1 **Urban organic aerosol composition in Eastern China differs from North to South: Molecular**

2 **insight from a liquid chromatography-Orbitrap mass spectrometry study**

3 Kai Wang[1,2,4], Ru-Jin Huang[1], Martin Brüggemann[3], Yun Zhang[2], Lu Yang[1], Haiyan Ni[1], Jie Guo[1],

4 Meng Wang[1], Jiajun Han[5], Merete Bilde[4], Marianne Glasius[4], and Thorsten Hoffmann[2]

5 [1]State Key Laboratory of Loess and Quaternary Geology (SKLLQG), Center for Excellence in

6 Quaternary Science and Global Change, and Key Laboratory of Aerosol Chemistry and Physics,

7 Institute of Earth and Environment, Chinese Academy of Sciences, Xi'an 710061, China

8 [2]Institute of Inorganic and Analytical Chemistry, Johannes Gutenberg University Mainz,

9 Duesbergweg 10−14, Mainz 55128, Germany

10 [3]Atmospheric Chemistry Department (ACD), Leibniz Institute for Tropospheric Research

11 (TROPOS), Permoserstraße 15, 04318 Leipzig, Germany

12 [4]Department of Chemistry, Aarhus University, Langelandsgade 140, DK-8000 Aarhus C, Denmark

13 [5]Department of Chemistry, University of Toronto, 80 St. George Street, M5S3H6 Toronto, Canada

14 Corresponding Author: Ru-Jin Huang (rujin.huang@ieecas.cn) and Thorsten Hoffmann

15 (t.hoffmann@uni-mainz.de)

Abstract:

Air pollution by particulate matter in China affects human health, the ecosystem and the climate. However, the chemical composition of particulate aerosol, especially of the organic fraction, is still not well understood. In this study, particulate aerosol samples with a diameter of $\leqslant 2.5$ μm ($PM_{2.5}$) were collected in January 2014 in three cities located in Northeast, East and Southeast China, namely Changchun, Shanghai and Guangzhou. Organic aerosol (OA) in the $PM_{2.5}$ samples was analyzed by ultrahigh performance liquid chromatography (UHPLC) coupled to high-resolution Orbitrap mass spectrometry in both negative mode (ESI−) and positive mode electrospray ionization (ESI+). After non-target screening including the assignment of molecular formulas, the compounds were classified into five groups based on their elemental composition, i.e., CHO, CHON, CHN, CHOS and CHONS. The CHO, CHON and CHN groups present the dominant signal abundances of 81−99.7% in the mass spectra and the majority of these compounds were assigned to mono- and polyaromatics, suggesting that anthropogenic emissions are a major source of urban OA in all three cities. However, the chemical characteristics of these compounds varied between the different cities. The degree of aromaticity and the number of polyaromatic compounds were substantially higher in samples from Changchun, which could be attributed to the large emissions from residential heating (i.e. coal combustion) during winter time in Northeast China. Moreover, the ESI− analysis showed higher H/C and O/C ratios for organic compounds in Shanghai and Guangzhou compared to samples from Changchun, indicating that OA undergoes more intense photochemical oxidation processes in lower latitude regions of China and/or is affected to a larger degree by biogenic sources. The majority of sulfur-containing compounds (CHOS and CHONS) in all cities were assigned to aliphatic compounds with low degrees of unsaturation and aromaticity. Here again, samples from Shanghai and Guangzhou show a greater chemical similarity but differ largely from those from Changchun. It should be noted that the conclusions drawn in this study are mainly based on comparison of molecular formulas weighted by peak abundance, and thus, are associated with inherent uncertainties due to different ionization efficiencies for different organic species.

1. Introduction

In the last decades, China has experienced rapid industrialization and urbanization accompanied by severe and persistent particulate air pollution (Huang et al., 2014; Sun et al., 2014; Ding et al., 2016; Song et al., 2018; Shi et al., 2019; Xu et al., 2019). These particulate air pollution extremes can not only influence the regional air quality and human health in China, but also lead to a global

59 environmental problem due to long-distance transport of pollutants. To better understand the effects

60 of air pollution on air quality and human health, chemical characterization of fine particle

61 (particulate matter with an aerodynamic diameter of less than 2.5 μm, or $PM_{2.5}$) is crucial. However,

62 the chemical composition of $PM_{2.5}$ in China is still poorly understood due to a wide variety of

63 natural and anthropogenic sources as well as complex multiphase chemical reactions (Lin et al.,

64 2012a; Huang et al., 2014; Ding et al., 2016; Wang et al., 2017; Wang et al., 2018; An et al., 2019;

65 Tong et al., 2019; Wang et al., 2019a; Wang et al., 2019b). In particular, compared to the fairly

66 well understood nature of the inorganic fraction of aerosol, the organic fraction, also named organic

67 aerosol (OA), is considerably less understood in terms of chemical composition, corresponding

68 precursors, sources and formation mechanisms (Huang et al., 2017).

69 During pollution events in China, OA accounts for as high as more than 50% of the total mass of

70 fine particle (An et al., 2019). Chemical compounds in OA cover a large complexity of species

71 including alcohols, aldehydes, carboxylic acids, imidazoles, organosulfates, organonitrates and

72 polycyclic aromatic hydrocarbons (PAHs) (Lin et al., 2012a; Rincón et al., 2012; Kourtchev et al.,

73 2014; Wang et al., 2018; Elzein et al., 2019; Wang et al., 2019a). Thus, the capacity of traditional

74 analytical techniques is limited to identify the compounds in OA and the majority (> 70%) of OA

75 has not been identified yet as specific compounds (Hoffmann et al., 2011). The insufficient

76 knowledge of chemical composition of OA hinders a better understanding of the sources, formation

77 and atmospheric processes of air pollution in China.

