Supporting information

Characterization of chromophoric water-soluble organic matter in urban, forest and marine aerosols by HR-ToF-AMS analysis and excitation–emission matrix spectroscopy

*Qingcai Chen[1], Yuzo Miyazaki [2], Kimitaka Kawamura[2,3], Kiyoshi Matsumoto[4], Sean Coburn[5,6,7], Rainer Volkamer[5,6], Yoko Iwamoto[8,9], Sara Kagami[1], Yange Deng[1], Shuhei Ogawa[1], Sathiyamurthi Ramasamy[10], Shungo Kato[11], Akira Ida[10,12], Yoshizumi Kajii[10,13], Michihiro Mochida[*1]*

[1] Graduate School of Environmental Studies, Nagoya University, Nagoya 464-8601, Japan
[2] Institute of Low Temperature Science, Hokkaido University, Sapporo 464-8601, Japan
[3] Now at Chubu Institute for Advanced Studies, Chubu University, Kasugai 487-8501, Japan
[4] Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, Kofu 400-8510, Japan
[5] Department of Chemistry & Biochemistry, University of Colorado, Boulder, CO 80309-0215, USA
[6] Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, CO 80309-0215, USA
[7] Now at Department of Mechanical Engineering, University of Colorado Boulder, CO 80309-0215, USA
[8] Institute for Advanced Research, Nagoya University, Nagoya 464-8601, Japan
[9] Now at Faculty of Science Division I, Tokyo University of Science, Tokyo 162-8601, Japan
[10] Graduate School of Global Environmental Studies, Kyoto University, Kyoto 606-8501, Japan
[11] Department of Applied Chemistry, Tokyo Metropolitan University, Tokyo 192-0397, Japan
[12] Now at Seraku Co. Ltd., Tokyo 160-0023, Japan
[13] Center for Regional Environmental Research, National Institute for Environmental Studies, Tsukuba 305-8506, Japan

**Corresponding author e-mail: mochida.michihiro@g.mbox.nagoya-u.ac.jp*
Phone/fax: 052-788-6157
Address: Department of Earth and Environmental Sciences, Graduate School of Environmental Studies, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan

Contents:

S1. Definitions of acronyms

WSOM: water-soluble organic matter

WISOM: water-insoluble organic matter

HULIS: humic-like substances

PLOM: protein-like organic matter

BSOA: biogenic secondary organic aerosol

BBOA: biomass burning organic aerosol

HOOA: highly oxygenated organic aerosol

LOOA: less oxygenated organic aerosol

EEM: excitation-emission matrix

MAE: mass absorption efficiency

NFV: fluorescence volume normalized by organic carbon concentration

NMF: non-negative matrix factorization

ALS: alternating least-squares algorithm

MULT: multiplicative update algorithm

S2. Sampling locations

Aerosol samples were collected at an urban site in Nagoya, Japan (Nagoya University; 35.15°N, 136.97°E), a forest site in Kii Peninsula, Japan (Wakayama Forest Research Station (WFRS), Kyoto University; 34.06°N, 135.52°E) and over the eastern equatorial Pacific during the TORERO/KA-12-01 cruise. The characteristics of the sampling locations are as follows. Nagoya is the central city of the Chukyo metropolitan area (population of the area: 9 million). The sampling site

in Nagoya is surrounded by residential areas, and industrial areas are located to the south of this site, near the coastlines facing Ise Bay and Mikawa Bay (http://nrb-www.mlit.go.jp/webmapc/mapmain.html). The Pacific Ocean is to the south of the metropolitan area. In Kii Peninsula, where the forest aerosols were collected, coniferous trees of *Cryptomeria japonica* and *Chamaecyparis obtusa* are distributed broadly and broad-leaf trees such as *Quercus serrata* are also distributed.[1] During the atmospheric observation of volatile organic compounds (VOCs) from 28 July to 8 August 2014, the mean ± standard deviation of the concentrations of isoprene and the sum of twelve monoterpenes (α-pinene, camphene, β-pinene, myrcene, α-phellandrene, Δ-3-carene, α-terpinene, p-cymene, limonene, ocimene, γ-terpinene, and terpinolene) at the studied site were 779 ± 430 and 563 ± 196 pptv during the daytime (06:00-18:00) and 205 ± 250 and 699 ± 281 pptv during the nighttime (18:00-06:00), respectively.[2] The ratios of daily average concentrations of isoprene to those of the sum of monoterpenes were in the range of 0.6-1.2, which indicated that their daily average concentrations were similar regardless of the day. The nearest megacity of Osaka is located approximately 70 km north of the WFRS forest site, and the coastline facing the North Pacific is 60 km southeast of the sampling site. The ship track and the sampling locations during the TORERO/KA-12-01 cruise are presented in Figure S13. The sampling locations (from the start to the end of the collection of respective samples) were approximately distributed in the ranges of 15°N–10°S and 90°W–130°W.[3,4] The distances from the sampling locations over the Pacific to the American continent were up to ~3000 km.

S3. Analysis of WSOC, amino acids and BSOA tracers

Quantification of WSOC was performed using a TOC analyzer (model TOC-V_{CSH}, Shimadzu, Japan). A portion of a filter on which the aerosol was collected (7.07 cm^2 for the urban samples, 3.14

cm^2 for the forest samples collected in 2010, 14.13 cm^2 for the forest samples collected in 2014, and 19.63 cm^2 for the marine samples) was extracted with water (15 mL for the forest samples collected in 2010 and 20 mL for other samples) by ultrasonication. The extract solution was filtrated with a disc filter (Millex-GV, Millipore, Billerica, MA, USA), followed by analysis using a TOC analyzer.

Amino acids both in free and combined forms in water extracts from the urban aerosol samples were quantified using the pre-column derivatization HPLC method described by Koop et al.,[5] with some modification. Briefly, aerosol components on a filter cut of 13.2 cm^2 was extracted with 3 ml of water by ultrasonication, and the extract solution was filtered with a PTFE filter (Dismic 13-HP, Toyo Roshi Kaisha Ltd.). Free form amino acids in the 200 μL of water extract were reacted with 100 μL of 0.1 M phenylisothiocyanate and 100 μL of 1 M triethylamine in acetonitrile at 40 °C for 20 min to form phenylisothiocyanate derivatives and were then quantified using an HPLC system with an ODS column (Shim-pack VP-ODS, Shimadzu Corp.) and a UV-Vis detector (SPD-20A, Shimadzu corp.). Combined amino acids in the water extract were hydrolyzed using 6 M hydrochloric acid at 110 °C for 24 h. The quantified free form amino acids were aspartic acid, glutamic acid, serine, glycine, histidine, arginine, threonine, alanine, proline, tyrosine, valine, methionine, cysteine, isoleucine, leucine, phenylalanine, lysine, asparagine, citrulline and tryptophan/ornithine. Those in combined form were aspartic acid, glutamic acid, serine, glycine, histidine, arginine, threonine, alanine, proline, tyrosine, valine, methionine, cysteine, isoleucine, leucine, phenylalanine and lysine.

Biogenic molecular tracers in the forest samples (2-methylthreitol, 2-methylerythritol, 2–methylglyceric acid, pinic acid, pinonic acid, and 3-hydroxyglutaric acid in the samples collected in 2010 and 2014, and 3-methyl-1,2,3-butanetricarboxylic acid(3-MBTCA) in the samples collected

in 2014) were determined using capillary GC (Model 6890 for 2010 samples and model 7890A for 2014 samples, Agilent) coupled to a mass-selective detector (Model 5973 for 2010 samples and model 5975C for 2014 samples, Agilent) in the manner analogous to that described by Fu et al. and Mochizuki et al..[6,7] Briefly, aerosol components on a portion of a filter (3.5-5 cm^2 and 19.63 cm^2 for the samples collected in 2010 and 2014, respectively) were extracted with dichloromethane/methanol, and the extract was then reacted with 50 μl of N,O-bis-(trimethylsilyl) trifluoroacetamide with 1 % trimethylsilyl chloride and 10 μL of pyridine at 70 °C for 3 h to form trimethylsilyl (TMS) esters or TMS ethers. The TMS derivatives were analyzed using GC-MS, and the tracer compounds were quantified.

Table S1. Filter samples used and the extraction conditions for the EEM analysis and HR-AMS analysis.

Sampling location	Sampling period	Samples			Blank filters	Extraction	
		Number	ID	Sampling time (h)		Filter (cm^2)	Water (g)
Urban (Nagoya)	11 August 2013–14 September 2013	17	1–17 (Summer)	47–48	5	1 × 9.1	10
	4 March 2014–24 March 2014	10	18 –27 (Winter)	47–48	5	1 × 9.1	10*
Forest (WFRS, Kii Peninsula)	20 August 2010–30 August 2010	5	28–32 (WKYM2010QFF)	43–48	3	2 × 9.1	5
	28 July 2014–8 August 2014	11	33–43 (WKYM2014QFF)	23–24	3	3 × 9.1	5
Ocean (eastern equatorial Pacific)	1 February 2012–29 February 2012	20	44–63 (TRR)	11–30	3	2 × 9.1	4

* For the analysis of five extracts (sample ID: 18, 22, 24, 25 and 27) using EEM spectroscopy, the extracts were diluted 3 times with Mill-Q water.

