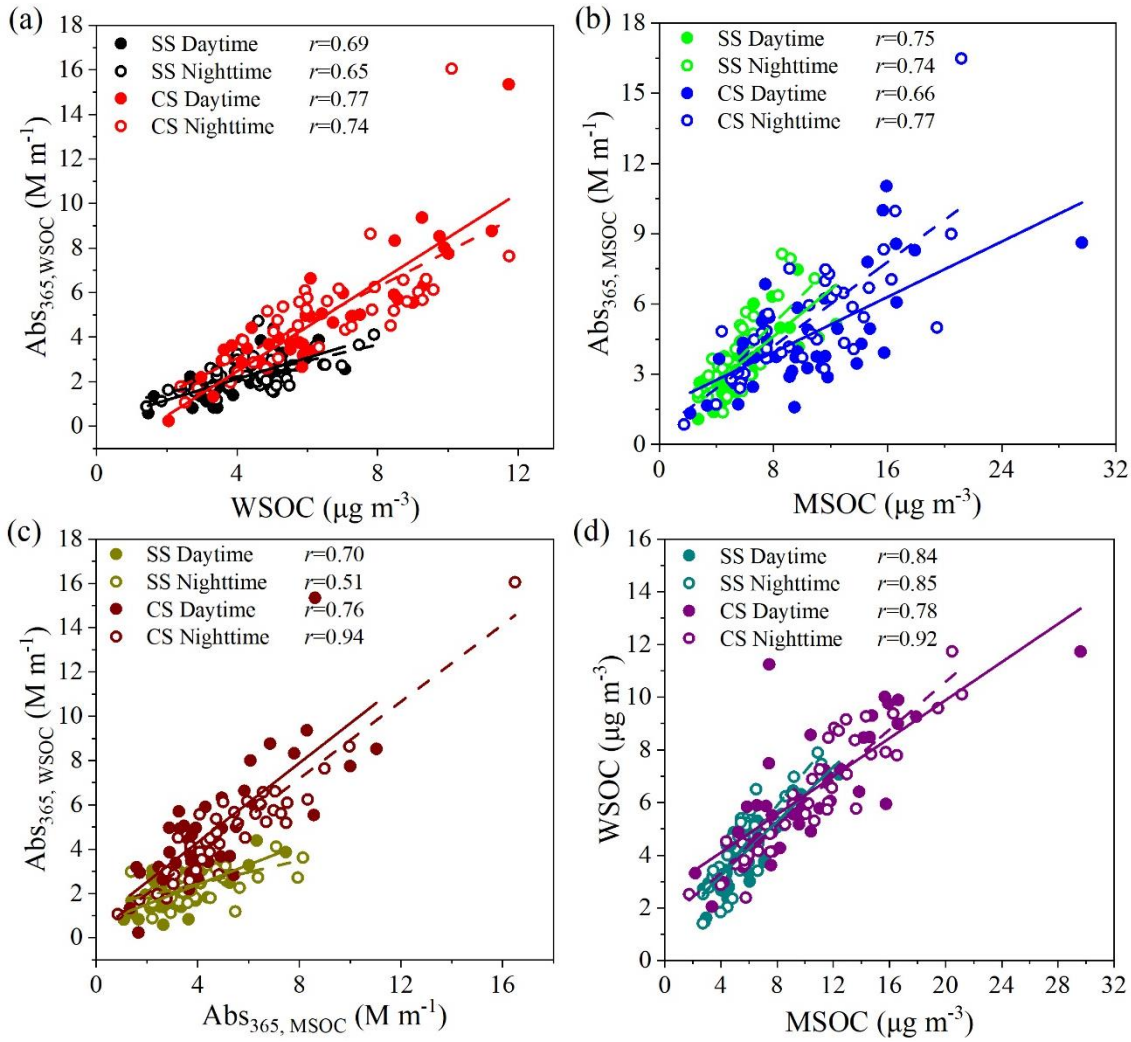


Figure S5. Scatter plots of: (a) light absorption at 365 nm of WSOC ($Abs_{365,\ WSOC}$) versus WSOC concentrations; (b) light absorption at 365 nm of MSOC ($Abs_{365,\ MSOC}$) versus MSOC concentrations; (c) WSOC versus MSOC, and (d) $Abs_{365,\ WSOC}$ versus $Abs_{365,\ MSOC}$.