78 Recently, ultrahigh resolution mass spectrometry (UHRMS), such as Fourier transform ion

79 cyclotron resonance mass spectrometry (FTICR−MS) and Orbitrap−MS, coupled with soft

80 ionization sources (e.g., electrospray ionization (ESI) and atmospheric pressure chemical ionization

81 (APCI)) have been introduced to elucidate the molecular composition of OA (Nizkorodov et al.,

82 2011; Lin et al., 2012a; Lin et al., 2012b; Rincón et al., 2012; Noziere et al., 2015; Kourtchev et al.,

83 2016; Tong et al., 2016; Tu et al., 2016; Brüggemann et al., 2017; Wang et al., 2017; Fleming et

84 al., 2018; Laskin et al., 2018; Song et al., 2018; Wang et al., 2018; Brüggemann et al., 2019;

85 Daellenbach et al., 2019; Ning et al., 2019; Wang et al., 2019a). Due to the two outstanding features

86 of high resolving power and high mass accuracy, UHRMS can give precise elemental compositions

87 of individual organic compounds. However, UHRMS studies on Chinese urban OA are very limited.

88 Wang et al. (Wang et al., 2017) characterized OA in Shanghai and showed variations in chemical

89 composition among different months and between daytime and nighttime. Our recent Orbitrap MS

90 study (Wang et al., 2018) showed that wintertime OA in $PM_{2.5}$ collected in Beijing, China and

91 Mainz, Germany were very different in terms of chemical composition. In contrast, for summertime

92   OA from Germany and China, Brüggemann et al. (2019) found similar compounds and

93   concentrations of terpenoid organosulfates in PM_{10}, demonstrating that biogenic emission can

94   significantly affect OA composition at both locations. Ning et al. (2019) analyzed the OA collected

95   in a coastal Chinese city (Dalian) and found that more organic compounds were identified in haze

96   days compared to non-haze days. Nonetheless, since severe particulate pollution in China occurs

97   on a large-scale, more UHRMS studies are needed to fully elucidate the chemical composition of

98   OA in different Chinese cities.

99   In this study, $PM_{2.5}$ aerosol samples were collected in three Chinese cities, i.e., Changchun,

100   Shanghai and Guangzhou, and their organic fraction was analyzed using ultra-high-performance

101   liquid chromatography (UHPLC) coupled with Orbitrap−MS. The Chinese cities of Changchun,

102   Shanghai and Guangzhou are located in the Northeast, East and Southeast of China, which are

103   major populated regions in China with a population of 7.5, 24 and 15 million, respectively. The

104   geographic locations of these three cities cover a large latitude spanning from 23.12° N to 43.53° N

105   resulting in different meteorological conditions, including intensity and duration of sunlight,

106   average daily temperature and monsoon climate. In addition, the industrial structure, energy

107   consumption and energy sources in these three cities are different, such as much more heavy

108   industries (e.g., coal chemical industry and steelworks) in Northeast China (Zhang, 2008), which

109   can cause difference in anthropogenic emissions, and can therefore influence the chemical

110   composition of urban OA. Moreover, OA is strongly affected by residential coal combustion during

111   winter in Northeast China (Huang et al., 2014; An et al., 2019). Therefore, this study presents a

112   comprehensive overview of chemical composition of OA in three representative Chinese cities

113   during pollution episodes, which eventually can improve our understanding of OA effects on

114   climate and public health and also provide a chemical database for haze mitigation strategies in

115   China.

116   **2. Experimental**

117   **2.1 $PM_{2.5}$ samples**

118   Three 24−h integrated urban $PM_{2.5}$ samples were collected during severe haze pollution events with

119   daily average $PM_{2.5}$ mass concentration higher than 115 μg m^{-3} in each of the three Chinese cities:

120   Changchun (43.54° N, 125.13° E, 1.5 m above the ground), Shanghai (31.30° N, 121.50° E, 20 m

121   above the ground) and Guangzhou (23.07° N, 113.21° E, 53 m above the ground), which are located

122   in the Northeast, East and Southeast regions of China, respectively (see Fig. 1). Samples in

123   Changchun were collected on 4, 24 and 29 of January 2014 with $PM_{2.5}$ mass concentrations of

124 185−222 µg m^{-3}, samples in Shanghai were collected on 1, 19 and 20 of January 2014 with PM$_{2.5}$

125 mass concentrations of 159−172 µg m^{-3} and samples in Guangzhou were collected on 5, 6 and 11

126 of January 2014 with PM$_{2.5}$ mass concentrations of 138−152 µg m^{-3}. Further details (e.g., the daily

127 average concentrations of PM$_{2.5}$, SO$_2$, NO$_2$, CO and O$_3$, the average temperature and the daily solar

128 radiation value during sampling dates) are presented in Table S1, the 48 hours back trajectories of

129 air arriving at the three sampling sites during the sampling periods are shown in Fig. S1. All PM$_{2.5}$

130 samples were collected on prebaked quartz-fiber filters (20.3×25.4 cm) using a high-volume PM$_{2.5}$

131 sampler at a flow rate of 1.05 m^3 min^{-1} (Tisch Environmental, USA) and at each sampling site field

132 blanks were taken. After sample collection, filters were stored at −20 °C until analysis.

133 **2.2 Sample analysis**

134 Detailed description on the filter sample extraction and UHPLC−Orbitrap MS analysis can be found

135 in our previous studies (Wang et al., 2018; Wang et al., 2019a). Briefly, a part of the filters (around

136 1.13 cm^2, corresponding to about 600 µg particle mass in each extracted filter) was extracted three

137 times with 1.0−1.5 mL of acetonitrile-water (8/2, v/v) in an ultrasonic bath. The extracts were

138 combined, filtered through a 0.2 µm Teflon syringe filter and evaporated to almost dryness under

139 a gentle nitrogen stream. Finally, the residue was redissolved in 1000 µL acetonitrile-water (1/9,

140 v/v) to reach the total particulate mass concentration of around 600 µg mL^{-1} for the following

141 analysis.

142 Compared to the direct infusion method applied in other UHRMS studies (Lin et al., 2012a; Lin et

143 al., 2012b; Rincón et al., 2012; Kourtchev et al., 2016; Fleming et al., 2018), the UHPLC technique

144 was used in this study, which could separate and concentrate the compounds before they entered

145 the ion source, reducing the ionization suppression and increasing the sensitive of the measurement.

