1 *Supplement of*

2 **Molecular signatures and formation mechanisms of particulate matter**

3 **(PM) water-soluble chromophores from Karachi (Pakistan) over**

4 **South Asia**

5 Jiao Tang[1], Jun Li[1], Shizhen Zhao[1], Guangcai Zhong[1], Yangzhi Mo[1], Hongxing Jiang [2],

6 Bin Jiang[1], Yingjun Chen[2], Jianhui Tang[3], Chongguo Tian[3], Zheng Zong[4], Jabir Hussain

7 Syed[5], Jianzhong Song[1], Gan Zhang[1]

9 [1]State Key Laboratory of Organic Geochemistry and Guangdong province Key

10 Laboratory of Environmental Protection and Resources Utilization, Guangdong-Hong

11 Kong-Macao Joint Laboratory for Environmental Pollution and Control, Guangzhou

12 Institute of Geochemistry, Chinese Academy of Sciences, Guangzhou, 510640, China

13 [2]Shanghai Key Laboratory of Atmospheric Particle Pollution and Prevention (LAP[3]),

14 Department of Environmental Science and Engineering, Fudan University, Shanghai

15 200433, China

16 [3]Key Laboratory of Coastal Environmental Processes and Ecological Remediation,

17 Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai,

18 264003, China

19 [4]Department of Civil and Environmental Engineering, Hong Kong Polytechnic

20 University, Hong Kong, 999077, China

21 [5]Department of Meteorology, COMSATS University Islamabad (CUI), Islamabad,

22 45550, Pakistan

24   Corresponding Authors:

25   Jun Li: junli@gig.ac.cn; Gan Zhang: zhanggan@gig.ac.cn

Contents of this file

Texts S1 to S7; Tables S1 to S8; Figures S1 to S14; and References

Text S1. Extraction and analysis of reproducibility and blank levels

The reproducibility and the blank levels of the analysis of the WSOC by carbon mass analysis, the UV-visible absorption spectrophotometer, and the EEM fluorescence spectrophotometer were assessed by the duplicate extraction and analysis of an identical aerosol sample and a blank filter. The concentration of WSOC and WSTN was quantified using a TOC analyzer (Vario TOC cube; Elementar). The relative standard deviation (RSD) of WSOC mass by duplicate extraction and duplicate analyses was less than 1.8 % and 2.6 %, respectively. The RSD for UV-vis spectra of the overall wavelength (200–800 nm) and EEM spectra at selected Em/Ex = 411/239 nm was 0–8.1 % and 2.9 %–4.6 %, respectively.

The error of WSOC concentration in three blank samples was 7.7 %. No obvious peaks were observed in the blank EEMs, which were also subtracted from the sample EEMs. In addition, the fluorescence spectrum of samples was measured with their absorbance lower than 1.

Text S2. PARAFAC analysis and determination of the optical parameters

The contribution of solvents was subtracted from the extract spectra. To obtain WSOC quality metrics from EEMs, PARAFAC analysis was conducted using the drEEM Toolbox in MATLAB version R2016a (http://models.life.ku.dk/drEEM, last access: June 2014) (Murphy et al., 2013). The measured absorbance that was not greater than 1 at 254 nm was appropriately used to correct the EEM for inner-filter effects (IFEs) (Kothawala et al., 2013; Murphy et al., 2013). Each EEM was normalized to the Raman peak area of purified water collected on the same day to correct fluorescence in Raman units (RU) at an excitation of 350 nm (Lawaetz and Stedmon, 2009), and signals of first-order Rayleigh and Raman scattering, as well as second-order Rayleigh scattering in the EEM spectra, were removed using the interpolation (Bahram et al., 2006). The

54   non-negativity constraint is necessary to obtain reasonable spectra according to the

55   tutorials, and a 2–7 component PARAFAC model based on 96 EEMs was analyzed

56   along with the residual analysis, visual examination of the component spectra, core

57   consistency, and validated using split-half analysis to ensure that they are representative

58   of plausible fluorophores (Murphy et al., 2013).The three-component PARAFAC

59   solutions passed the split-half analysis with the split style of "S$_4$C$_6$T$_3$" (Fig. S12). A

60   critical prerequisite to successful PARAFAC modeling is a data set encompassing

61   enough compositional variability and lacking excessive noise (Murphy et al., 2018).

62   Once the modeling constraints and criteria have been decided, we ultimately adopt only

63   the model that represents the least-squares (minimum error) solution.

64   ***Optical indices***

65   The specific UV-vis absorbance at 254 nm was successfully used to characterize

66   the chemical properties of organic fractions in atmospheric samples (Li et al., 2018),

67   which were obtained by the normalization of UV−vis absorbance at 254 nm by the

68   organic carbon in the solution, and the equation was listed as follows (Weishaar et al.,

69   2003):

70                      $\text{SUVA} = A/bc$ (1)

71   Here, SUVA is the specific UV-Vis absorbance (m^2 g^{-2} C), A is the absorbance at

72   254 nm, b (m) is the cell path length, and c (mg C L^{-1}) is the WSOC concentration in

73   the solutions.

74   S$_R$ was used as a proxy for molecular weight in a broad range of samples, by

75   calculating the ratio of the slope of the shorter wavelength region (275–295 nm) to that

76   of the longer wavelength region (330–400 nm) (Helms et al., 2008; Twardowski et al.,

77   2004). The S$_R$ was derived from absorption spectra by fitting the absorption data to the

78   following equations:

79                      $\alpha = 2.303\ A / L$ (2)

80                      $\alpha_\lambda = \alpha_{\lambda_{ref}} e^{-s(\lambda - \lambda_{ref})}$ (3)

81                      $S_R = S_{275\text{-}295} / S_{330\text{-}400}$ (4)

82   Here, α = the Napierian absorption coefficient (m^{-1}), A = absorbance, L = path

83   length (m), λ= wavelength (nm), and λ_{ref} = reference wavelength (nm).

84   In addition, fluorescence indices based on intensity ratios that provide insight into

85   the origins of dissolved BrC, such as the humification index (HIX) (the ratio of average

86   emission intensity in the 435−480-nm range to that in the 300−345-nm range following

87   excitation at 254 nm, which was used to reflect the degree of humification) (Zsolnay et

88   al., 1999), the biological index (BIX) (the ratio of emission intensities at 380 and 430

89   nm following excitation at 310 nm, reflecting autochthonous biological activity in water

90   samples) (Huguet et al., 2009), and fluorescence index (FI) (the ratio of emission

91   intensities at 470 and 520 nm following excitation at 370 nm, reflecting the possibility

92   of microbial origin and for examining differences in precursor organic materials) (Lee

93   et al., 2013; Murphy et al., 2018).

94   ***The calculation of MAE***

95   The light absorption coefficient (Abs$_\lambda$, Mm^{-1}) was calculated using the following

96   equation (Yan et al., 2015):

97      $$Abs_\lambda = \frac{(A_\lambda - A_{700}) \times V_1 \times a \times \ln(10)}{V_a \times a_1 \times l} \qquad (5)$$

98   Here, A$_\lambda$ is the value of light absorption at the given wavelength given by the

99   spectrophotometer; V_1 and a_1 are the volume of ultra-pure deionized water for

100   extraction and area of the extracted filter; V_a is the volume of sampling air; l is the

101   optical path length. In addition, absorption Ångström exponent (AAE) represents the

102   wavelength dependence of absorption is calculated (Abs = k • λ^{-AAE}), and K is constant.

103   Mass absorption efficiency (MAE, m^2 g^{-1} C) can be obtained as follows:

104   MAE$_\lambda$=Abs$_\lambda$/C_i (6)

105   Here, C_i ($\mu g\ C\ m^{-3}$) is the concentration of WSOC after conversion to the

106   atmosphere.

107   **Text S3. Evaluation of the SPE method**

108   It is necessary to avoid the interference of inorganic constituents for the molecular

109   characterization of WSOC using FT-ICR MS. The SPE method developed effectively

110   removed inorganic salts and enriched the organic material from dilute aqueous solutions

111   (Varga et al., 2001). The cartridges were conditioned and equilibrated with 4 mL of

112   methanol (HPLC grade) and 5 mL of ultrapure deionized water, respectively. An aliquot

113   of 20 mL extract was acidified to pH = 2 using hydrochloric acid and loaded on an SPE

114   cartridge. However, some organic materials, such as low-molecular-weight organic

115   acids, and sugars were not retained by the SPE cartridge and appeared in the effluent

116   solution (Lin et al., 2010). Thus, to assess the efficiency of SPE, we measured the total

117   organic carbon masses, UV–vis absorption, and EEM spectra of the original samples

118   and eluents, which ensured that the majority of organics were preserved on the

119   cartridges. In this step, the WSOC mass of eluents accounted for 67 %±3.3 % of original

120   WSOC samples on average. In addition, the light absorbance at 280 nm of eluents

121   accounted for 80 %±6.3 % of original samples on average, indicating most of the light-

122   absorbing organics were retained on cartridges. In fluorescence spectroscopic

123   measurements average of 97 %±7.7% of the total fluorescence intensity at Em/Ex =

124   411/239 nm was detected in the eluate, suggesting the fluorescent compounds were

125   almost entirely retained in the columns. Normally, the low-molecular-weight organic

126   acids, and sugars in the effluent were not in the detected m/z range of 150–800.

127   The ammonia added to methanol as elution was to drop the percentage of the

128   irreversibly absorbed carbon mass (Lin et al., 2010). In addition, ammonia was selected

129   for its volatility so that it could be removed from the eluate by evaporation to dryness.

130   **Text S4. FT-ICR MS data processing**

131   The mass spectra were externally calibrated with arginine clusters using a linear

132   calibration and then internally recalibrated with typical O_6S_1 class species peaks using

133   quadratic calibration in DataAnalysis version 4.4 (Bruker Daltonics). Custom software

134   was applied to calculate all mathematically possible formulas for all ions with a signal-

135   to-noise ratio > 4 using a mass tolerance of ± 1.0 ppm. The formula assigned to peaks

136   with elemental combinations of $^{12}C_{1-80}{}^{1}H_{1-200}{}^{16}O_{0-50}{}^{14}N_{0-4}{}^{32}S_{0-2}$ in the ESI– and

137    $^{12}C_{1-80}{}^{1}H_{1-200}{}^{16}O_{0-50}{}^{14}N_{0-4}{}^{32}Na_{0-1}$ in the ESI+ mode was identified. Identified formulas

138   with isotopomers (i.e., ^{13}C, ^{18}O, or ^{34}S) were not discussed in this study.

139   The double bond equivalent (DBE) values of the neutral formulas were calculated

140   using as follows (Lin et al., 2012):

$$DBE = (2c + 2 - h + n)/2 \quad (7)$$

142   The modified aromaticity index (AI_{mod}) values were obtained from the equation

143   (Koch and Dittmar, 2006):

$$AI_{mod} = (1 + c - 0.5o - s - 0.5h)/(c - 0.5o - s - n) \quad (8)$$

145   The carbon oxidation state ($\overline{OS_C}$) was previously observed to increase with

146   oxidation for atmospheric organic aerosol and was strongly linked to aerosol volatility

147   (Kroll et al., 2011). The $\overline{OS_C}$ of each formula was calculated using the following

148   equation:

$$\overline{OS}_C = -\Sigma_i \, os_i \frac{n_i}{n_c} \quad (9)$$

150   where the summation is over all non-carbon elements, OS_i is the oxidation state

151   associated with element i, and n_i/n_C is the molar ratio of element i to carbon.