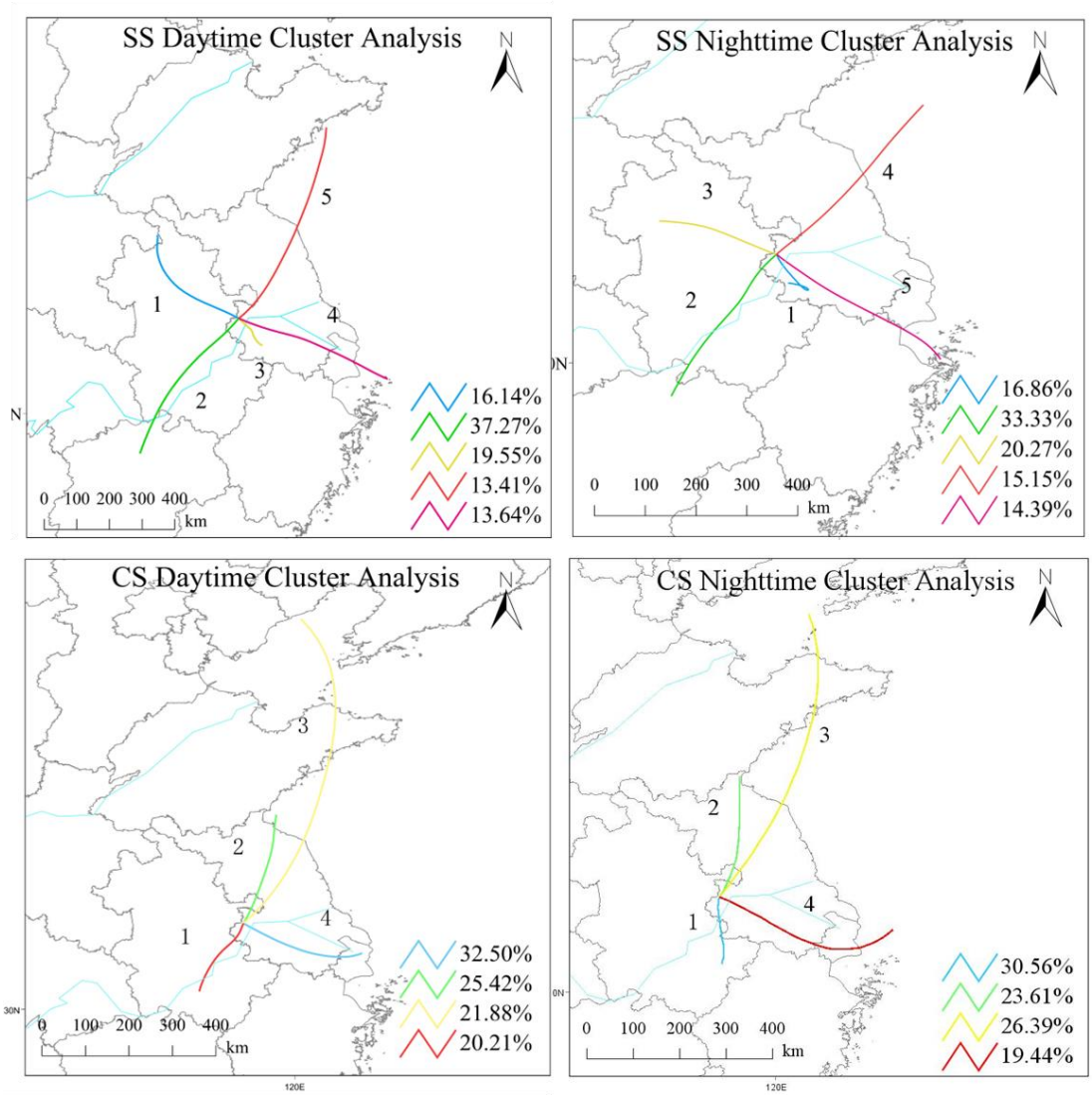


Figure S6. Clustered backward trajectories of different sampling periods.

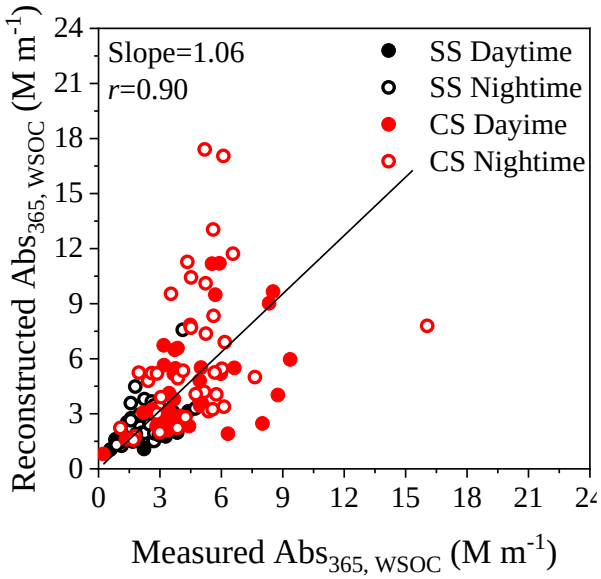


Figure S7. Scatter plot of the reconstructed Abs_{365} from the multilinear regression versus measured Abs_{365}.