Table S2. The Pearson's correlation coefficients (r) and the significance levels (p, two-sided t-test) from the correlation analysis between the relative intensity of the three and seven PARAFAC components (n = 63).

		7CM-C1	7CM-C3	7CM-C2	7CM-C4	7CM-C5	7CM-C6	7CM-C7
3CM-C1	*r*	0.858**	0.795**	-0.698**	-0.230	-0.413**	-0.521**	-0.461**
	p	<0.001	<0.001	<0.001	0.072	<0.001	<0.001	<0.001
3CM-C2	*r*	-0.901**	-0.790**	0.887**	0.526**	0.529**	0.246	0.018
	p	<0.001	<0.001	<0.001	<0.001	<0.001	0.054	0.889
3CM-C3	*r*	-0.254*	-0.317*	-0.070	-0.446**	-0.048	0.699**	0.980**
	p	0.046	0.012	0.588	<0.001	0.713	<0.001	<0.001

* $0.01 \leq p < 0.05$; ** $p < 0.01$.

Table S3. The Pearson's correlation coefficients (r) and the significance levels (p, two-sided t-test) from the correlation analysis between the relative intensity of the PARAFAC components and the NMF factors for the urban and forest samples (n = 43).

		7CM-C1	7CM-C3	7CM-C5	7CM-C2	7CM-C4	7CM-C6	7CM-C7
HOOA-1	*r*	0.386*	0.297	0.077	-0.489**	-0.342*	0.229	-0.478**
	p	0.013	0.060	0.634	0.001	0.029	0.149	0.002
HOOA-2	*r*	-0.703**	-0.022	0.153	0.631**	-0.003	-0.051	0.757**
	p	<0.001	0.892	0.340	<0.001	0.986	0.754	<0.001
LOOA-1	*r*	0.339*	0.209	0.041	-0.401**	-0.408**	0.297	-0.415**
	p	0.030	0.190	0.799	0.009	0.008	0.059	0.007
LOOA-2	*r*	-0.300	-0.313*	-0.107	0.420**	0.478**	-0.325*	0.395*
	p	0.056	0.047	0.504	0.006	0.002	0.038	0.011

* $0.01 \leq p < 0.05$; ** $p < 0.01$.

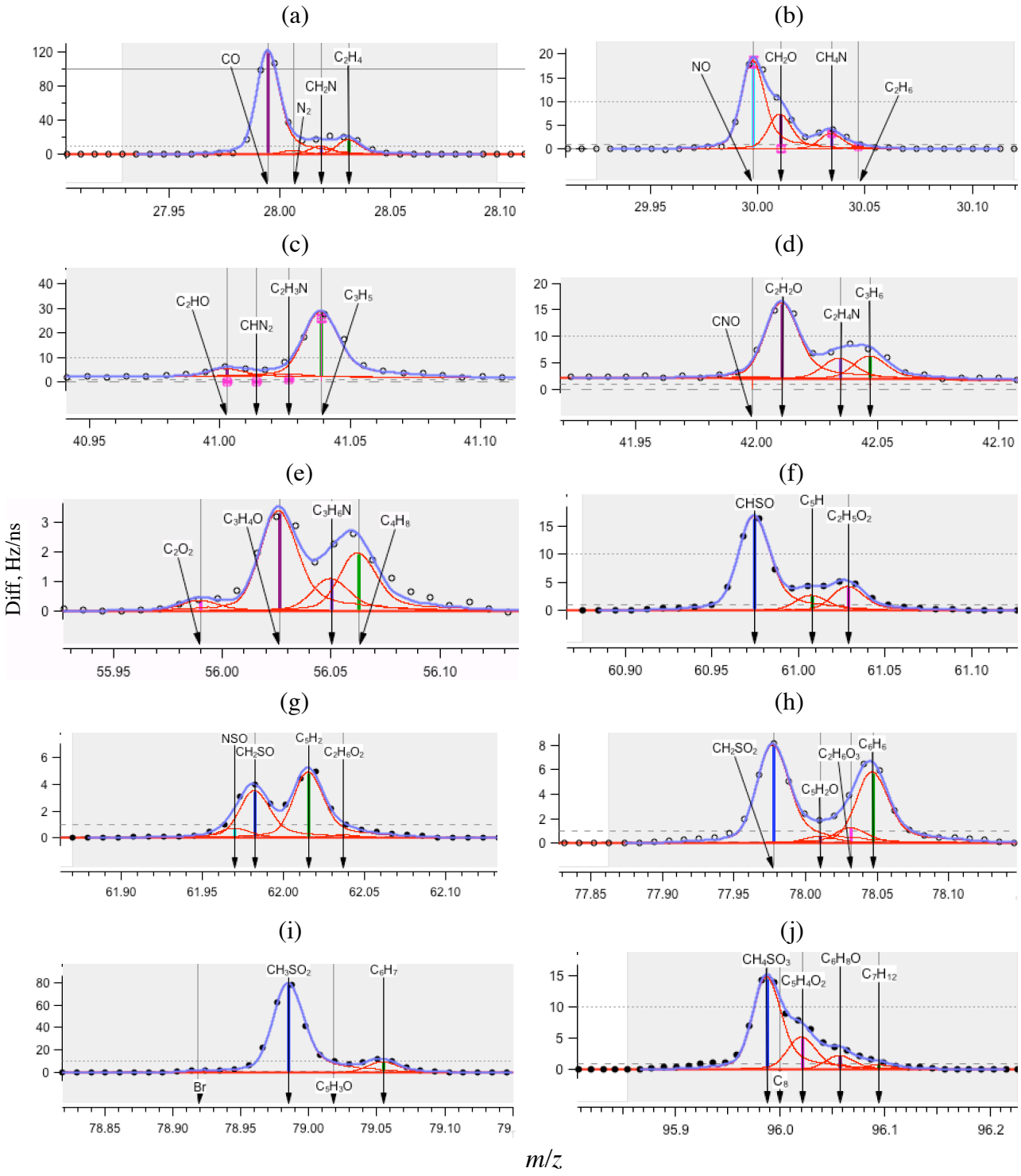

Figure S1. (a–e) Curve fittings for mass spectral peaks from ions, including N-containing ions, in the average HR-AMS spectra of water-soluble matter in all samples. (f–j) Curve fittings for mass spectral peaks from ions, including S-containing ions, in the average HR-AMS spectra of water-soluble matter in marine samples.

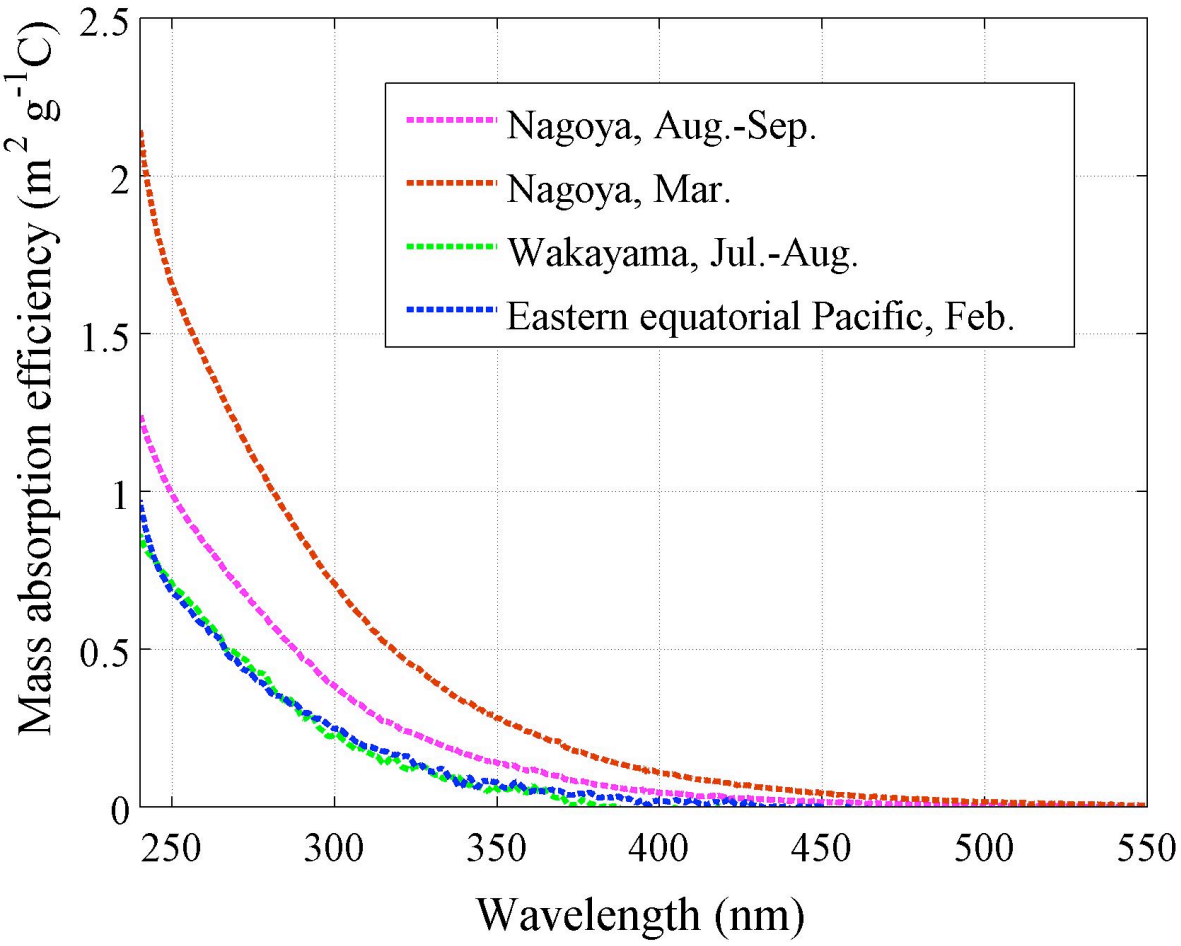

Figure S2. The average UV-visible absorbance spectra of the water-soluble extracts from urban, forest and marine aerosol samples.