152   We calculated the nominal oxidation state of carbon (Kellerman et al., 2015; Lv et

153   al., 2016):

$$NOSC = 4 - \left[\frac{4c + h - 3n - 2o - 2s}{c}\right] \quad (10)$$

155   where $c, h, n, o,$ and s refer to the number of atoms per formula of carbon, hydrogen,

156   nitrogen, oxygen, and sulfur, respectively.

157   **Text S5. Spearman's rank correlation analysis**

158   Spearman's rank correlation analysis between relative FT-ICR MS peak intensities

159   and percent PARAFAC components and optical parameters was performed to identify

160   molecular families associated with PARAFAC components and optical parameters.

161   Formulas that appeared in fewer than two samples were not considered; thus, a total of

162   7687 ESI– formulas and 6564 ESI+ formulas were included. The Spearman correlation

163   coefficient is defined as the Pearson correlation coefficient between the rank variables.

164   For an n of 12 samples, the 12 raw scores X_i (i.e. percent PARAFAC components,

165   optical indices, Abs_{365}), Y_i (i.e. relative FT-ICR MS intensities) are converted to ranks

166    $R(X_i)$, $R(Y_i)$, and Spearman's coefficient (r_s) is computed. To assess the significance of

167   the r_s, a Student's t test was performed. For an n of 12 samples, the threshold of a r_s

168   was computed to be significant at the α = 0.05 confidence level, using the following

169   equation:

170    $$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad (11)$$

171   where t = 2.201 for n = 12. Therefore, a r_s of 0.57 was calculated to be significant

172   at the 95% confidence limit. And, the molecular formulas associated with PARAFAC

173   components and optical parameters across the 12 samples with Spearman's $r_s \geq 0.57$

174   were assigned to each PARAFAC component and optical parameter (Stubbins et al.,

175   2014; Singer et al., 2012). Note that many formulas were not fluorophores or

176   chromophores but had similar sources or biogeochemical processes with the

177   corresponding fluorophores or chromophores.

178   **Text S6. Redundancy Analysis (RDA)**

179   RDA is a method combining regression and principal component analysis (PCA).

180   It is a direct extension of multiple regression analysis to model multivariate response

181   data. In this study, we assumed that the indices derived from optical measurement (FI,

182   BIX, HIX, A_{254}, $SUVA_{254}$, and S_R) are representative of the important chemical

183   properties of WSOC. Hence, to explore the relationships between these chemical

184   characteristics and fluorescent components, we conducted an RDA model by using the

185   fluorescent component as response variables and these optical indices as explanatory

186   variables. The RDA was preceded using the "vegan" package in R (Oksanen et al.,

187   2022). Before analysis, we used the **decorana()** function to identify whether the model

188   is suitable for RDA. We checked the length of the first DCA axis (which is scaled in

189   units of standard deviation, S.D.), and the value was 0.29, which was far low than 3

190   S.D, indicating a homogeneous dataset for which linear methods are suitable (Ramette,

191   2007); hence, we select the RDA model. We began this process while exploring the data:

192   our response data is Hellinger-transformed, and our explanatory variables are centered

193   and standardized. The RDA result indicated that 69.5 % of the response variables could

194   be explained by the optical indices.

195   In addition, we performed a permutation test for the RDA results, including a

196   global RDA significance, all canonical axis significance, and explanatory variables

197   significance. The permutation test result for all and each canonical axis are shown in

198   Table S7, showing that our full model is statistically significant ($p = 0.001$), and the

199   selected two canonical axes resulting from the RDA is also statistically significant ($p =$

200   0.001). In addition, every variable included in this model is significant as well ($p =$

201   0.001), except for $SUVA_{254}$ ($p = 0.010$). In addition, it is reasonable to look at the

202   residual variation in the data by using the Kaiser-Guttman criterion. There may still be

203   some variation in these data that has not been explained by our set of explanatory

204   variables.

205   Sometimes, we also wish to reduce the number of explanatory variables for

206   seeking parsimony and the possible strong linear dependencies (correlations) among

207   the explanatory variables in the RDA model, which could render the regression

208   coefficients of the explanatory variables in the model unstable. Linear dependencies

209   can be explored by using computing the variables' variance inflation factors (VIF).

210   Normally, VIF values above 20 indicate strong collinearity. Ideally, VIFs above 10

211   should be at least examined and avoided if possible. In the current study, the VIF values

212   of all explanatory variables are lower than 10, indicating no obvious collinearity. In

213   addition, model simplification to reduce the number of variables indicated explanatory

214   variables can be selected by the forward, backward, or stepwise selection that removes

215   non-significant explanatory variables. In RDA, the forward selection is the method

216   most often applied because it works even in cases where the number of explanatory

217   variables is larger. Even though we had only six explanatory variables, we used the

218 method of forward selection with functions **ordiR2step()**. The decision to retain a

219 variable or to stop the procedure was made on two stopping criteria: a pre-selected

220 significance level α, and the adjusted coefficient explanatory variables (R^2_{adj}) of the

221 global model. In forward selection, variables are added in order of decreasing F-values

222 (or increasing R^2_{adj}), each addition being tested by permutation, and stops when the

223 permutational probability exceeds the predefined α significance level (or R^2_{adj} of the

224 global model). These analyses show that the response variables would be to settle for a

225 model containing the six explanatory variables (HIX, S_R, BIX, FI, A_{254}, and $SUVA_{254}$)

226 (Table S8).

227 **Text S7. Atmospheric PARAFAC components differ from the aquatic**

228 **environment**

229 The EEM-based method was initially used to characterize the DOM fraction in the

230 aquatic environment, and further extended to the atmosphere. Therefore, the name and

231 the possible fate of atmospheric fluorescent components were previously referred to as

232 that aquatic DOM. However, multiple studies have shown that the atmospheric

233 chromophores maybe not same as the similar character in other environments (Wu et

234 al., 2021; Jiang et al., 2022), yet direct evidence is lacking. In this study, we examined

235 molecular-level differences between the fluorescent components in the atmosphere and

236 aquatic environment. The molecular signatures of PARAFAC components of DOM in

237 rivers and streams as well as lakes have been determined by using Spearman's

238 correlation analysis (Stubbins et al., 2014; Kellerman et al., 2015). The VK space of

239 molecular families assigned to PARAFAC components and excitation and emission

240 maxima in our study along with those in rivers and streams, and lakes, were shown in

241 Fig. S14. Several PARAFAC components from aquatic DOM samples were more

242 inclined to group within the region of aromatic compounds, exhibiting lower average

243 H/C ratios than those from aerosol samples, inconsistent with their emission maxima.

244 For example, P6 (in rivers and streams), C6 (in lakes), and C3 in this study (Fig. S14)

245 are all defined as protein-like components with similar excitation and emission maxima,

246 yet P6 formulas had lower H/C ratios and higher O/C ratios. C2 in aerosol samples had

247 a higher emission/excitation maximum than C1 (in lakes), conversely, C2 (in aerosol)

248 formulas grouped in the region of highly unsaturated and phenolic compounds, yet C1

249 (in lakes) formulas grouped in aromatic compounds region, implying although similar

250 fluorescence pattern and Ex/Em maximum of PARAFAC components, their chemical

251 properties vary in different environments, such as an unsaturated degree. The

252 phenomenon indicated that the fluorescent components may be not only affected by the

253 chemical structure, but other factors, such as processes, and origins, could be

254 responsible for the pattern or location of EEM spectra. Optical indices such as $SUVA_{254}$,

255 A_{254}, and HIX-assigned formulas in lake DOM prefer the bottom right corner of the VK

256 diagram compared to that in aerosols (detail in Fig. S13), strongly suggesting they had

257 different chemical properties in various environments. Although fluorescence-based

258 methods have been extended to OA, this study provided molecular-related evidence that

259 the molecular basis of fluorophores and optical parameters in atmospheric WSOC differ

260 from the characteristics of aquatic DOM.

261

262

263 **Table S1**. Summary of the climate, the concentration of total suspended particulate (TSP), WSOC,
264 WSTN, and ions, and optical characteristics of the aerosol samples in Karachi (Pakistan) during
265 different seasons.

	Annual (n=96)	Pre-monsoon (n=28)	Monsoon (n=26)	Post-monsoon (n=14)	Winter (n=28)
Meteorology					
Temperature (°C)	27±4.3	29±2.5	30±1.6	27±3.4	22±3.7
Pressure (Pa)	755±4.7	755±3.0	750±2.3	758±2.6	760±1.8
Relative humidity (%)	61±17	63±13	73±5.8	59±19	48±19
Wind speed (m s^{-1})	3.5±2.0	3.9±1.4	4.9±1.3	2.0±1.1	2.6±2.3
Visibility (km)	3.9±0.4	4.0±0	4.0±0	3.8±0.38	3.6±0.70
Concentrations					
TSP ($\mu g\ m^{-3}$)	417±141	468±135	390±133	378±120	412±154
WSOC ($\mu g\ m^{-3}$)	5.7±2.8	4.7±1.6	3.9±1.6	6.9±3.2	7.8±2.8
WSTN ($\mu g\ m^{-3}$)	2.1±1.1	1.9±0.46	1.2±0.52	2.7±0.8	2.8±1.5
*NO_3^- ($\mu g\ m^{-3}$)	5.6±4.5	4.2±1.1	2.6±1.5	7.3±3.3	8.8±6.4
*NH^+ ($\mu g\ m^{-3}$)	1.2±0.55	1.2±0.44	0.87±0.47	1.5±0.35	1.5±0.60
*SO_4^{2-} ($\mu g\ m^{-3}$)	17±6.0	15±2.9	17±5.7	18±4.4	18±8.6
Optical variables					
FI	1.7±0.18	1.7±0.14	1.7±0.23	1.7±0.19	1.7±0.15
BIX	0.80±0.15	0.74±0.10	0.77±0.17	0.92±0.17	0.83±0.13
HIX	3.6±1.21	4.1±0.92	3.4±1.5	2.8±0.65	3.7±1.2
S_R	1.3±0.21	1.4±0.18	1.4±0.22	1.3±0.16	1.2±0.17
A_{254} (m^{-1})	0.07±0.05	0.04±0.02	0.04±0.03	0.12±0.05	0.12±0.05
$SUVA_{254}$ ($m^2\ g^{-1}$ C)	1.7±0.60	1.5±0.42	1.2±0.37	2.0±0.51	2.1±0.56
Abs_{365} (Mm^{-1})	5.1±4.2	3.3±1.7	2.5±2.0	6.9±5.3	8.5±4.4
MAE_{365} ($m^2\ g^{-1}$ C)	0.81±0.40	0.69±0.27	0.59±0.27	0.92±0.36	1.1±0.46
AAE	4.5±0.69	4.4±0.66	4.4±0.85	4.6±0.61	4.5±0.59
C1%	57±5%	59±5%	55±6%	58±2%	58±4%
C2%	23±4%	22±3%	21±5%	21±3%	25±6%
C3%	20±6%	19±4%	24±7%	21±4%	18±6%

266   *described in our previous study (Zong et al., 2020)

267

268   **Table S2**. The MAE$_{365}$ and AAE of WSOC reported in recent studies across South Asia.