146 In addition, it can provide separation of some compounds and information of retention time of the

147 compounds, which is useful for the identification of the compounds and the separation of isomers.

148 The analytes were separated using a Hypersil Gold column (C18, 50 x 2.0 mm, 1.9 µm particle size)

149 with mobile phases consisting of (A) 0.04% formic acid and 2% acetonitrile in MilliQ water and

150 (B) 2% water in acetonitrile. Gradient elution was applied with the A and B mixture at a flow rate

151 of 500 µL min^{-1} as follows: 0−1.5 min 2% B, 1.5−2.5 min from 2% to 20% B (linear), 2.5−5.5 min

152 20% B, 5.5−6.5 min from 20% to 30% B (linear), 6.5−7.5 min from 30% to 50% B (linear), 7.5−8.5

153 min from 50% to 98% B (linear), 8.5−11.0 min 98% B, 11.0−11.05 min from 98% to 2% B (linear),

154 and 11.05−11.1 min 2% B. The Q Exactive Hybrid Quadrupole-Orbitrap MS was equipped with a

155 heated ESI source at 120 °C, applying a spray voltage of −3.3 kV and 4.0 kV for negative ESI mode

156 (ESI−) and positive ESI mode (ESI+), respectively. The mass scanning range was set from m/z 50

157 to 500 with a resolving power of 70,000 @ m/z 200. The Orbitrap MS was externally calibrated

158 before each measurement sequence using an Ultramark 1621 solution (Sigama−Aldrich, Germany)

159 providing mass accuracy of the instrument lower than 3 ppm. Each sample was measured in

160 triplicate with an injection volume of 10 μL.

161 **2.3 Data processing**

162 A non-target peak picking software (SIEVE®, Thermo Fisher Scientific, Germany) was used to find

163 significant peaks in the LC-MS dataset and to calculate all mathematically possible chemical

164 formulas for ions signals with a sample-to-blank abundance ratio $\geqslant 10$ using a mass tolerance of $\pm$

165 2 ppm. The permitted maximum elemental number of atoms was set as follows: ^{12}C (39), ^{1}H (72),

166 ^{16}O (20), ^{14}N (7), ^{32}S (4), ^{35}Cl (2) and ^{23}Na (1) (Kind and Fiehn, 2007; Lin et al., 2012a; Wang et

167 al., 2018). To remove the chemically unreasonable formulas, further constraint was applied by

168 setting H/C, O/C, N/C, S/C and Cl/C ratios in the ranges of 0.3−3, 0−3, 0−1.3, 0−0.8 and 0−0.8

169 (Kind and Fiehn, 2007; Lin et al., 2012a; Rincón et al., 2012; Wang et al., 2018; Zielinski et al.,

170 2018), respectively. For chemical formula $C_cH_hO_oN_nS_sCl_x$, the double bond equivalent (DBE) was

171 calculated by the equation: $DBE = (2c + 2 - h - x + n) / 2$. The aromaticity equivalent (X_C) as a

172 modified index for aromatic compounds was obtained using the equation: $Xc = [3(DBE - (p \times o +$

173 $q \times n)) - 2] / [DBE - (p \times o + q \times n)]$, where p and q, respectively, refer to the fraction of oxygen

174 and sulfur atoms involved in the π-bond structure of a compound. As such the values of p and q

175 vary between compound categories (Yassine et al., 2014). For example, carboxylic acids and esters

176 are characterized using $p = q = 0.5$, while $p = q = 1$ and $p = q = 0$ are used for carbonyl and hydroxyl,

177 respectively. Since it is impossible to identify the structurers of the hundreds of formulas observed

178 in this study, we cannot know the exact values of p and q in an individual compound. Therefore, in

179 this study, $p = q = 0.5$ was applied for compounds detected in ESI− as carboxylic compounds are

180 preferably ionized in negative mode. However, because of the high complexity of the mass spectra

181 in ESI+, $p = q = 1$ was used in ESI+ to avoid an overestimation of the amount of aromatics.

182 Moreover, for $DBE \leq (p \times o + q \times n)$ or $X_C \leq 0$, X_C was defined as zero. Furthermore, in ESI−, for

183 odd numbers of $(p \times o + q \times n)$, the value of $(p \times o + q \times n)$ was rounded down to the lower integer.

184 $Xc \geqslant 2.50$ and $Xc \geqslant 2.71$ have been suggested as unambiguous minimum criteria for the presence

185 of monoaromatics and polyaromatics, respectively (Yassine et al., 2014).

186 Comparing the peak abundance has been used in recent UHRMS studies (Wang et al., 2017;

187 Fleming et al., 2018; Song et al., 2018; Ning et al., 2019) to illustrate the relative importance of

188   specific types of compounds. However, it should be noted that different organic compounds have

189   different signal response in the mass spectrometer due to the differences in ionization and

190   transmission efficiencies (Schmidt et al., 2006; Leito et al., 2008; Perry et al., 2008; Kruve et al.,

191   2014). Therefore, uncertainties may exist when comparing the peak areas among compounds. In

192   this work, we assume that all organic compounds have the same peak abundance response in the

193   mass spectrometer. The peak abundance-weighted average molecular mass (MM), elemental ratios,

194   DBE, and Xc for formula $C_cH_hO_oN_nS_sCl_x$ were calculated using following equations:

195    $MM_{avg} = \sum (MM_i \times A_i) / \sum A_i$

196    $O/C_{avg} = \sum (O/C_i \times A_i) / \sum A_i$

197    $H/C_{avg} = \sum (H/C_i \times A_i) / \sum A_i$

198    $DBE_{avg} = \sum (DBE_i \times A_i) / \sum A_i$

199    $Xc_{avg} = \sum (Xc_i \times A_i) / \sum A_i$

200   where A_i is the peak abundance for each individual compound i.