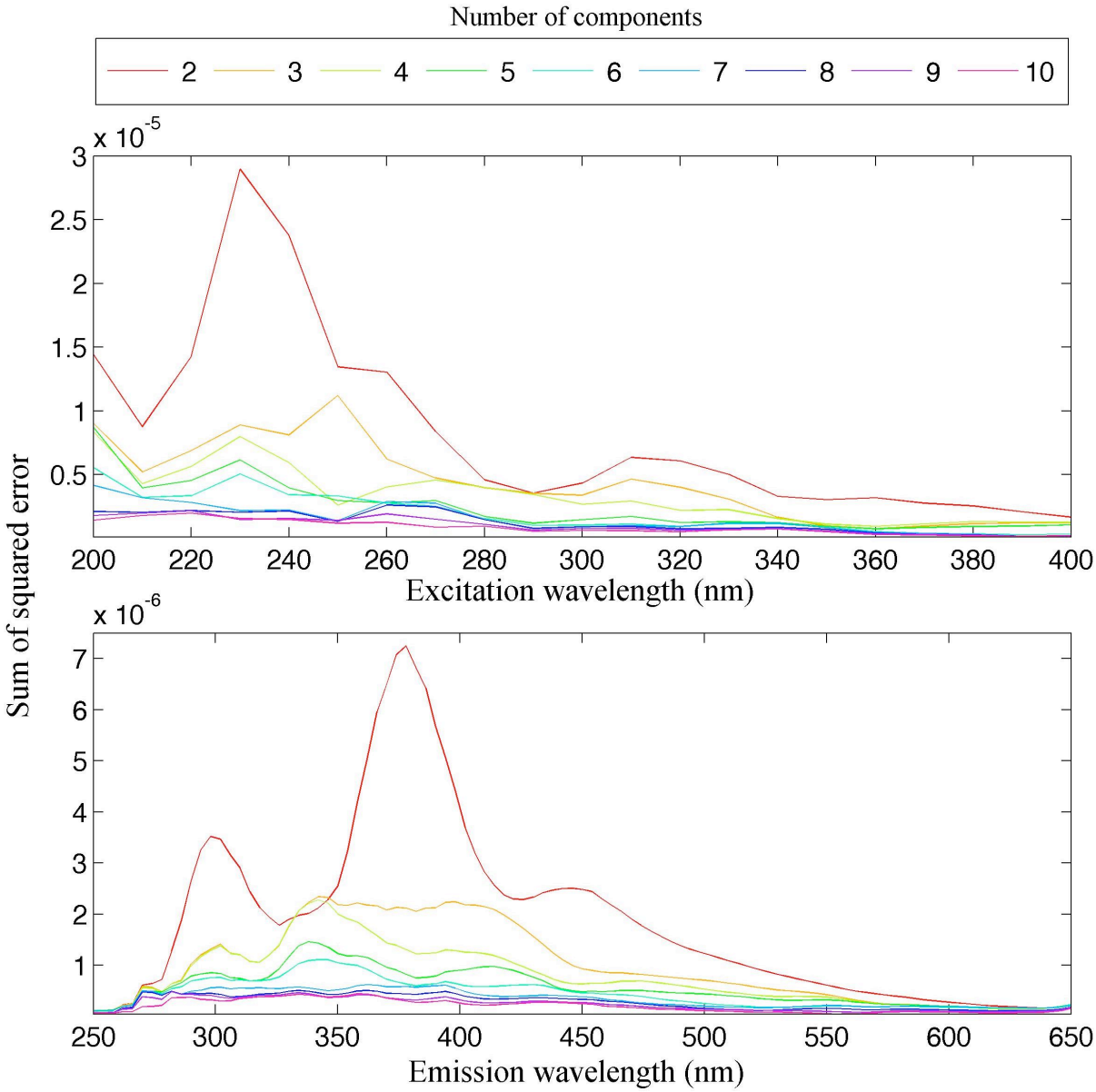

Figure S3. Residual analysis of two- to ten-component PARAFAC models for all EEMs.

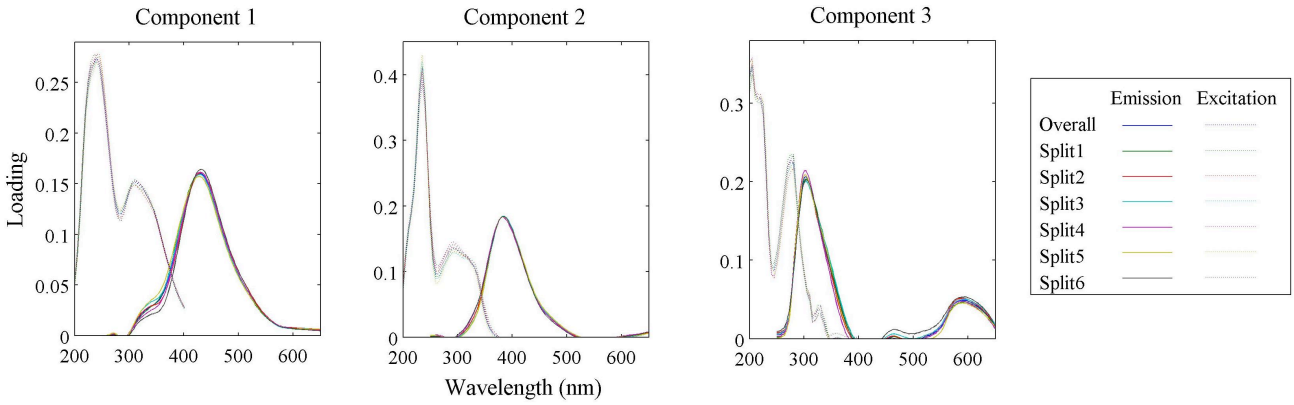

Figure S4. Split half analysis of three-component PARAFAC model with the split style of '$S_4C_6T_3$' for all EEMs.

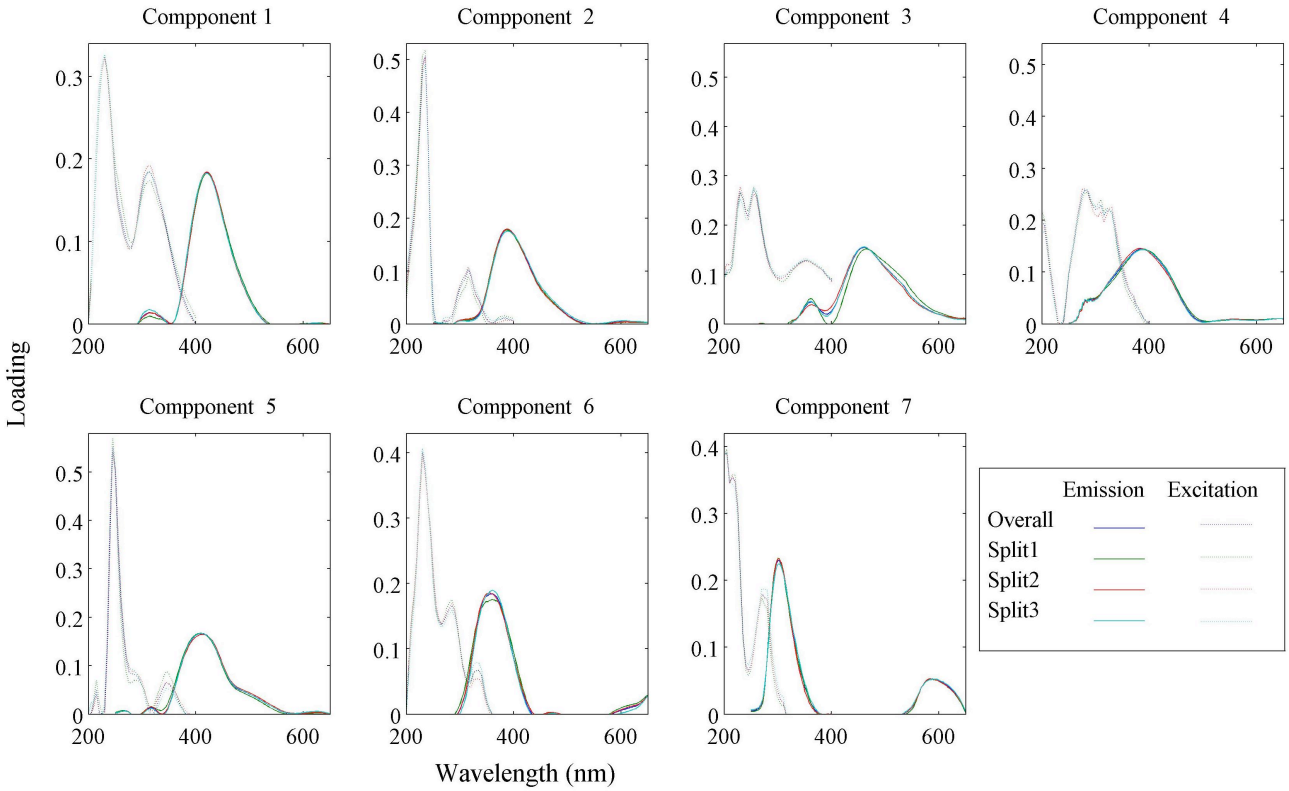

Figure S5. Split analyses of seven-component PARAFAC model with the split style of '$S_3C_3T_3$' (number of groups after splitting: 3 (A, B, and C), number of combinations: 3 (AB, AC, and BC), number of tests: 3 (AB vs. AC, AB vs. BC, and AC vs. BC)) for all EEMs.