Site	Size	Sampling period	AAE	Fitting wavelength	MAE$_{365}$ (m^2 g^{-1} C)	References
Karachi, Pakistan	TSP	2016−2017	4.5±0.69	330−400	0.81±0.40	In this study
New Delhi	PM$_{2.5}$		3.2±0.22		2.5±0.3	
Bhola Island in the delta of Bay of Bengal (BCOB)	PM$_{2.5}$	2016 winter	5.4±0.30	330−400	1.4±0.2	Dasari et al. (2019)
Northern region of Maldives (MCOH), Indian Ocean	PM$_1$		6.9±0.26		0.4±0.1	
MCOH	PM$_{2.5}$	2012 winter	7.2±0.7	330−400	0.46±0.18	Bosch et al. (2014)
Kharagpur, Indo-Gangetic Plain (IGP)	PM$_{2.5}$	2009−2010 winter	6±1.1	330−480	0.78±0.24	Srinivas and Sarin (2014)
Bay of Bengal (IGP outflow)	PM$_{10}$/TSP	2008−2009 winter	9.1±2.5	330−480	0.45±0.14	Srinivas and Sarin (2013)
Godavari, Nepal	PM$_{10}$	2012−2014	5.23±0.52	300-400	0.77±0.23	Wu et al. (2019)
New Delhi, India	PM2.5	2010 winter	5.1±2	330-400	1.6±0.5	Kirillova et al. (2014)
Kanpur, IGP	PM$_{2.5}$	2014−2015 winter	4.4±0.4	300−700	2.5±0.5	Choudhary et al. (2021)
		2014−2015 monsoon	3.8±0.6	300−700	1.3±0.3	
Patiala, IGP	PM$_{2.5}$	2011 winter	5.1±1.9	300−700	1.3±0.7	Srinivas et al. (2016)
Kanpur, IGP	PM$_1$	Wintertime	2.8±0.6	300−700	1.8±0.2	Choudhary et al. (2018)
Kathmandu, Nepal	TSP	2013−2018	4.6±0.8	300−400	1.4±0.3	Chen et al. (2020)

269

270   **Table S3.** The number of formulas, average molecular weight (MW), and structural groupings of molecular formulas determined by FT-ICR MS in 12 aerosol samples.

271   The table also showed the assigned molecular formulas ($r \geq 0.57$) to three fluorescence components (C1–3), and those which did not correlate with any of the PARAFAC

272   components.

	All ESI– compounds	no assigned molecules	C1-	C2-	C3-	All ESI+ compounds	no assigned molecules	C1+	C2+	C3+
no. total formulas	7687	5993	134	1328	284	6564	5054	208	1182	162
(% of number)	(100 %)	(78 %)	(2 %)	(17 %)	(4 %)	(100 %)	(77 %)	(3 %)	(18 %)	(2 %)
no. formula without N	3591	2865	49	495	201	2677	2138	82	374	101
(% of number)	(100 %)	(80 %)	(1 %)	(14 %)	(6 %)	(100 %)	(80 %)	(3 %)	(14 %)	(4 %)
no. formula with N	4096	3128	85	883	83	3887	2916	126	808	61
(% of number)	(100 %)	(76 %)	(2 %)	(22 %)	(2 %)	(100 %)	(75 %)	(3 %)	(21 %)	(2 %)
aliphatic compounds	2829	2522	45	138	135	3209	2690	104	358	74
(% of number)	(100 %)	(89 %)	(2 %)	(5 %)	(5 %)	(100 %)	(84 %)	(3 %)	(11 %)	(2 %)
highly unsaturated and phenolic compounds	3538	2681	47	689	136	2276	1495	82	657	60
(% of number)	(100 %)	(76 %)	(1 %)	(19 %)	(4 %)	(100 %)	(66 %)	(4 %)	(29 %)	(3 %)
aromatic compounds	817	552	16	248	10	647	533	13	85	21
(% of number)	(100 %)	(68 %)	(2 %)	(30 %)	(1 %)	(100 %)	(82 %)	(2 %)	(13 %)	(3 %)
condensed aromatic compounds	503	238	26	253	3	432	336	9	82	7
(% of number)	(100 %)	(47 %)	(5 %)	(50 %)	(1 %)	(100 %)	(78 %)	(2 %)	(19 %)	(2 %)
CHO compounds	2063	1483	30	449	118	541	378	34	122	14
(% of number)	(100 %)	(72 %)	(1 %)	(22 %)	(6 %)	(100 %)	(70 %)	(6 %)	(23 %)	(3 %)
CHON compounds	3191	2275	76	823	49	3603	2654	121	792	58

(% of number)	(100 %)	(71 %)	(2 %)	(26 %)	(2 %)	(100 %)	(74 %)	(3 %)	(22 %)	(2 %)
CHOS compounds	1528	1382	19	46	83					
(% of number)	(100 %)	(90 %)	(1 %)	(3 %)	(5 %)					
CHONS compounds	905	853	9	10	34					
(% of number)	(100 %)	(94 %)	(1 %)	(1 %)	(4 %)					
CHONa+ compounds						2136	1760	48	252	87
(% of number)						(100 %)	(82 %)	(2 %)	(12 %)	(4 %)
CHN+ compounds						284	262	5	16	3
(% of number)						(100 %)	(92 %)	(2 %)	(6 %)	(1 %)
Average MW	390±99	394±100	400±112	380±91	361±84	430±140	442±142	443±127	367±110	474±142
DBE	7.4±3.9	6.9±3.9	8.1±3.0	9.3±3.6	6.3±3.4	7.5±4.9	7.4±5.1	7.6±4.2	7.3±3.7	8.4±6.9
AI_{mod}	0.25±0.26	0.22±0.24	0.31±0.33	0.39±0.29	0.19±0.17	0.25±0.24	0.24±0.25	0.22±0.20	0.27±0.22	0.25±0.23
O/C	0.46±0.24	0.44±0.24	0.48±0.18	0.59±0.20	0.30±0.12	0.29±0.20	0.26±0.19	0.28±0.16	0.39±0.19	0.24±0.20
H/C	1.3±0.40	1.4±0.39	1.2±0.43	1.1±0.35	1.5±0.32	1.5±0.39	1.5±0.39	1.5±0.33	1.4±0.33	1.5±0.43
NOSC	-0.23±0.72	-0.33±0.69	-0.059±0.68	0.36±0.52	-0.77±0.35	-0.72±0.53	-0.79±0.50	-0.82±0.43	-0.37±0.54	-0.91±0.44
$\overline{OS}_C$	-0.43±0.61	-0.53±0.57	-0.25±0.59	0.12±0.47	-0.89±0.33	-0.91±0.46	-0.98±0.43	-0.93±0.44	-0.60±0.50	-1.0±0.39
N/C	0.050±0.062	0.045±0.057	0.055±0.062	0.076±0.078	0.021±0.039	0.064±0.084	0.062±0.084	0.038±0.046	0.079±0.086	0.043±0.071

273

274 **Table S4.** The formulas assigned to C1–3 with 0N, 1N, 2N, 3N, and 4N. ESI– and ESI+ refer to
275 formulas detected in the negative and positive modes, respectively.

	Ion sources	Total	0N	1N	2N	3N	4N
C1	ESI–	134	49	66	9	10	0
		(100 %)	(37 %)	(49 %)	(7 %)	(7 %)	(0 %)
	ESI+	208	82	87	15	23	1
		(100 %)	(39 %)	(42 %)	(7 %)	(11 %)	(0 %)
C2	ESI–	1328	495	328	282	213	10
		(100 %)	(37 %)	(25 %)	(21 %)	(16 %)	(1 %)
	ESI+	1182	374	316	311	162	19
		(100 %)	(32 %)	(27 %)	(26 %)	(14 %)	(2 %)
C3	ESI–	284	201	74	9	0	0
		(100 %)	(71 %)	(26 %)	(3 %)	(0 %)	(0 %)
	ESI+	162	101	18	5	26	12
		(100 %)	(62 %)	(11 %)	(3 %)	(16 %)	(7 %)

276

277 **Table S5.** CHO molecules detected in secondary organic aerosol (SOA) from limonene ozonolysis
278 (Kundu et al., 2012) that overlapped with CHO formulas assigned to C1–3, respectively. The
279 number of CHO (neutral molecules) in this study contains the sum of [CHO–H]$^-$, [CHO+H]$^+$, and
280 [CHO+Na]$^+$, and the overlapped molecules only count once.

Types	CHO formulas	m/z
Limonene ozonolysis SOA	1197 (100 %)	517±178
C1	46 (42 %)	438±121
C2	244 (34 %)	364±90
C3	11 (5.3 %)	474±77

281
282

283 **Table S6.** List of identified CHO products of naphthalene photooxidation, biomass-burning-
284 emission, and aqueous-phase reactions of phenols detected in this study.

Molecular formula	C1	C2	C3	Compound type	References
$C_8H_6O_2$		1			
$C_8H_6O_5$		1			
$C_{10}H_6O_6$		1			
$C_{10}H_8O_5$		1			
$C_{10}H_8O_6$		1		Naphthalene	(Siemens et
$C_{11}H_{12}O_7$		1		photooxidation	al., 2022)
$C_{17}H_{10}O_6$		1			
$C_{18}H_{10}O_6$		1			
$C_{18}H_{12}O_{10}$		1			
$C_{19}H_{18}O_{10}$		1			
$C_5H_8O_3$		1			
$C_6H_{12}O_6$		1			
$C_7H_{12}O_2$		1			
$C_8H_8O_6$		1			
$C_8H_8O_8$		1			
$C_9H_{18}O_4$		1			
$C_9H_6O_4$	1	1			
$C_9H_6O_5$		1			
$C_9H_6O_6$	1	1			
$C_9H_6O_7$		1			
$C_{13}H_8O_7$		1			
$C_{13}H_8O_8$		1			
$C_{13}H_8O_9$		1			
$C_{14}H_{14}O_6$		1			(Mo et al.,
$C_{14}H_{14}O_7$		1		Biomass burning	2022;Smith et
$C_{14}H_{14}O_{10}$		1		emission	al., 2009)
$C_{14}H_{14}O_{11}$		1			
$C_{14}H_{14}O_{12}$		1			
$C_{16}H_{18}O_{10}$		1			
$C_{16}H_{18}O_{11}$		1			
$C_{16}H_{18}O_{12}$		1			
$C_{16}H_{18}O_7$		1			
$C_{16}H_{18}O_9$		1			
$C_{16}H_{24}O_7$		1			
$C_{19}H_{26}O_3$		1			
$C_{20}H_{26}O_2$		1			
$C_{20}H_{26}O_3$		1			
$C_{20}H_{28}O_2$			1		
$C_{20}H_{30}O_2$		1			

Formula			Source	Reference
$C_{24}H_{44}O_4$		1		
$C_{24}H_{46}O_4$		1		
$C_{12}H_{10}O_7$	1			
$C_{12}H_{12}O_7$	1			
$C_{12}H_{12}O_6$	1			
$C_{15}H_{14}O_6$	1		Aqueous-phase reactions of phenols	(Yu et al., 2014)
$C_{15}H_{14}O_8$	1			
$C_{15}H_{16}O_9$	1			
$C_{15}H_{18}O_7$	1			
$C_{16}H_{18}O_9$	1			

285

286

287 **Table S7.** Permutation tests of RDA results of a global test first, followed by a test of the canonical
288 axes, and the explanatory variables. The number of permutations is 999.