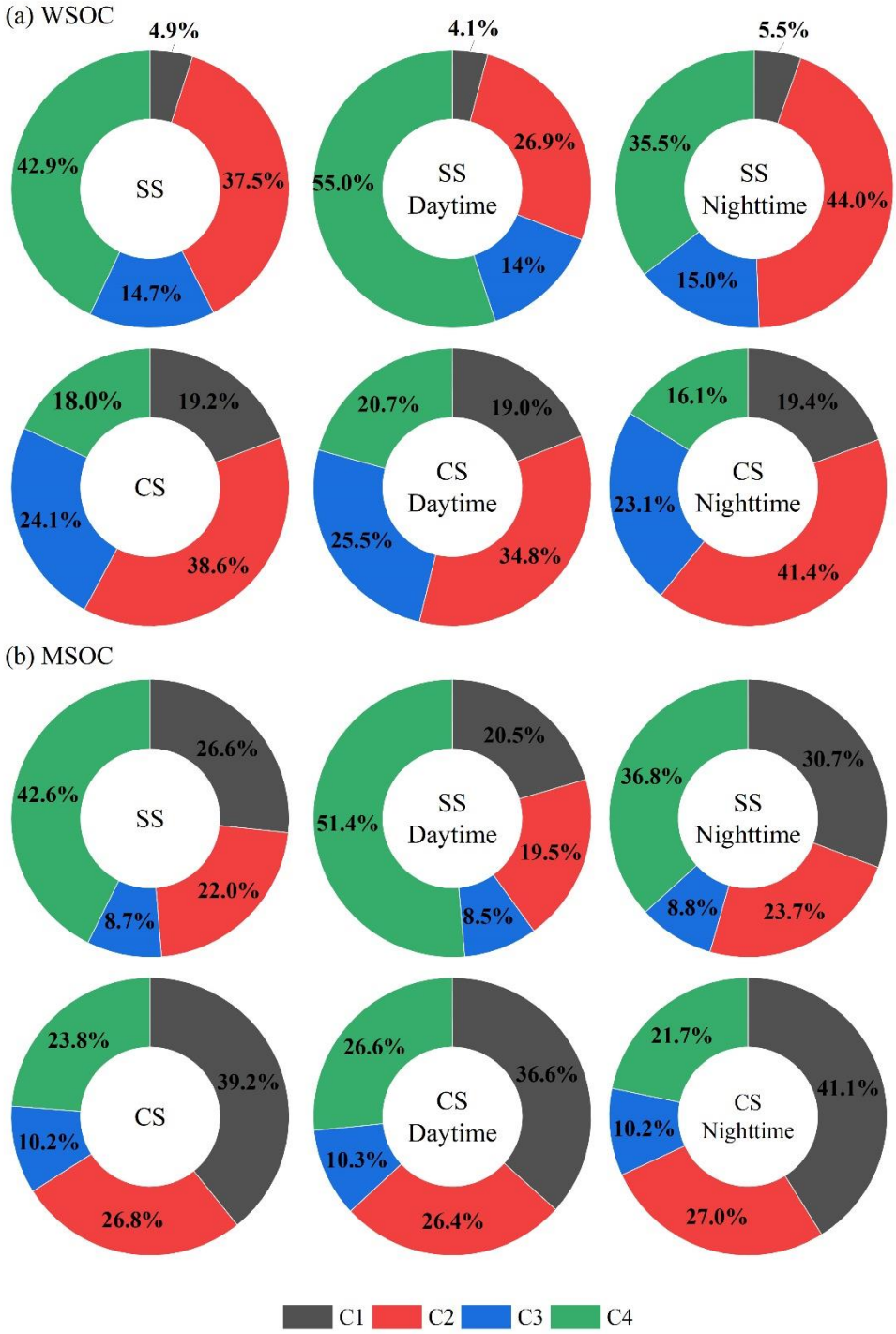


Figure S8. Average contributions of the PARAFAC-derived fluorescent components of (a) WSOC and (b) MSOC during different periods.

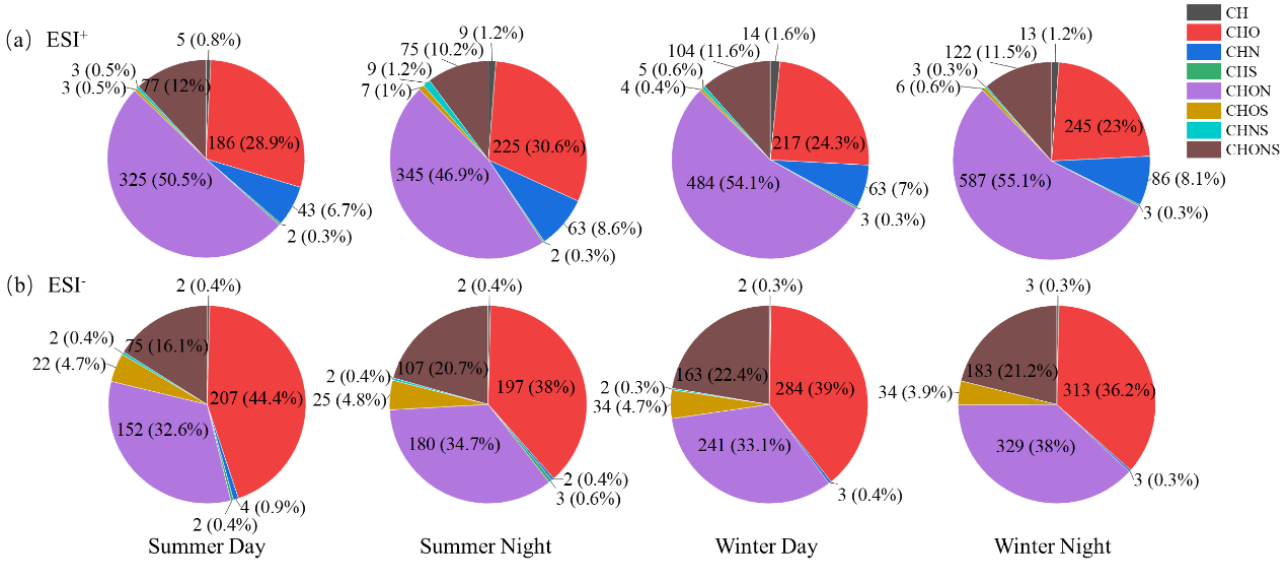


Figure S9. Numbers and number fractions of different types of compounds identified during different periods. (a) ESI^+ mode, (b) ESI^- mode.