201   **3. Results and discussion**

202   **3.1 General characteristics**

203   The main purpose of this study was to tentatively identify and compare the chemical composition

204   of organic compounds in the $PM_{2.5}$ samples collected in the three Chinese cities: Changchun,

205   Shanghai and Guangzhou during pollution episodes. To reduce the uncertainty caused by the

206   variability between the samples collected at each location, only organic compounds measured in

207   all three samples of each city are used for intercity comparison in this study. The number of organic

208   compounds and molecular formulas detected in each city, the peak abundance-weighted average

209   values of molecular mass (MM_{avg}), elemental ratios, DBE, Xc and the isomer number fraction

210   (meaning the percentage of formula numbers that have isomers among all assigned formulas) for

211   each subgroup are listed in Table 1. It should be noted that in this study we focus solely on organic

212   compounds with elevated signal abundances, and thus, presumably rather high concentrations. In

213   contrast to our previous study (Wang et al., 2018), compounds with low concentrations were

214   excluded by increasing the reconstitution volume from 500 μL to 1000 μL, reducing the sample

215   injection volume from 20 μL to 10 μL, and increasing the sample-to-blank ratio from 3 to 10 during

216   data processing.

217 Overall, 416−769 (assigned to 272−415 molecular formulas) and 687−2943 (assigned to 383−679

218 molecular formulas) organic compounds in different samples were determined in ESI− and ESI+,

219 respectively. The largest number of organic compounds was observed in Changchun samples in

220 both ESI− and ESI+, indicating that OA collected during winter season in Northeast China was

221 more complex compared to urban OA in East and Southeast China. This increased number of

222 compounds can possibly be explained by the large residential coal combustion emissions in winter

223 in North China (Huang et al., 2014; Song et al., 2018; An et al., 2019). In addition, ambient

224 temperatures were lowest during the sampling period in Changchun (i.e., −14 ℃ to −9 ℃, Table

225 S1), which likely led to a decreased boundary layer height and therefore enhanced accumulation of

226 pollutants and enhanced formation of secondary organic aerosol through for example gas-to-

227 particle partitioning.

228 As shown in Table 1, the abundance-weighted average values of MM_{avg} and O/C ratio of the total

229 assigned formulas for Changchun samples dectected in negative mode (Changchun−) are 169 and

230 0.58, respectively, which are lower than those for Shanghai− (MM_{avg} = 176 and O/C = 0.69) and

231 for Guangzhou− (MM_{avg} = 183 and O/C = 0.74). On the contrary, the aromaticity equivalent Xc for

232 organics detected in Changchun−, Xc(Changchun–) = 2.13, is substantially higher than that for

233 Shanghai−, Xc(Shanghai–) = 1.92, and Guangzhou− , Xc(Guangzhou–) = 1.65. These observations

234 indicate that urban OA in Northeast China features a lower degree of oxidation and a higher degree

235 of aromaticity compared to urban OA in East and Southeast China. Furthermore, the relative peak

236 abundance fraction of compounds with O/C ≥ 0.6, which are considered as highly oxidized

237 compounds (Tu et al., 2016), is 31% in Changchun−, and higher in Shanghai– (46%) and

238 Guangzhou– (51%). The different chemical composition of the samples is probably caused by the

239 rather low ambient temperatures and decreased photochemical processing of organic compounds

240 in Northeast China (indicated by the lower solar radiation in Northeast China, see Table S1),

241 slowing down oxidation processes and leading to a larger number of PAHs, which are mainly

242 emitted from coal burning (Huang et al., 2014; Song et al., 2018) or by different

243 biogenic/anthropogenic precursors. In addition, long-range transport of air masses (see the 48 hours

244 back trajectories in Fig. S1) may have a certain effect on the chemical properties of aerosol samples

245 collected in the three cities.

246 Figure 1 shows the reconstructed mass spectra of organic compounds detected in ESI− and ESI+.

247 A major fraction organic species detected in ESI− are attributed to CHO− and CHON−, accounting

248 for 30−42% and 39−55% in terms of peak abundance, respectively, and comprising 39−45% and

249 23−33% in terms of peak numbers, respectively. This is consistent with previous studies on Chinese

urban OA by Wang et al. (2017 and 2018) and Brüggemann et al. (2019). Comparing the organic compounds detected in ESI− for the three cities, 120 formulas were observed in all cities as common formulas (which refer to the compounds detected in all cities with the same molecular formulas and with the same retention times (retention time difference ≤ 0.1 min)) (Fig. 2a), accounting for 29−44% and 57−71% of all assigned formulas in terms of formula numbers and peak abundance, respectively. Despite the above-mentioned differences in chemical composition for OA from Changchun compared to OA from Shanghai and Guangzhou, these results demonstrate that still a large number of common organic compounds exist in Chinese urban OAs collected in different cities, in particular for organics with higher signal abundances. Furthermore, as shown by the pie chart in Fig. 2b, these common formulas are dominated by CHON− and CHO−, accounting for 62% and 30% of the total common formulas in terms of peak abundance, respectively.

As it is commonly known, ESI exhibits different ionization mechanisms in negative and positive ionization modes. While ESI− is especially sensitive to deprotonatable compounds (e.g., organic acids), ESI+ is more sensitive to protonatable compounds (e.g., organic amines) (Ho et al., 2003). Due to the different ionization mechanisms, clear differences were observed in the mass spectra (Fig. 1) and chemical characteristics (Table 1) from ESI− and ESI+ measurements. For example, CHO compounds were preferentially detected in ESI−, accounting for a relatively large fraction of 30−42% of all detected compounds in terms of peak abundance, compared to merely 4−13% for such CHO compounds in ESI+. In contrast, CHN compounds were only observed in ESI+, yielding a rather large peak abundance fraction of 40−71%. In particular, as can be seen in Fig.1, several peaks of CHN+ compounds in Shanghai+ and Guangzhou+ have much higher abundance compared to other organic species, probably due to their high concentrations and/or high ionization efficiencies in the positive mode. This observation indicates that most CHO compounds with high concentrations are probably organic acids, whereas the majority of CHN compounds likely belong to the group of organic amines, which is in good agreement with previous studies (Lin et al., 2012a; Wang et al., 2017; Wang et al., 2018). Organic compounds in ESI+ are dominated by CHN+ and CHON+ compounds in terms of both peak numbers and peak abundance and these compounds are characterized by rather high H/C ratio and low O/C ratios (Table 1), indicating a low degree of oxidation. The Venn diagram presented for ESI+ measurements in Fig. 2a shows that out of a total of 383−679 formulas, 129 formulas were found in samples from all three cities. Such common formulas, thus, account for 19−34% and 30−75% of all assigned formulas in terms of formula numbers and peak abundance, respectively. Among these common formulas, CHN+ and CHON+ exhibit the highest abundance fractions of 72% and 26%, respectively (Fig. 2b).