Model: rda (formula = EEM_hel ~ FI + BIX + HIX + S_R + A_{254} + $SUVA_{254}$, data = env, scale = FALSE)

Global model

	Df	Variance	F	Pr(>F)
Model	6	0.0055059	37.163	0.001***
Residual	89	0.0021977		

Forward tests for axes

	Df	Variance	F	Pr(>F)
RDA1	1	0.0048303	202.2108	0.001***
RDA2	1	0.0006743	28.2299	0.001***
RDA3	1	0.0000012	0.0521	1.000
Residual	92	0.0021977		

Terms added sequentially (first to last)

	Df	Variance	F	Pr(>F)
FI	1	0.00226944	91.9071	0.001***
BIX	1	0.00073770	29.8754	0.001***
HIX	1	0.00101865	41.2529	0.001***
S_R	1	0.00092849	37.6019	0.001***
A_{254}	1	0.00041732	16.9006	0.001***
$SUVA_{254}$	1	0.00013430	5.4387	0.010**
Residual	89	0.00219765		

289 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
290

291 **Table S8.** Forward selection using vegan's **ordiR2step()** function using the two criteria (the
292 predefined α level and R^2_{adj} of the global model).

Step: R^2.adj = 0

Call: EEM_hel ~ 1

	R^2.adjusted			
<All variables>	0.69548999			
+ HIX	0.4689567			
+ BIX	0.30136019			
+ FI	0.28709195			
+ S_R	0.19795852			
+ A_{254}	0.03880186			
+ SUVA$_{254}$	0.01804164			
<none>	0			
	Df	AIC	F	Pr(>F)
+ HIX	1	-525.92	84.893	0.001***

Step: R^2.adj = 0.4689567

Call: EEM_hel~ HIX

	R^2.adjusted			
<All variables>	0.69549			
+ S_R	0.5885705			
+ A_{254}	0.532254			
+ FI	0.5017867			
+ SUVA$_{254}$	0.4943887			
+ BIX	0.4734882			
<none>	0.4689567			
	Df	AIC	F	Pr(>F)
+ SR	1	-549.45	28.328	0.001***

Step: R^2.adj = 0.5885705

Call: EEM_hel ~ HIX + S_R

	R2.adjusted			
<All variables>	0.69549			
+ A_{254}	0.6414914			
+ FI	0.6241883			
+ SUVA$_{254}$	0.6146478			
+ BIX	0.5945623			
<none>	0.5885705			
	Df	AIC	F	Pr(>F)
+ A_{254}	1	-561.71	14.728	0.001***

Step: R^2.adj = 0.6414914

Call: EEM_hel ~ HIX + S_R + A_{254}

	R^2.adjusted
<All variables>	0.69549

		Df	AIC	F	Pr(>F)
+ FI	0.661776				
+ BIX	0.6588105				
+ SUVA$_{254}$	0.6521977				
<none>	0.6414914				
	Df		AIC	F	Pr(>F)
+ FI	1		-566.35	6.5176	0.007**

Step: R^2.adj = 0.661776

Call: EEM_hel ~ HIX + S$_R$ + A$_{254}$ + FI

	R^2.adjusted	Df	AIC	F	Pr(>F)
<All variables>	0.69549				
+ BIX	0.680472				
+ SUVA254	0.6667591				
<none>	0.661776				
	Df		AIC	F	Pr(>F)
+ BIX	1		-570.87	6.3246	0.011*

Step: R^2.adj = 0.680472

Call: EEM_hel ~ HIX + S$_R$ + A$_{254}$ + FI + BIX

	R2.adjusted	Df	AIC	F	Pr(>F)
<All variables>	0.69549				
+ SUVA254	0.69549				
<none>	0.680472				
	Df		AIC	F	Pr(>F)
+ SUVA254	1		-574.56	5.4387	0.016*

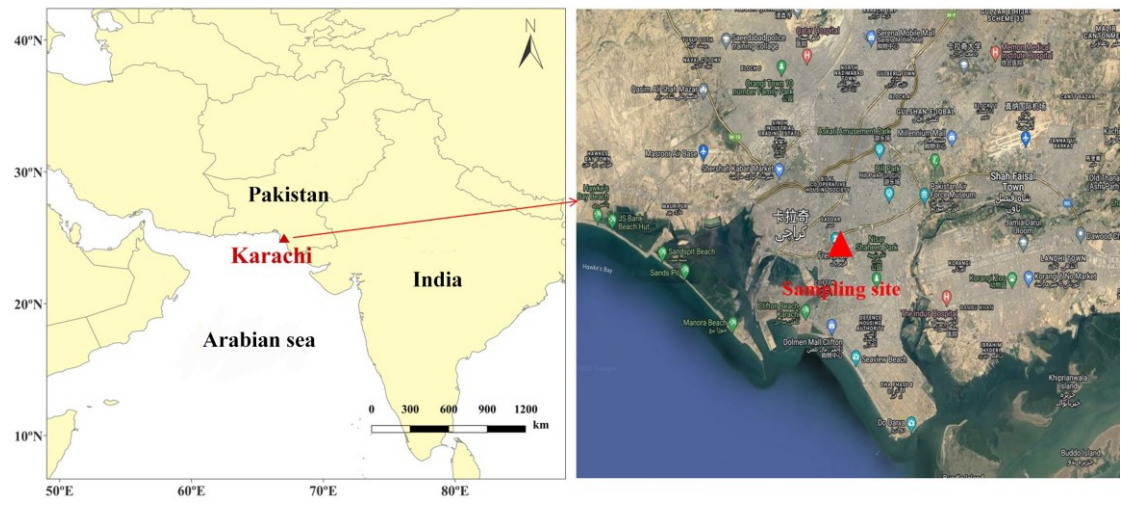
Step: R^2.adj = 0.69549

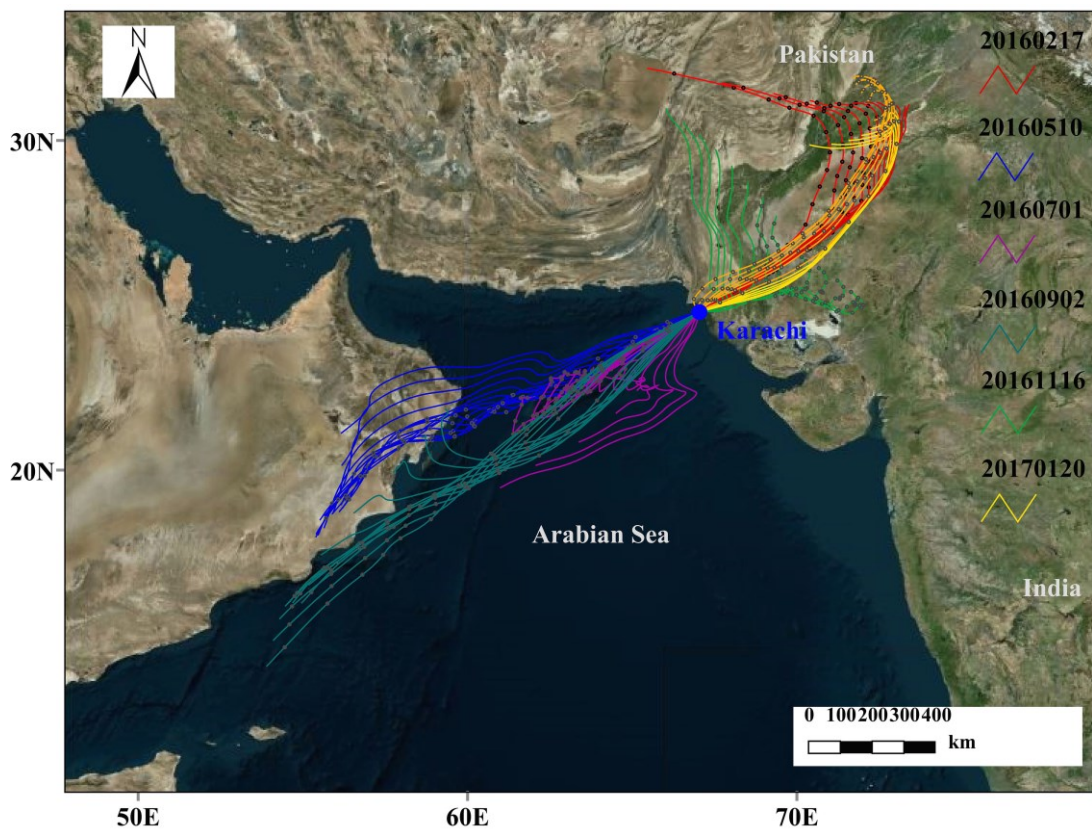
Call: EEM_hel ~ HIX + S$_R$ + A$_{254}$ + FI + BIX + SUVA$_{254}$