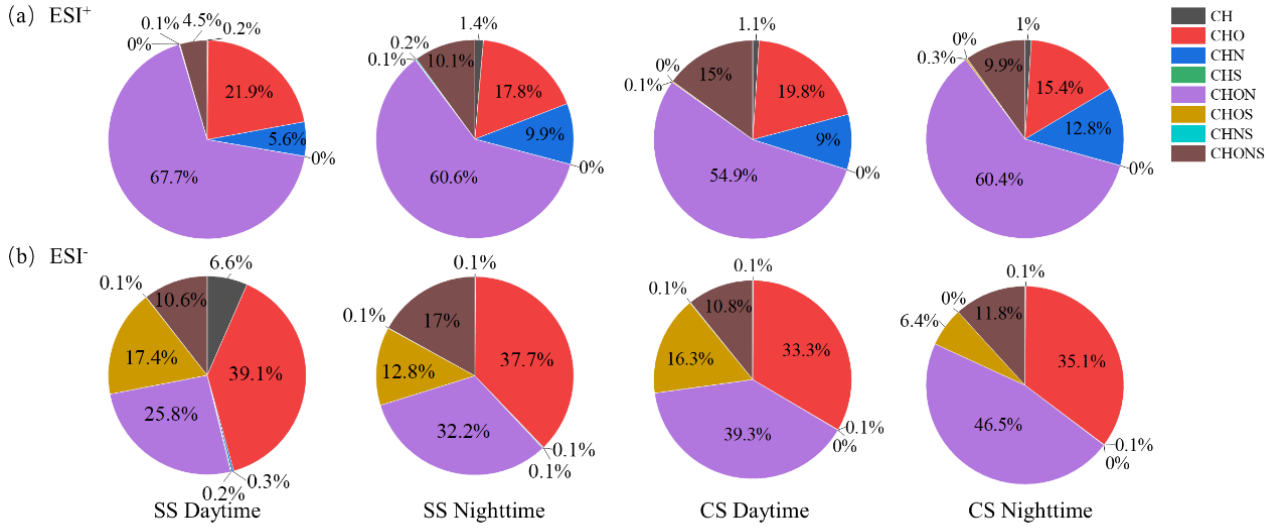


Figure S10. Contributions of the signal relative abundance of different types of compounds identified during different periods. (a) ESI^+ mode, and (b) ESI^- mode.

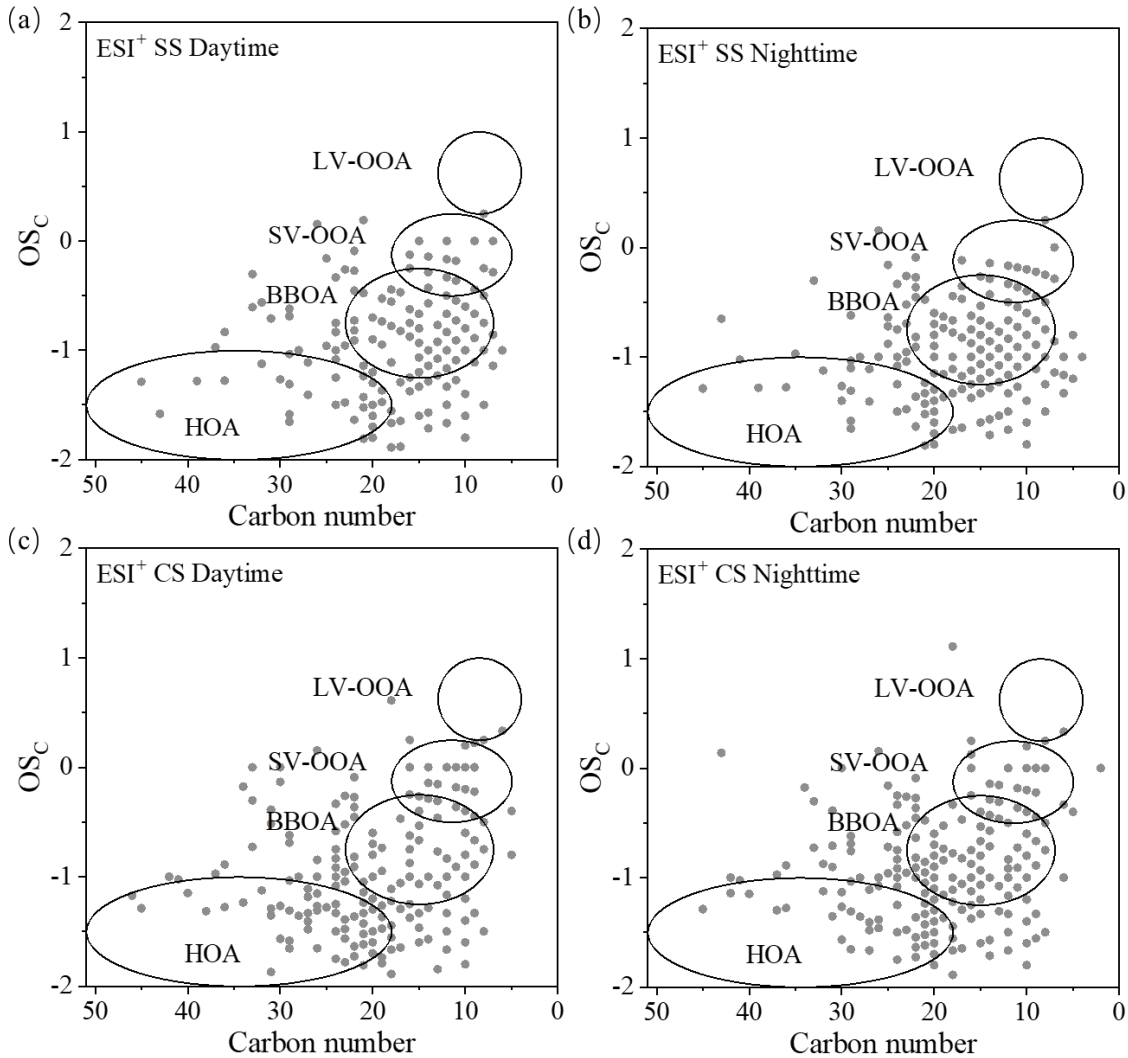


Figure S11. Scatter plots of the carbon oxidation state (OSc) versus carbon number for all CHO compounds in ESI^+ mode during diffeent periods. (a) SS daytime, (b) SS nighttime, (c) CS daytime, and (d) CS nighttime. The circled areas represent those from fossil fuel combustion hydrocarbon-like OA (HOA), biomass burning OA (BBOA), semi-volatile oxygenated OA (SV-OOA) and low-volatility oxygenated OA (LV-OOA) (Kroll et al., 2011).