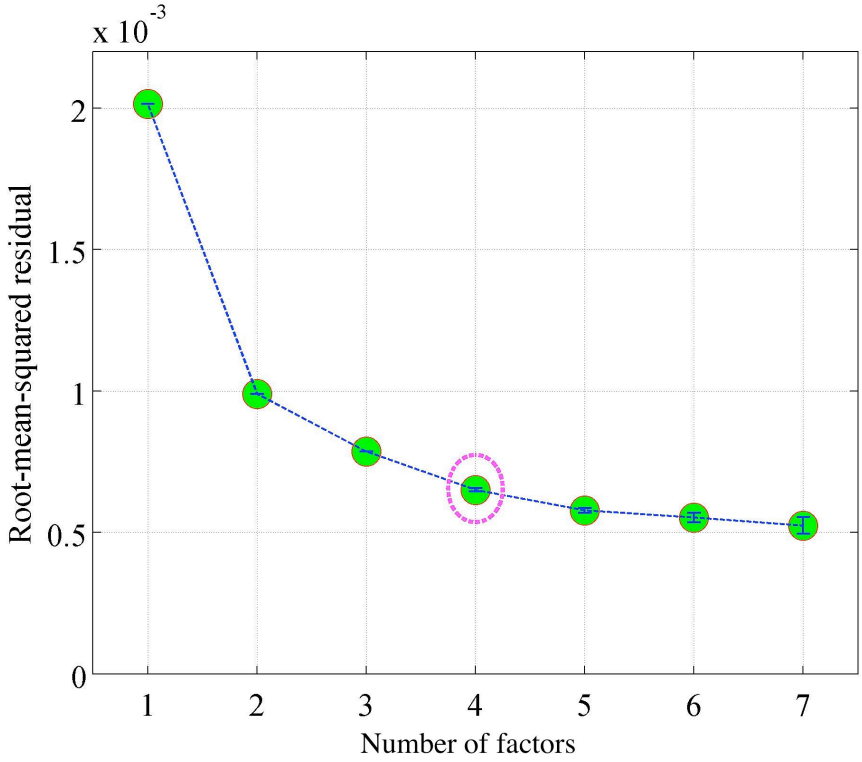

Figure S6. Residual analysis of one- to seven-factor NMF models for the HR-AMS data. The bars indicate the standard deviation.

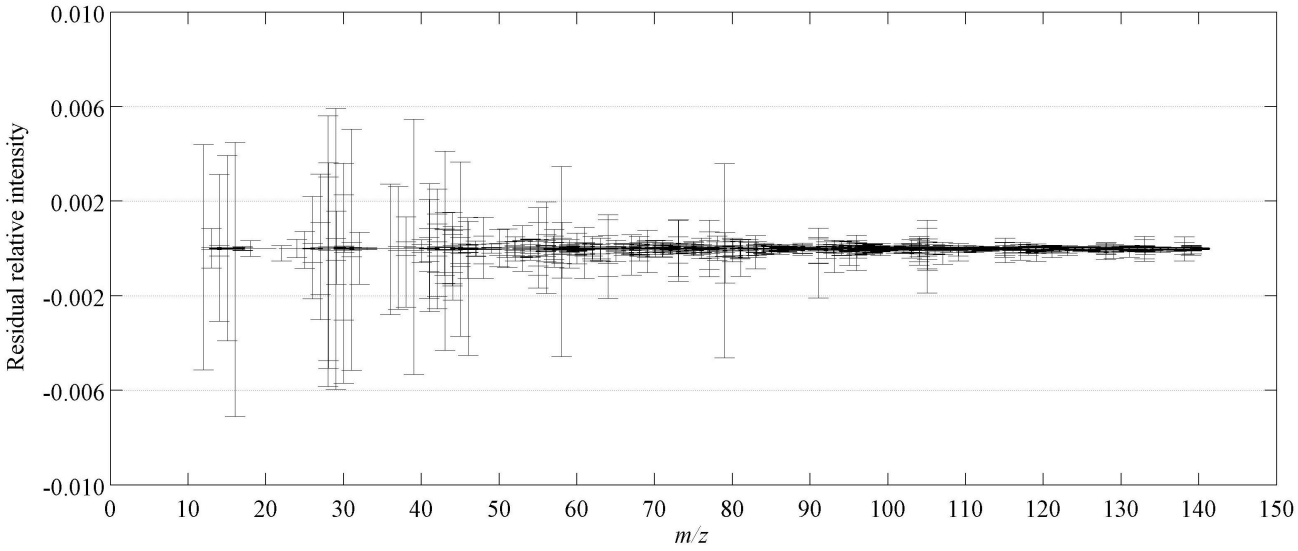

Figure S7. Residuals of the relative intensity of ions in the AMS spectra in the four-factor NMF model. The bars indicate the standard deviation (n = 61).

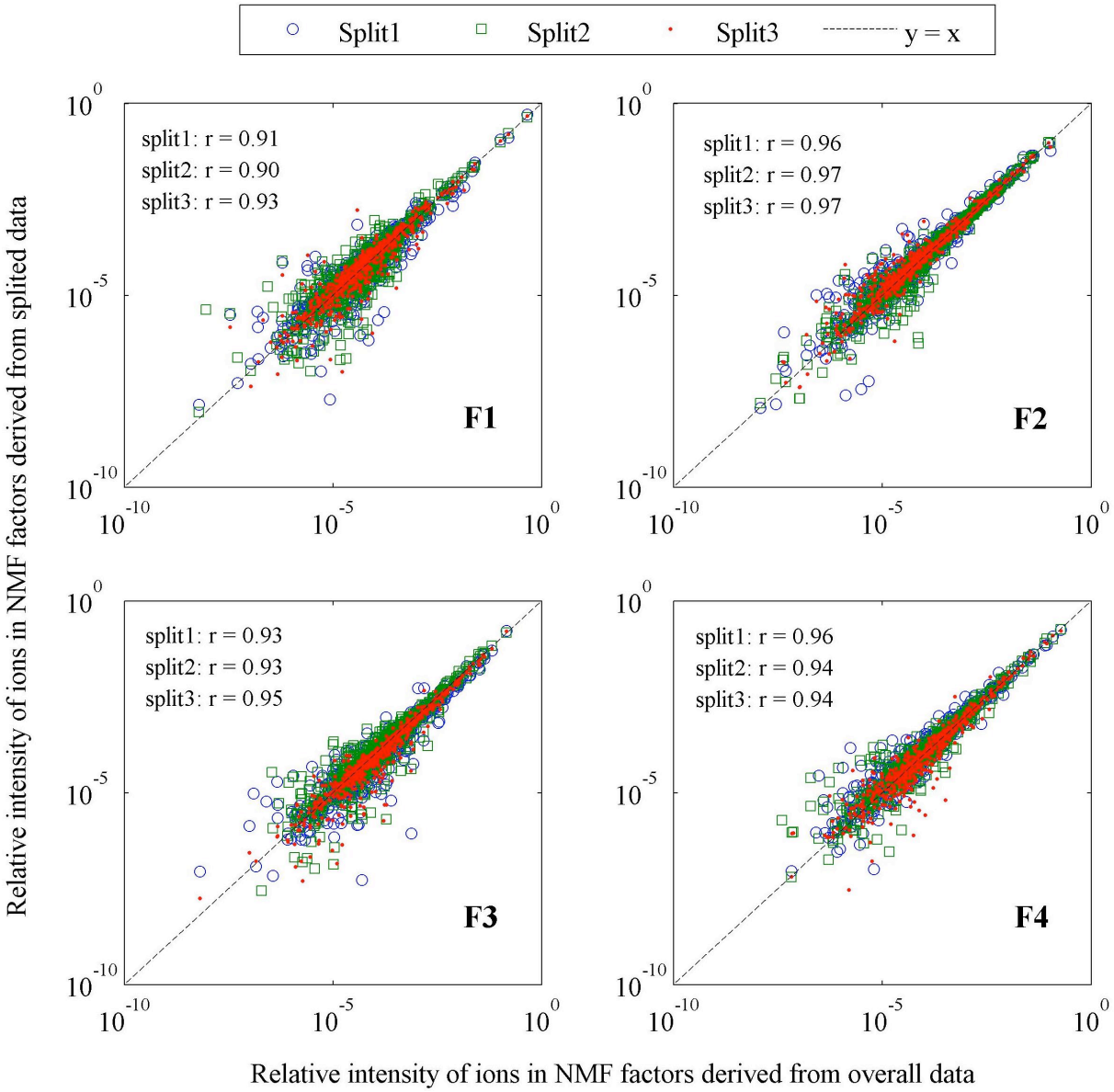

Figure S8. Plots of the relative intensity of ions in the NMF factors derived from the overall data versus that derived from three subsets of data, which were prepared by splitting of the overall data into three, followed by merging two of the three in three different manners (i.e., XY, XZ and YZ from X, Y and Z). The corresponding Pearson correlation coefficients for the logarithm of the relative intensity of ions are given in the panels. These correlation coefficients are higher than those between the logarithm of the relative intensity of ions in the four NMF factors derived from the overall data (F1–F4) (0.67–0.78).