293 Signif. codes: 0'***' 0.001'**' 0.01'*' 0.05'.' 0.1'' 1

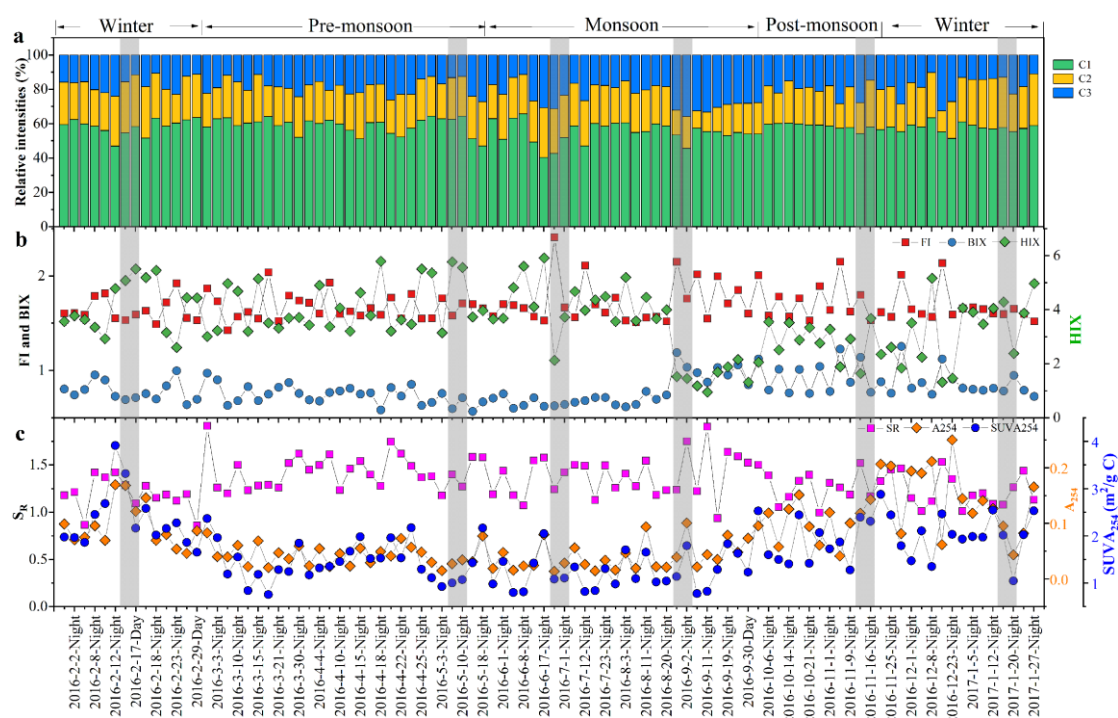
294

295

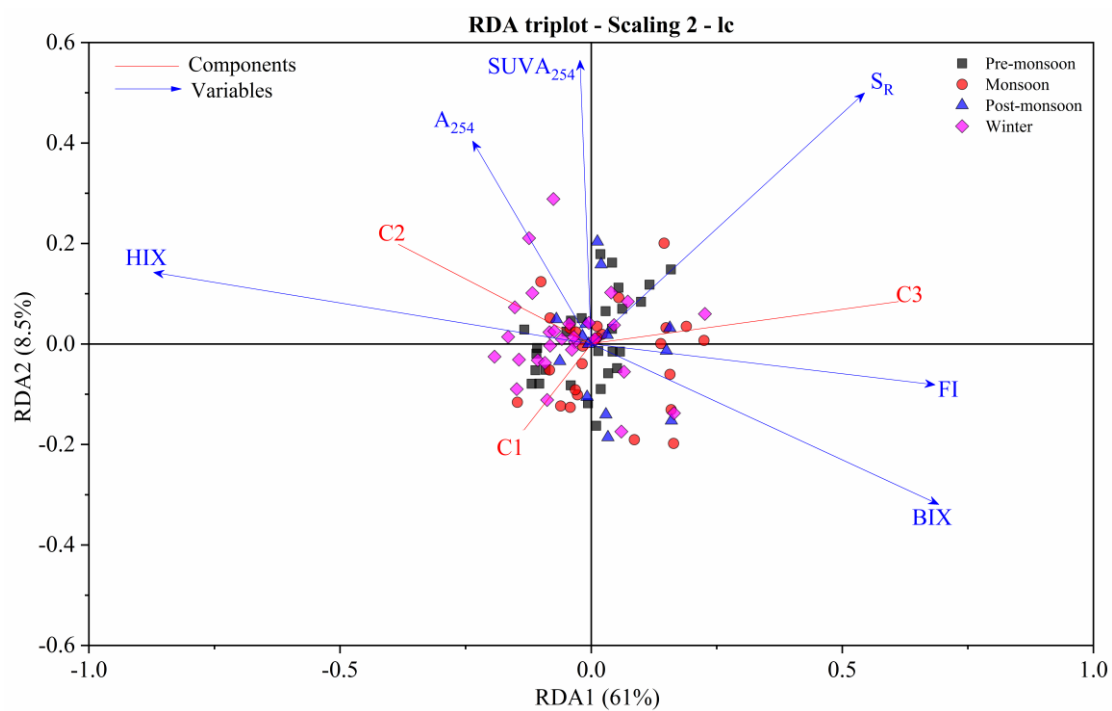
296

297 **Figure S1**. Detailed location of sampling site in this study. The satellite image at the right was
298 derived from Google Maps (Image © Google Maps 2019).

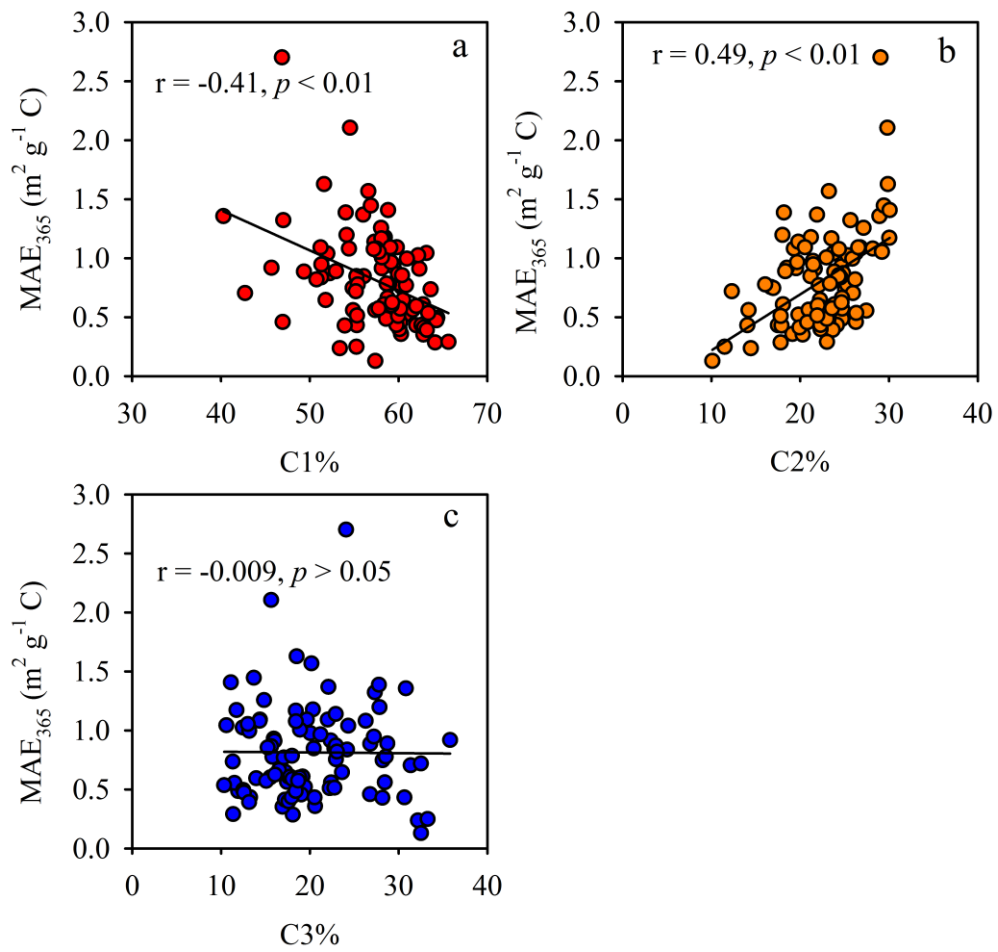
299

300

301 **Figure S2.** The 72 h back air-mass trajectories at Karachi from Pakistan for the selected samples
302 via FT-ICR MS analysis on February 16, May 10, July 1, September 2, November 16, 2016, and
303 January 20, 2017, correspond to different seasons. The trajectories with black dots represent the
304 corresponding night, otherwise, it is the day. The air-mass trajectories were analyzed by the
305 HYSPLIT model. The base map was derived from Bing Maps (© 2023 Microsoft).

306

307

308 **Figure S3**. The relative abundances of PARAFAC components (C1, C2, and C3), **a**; the fluorescence
309 index (FI), biological index (BIX), humification index (HIX) values, **b**; and the spectra slope (S_R),
310 absorption at 254 nm (A_{254}), specific ultraviolet absorbance at 254 nm (SUVA$_{254}$), **c**; in the WSOC
311 over Karachi, Pakistan. The shade regions refer to the selected samples for FT-ICR MS analysis.
312

313

314

315 **Figure S4.** RDA triplot in Scaling 2, Hellinger-transformed PARAFAC component (red line)
316 constrained by significant six optical parameters variables (blue arrow) ($p < 0.01$). The dots
317 represent the fitted site scores, and different colored dots refer to season variables.
318
319

320
321 **Figure S5.** Correlation of MAE_{365} and relative abundance of PARAFAC components (C1–3, a–c).
322

323

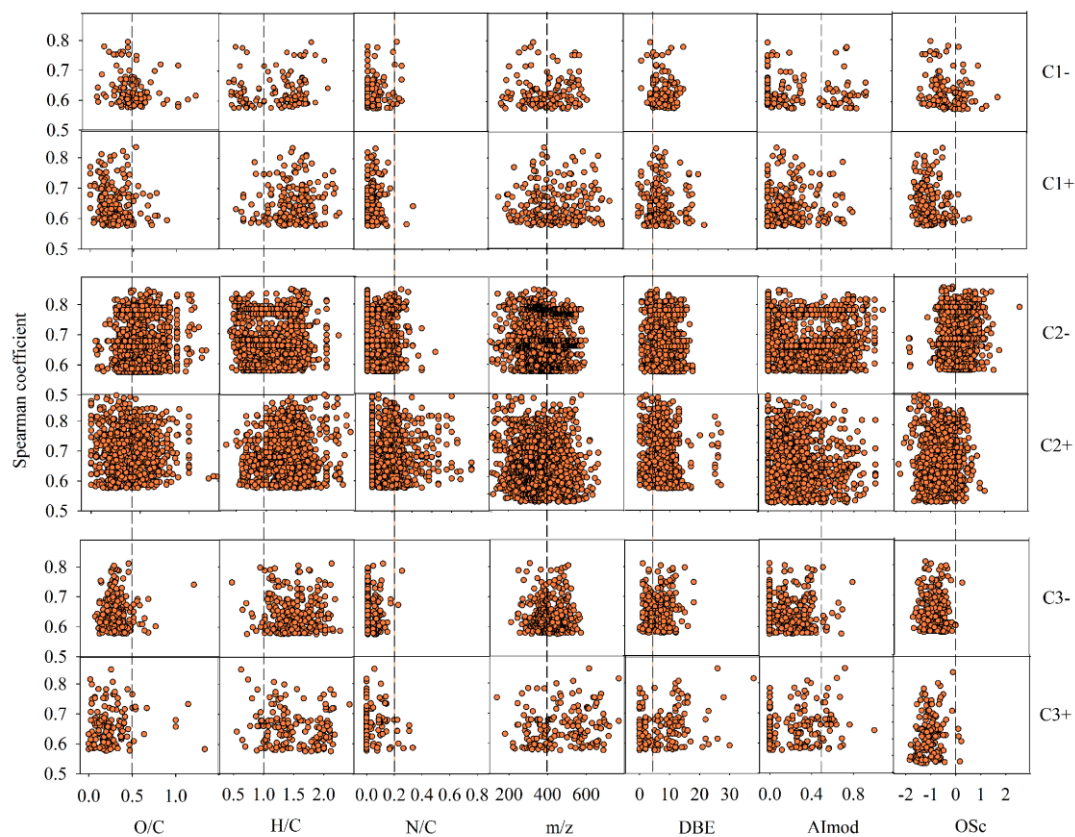



Figure S6. The strength and spread of significant Spearman correlations ($r_s > 0.57$; $p < 0.05$) along compositional gradients for each fluorescence component (C1, C2, and C3). The dashed line, for reference, is located at O/C = 0.5, H/C = 1.0, N/C=0.2, m/z = 400, DBE = 4, AI_{mod} = 0.5, and carbon oxidation state (OSc) = 0.