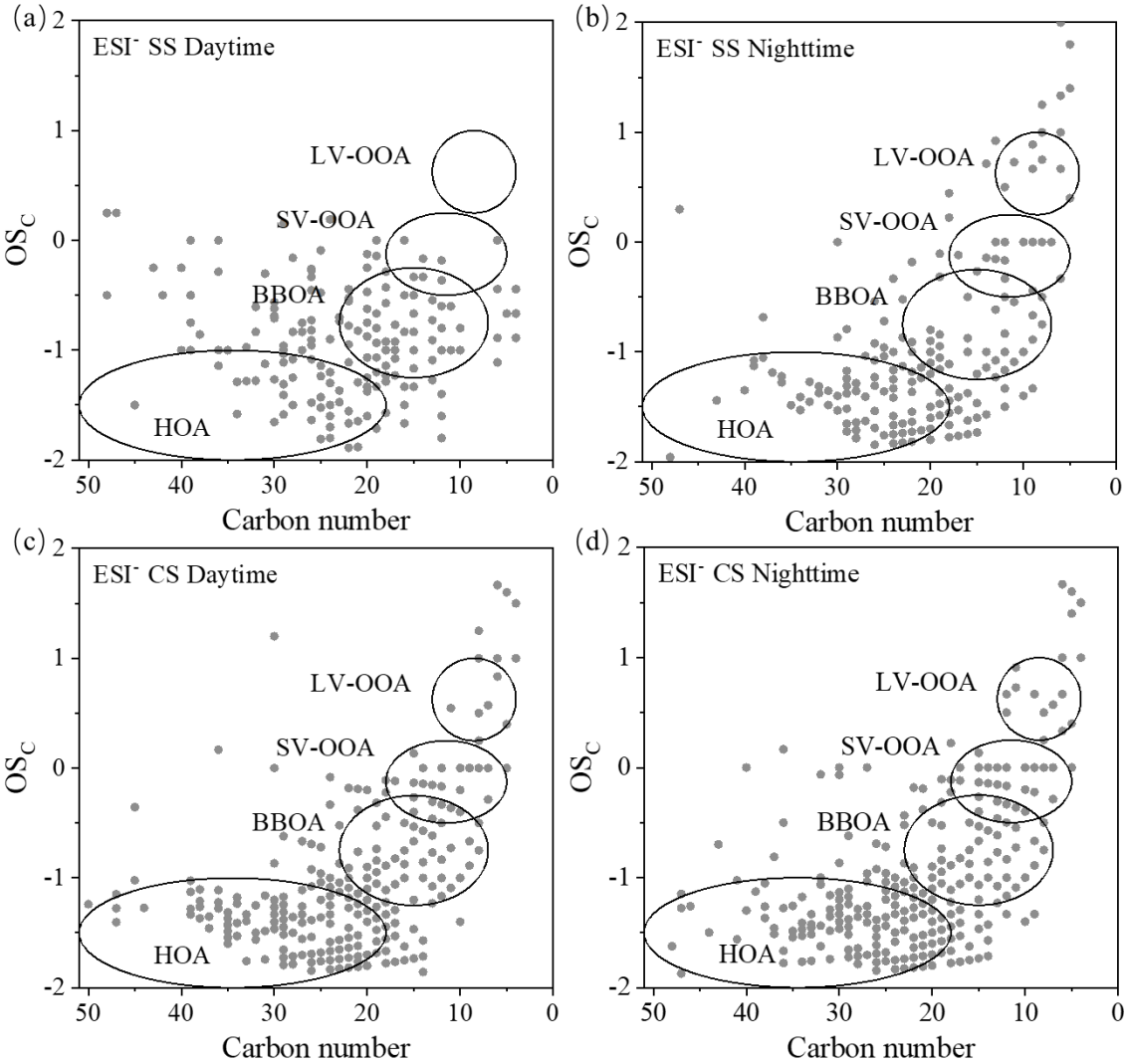


Figure S12. Scatter plots of the carbon oxidation state (OSc) versus carbon number for all CHO compounds in ESI^- mode during diffeent periods. (a) SS daytime, (b) SS nighttime, (c) CS daytime, and (d) CS nighttime. Meanings of the circled areas are same as those described in Figure S11.