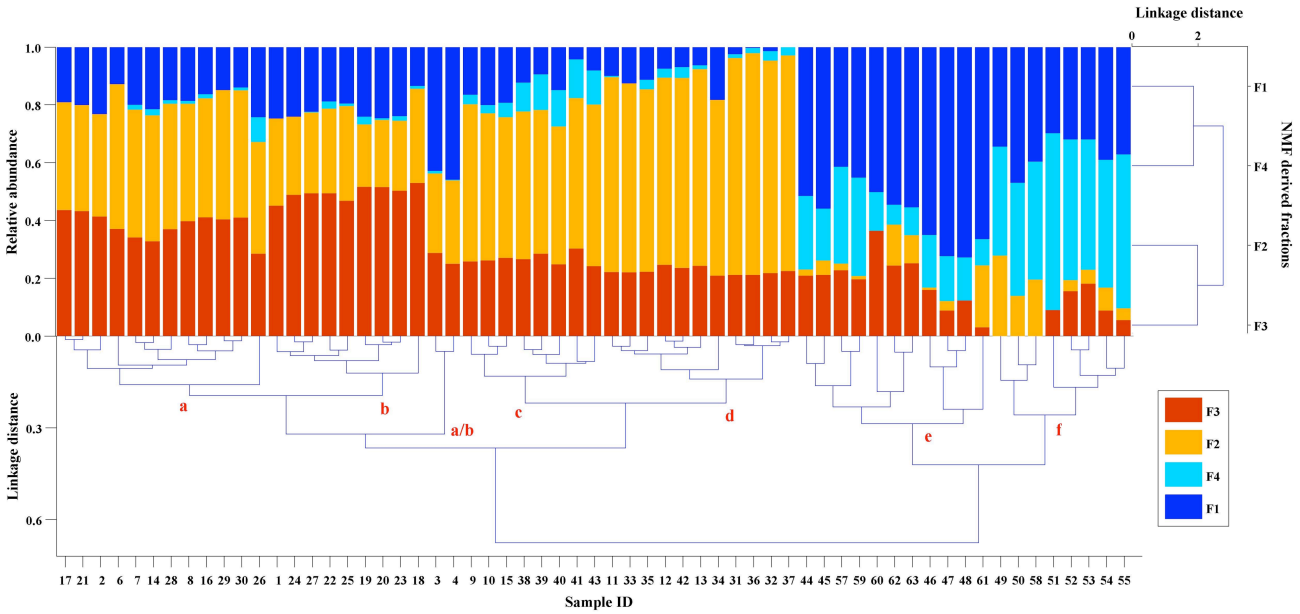

Figure S9. Hierarchical cluster analysis based on the relative contributions of NMF factors (F1: HOOA-1; F2: FOO-2; F3: LOOA-1; F4: HOOA-2) to respective samples. The sample IDs correspond to different types of aerosols as follows: ID 1-17: the urban aerosols in August and September; ID 18-27: the urban aerosols in March; ID 28-43: the forest aerosols; ID 44-63: the marine aerosols.

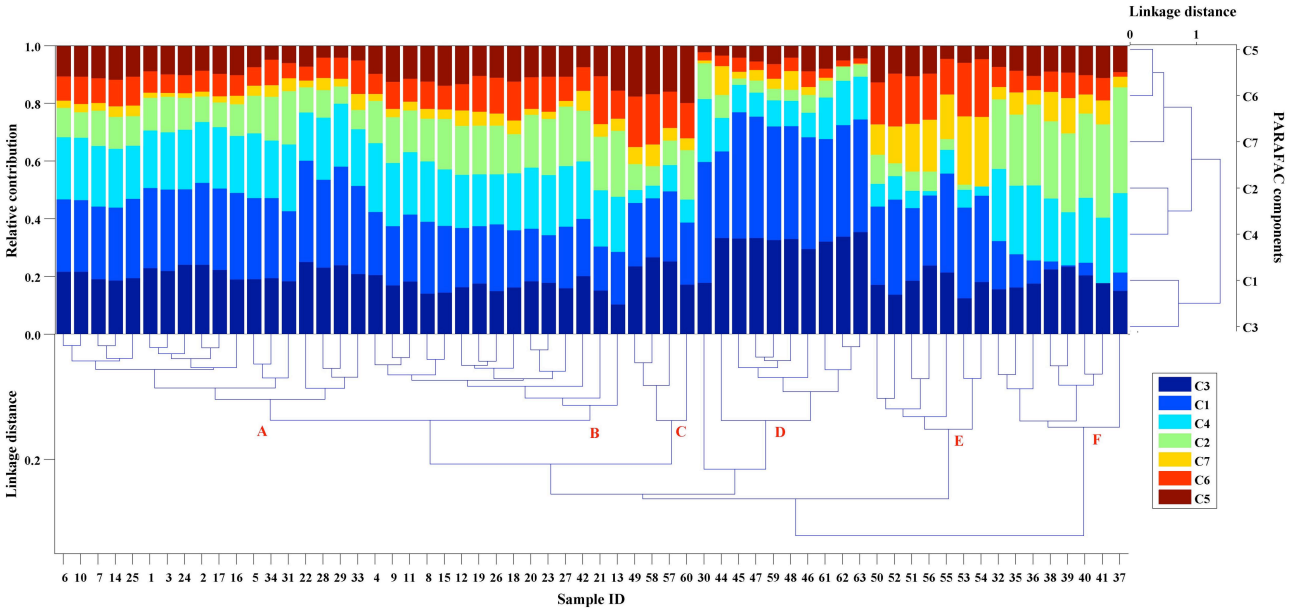

Figure S10. Hierarchical cluster analysis based on the relative contributions of PARAFAC components (C1 and C3: associated with HULIS-1; C2, C4 and C5: associated with HULIS-2; C6 and C7: associated with PLOM) to respective samples. The sample IDs correspond to different types of aerosols as follows: ID 1-17: the urban aerosols in August and September; ID 18-27: the urban aerosols in March; ID 28-43: the forest aerosols; ID 44-63: the marine aerosols.