329

330

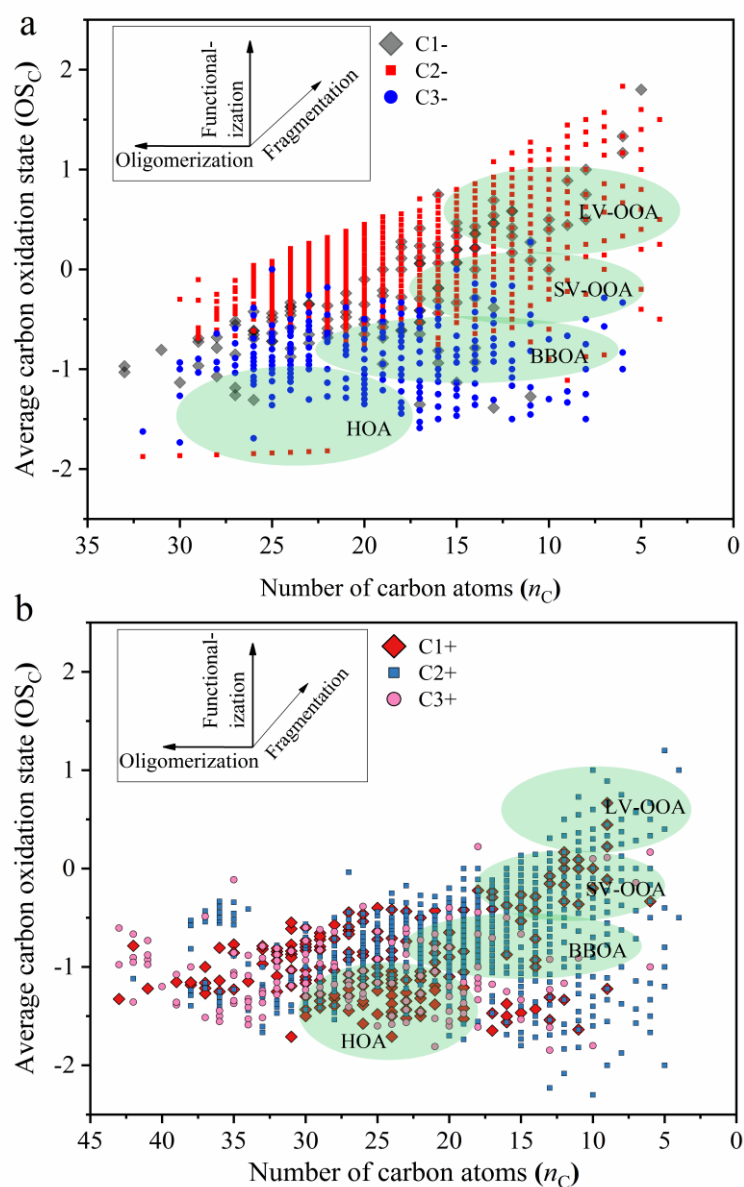
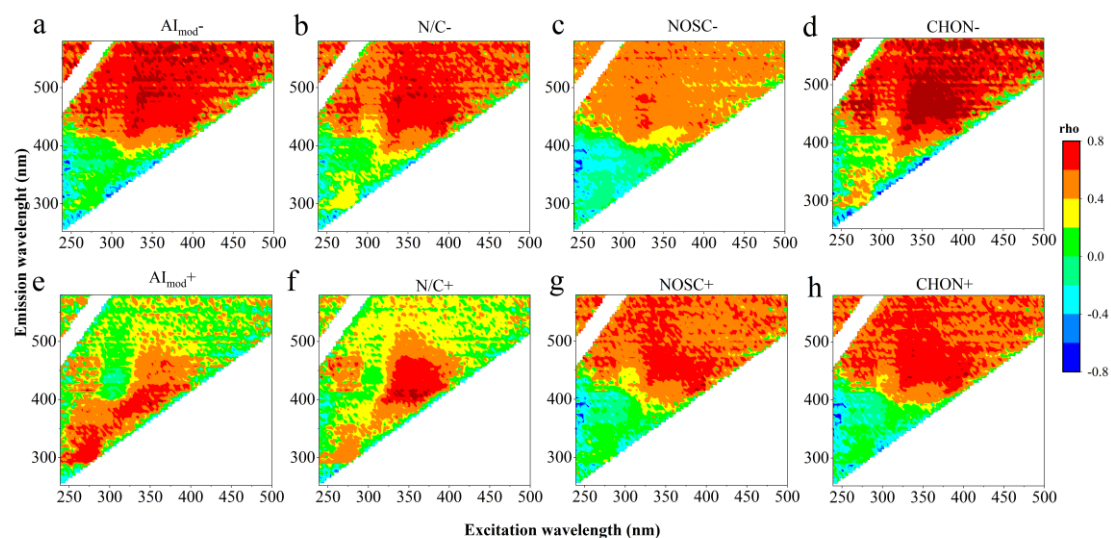
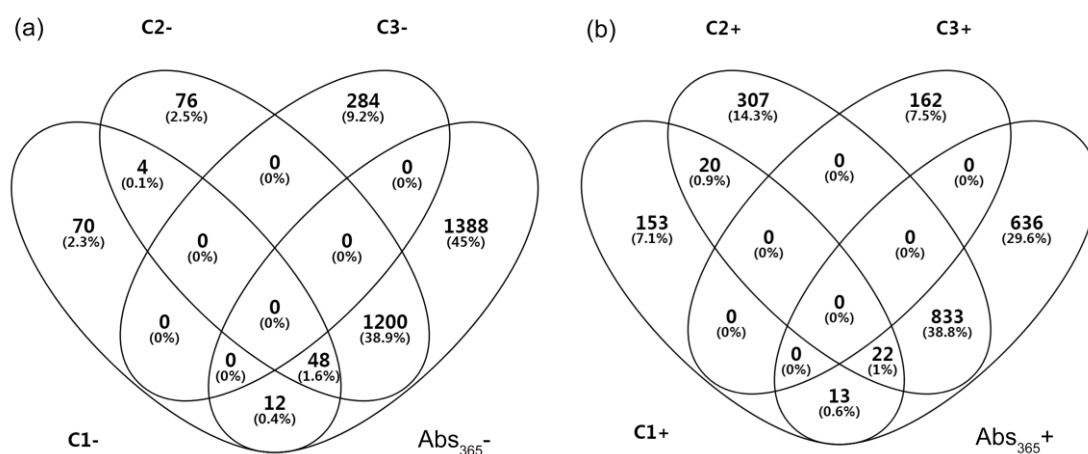
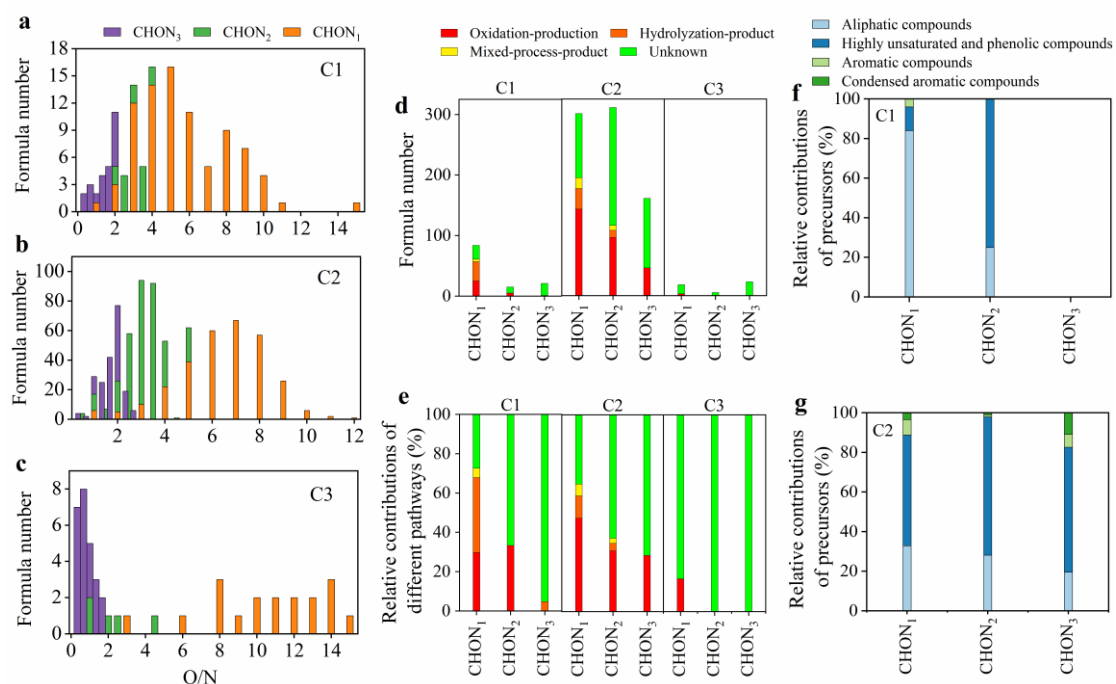



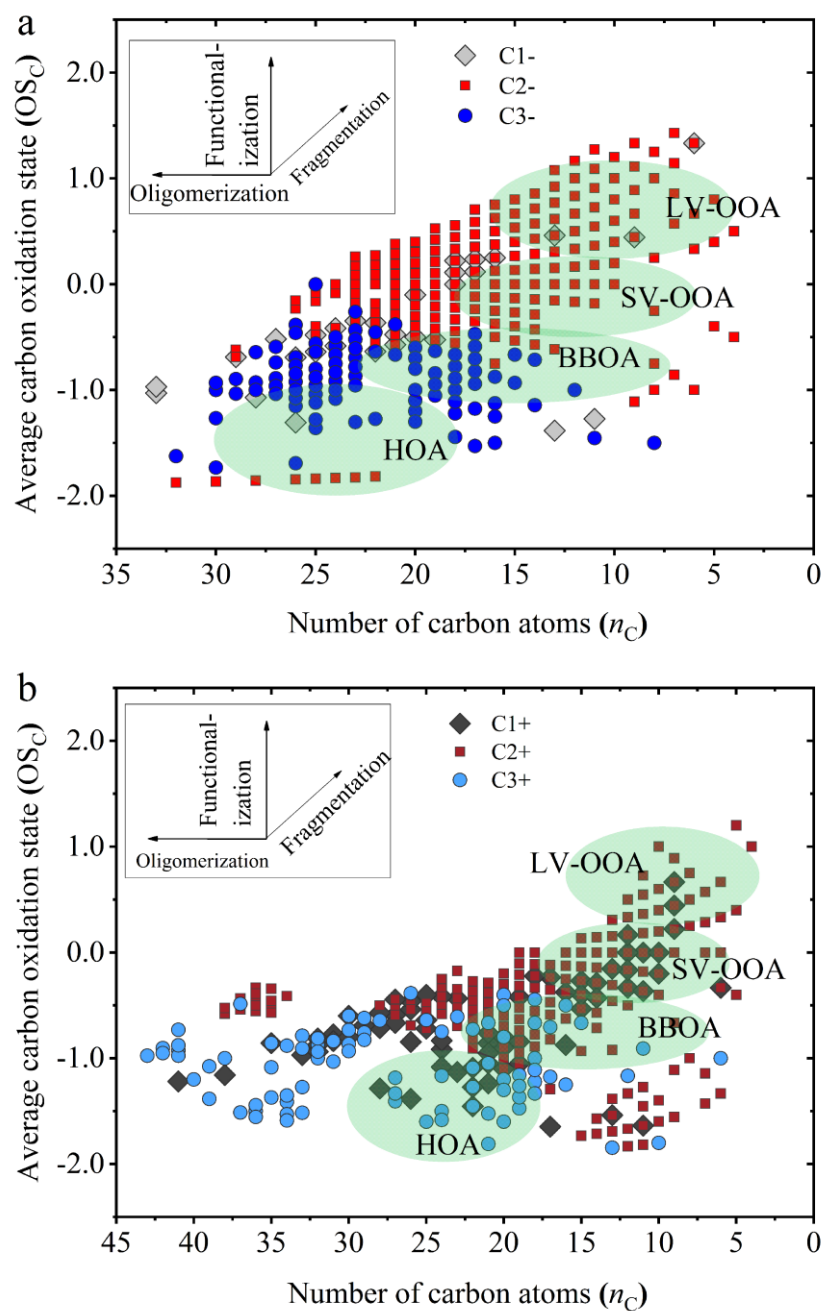
Figure S7. Location in n_c-$\overline{OS}_C$ space of fluorescent components-assigned molecular formulas determined by FT-ICR MS. The range in black circles indicated locations of different organic aerosol classes (Kroll et al., 2011). Hydrocarbon-like organic aerosol (HOA) and biomass-burning organic aerosol (BBOA) correspond to primary particulate matter directly emitted into the atmosphere. Semi-volatile and low-volatility oxidized organic aerosol (SV-OOA and LV-OOA) correspond to 'fresh' and 'aged' secondary aerosol. Inset: vectors corresponding to key classes of reactions of atmospheric organics: functionalization (addition of polar functional groups), fragmentation (cleavage of C–C bonds), and oligomerization (covalent association of two organic species).