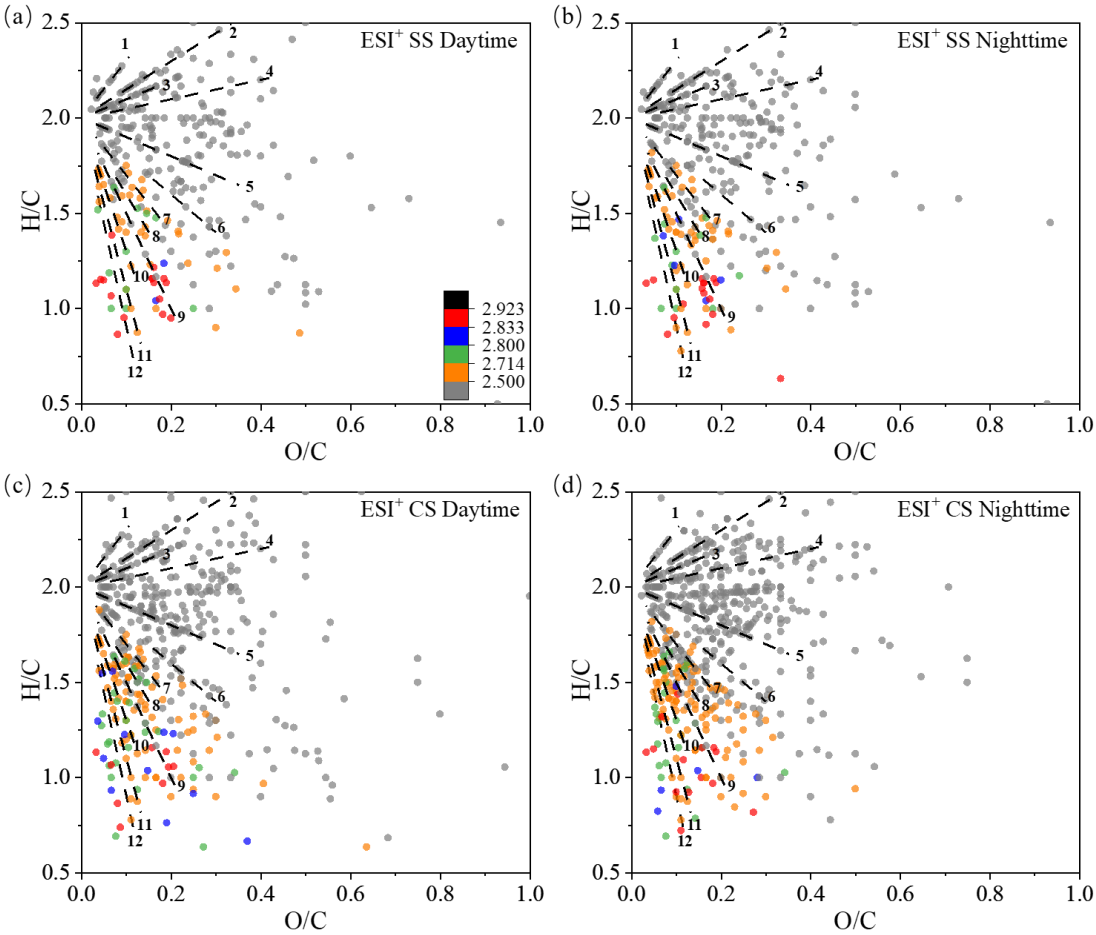


Figure S13. Van Krevelen diagram for CHON compounds detected in ESI^+ mode during different periods. (a) SS daytime, (b) SS nighttime, (c) CS daytime, and (d) CS nighttime. The markers with different colors represent aliphatic compounds (Xc < 2.50), aromatic benzene ring structures (2.50 ≤ Xc < 2.71), naphthalene ring structures (2.71 ≤ Xc < 2.80), anthracene ring structures (2.80 ≤ Xc < 2.83), and pyrene ring structures (2.83 ≤ Xc < 2.92), respectively (Mao et al., 2022); Different dash lines represent different series of compounds.