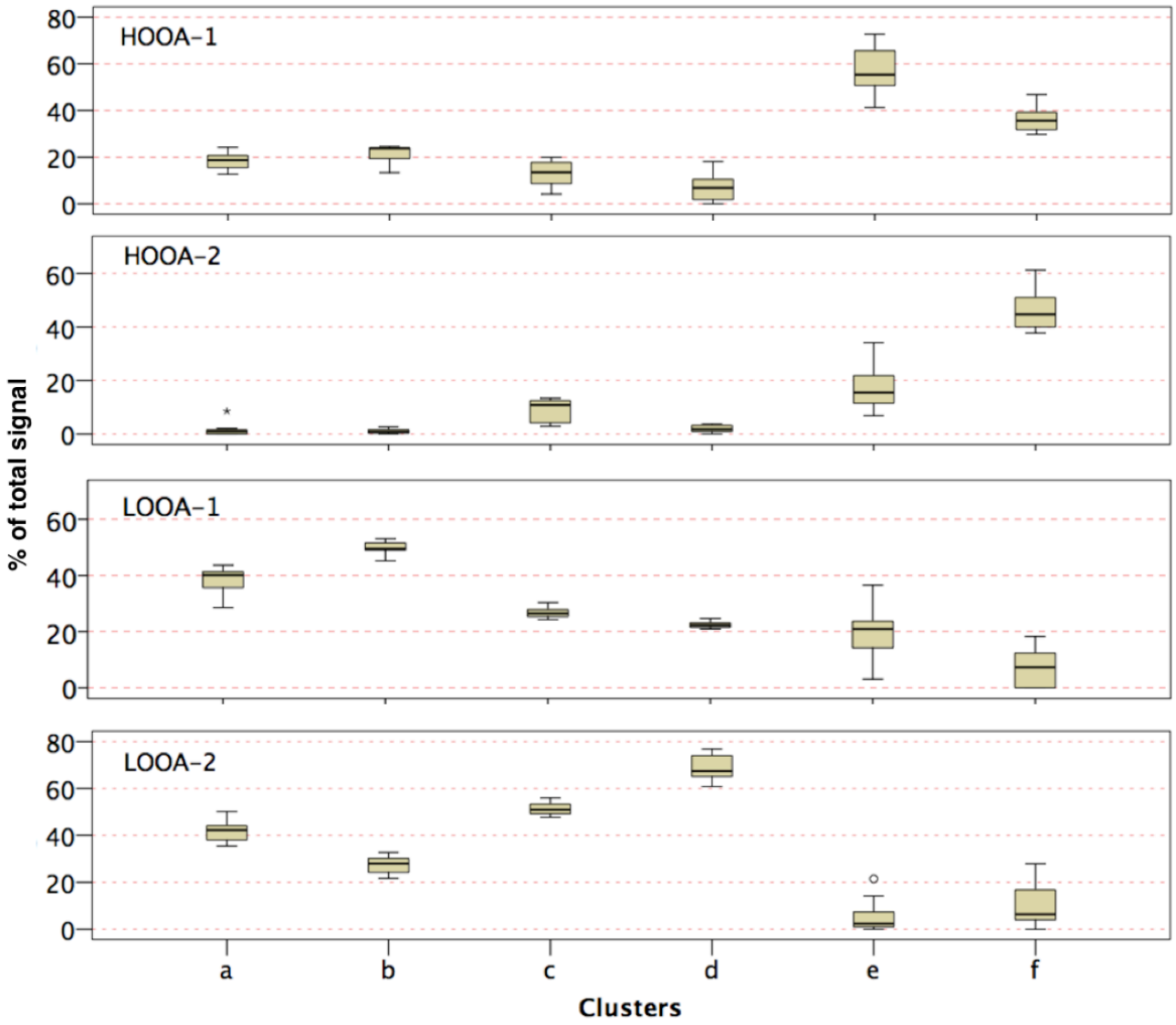

Figure S11. Boxplots of the relative contributions of four NMF factors to respective clusters. The boxes show the 25th, 50th, and 75th percentiles, and the whiskers show the ranges from minimum to maximum. The circle and the star represent mild and extreme outliers, respectively.

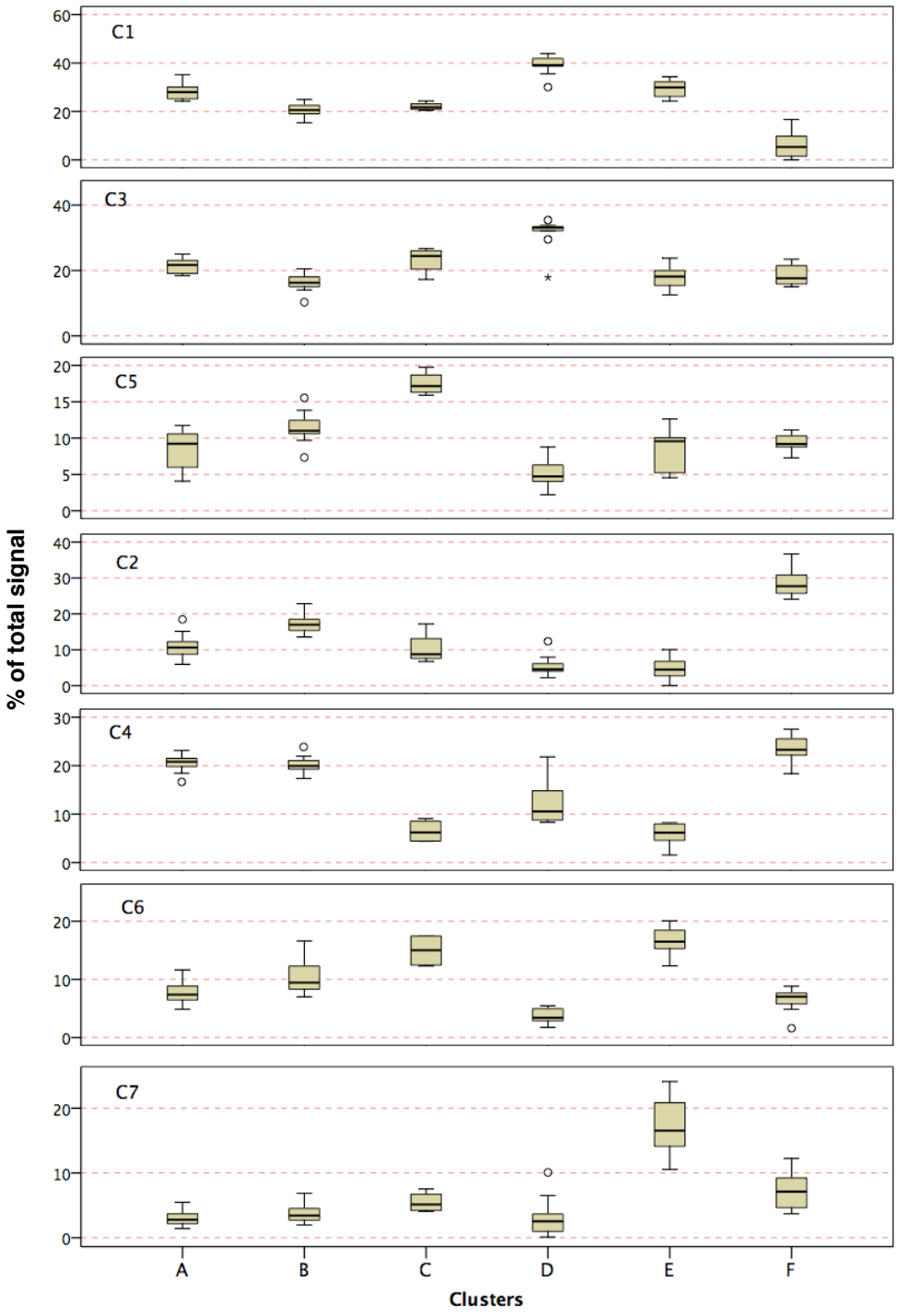

Figure S12. Boxplots of the relative contributions of seven PARAFAC components to respective clusters. The boxes show the 25th, 50th, and 75th percentiles, and the whiskers show the ranges from minimum to maximum. The circle and the star represent mild and extreme outliers, respectively.

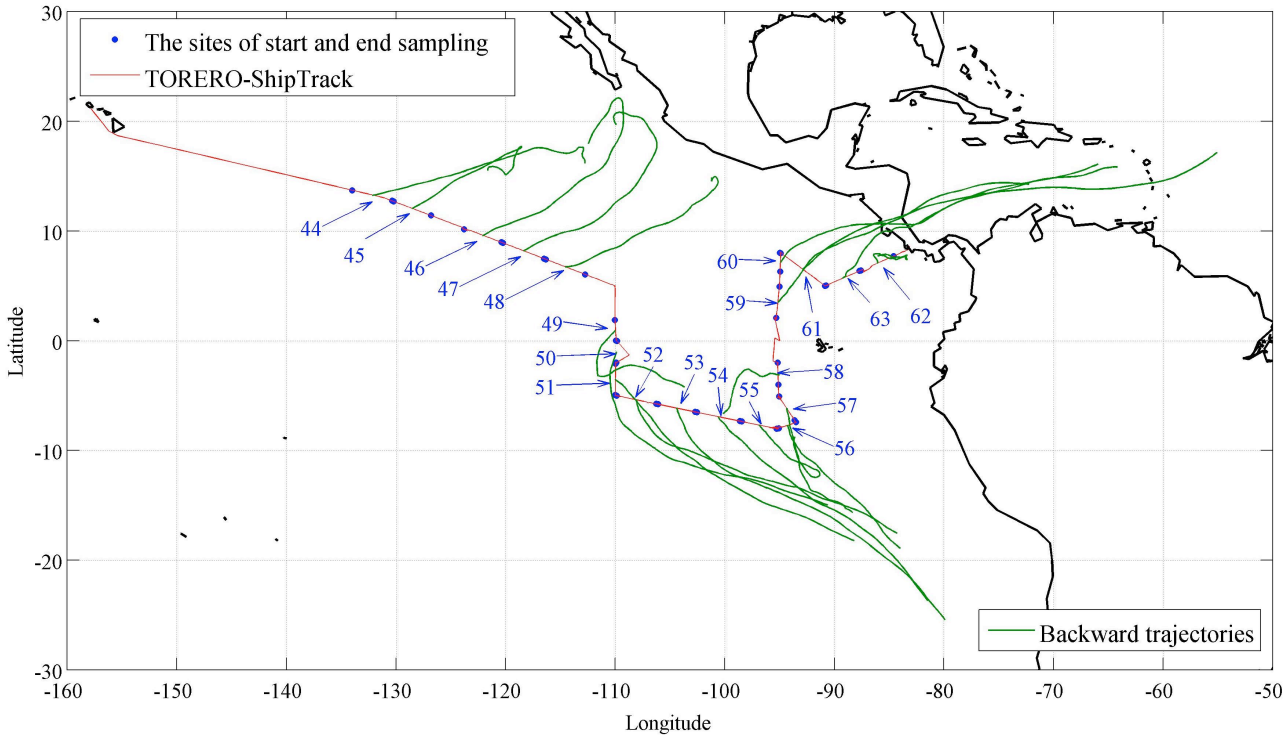

Figure S13. The ship tracks and the locations at the start and end of the collection of each sample during the TORERO/KA-12-01 cruise. Ten-day backward trajectories of air masses in the middle time of each sampling period, were also plotted (start height: 500 m above mean sea level). Analysis on air mass trajectories is also presented by Coburn et al.[3] The backward trajectories were calculated using the Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT) model version 4.[8-10]