283   In the following, we will compare and discuss the chemical properties in detail for the three cities,

284   including degrees of oxidation, unsaturation and aromaticity of each organic compound class (i.e.,

285   CHO, CHON, CHN, CHOS and CHONS). It should be noted that the chlorine-containing

286   compounds were not discussed in this study due to the very low MS signal abundance. In addition,

287   since peak abundances for the formula can vary by orders of magnitude, the area of the circles

288   presented in the Figure 3 and Figures 5–7 is proportional to the fourth root of the peak abundance

289   of each formula to reduce the size difference of the circles. For a more detailed comparison, figures

290   with the circle size related to the absolute peak abundances are presented in the SI.

291   **3.2 CHO compounds**

292   CHO compounds have been widely observed in urban OA, accounting for a substantial fraction

293   (8−67%) of OA (Rincón et al., 2012; Tao et al., 2014; Wang et al., 2017; Wang et al., 2018).

294   Previous studies have shown that a large fraction of CHO compounds in urban OA is composed of

295   organic acids, containing deprotonatable carboxyl functional groups, which are detected

296   preferentially in negative ionization mode when using ESI−MS. As shown in Table 1, a total of

297   346, 164, and 196 CHO− compounds were detected in ESI− in the OA samples collected in

298   Changchun, Shanghai and Guangzhou, accounting for 30%, 40% and 42% of the overall peak

299   abundance in each sample, respectively. Out of all assigned formulas, 47 common CHO− formulas

300   were observed for all cites, accounting for 35−52% and 42−68% of all identified CHO− formulas

301   in terms of formula numbers and peak abundance, respectively.

302   Despite this similarity, OA samples from Changchun− (i.e. in negative ionization mode) exhibit

303   certain differences compared to samples from Shanghai− and Guangzhou−. The average H/C

304   values for CHO− compounds are in a similar range for the three locations (i.e., 0.96−1.10), however,

305   the average O/C values for O/C(Shanghai−) = 0.59 and O/C(Guangzhou−) = 0.65 are rather high

306   compared to the average O/C ratio for Changchun−, O/C(Changchun–) = 0.41. Furthermore, the

307   relative peak abundance fraction of CHO− compounds with O/C ≥ 0.6, which are considered as

308   highly oxidized compounds (Tu et al., 2016), is 14% in Changchun and somewhat higher in

309   Shanghai– (34%) and Guangzhou– (45%). Altogether, these results indicate that CHO− compounds

310   in urban OA from East and Southeast China experienced more intense oxidation and aging

311   processes and/or were affected to a larger degree by biogenic sources.

312   Similarly, as shown in Fig. 3, the abundance-weighted average molecular formulas for CHO−

313   compounds in Changchun−, Shanghai– and Guangzhou− are $C_{8.58}H_{7.86}O_{3.22}$ (MM_{avg}(Changchun–)

314   = 162), $C_{8.01}H_{7.27}O_{4.22}$ (MM_{avg}(Shanghai–) = 171) and $C_{7.70}H_{8.04}O_{4.48}$ (MM_{avg}(Guangzhou–) = 172),

315   respectively. Again, these average formulas show that CHO− in Shanghai− and Guangzhou−

316   experienced more intense oxidation processes and/or were affected to a larger degree by biogenic

317   precursors, indicated by the larger abundance-weighted MM_{avg} with a higher degree of oxygenation.

318   In contrast, CHO− compounds from OA samples in Changchun− exhibit a lower abundance-

319   weighted MM_{avg} with a decreased oxygen content.

320   Besides oxygenation, the aromaticity of the detected CHO− compounds exhibits remarkable

321   differences in these three cities. In all cities, the CHO– compounds with high peak abundance were

322   mainly assigned to monoaromatics with $2.5 \leqslant Xc < 2.7$ (purple circles in Fig. 3) in the region of

323   7−12 carbon atoms per compound and DBE values of 5−7. The relative peak abundance fraction

324   of monoaromatics in total CHO– compounds is 67% in Changchun, which is higher compared to

325   64% in Shanghai and 49% in Guangzhou. In addition, 14% of CHO− compounds in Changchun

326   were identified as polyaromatic compounds with $Xc \geqslant 2.7$ (red circles in Fig. 3), which is higher

327   than the 8% in Shanghai and 4% in Guangzhou. These observations indicate that CHO− compounds

328   in the three Chinese cities are highly affected by aromatic precursors (e.g., benzene, toluene and

329   naphthalene), in particular for the Changchun aerosol samples.

330   Besides the monoaromatics and polyaromatics, the rest of the detected CHO− compounds were

331   assigned to aliphatic compounds with an Xc lower than 2.5 (grey circles in Fig. 3). Interestingly,

332   these aliphatic compounds account for about 47% of all CHO− compounds for Guangzhou−

333   samples in terms of peak abundance, whereas samples from Changchun− and Shanghai− exhibit

334   only rather small fractions of such CHO– compounds, i.e., 19% and 28%, respectively. Such

335   aliphatic compounds are commonly derived from biogenic precursors (Kourtchev et al., 2016) and

336   vehicle emission (Tao et al., 2014; Wang et al., 2017) and/or generated by intense oxidation

337   processes of aromatic precursors, indicating the different biogenic and anthropogenic emission

338   sources and chemical reaction processes for OAs in the three cities.