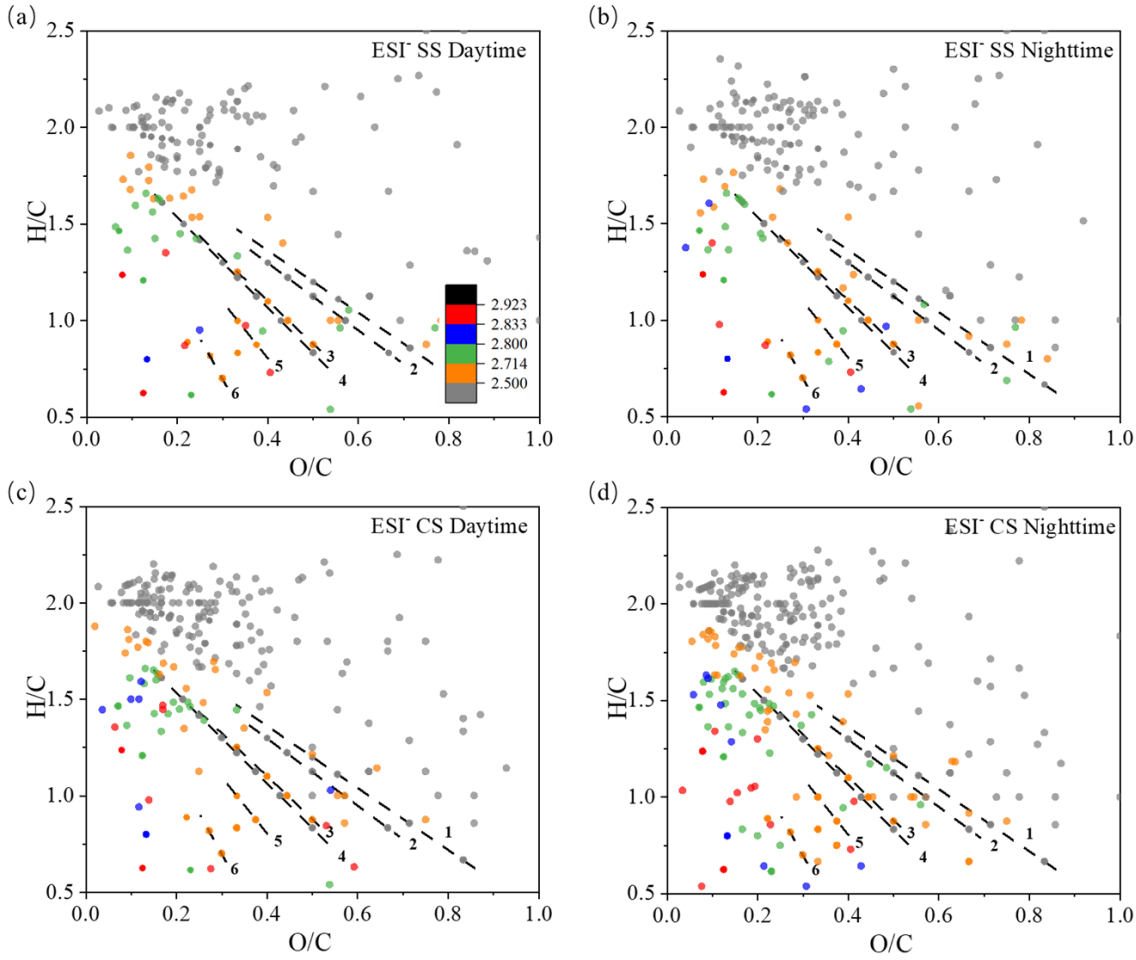


Figure S14. Van Krevelen diagram for CHON compounds detected in ESI⁻ mode during different periods. (a) SS daytime, (b) SS nighttime, (c) CS daytime, and (d) CS nighttime. Meanings of the clored values and dash lines are same as those described in Figure S13.

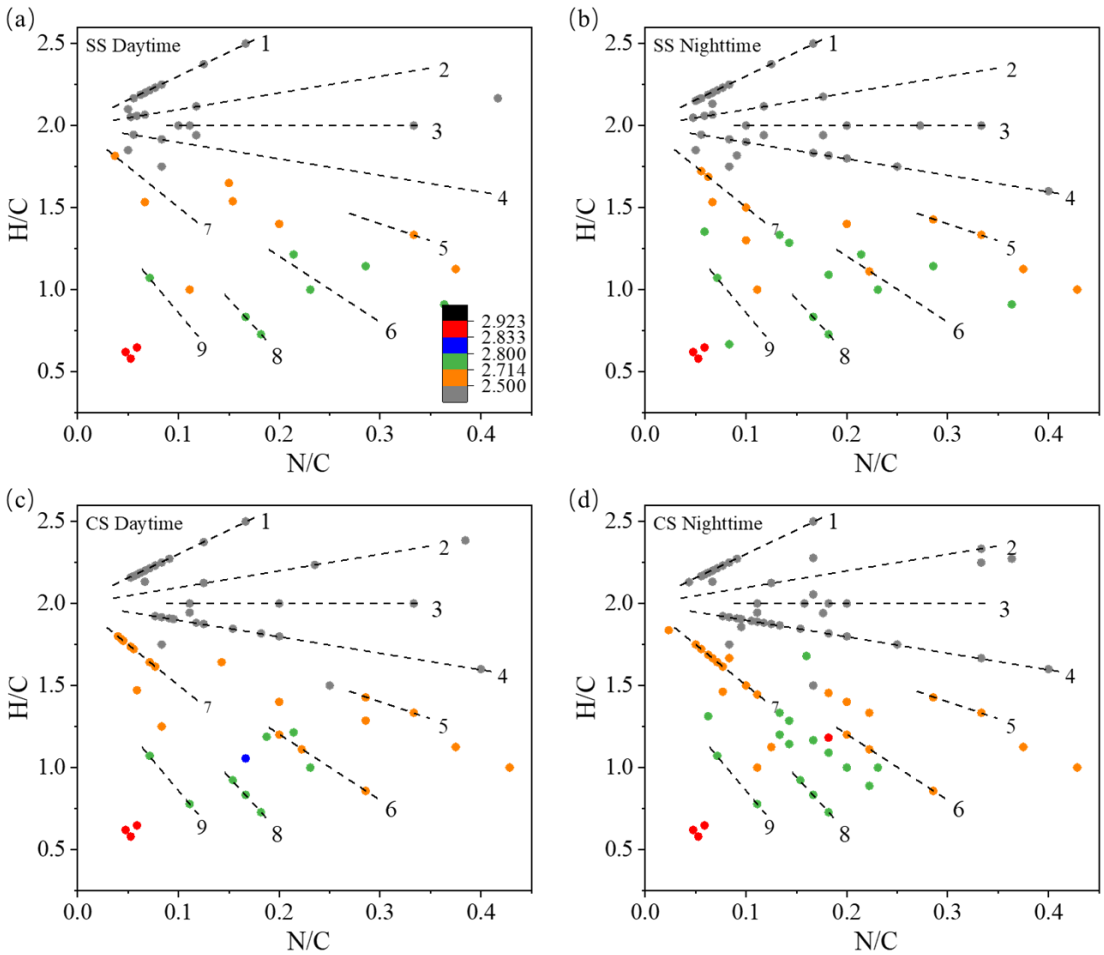


Figure S15. Van Krevelen diagram for CHN compounds detected in ESI^+ mode during different periods. (a) SS daytime, (b) SS nighttime, (c) CS daytime, and (d) CS nighttime. Meanings of the clored values and dash lines are same as those described in Figure S13.

References

Al-Abadleh, H. A., Motaghedi, F., Mohammed, W., Rana, M. S., Malek, K. A., Rastogi, D., Asa-Awuku, A. A., and Guzman, M. I.: Reactivity of aminophenols in forming nitrogen-containing brown carbon from iron-catalyzed reactions, Commun. Chem., 5, 112, https://doi.org/10.1038/s42004-022-00732-1, 2022.