340

341 **Figure S8**. Relationships between the normalized fluorescence intensity at each excitation-emission
342 pair and molecular-level derived compositional characteristics. Spearman correlations were
343 calculated between the fluorescence intensity normalized to the total fluorescence intensity at each
344 excitation-emission wavelength pair and the intensity-weighted average of molecular characteristics
345 derived from FT-ICR-MS detected compounds. Molecular characteristics included AI_{mod}, N/C,
346 NOSC, and CHON. AI_{mod}- and AI_{mod}+ refer to the aromaticity index of compounds detected in the
347 ESI– and ESI+ modes, respectively, and the others are defined similarly. Associations with rho >
348 0.57 were considered significant (t-test).

349

350

351

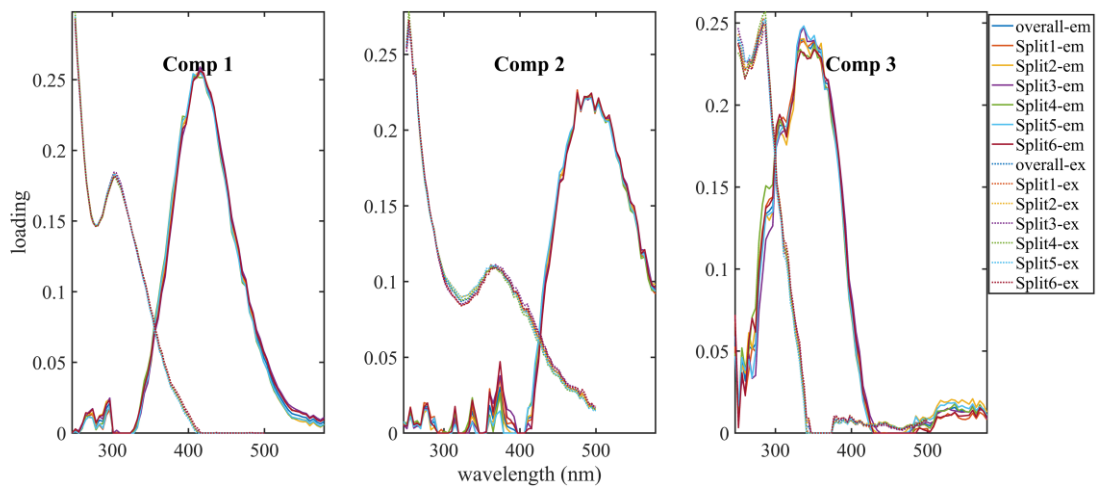
352 **Figure S9**. Venn diagrams of C1–3- and BrC-assigned formulas in the ESI– (a) and ESI+ (b) modes,
353 respectively. The areas of overlap are the common elements in both, three, and four species. The
354 BrC-assigned formulas were obtained similarly to that of PARAFAC components. Note that we
355 selected the light absorption coefficient at 365 nm (Abs$_{365}$, detail in Text S2) as a proxy of BrC,
356 because of the light absorption coefficient at other wavelengths (250 nm, 280 nm, 400 nm, 500 nm,
357 and 550 nm) were strongly correlated with Abs$_{365}$ (r = 0.98~0.99, $p < 0.000$).
358

359
360 **Figure S10**. O/N ratios distribution of CHON+ compounds assigned to C1 (a), C2 (b), and C3 (c),
361 and the corresponding formula number distribution of different pathways for CHON+ molecules
362 assigned to C1–3 (d) and the relative contributions of different pathways (e). In addition, panels (f)
363 and (g) refer to the relative contributions of precursor types to the oxidation-product pair for $CHON_{1-3}$
364 $_3$ of C1 and C2, respectively.

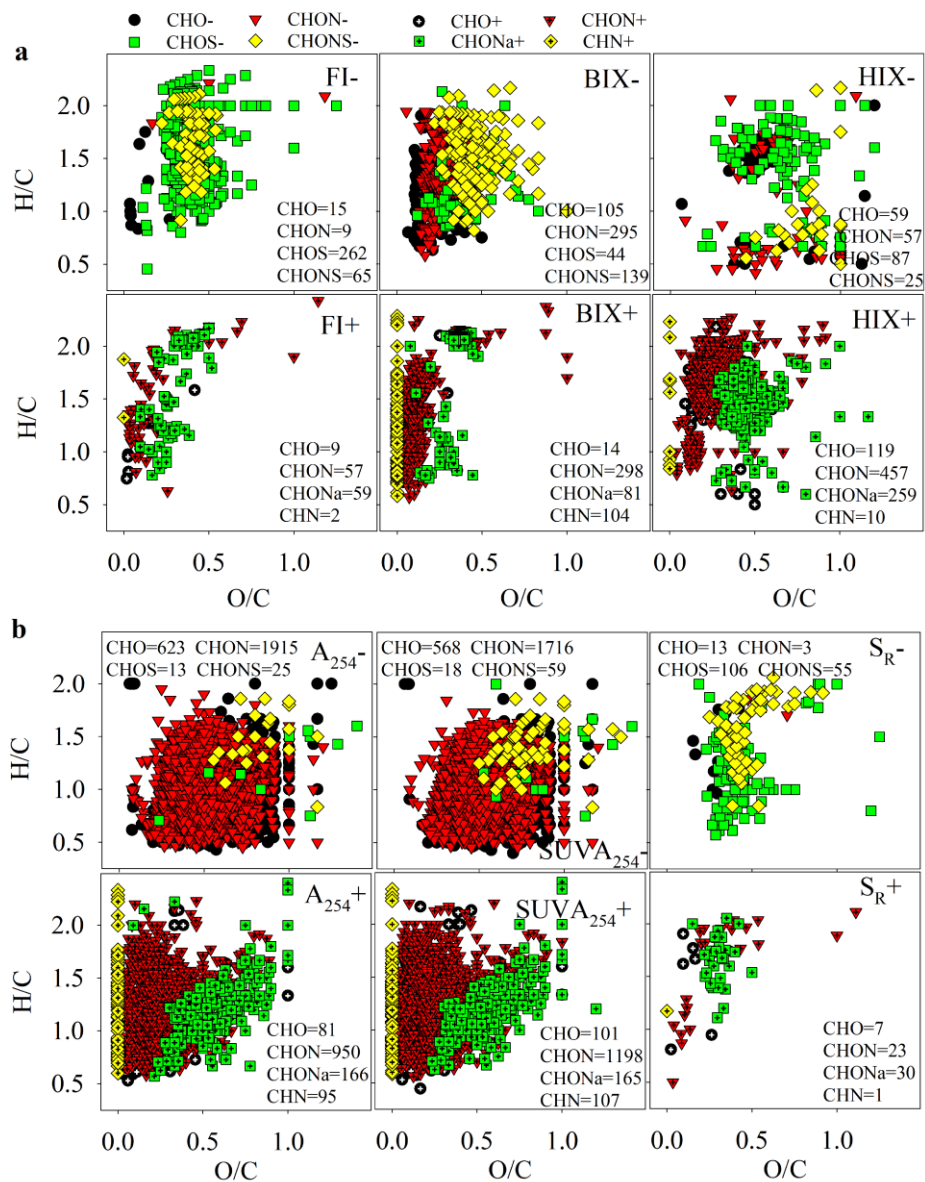
365

366 **Figure S11.** Location in n_c-$\overline{OS}_C$ space of CHO compounds assigned to PARAFAC components.

367

368

369

370

371 **Figure S12.** Split-half analysis of 3-component PARAFAC model with the split style 'S$_4$C$_6$T$_3$' for

372 all EEMs of WSOC from Karachi, Pakistan.

373

374 **Figure S13**. van Krevelen diagrams of FT-ICR MS-identified compounds assigned to fluorescent

375 indices (FI, BIX, and HIX, **a**), and light absorption indices (A_{254}, $SUVA_{254}$, and S_R, **b**), respectively.

376 FI- and FI+ refer to the molecules that are assigned to FI in the ESI– and ESI+ mode, respectively,

377 and the others are defined analogously.

378

379

380

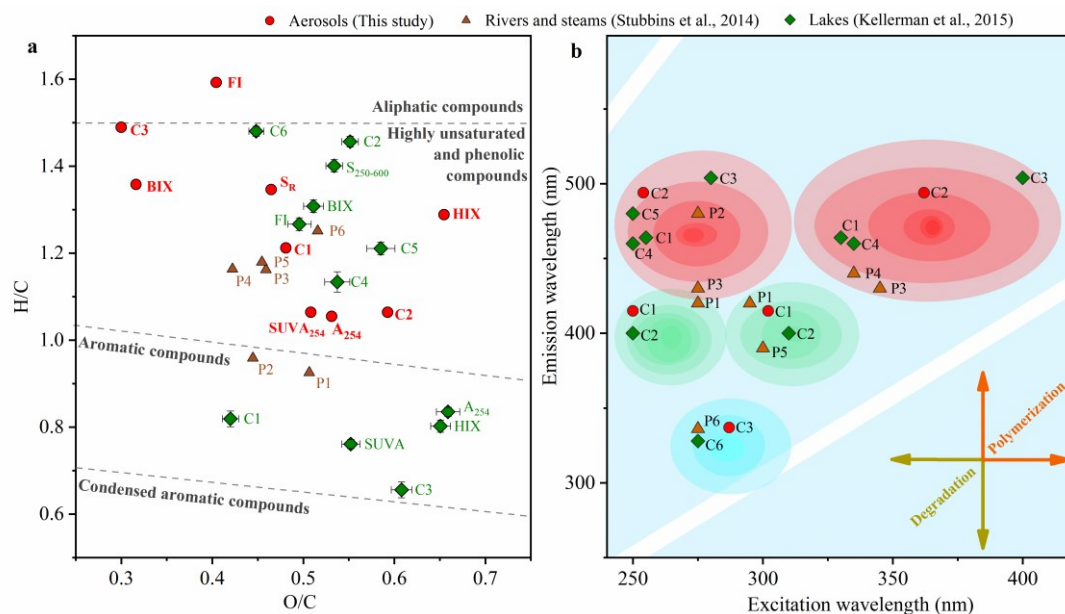



Figure S14. (**a**) Comparison of average chemical characteristics of molecules assigned to C1, C2, and C3, as well as optical indices of WSOC in the aerosols (marked as a red dot), DOM from rivers and streams (brown triangle, fluorescent component P1–6) and lakes (green diamond, fluorescent component C1–6, and FI, BIX, HIX, SUVA, A_{254}, and $S_{250-600}$); (**b**) the corresponding excitation and emission maxima in the EEM spectra (Stubbins et al., 2014; Kellerman et al., 2015). Note that the fluorescent components associated with molecular data (**a**) of DOM in rivers and streams were re-extracted from its supplement and the error bars were not shown. The error bars in this study were shown in Table S3. The range of the red circle, green circle, and blue circle (**b**) represent the range of HULIS-1, HULIS-2, and PLOM fluorescence, respectively, and the arrows indicate polymerization and degradation defined according to the study (Chen et al., 2016).