339   In addition, through the analysis of individual formulas, we find that for the Changchun− samples,

340   formulas of $C_8H_6O_4$, $C_7H_6O_2$, $C_7H_6O_3$, $C_8H_8O_2$, and $C_8H_8O_3$ with DBE values of 6, 5, 5, 5, and 5

341   dominate the assigned CHO formulas with respect to peak abundance. According to previous

342   studies, $C_8H_6O_4$, $C_7H_6O_2$ and $C_7H_6O_3$ are suggested to be phthalic acid, benzoic acid and

343   monohydroxy benzoic acid, respectively, which are derived from naphthalene (Kautzman et al.,

344   2010; Riva et al., 2015; Wang et al., 2017; He et al., 2018; Huang et al., 2019).$C_8H_8O_2$ is likely 4-

345   hydroxy acetophenone, which could be derived from estragole (Pereira et al., 2014), while $C_8H_8O_3$

346   is suggested to be either 4-methoxybenzoic acid generated from estragole (Pereira et al., 2014) or

347   vanillin emitted from biomass burning (Li et al., 2014). For the Shanghai− samples, besides $C_8H_6O_4$,

348    $C_7H_6O_3$ and $C_7H_6O_2$, formulas of $C_6H_8O_7$ and $C_9H_8O_4$ with DBE values of 3 and 6 were observed

349   with high peak abundances. $C_6H_8O_7$ was identified as citric acid in the pollen sample and mountain

350   particle sample in previous studies (Fu et al., 2008; Wang et al., 2009; Jung and Kawamura, 2011)

351   and $C_9H_8O_4$ are probably homophthalic acid derived from e.g. estragole (Pereira et al., 2014). For

352   the Guangzhou− samples, besides the formulas of $C_8H_6O_4$ and $C_6H_8O_7$ discussed above, $C_4H_6O_4$

353   and $C_4H_6O_5$ with low DBE values of two were detected with high abundances and are suggested to

354   be succinic acid and malic acid, respectively (Claeys et al., 2004; Wang et al., 2017).

355   **3.3 CHON compounds**

356   A large amount of nitrogen-containing organic compounds was detected in these three cities,

357   accounting for 39−55% and 25−47% of total peak abundance detected in ESI− and ESI+,

358   respectively. Out of all assigned formulas, 45 common CHON− and 62 common CHON+ formulas

359   were observed in all cities, accounting for 65−82% and 25−44% of all CHON compounds detected

360   in ESI− and ESI+ in terms of peak abundance, respectively. It indicates that a large amount of

361   CHON compounds in all three Chinese cities show similar properties of chemical composition.

362   The CHON compounds were further classified into different subgroups according to their O/N

363   ratios (Fig. 4 for CHON− and Fig. S3 for CHON+) or according to the number of nitrogen atoms

364   in their molecular formulas (see Fig. S4 for CHON− and S5 for CHON+). As shown in Fig. 4, the

365   majority (84−96% in terms of peak abundance) of CHON− compounds exhibited O/N ratios ⩾ 3,

366   allowing the assignment of one nitro (−NO₂) or nitrooxy (−ONO₂) group for these formulas, which

367   are preferentially ionized in ESI− mode (Lin et al., 2012b; Wang et al., 2017; Song et al., 2018;

368   Wang et al., 2018). CHON− formulas with O/N ratios ⩾ 4 suggest the presence of further

369   oxygenated functional groups, such as a hydroxyl group (−OH) or a carbonyl group (C=O). In

370   terms of peak abundance, 59% of CHON− compounds observed in Guangzhou− exhibited formulas

371   with O/N ratios ⩾ 4, which is higher than 51% in Changchun− and 45% in Shanghai−, indicating

372   that CHON− compounds in Southeast China show a higher degree of oxidation compared to those

373   in Northeast and East China. Not surprisingly, CHON+ compounds generally exhibit lower O/N

374   ratios (Fig. S3), as they probably contain reduced nitrogen functional group (e.g., amines) which

375   are preferably detected in ESI+. As shown in Fig. S3, CHON+ compounds with O/N ratio of 1 are

376   dominant in Changchun+, whereas CHON+ compounds in Shanghai+ and Guangzhou+ show a

377   broader range of O/N ratios from 1 to 3. Moreover, the average O/C ratios (0.27−0.45) in Shanghai+

378   and Guangzhou+ (Table 1) are much greater than that (0.19) in Changchun+. Consistent with the

379   observations for CHO compounds, these results indicate again that CHON+ compounds in the OA

380   of East and Southeast China experienced more intensive photooxidation and/or were affected to a

381   larger degree by biogenic precursors.

382   Figure 5 shows the DBE versus C number of CHON− compounds for the three cities. The majority

383   of CHON− compounds lie in the region of 5−15 C atoms and 3−10 DBEs. 67% of CHON−

384   compounds in terms of peak abundance were assigned to mono or polyaromatics in Shanghai−,

385   which is higher than 52% in Guangzhou− and 55% in Changchun−. It indicates that CHON−

386   compounds are dominated with aromatic compounds in all cities, while relatively higher peak

387   abundance weighted fraction of aromatic CHON− compounds were observed in Shanghai. The

388   peak abundance-weighted average molecular formulas for CHON− compounds in Changchun−,

389   Shanghai− and Guangzhou− are $C_{7.10}H_{6.76}O_{3.56}N_{1.03}$, $C_{7.07}H_{6.03}O_{3.80}N_{1.24}$ and $C_{7.12}H_{6.36}O_{3.99}N_{1.24}$,

390   respectively, showing that CHON− formulas in Shanghai− and Guangzhou− contain more O and

391   N atoms on average than those for Changchun−. Formulas of $C_6H_5O_3N_1$, $C_6H_5O_4N_1$, $C_7H_7O_3N_1$,

392    $C_7H_7O_4N_1$, $C_8H_9O_3N_1$, and $C_8H_9O_4N_1$ were detected with the highest abundance in all cities. These

393   molecular formulas are in line with nitrophenol or nitrocatechol analogs, which have been identified

394   in a previous urban OA study (Wang et al., 2017). Furthermore, these nitrooxy-aromatic

395   compounds were shown to enhance light absorbing properties of OA (Laskin et al., 2015; Lin et al.,