Aurell, J., Gullett, B. K., and Tabor, D.: Emissions from southeastern U.S. Grasslands and pine savannas: Comparison of aerial and ground field measurements with laboratory burns, Atmos. Environ., 111, 170-178, https://doi.org/10.1016/j.atmosenv.2015.03.001, 2015.

Chen, K. P., Raeofy, N., Lum, M., Mayorga, R., Woods, M., Bahreini, R., Zhang, H. F., and Lin, Y. H.: Solvent effects on chemical composition and optical properties of extracted secondary brown carbon constituents, Aerosol Sci. Technol., 56, 917-930, https://doi.org/10.1080/02786826.2022.2100734, 2022.

Huang, R. J., Yang, L., Shen, J. C., Yuan, W., Gong, Y. Q., Ni, H. Y., Duan, J., Yan, J., Huang, H. B., You, Q. H., and Li, Y. J.: Chromophoric fingerprinting of brown carbon from residential biomass burning, Environ. Sci. Technol. Lett., 9, 102-111, https://doi.org/10.1021/acs.estlett.1c00837, 2022.

Kourtchev, I., Godoi, R. H. M., Connors, S., Levine, J. G., Archibald, A. T., Godoi, A. F. L., Paralovo, S. L., Barbosa, C. G. G., Souza, R. A. F., Manzi, A. O., Seco, R., Sjostedt, S., Park, J. H., Guenther, A., Kim, S., Smith, J., Martin, S. T., and Kalberer, M.: Molecular composition of organic aerosols in central Amazonia: an ultra-high-resolution mass spectrometry study, Atmos. Chem. Phys., 16, 11899-11913, https://doi.org/10.5194/acp-16-11899-2016, 2016.

Kroll, J. H., Donahue, N. M., Jimenez, J. L., Kessler, S. H., Canagaratna, M. R., Wilson, K. R., Altieri, K. E., Mazzoleni, L. R., Wozniak, A. S., Bluhm, H., Mysak, E. R., Smith, J. D., Kolb, C. E., and Worsnop, D. R.: Carbon oxidation state as a metric for describing the chemistry of atmospheric organic aerosol, Nat. Chem., 3, 133-139, https://doi.org/10.1038/nchem.948, 2011.

Kuang, Y., Shang, J., and Chen, Q. C.: Effect of ozone aging on light absorption and fluorescence of brown carbon in soot particles: The important role of polycyclic aromatic hydrocarbons, J. Hazard. Mater., 413, 125406, https://doi.org/10.1016/j.jhazmat.2021.125406, 2021.

Kuang, Y., Shang, J., Sheng, M. S., Shi, X. D., Zhu, J. L., and Qiu, X. H.: Molecular Composition of Beijing PM(2.5) Brown Carbon Revealed by an Untargeted Approach Based on Gas Chromatography and Time-of-Flight Mass Spectrometry, Environ. Sci. Technol., 57, 909-919, https://doi.org/10.1021/acs.est.2c05918, 2023.

Le Person, A., Lacoste, A., and Cornard, J.: Photo-degradation of trans-caffeic acid in aqueous solution and influence of complexation by metal ions, J. Photochem. Photobiol., A, 265, 10-19, https://doi.org/10.1016/j.jphotochem.2013.05.004, 2013.

Li, M., Wang, X. F., Lu, C. Y., Li, R., Zhang, J., Dong, S. W., Yang, L. X., Xue, L., Chen, J. M., and Wang, W. X.: Nitrated phenols and the phenolic precursors in the atmosphere in urban Jinan, China, Sci. Total Environ., 714, 136760, https://doi.org/10.1016/j.scitotenv.2020.136760, 2020.

Ma, Y. L. and Hays, M. D.: Thermal extraction–two-dimensional gas chromatography–mass spectrometry with heart-cutting for nitrogen heterocyclics in biomass burning aerosols, J. Chromatogr. A, 1200, 228-234, https://doi.org/10.1016/j.chroma.2008.05.078, 2008.

Mao, J. F., Cheng, Y., Bai, Z., Zhang, W., Zhang, L. Y., Chen, H., Wang, L. N., Li, L., and Chen, J. M.: Molecular characterization of nitrogen-containing organic compounds in the winter North China Plain, Sci Total Environ, 838, 156189, https://10.1016/j.scitotenv.2022.156189, 2022.

Negron-Encarnacion, I. and Arce, R.: Light-induced transformations of aza-aromatic pollutants adsorbed on models of atmospheric particulate matter: Acridine and 9(10-H) acridone, Atmos. Environ., 41, 6771-6783, https://doi.org/10.1016/j.atmosenv.2007.04.062, 2007.

Tong, H. J., Kourtchev, I., Pant, P., Keyte, I. J., O'Connor, I. P., Wenger, J. C., Pope, F. D., Harrison, R. M., and Kalberer, M.: Molecular composition of organic aerosols at urban background and road tunnel sites using ultra-high resolution mass spectrometry, Faraday Discuss., 189, 51-68, https://doi.org/10.1039/C5FD00206K, 2016.

Yassine, M. M., Harir, M., Dabek-Zlotorzynska, E., and Schmitt-Kopplin, P.: Structural characterization of organic aerosol using Fourier transform ion cyclotron resonance mass spectrometry: Aromaticity equivalent approach, Rapid Commun. Mass Spectrom., 28, 2445-2454, https://doi.org/doi.org/10.1002/rcm.7038, 2014.

Zhang, S. Y.: Detection of scopolamine in nori fruit juice and its traditional chinese medicine compound health products by high performance liquid chromatography, World Latest Medicine Information, 18, 91-92, https://doi.org/10.19613/j.cnki.1671-3141.2018.82.066, 2018.