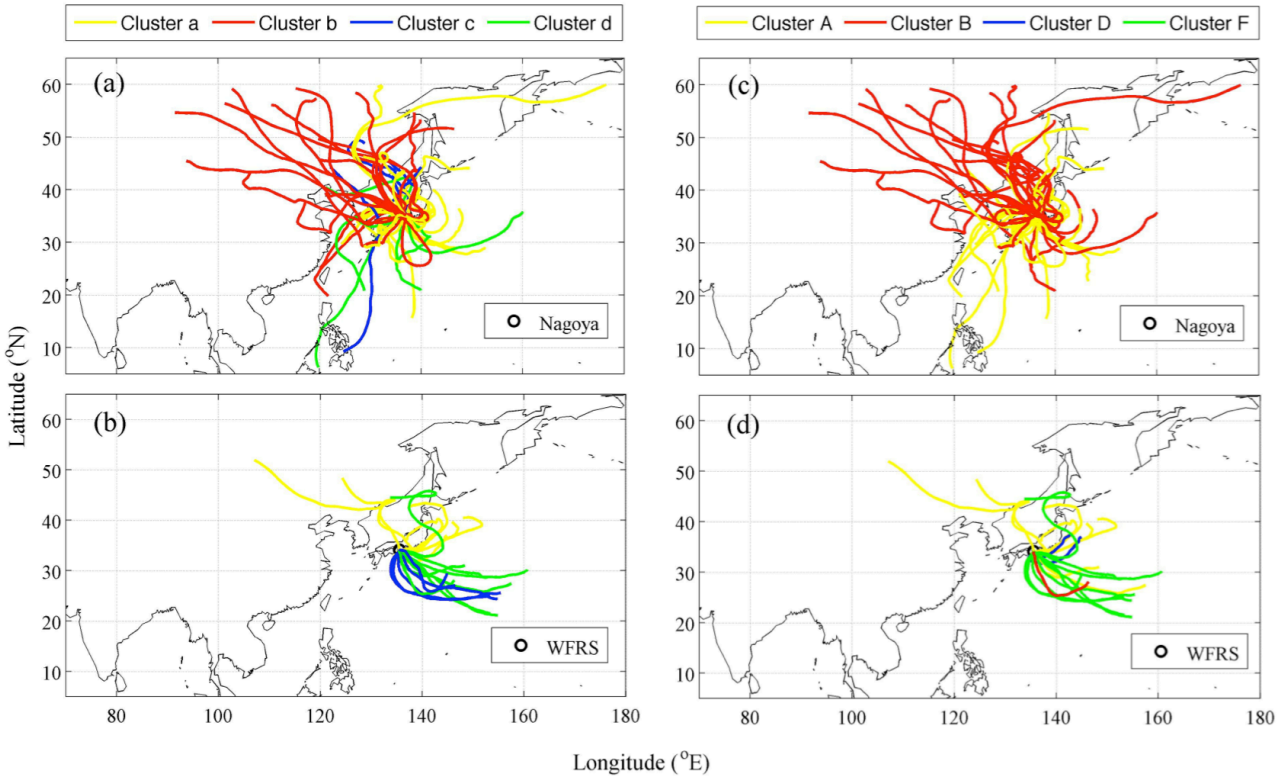

Figure S14. Ten-day backward trajectories of air masses at the middle time of respective sampling periods for the (a and c) urban and (b and d) forest aerosol samples (start height: 500 m above ground level). The colors of the trajectories represent the clusters of samples that correspond to respective trajectories of air masses. The clusters are from the hierarchical cluster analysis based on (a and b) the relative contributions of NMF factors and (c and d) the PARAFAC components for different samples. The backward trajectories were calculated using the Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT) model version 4.[8-10]

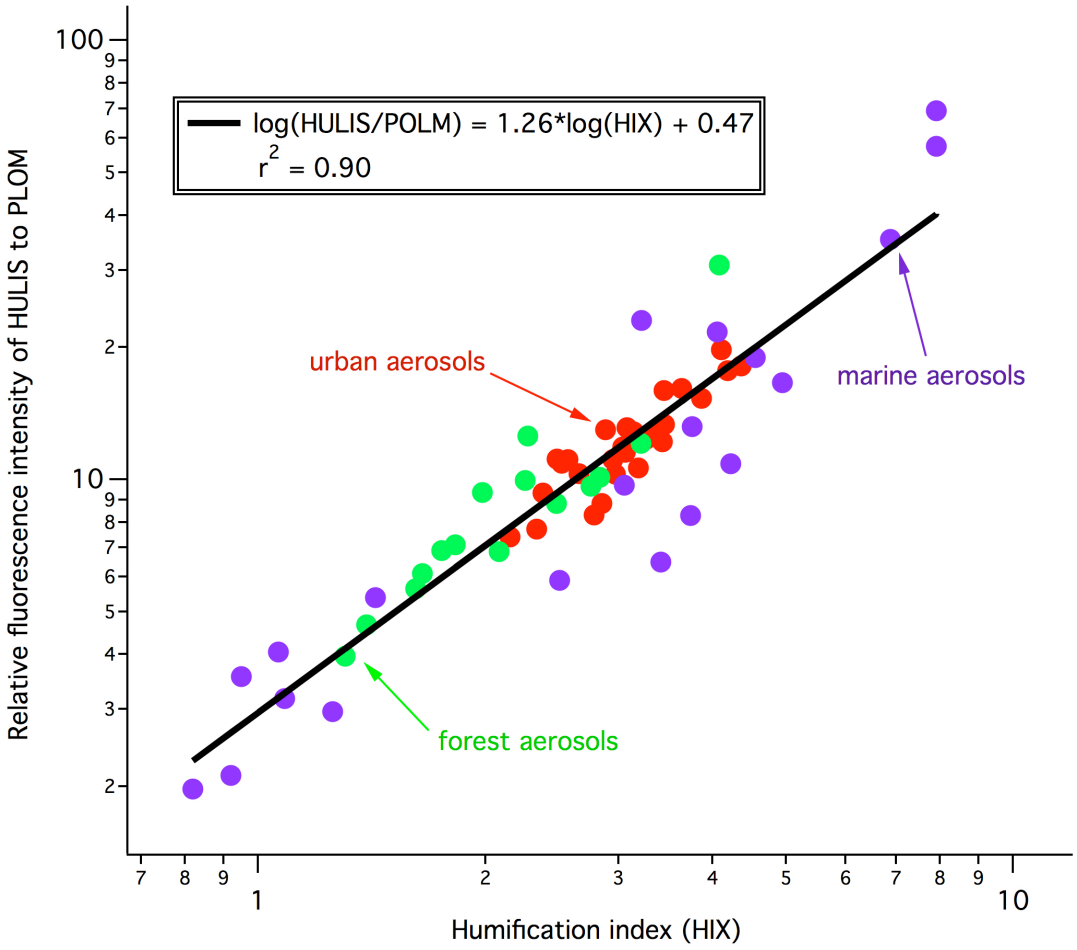

Figure S15. Plots of the relative fluorescence intensity of HULIS to PLOM versus the humification index for the urban, forest and marine aerosols.