396   2015). In addition, it should be noted that the Xc values for $C_6H_5O_4N_1$, $C_7H_7O_4N_1$ and $C_8H_9O_4N_1$

397   were calculated to be lower than 2.5, suggesting that the fraction of aromatics in CHON−

398   compounds was underestimated. This is because that for nitrocatechol analogs with formulas of

399    $C_6H_5O_4N_1$, $C_7H_7O_4N_1$ and $C_8H_9O_4N_1$, only one oxygen atom is involved in the π-bond structure

400   corresponding to the p value of 0.25 in the Xc calculation equation, which is lower than the p value

401   of 0.5 applied for the Xc calculation in this study. The diagram of DBE versus C number for

402   CHON+ compounds observed in the three locations (presented in Fig. S7 in SI) shows that more

403   aromatic CHON+ compounds with relatively lower degree of oxidation were assigned in

404   Changchun+ samples compared to Shanghai+ and Guangzhou+ samples.

405   **3.4 CHN+ compounds**

406   696 CHN+ compounds were detected in Changchun+ samples in ESI+, which is higher than in

407   Shanghai+ (253) and Guangzhou (205). These CHN+ compounds are likely assignable to amines

408   according to previous studies (Rincón et al., 2012; Wang et al., 2017; Wang et al., 2018). The

409   number of CHN+ compounds accounts for 24%, 36% and 30% of the total organic compounds in

410   Changchun+, Shanghai+ and Guangzhou+, respectively, whereas the peak abundance of these

compounds accounts for 40%, 71% and 62%, respectively. The majority (> 97% in terms of peak abundance) of CHN+ compounds have one or two nitrogen atoms in their molecular formulas (see Fig. S9). Comparing the CHN+ compounds for the three cities, 51 common CHN+ formulas were observed in all cities, which contribute to as much as 43−89% of the total abundance of CHN+ formulas. This large percentage indicates that CHN+ compounds with presumably high concentrations in Changchun+, Shanghai+ and Guangzhou+ exhibit similar chemical composition. However, again OA samples from Changchun show some distinct differences to samples from Guangzhou and Shanghai.

A van Krevelen diagram of CHN+ compounds detected in the three samples is shown in Fig. 6, illustrating H/C ratios as a function of N/C ratio. In this plot, major parts of the CHN+ compounds are found in a region, which is constraint by H/C ratios between 0.5 and 2 and N/C ratios lower than 0.5. Moreover, the pie charts show that the majority (83−87% in terms of peak abundance and 72−90% in terms of peak numbers) of these CHN+ compounds can be assigned to mono- and polyaromatics with Xc $\geq$ 2.5. In addition, as shown in Table 1, the average DBE and Xc values of CHN+ compounds are the highest among all organic species. These observations imply that CHN+ compounds exhibit the highest degree of aromaticity of all organics in the Chinese urban OA samples, which is consistent with previous studies (Lin et al., 2012b; Rincón et al., 2012; Wang et al., 2018). Polyaromatic compounds with Xc $\geq$ 2.7 are displayed in the lower left corner of the van Krevelen diagram, accounting for 41% in terms of peak abundance (48% in terms of peak numbers) of CHN+ compounds detected in Changchun+, but merely for 9−10% in terms of peak abundance (27−31% in terms of peak numbers) in Shanghai+ and Guangzhou+. For example, formulas of $C_{11}H_{11}N_1$ (Xc = 2.7), $C_{10}H_9N_1$ (Xc = 2.7), and $C_{12}H_{13}N_1$ (Xc = 2.7), which are assigned to be naphthalene core structure-containing compounds, have relatively higher abundance in Changchun+ than in Shanghai+ and Guangzhou+. Moreover, the average DBE and Xc values of CHN+ compounds (see Table 1) in Changchun+ are substantially higher than those in Shanghai+ and Guangzhou+, further indicating that CHN+ compounds in Changchun+ show a higher degree of aromaticity, which can be caused by large coal combustion emissions in the winter in Changchun. Remarkably, as can be seen in Fig. 6, the abundance of CHN+ compounds in Changchun+ distributes evenly among different individual CHN+ compounds, while in Shanghai+ and Guangzhou+ they are dominated by the formula of $C_{10}H_{14}N_2$ (the biggest purple circle in Fig. 6) with DBE value of 5, which probably has high concentration and/or high ionization efficiency in the positive ESI mode. According to a previous smog chamber study (Laskin et al., 2010), most CHN+ aromatics are probably generated from biomass burning through the addition of reduced

444   nitrogen (e.g., NH_3) to the organic molecules via imine formation reaction, indicating that biomass

445   burning probably made a certain contribution to the formation of CHN+ compounds observed in

446   the three urban OA samples in our study.

447   **3.5 CHOS− compounds**

448   In this study, 75−155 CHOS− compounds were observed, accounting for 10%, 12% and 14% of

449   the total peak abundance of all organics in Changchun−, Shanghai− and Guangzhou−, respectively.

450   Around 89−96% of these CHOS− compounds were found to fulfill the O/S $\geqslant$ 4 criterion allowing

451   the assignment of at least one $-OSO_3H$ functional group, and thus, a tentative classification to

452   organosulfates (OSs) (Lin et al., 2012a; Lin et al., 2012b; Tao et al., 2014; Wang et al., 2016; Wang

453   et al., 2017; Wang et al., 2018; Wang et al., 2019a). OSs were shown to affect the surface activity

454   and hygroscopic properties of the aerosol particles, leading to potential impacts on climate (Hansen

455   et al., 2015; Wang et al., 2019a). Out of all formulas, 23 common CHOS− formulas were detected

456   for the three sample locations, accounting for 28%, 58% and 52% of the CHOS− peak abundance

457   in Changchun−, Shanghai− and Guangzhou−, respectively. However, 40 common CHOS−

458   formulas were found between Shanghai− and Guangzhou−, accounting for 60−65% and 78−81%

459   in terms of the CHOS− formula numbers and peak abundance, respectively. This indicates that the

460   chemical composition of the major CHOS− compounds of Shanghai− and Guangzhou− are quite

461   similar, while they show substantial chemical differences for samples from Changchun−.