391

References:

Bahram, M., Bro, R., Stedmon, C., and Afkhami, A.: Handling of Rayleigh and Raman scatter for PARAFAC modeling of fluorescence data using interpolation, J. Chemom., 20, 99-105, https://doi.org/10.1002/cem.978, 2006.

Chen, Q., Miyazaki, Y., Kawamura, K., Matsumoto, K., Coburn, S., Volkamer, R., Iwamoto, Y., Kagami, S., Deng, Y., Ogawa, S., Ramasamy, S., Kato, S., Ida, A., Kajii, Y., and Mochida, M.: Characterization of Chromophoric Water-Soluble Organic Matter in Urban, Forest, and Marine Aerosols by HR-ToF-AMS Analysis and Excitation-Emission Matrix Spectroscopy, Environ. Sci. Technol., 50, 10351-10360, https://doi.org/10.1021/acs.est.6b01643, 2016.

Helms, J. R., Stubbins, A., Ritchie, J. D., Minor, E. C., Kieber, D. J., and Mopper, K.: Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter, Limnol. Oceanogr., 53, 955-969, https://doi.org/10.4319/lo.2008.53.3.0955, 2008.

Huguet, A., Vacher, L., Relexans, S., Saubusse, S., Froidefond, J. M., and Parlanti, E.: Properties of fluorescent dissolved organic matter in the Gironde Estuary, Org. Geochem., 40, 706-719, https://doi.org/10.1016/j.orggeochem.2009.03.002, 2009.

Jiang, H., Tang, J., Li, J., Zhao, S., Mo, Y., Tian, C., Zhang, X., Jiang, B., Liao, Y., Chen, Y., and Zhang, G.: Molecular Signatures and Sources of Fluorescent Components in Atmospheric Organic Matter in South China, Environ. Sci. Technol. Lett., 9, 913-920, https://doi.org/10.1021/acs.estlett.2c00629, 2022.

Kellerman, A. M., Kothawala, D. N., Dittmar, T., and Tranvik, L. J.: Persistence of dissolved organic matter in lakes related to its molecular characteristics, Nat. Geosci., 8, 454-U452, https://doi.org/10.1038/ngeo2440, 2015.

Koch, B., and Dittmar, T.: Koch, B. P. & Dittmar, T. From mass to structure: An aromaticity index for high-resolution mass data of natural organic matter. Rapid Commun. Mass Spectrom. 20, 926-932, Rapid Commun. Mass Spectrom., 20, 926-932, https://doi.org/10.1002/rcm.2386, 2006.

Kothawala, D. N., Murphy, K. R., Stedmon, C. A., Weyhenmeyer, G. A., and Tranvik, L. J.: Inner filter correction of dissolved organic matter fluorescence, Limnology and Oceanography: Methods, 11, 616-630, https://doi.org/10.4319/lom.2013.11.616, 2013.

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.

Kundu, S., Fisseha, R., Putman, A. L., Rahn, T. A., and Mazzoleni, L. R.: High molecular weight SOA formation during limonene ozonolysis: insights from ultrahigh-resolution FT-ICR mass spectrometry characterization, Atmos. Chem. Phys., 12, 5523-5536, https://doi.org/10.5194/acp-12-5523-2012, 2012.

Lawaetz, A. J., and Stedmon, C. A.: Fluorescence intensity calibration using the Raman scatter peak of water, Appl. Spectrosc., 63, 936-940, https://doi.org/10.1366/000370209788964548, 2009.

Lee, H. J., Laskin, A., Laskin, J., and Nizkorodov, S. A.: Excitation-emission spectra and

fluorescence quantum yields for fresh and aged biogenic secondary organic aerosols, Environ. Sci. Technol., 47, 5763-5770, https://doi.org/10.1021/es400644c, 2013.

Li, M., Fan, X., Zhu, M., Zou, C., Song, J., Wei, S., Jia, W., and Peng, P.: Abundances and light absorption properties of brown carbon emitted from residential coal combustion in China, Environ. Sci. Technol., 53, 595-603, https://doi.org/10.1021/acs.est.8b05630, 2018.

Lin, P., Huang, X.-F., He, L.-Y., and Zhen Yu, J.: Abundance and size distribution of HULIS in ambient aerosols at a rural site in South China, J. Aerosol Sci, 41, 74-87, https://doi.org/10.1016/j.jaerosci.2009.09.001, 2010.

Lin, P., Rincon, A. G., Kalberer, M., and Yu, J. Z.: Elemental composition of HULIS in the Pearl River Delta Region, China: results inferred from positive and negative electrospray high resolution mass spectrometric data, Environ. Sci. Technol., 46, 7454-7462, https://doi.org/10.1021/es300285d, 2012.

Lv, J., Zhang, S., Luo, L., and Cao, D.: Solid-phase extraction-stepwise elution (SPE-SE) procedure for isolation of dissolved organic matter prior to ESI-FT-ICR-MS analysis, Anal. Chim. Acta, 948, 55-61, https://doi.org/10.1016/j.aca.2016.10.038, 2016.

Mo, Y., Zhong, G., Li, J., Liu, X., Jiang, H., Tang, J., Jiang, B., Liao, Y., Cheng, Z., and Zhang, G.: The Sources, Molecular Compositions, and Light Absorption Properties of Water‐Soluble Organic Carbon in Marine Aerosols From South China Sea to the Eastern Indian Ocean, J. Geophys. Res.-Atmos., 127, https://doi.org/10.1029/2021jd036168, 2022.

Murphy, K. R., Stedmon, C. A., Graeber, D., and Bro, R.: Fluorescence spectroscopy and multi-way techniques. PARAFAC, Anal. Methods, 5, 6557-6566, https://doi.org/10.1039/c3ay41160e, 2013.

Murphy, K. R., Timko, S. A., Gonsior, M., Powers, L. C., Wunsch, U. J., and Stedmon, C. A.: Photochemistry Illuminates Ubiquitous Organic Matter Fluorescence Spectra, Environ. Sci. Technol., 52, 11243-11250, https://doi.org/10.1021/acs.est.8b02648, 2018.

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R., Solymos, P., Stevens, M. H. H., and Szoecs, E.: Package 'vegan'. R-package version 2.6-2. https://CRAN.R-project.org/package=vegan (accessed April 17, 2022), 2022.

Ramette, A.: Multivariate analyses in microbial ecology, FEMS Microbiol. Ecol., 62, 142-160, https://doi.org/10.1111/j.1574-6941.2007.00375.x, 2007.

Siemens, K., Morales, A., He, Q., Li, C., Hettiyadura, A. P. S., Rudich, Y., and Laskin, A.: Molecular Analysis of Secondary Brown Carbon Produced from the Photooxidation of Naphthalene, Environ. Sci. Technol., 56, 3340-3353, https://doi.org/10.1021/acs.est.1c03135, 2022.

Singer, G. A., Fasching, C., Wilhelm, L., Niggemann, J., Steier, P., Dittmar, T., and Battin, T. J.: Biogeochemically diverse organic matter in Alpine glaciers and its downstream fate, Nat. Geosci., 5, 710-714, https://doi.org/10.1038/ngeo1581, 2012.

Smith, J. S., Laskin, A., and Laskin, J.: Molecular Characterization of Biomass Burning Aerosols Using High-Resolution Mass Spectrometry, Anal. Chem., 81, 1512-1521, https://doi.org/10.1021/ac8020664, 2009.

Stubbins, A., Lapierre, J. F., Berggren, M., Prairie, Y. T., Dittmar, T., and del Giorgio, P. A.: What's in an EEM? Molecular signatures associated with dissolved organic fluorescence in boreal Canada, Environ. Sci. Technol., 48, 10598-10606, https://doi.org/10.1021/es502086e, 2014.

475 Twardowski, M. S., Boss, E., Sullivan, J. M., and Donaghay, P. L.: Modeling the spectral shape of
476 absorption by chromophoric dissolved organic matter, Mar. Chem., 89, 69-88,
477 https://doi.org/10.1016/j.marchem.2004.02.008, 2004.

478 Varga, B., Kiss, G., Ganszky, I., Gelencsér, A., and Krivácsy, Z.: Isolation of water-soluble organic
479 matter from atmospheric aerosol, Talanta, 55, 561-572, https://doi.org/10.1016/S0039-
480 9140(01)00446-5, 2001.

481 Weishaar, J. L., Aiken, G. R., Bergamaschi, B. A., Fram, M. S., Fujii, R., and Mopper, K.: Evaluation
482 of Specific Ultraviolet Absorbance as an Indicator of the Chemical Composition and Reactivity
483 of Dissolved Organic Carbon, Environ. Sci. Technol., 37, 4702-4708,
484 https://doi.org/10.1021/es030360x, 2003.

485 Wu, G., Fu, P., Ram, K., Song, J., Chen, Q., Kawamura, K., Wan, X., Kang, S., Wang, X., Laskin,
486 A., and Cong, Z.: Fluorescence characteristics of water-soluble organic carbon in atmospheric
487 aerosol, Environ. Pollut., 268, 115906, https://doi.org/10.1016/j.envpol.2020.115906, 2021.

488 Yan, C., Zheng, M., Sullivan, A. P., Bosch, C., Desyaterik, Y., Andersson, A., Li, X., Guo, X., Zhou,
489 T., Gustafsson, Ö., and Collett, J. L.: Chemical characteristics and light-absorbing property of
490 water-soluble organic carbon in Beijing: Biomass burning contributions, Atmos. Environ., 121,
491 4-12, https://doi.org/10.1016/j.atmosenv.2015.05.005, 2015.

492 Yu, L., Smith, J., Laskin, A., Anastasio, C., Laskin, J., and Zhang, Q.: Chemical characterization of
493 SOA formed from aqueous-phase reactions of phenols with the triplet excited state of carbonyl
494 and hydroxyl radical, Atmos. Chem. Phys., 14, 13801-13816, https://doi.org/10.5194/acp-14-
495 13801-2014, 2014.

496 Zong, Z., Tian, C., Li, J., Syed, J. H., Zhang, W., Fang, Y., Jiang, Y., Nasir, J., Mansha, M., Rizvi, S.
497 H. H., Shafiq, M., Farhan, S. B., and Zhang, G.: Isotopic Interpretation of Particulate Nitrate
498 in the Metropolitan City of Karachi, Pakistan: Insight into the Oceanic Contribution to NOx,
499 Environ. Sci. Technol., 54, 7787-7797, https://doi.org/10.1021/acs.est.0c00490, 2020.

500 Zsolnay, A., Baigar, E., Jimenez, M., Steinweg, B., and Saccomandi, F.: Differentiating with
501 fluorescence spectroscopy the sources of dissolved organic matter in soils subjected to drying,
502 Chemosphere, 38, 45-50, https://doi.org/10.1016/S0045-6535(98)00166-0, 1999.

503