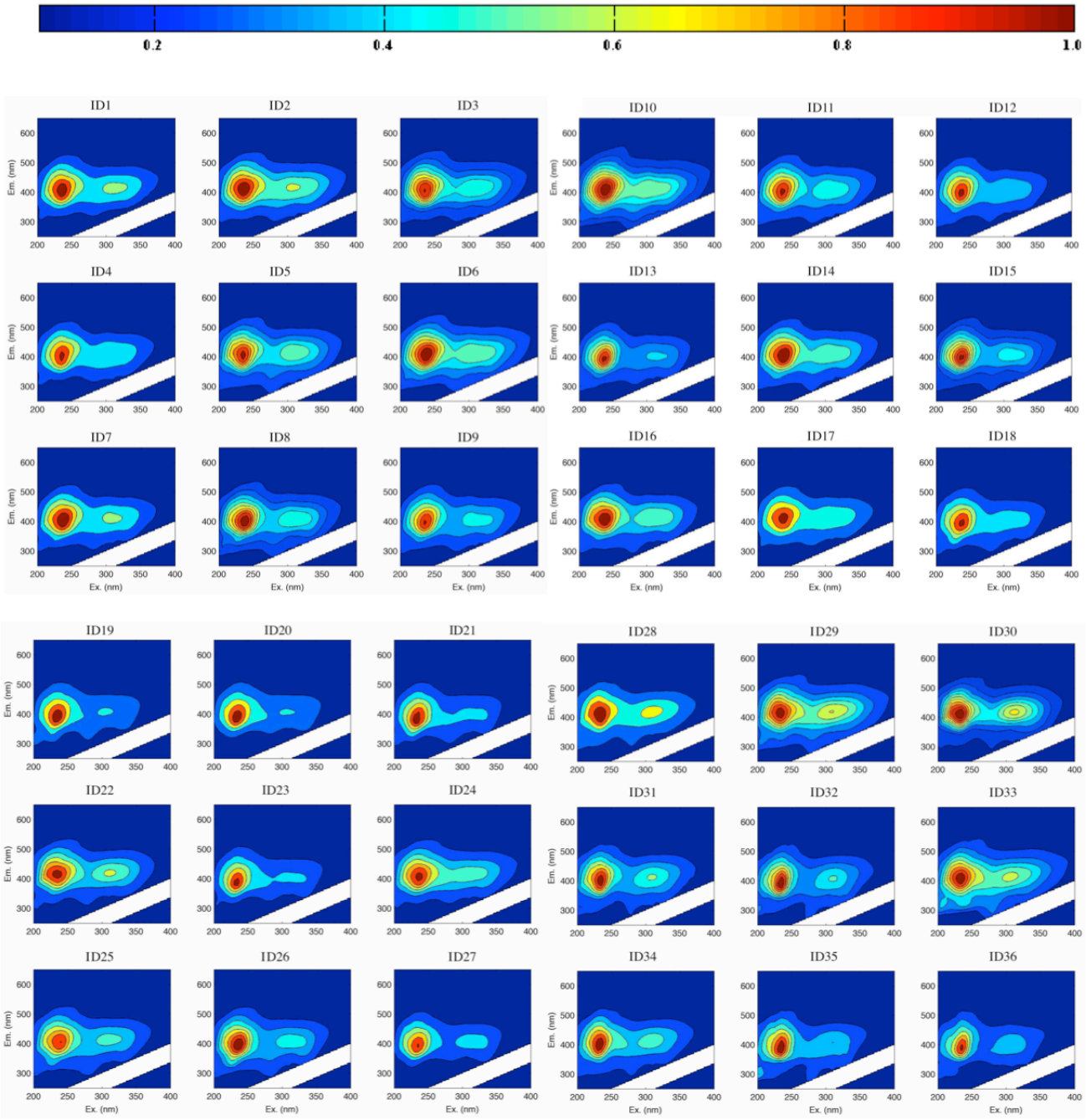

Figure S16. The EEM spectra of the water-soluble extracts (pre-processed for the PARAFAC analysis). The intensity was normalized to make the maximum to be unity. The sample IDs correspond to different types of aerosols as follows: ID 1-17: the urban aerosols in August and September; ID 18-27: the urban aerosols in March; ID 28-43: the forest aerosols; ID 44-63: the marine aerosols. Ex.: excitation wavelength; Em.: emission wavelength.

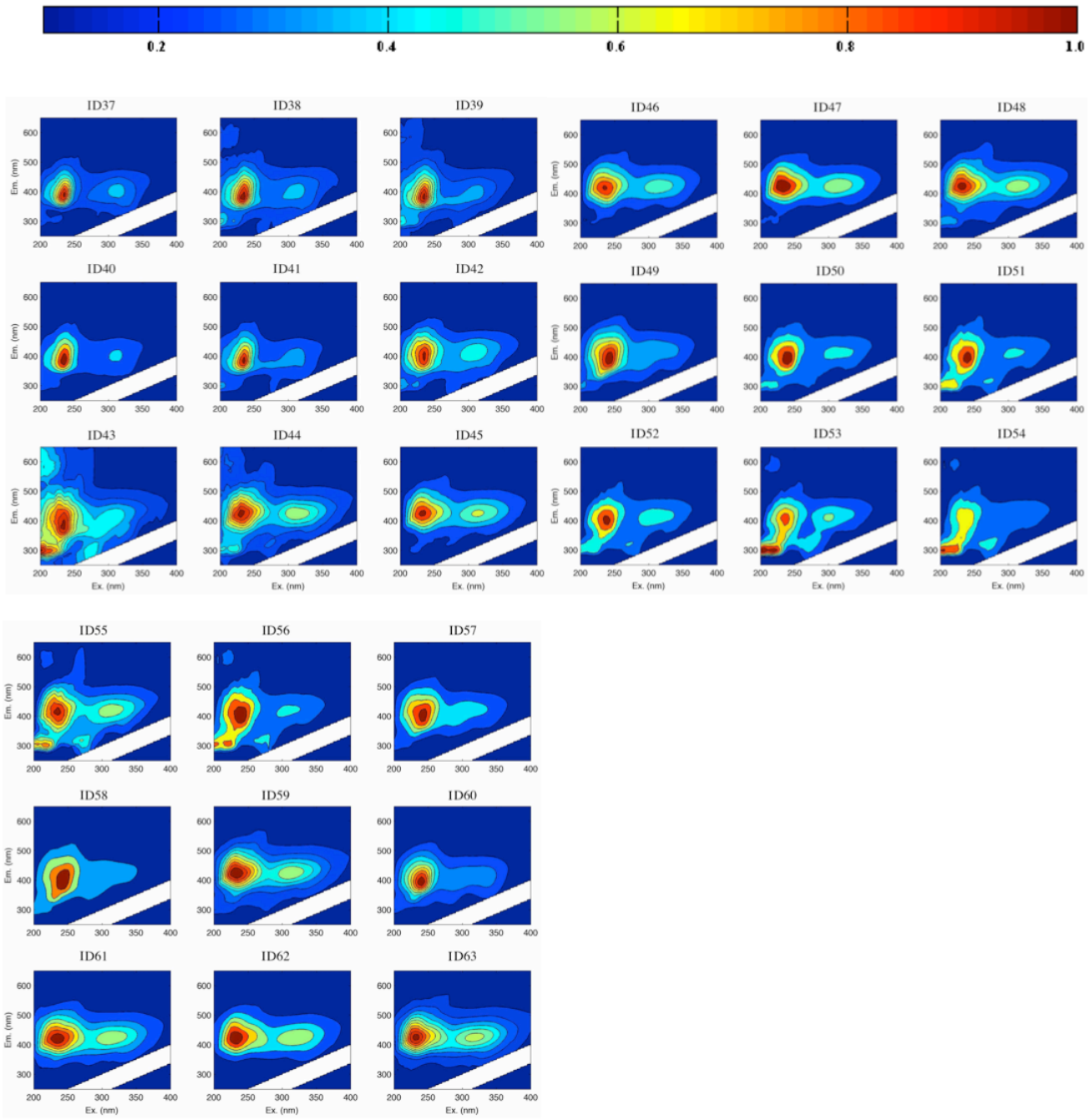

Figure S16. (continued)

REFERENCES

(1) Okumura, M. Estimation of Volatile Organic Compound Emissions from Forest Vegetation, *Ph.D. thesis* **2009**, Graduate School of Energy Science, Kyoto Univ., Kyoto, Japan.

(2) Sathiyamurthi, R,; Ida, A.; Jones, C.; Kato, S.; Tsurumaru, H.; Kishimoto, I.; Kawasaki, S.; Sadanaga, Y.; Nakashima, Y.; Nakayama, T.; Matsumi, Y.; Kagami, S.; Deng, Y.; Ogawa, S.; Kawana, K.; Mochida, M.; Kajii, Y. Total OH reactivity measurement in a BVOC dominated temperate forest during summer campaign, 2014. *Atmos. Environ.* **2016**, *131*, 41-54.

(3) Coburn, S.; Ortega, I.; Thalman, R.; Blomquist, B.; Fairall, C. W.; Volkamer, R. Measurements of diurnal variations and eddy covariance (EC) fluxes of glyoxal in the tropical marine boundary layer: description of the Fast LED-CE-DOAS instrument. *Atmos. Meas. Tech.* **2014**, *7*, 3579–3595.

(4) Miyazaki, Y.; Coburn, S.; Ono, K.; Ho, D. T.; Pierce, R. B.; Kawamura, K.; Volkamer, R. Contribution of dissolved organic matter to submicron water-soluble organic aerosols in the marine boundary layer over the eastern equatorial Pacific. *Atmos. Chem. Phys.* **2016**, *16* (12), 7695–7707, DOI: 10.5194/acp-16-7695-2016.

(5) Koop, D. R.; Morgan, E. T.; Tarr, G. E.; Coon, M. J. Purification and characterization of a unique isozyme of cytochrome P-450 from liver microsomes of ethanol-treated rabbits. *J. Biol. Chem.* **1982**, *257*, 8472– 8480.

(6) Fu, P. Q.; Kawamura, K.; Okuzawa, K.; Aggarwal, S. G.; Wang, G.; Kanaya, Y.; Wang, Z. Organic molecular compositions and temporal variations of summertime mountain aerosols over Mt. Tai, North China Plain. *J. Geophys. Res.* **2008**, *113*, D19107.

(7) Mochizuki, T.; Miyazaki, Y.; Ono, K.; Wada, R.; Takahashi, Y.; Saigusa, N.; Kawamura, K.; Tani, A. Emissions of biogenic volatile organic compounds and subsequent formation of secondary organic aerosols in a Larix kaempferi forest. *Atmos. Chem. Phys.* **2015,** *15*, 12029-12041.

(8) Draxler, R.R.; G.D. Hess. Description of the HYSPLIT_4 modeling system, NOAA Technical Memorandum ERL ARL-224, December, **1997**, 24 p.

(9) Draxler, R.R.; G.D. Hess. An overview of the HYSPLIT_4 modelling system for trajectories, dispersion and deposition. *Aust. Met. Mag.* **1998**, *47*, 295-308.

(10) Draxler, R.R.; Rolph, G.D. HYSPLIT (HYbrid Single-particle Lagrangian Integrated Trajectory) Model Access via NOAA ARL READY Website. NOAA Air Resources Laboratory, Silver Spring, MD, USA, **2003**, http://www.arl.noaa.gov/ready/hysplit4.html.