462   Figure 7 shows the DBEs as a function of carbon number for all CHOS− compounds detected for

463   the three cities. The CHOS− compounds exhibit a DBE range from 0 to 10 and carbon number

464   range of 2−15. However, the majority of CHOS− compounds with elevated peak abundances

465   concentrate in a region with rather low DBE values of 0−5. The average H/C ratios of CHOS−

466   compounds are in the range of 1.56−1.85, and thus, higher than for any other compound class,

467   whereas the average DBE values of 1.71−2.55 are the lowest among all classes. This indicates that

468   CHOS− compounds in the OA from the three Chinese cities are characterized by a low degree of

469   unsaturation. Moreover, the pie charts in Fig. 7 show that aliphatic compounds with Xc $\leqslant$ 2.5 are

470   dominant in CHOS− compounds with a fraction of 96−99% in terms of peak abundance, which is

471   substantially higher than that (13−48%) for CHO, CHON and CHN species. Aliphatic CHOS−

472   compounds with C $\leqslant$ 10 can be formed from biogenic and/or anthropogenic precursors (Hansen

473   et al., 2014; Glasius et al., 2018; Wang et al., 2019a), such as $C_2H_4O_6S_1$ (derived from glyoxal)

474   (Lim et al., 2010; McNeill et al., 2012), $C_3H_6O_6S_1$ (derived from isoprene) (Surratt et al., 2007) and

475   $C_8H_{16}O_4S_1$ (derived from α-pinene). However, more CHOS− compounds with C > 10 and with

476   DBEs lower than 1 are observed in Changchun−, such as $C_{14}H_{28}O_5S_1$, $C_{13}H_{26}O_5S_1$, $C_{12}H_{24}O_5S_1$,

477    $C_{11}H_{22}O_5S_1$ and $C_{11}H_{20}O_6S_1$. These high-carbon-number-containing CHOS− compounds are likely

478   formed from long-alkyl-chain compounds with less oxygenated functional groups, which were

479   previously suggested to be emitted from traffic (Tao et al., 2014) or derived from sesquiterpene

480   emissions (Brüggemann et al., 2019). However, as sesquiterpene emissions can be expected to be

481   very low in wintertime at Changchun, the presence of these compounds further underlines the

482   strong impact of anthropogenic emissions on CHOS− formation in Changchun−. In this study,

483   (O−3S)/C ratio was used instead of traditional O/C ratio to present the oxidation state of CHOS−

484   compounds, since the sulfate functional group contains three more oxygen atoms than common

485   oxygen-containing groups (e.g., hydroxyl and carbonyl), which makes no contribution to the

486   oxidation state of the carbon backbone of the CHOS− compounds. Comparing average values for

487   H/C, (O−3S)/C and DBEs of CHOS− for the three sample locations (see Table 1), we find that the

488   H/C ratios (1.85) and (O−3S)/C ratios (0.61−0.71) for Shanghai− and Guangzhou− samples are

489   larger than those for Changchun− samples (H/C = 1.56 and (O−3S)/C = 0.52), whereas the DBE

490   values (1.71−1.79) in Shanghai− and Guangzhou− are lower than those for Changchun− (2.55).

491   These observations indicate that CHOS− compounds in urban OA from Northeast China are less

492   oxidized but more unsaturated compared to those in East and Southeast China, likely due to

493   enhanced emissions from residential heating during winter in North China.

494   **3.6 CHONS compounds**

495   4-5% of the total organics detected in ESI− were identified as CHONS− compounds in terms of

496   peak abundance. In contrast, CHONS+ compounds account merely for 0.3−1% of all organics

497   detected in ESI+. The average MM_{avg} of the CHONS− compounds for the three sample locations

498   ranges from 214 to 293 Da, generally showing larger molecular masses than compounds of any

499   other class because of the likely presence of both nitrate and sulfate functional groups. In total, only

500   5 common CHONS− formulas were detected for all three sample locations, accounting for 4%, 21%

501   and 20% of the CHONS− peak abundance in Changchun−, Shanghai− and Guangzhou−,

502   respectively. As already observed for other compound classes, these percentages imply that the

503   CHONS− compounds in urban OA of Shanghai− and Guangzhou− exhibit a rather similar chemical

504   composition, whereas such compounds are different for Changchun−.

505   In the OA samples of Shanghai− and Guangzhou−, 78−87% of CHONS− compounds in terms of

506   peak abundance have 7 or more O atoms in their formulas, allowing the assignment of one −OSO_3H

507   and one -NO_3 functional groups in the molecular structures, thus, classifying them as potential

508 nitrooxy-organosulfates. In contrast to Shanghai− and Guangzhou−, only 26% of CHONS−
509 compounds were assigned to such nitroooxy-organosulfates for Changchun−, indicating that most
510 of the N atoms in the CHONS− compounds are present in a reduced oxidation state, e.g., in the
511 form of amines. The average DBE and Xc values of CHONS− compounds in Shanghai− and
512 Guangzhou− are 3.3−3.45 and 0.43−0.44, respectively. Again these values differ for the
513 Changchun− samples with an increased average DBE of 3.75 and an average Xc of 1.06, indicating
514 that CHONS- compounds in Changchun− possess on average a higher degree of unsaturation and
515 aromaticity compared to such compounds in Shanghai− and Guangzhou− samples. Interestingly,
516 the compound with formula $C_{10}H_{17}O_7NS$ has the highest relative peak abundance (32%) in
517 Shanghai− and Guangzhou−, whereas in Changchun− the compound with formula $C_2H_3O_4NS$ is
518 dominant. $C_{10}H_{17}O_7NS$ has previously been identified as mononitrate organosulfate generated from
519 α/β-pinene (Iinuma et al., 2007; Surratt et al., 2008; Lin et al., 2012b; Wang et al., 2017), while
520 $C_2H_3O_4NS$ may be assigned as a cyanogroup-containing sulfate. This observation is comparable to
521 our previous study (Wang et al., 2019a), which found that $C_{10}H_{17}O_7NS$ was dominant for CHONS−
522 compounds in low-concentration aerosol samples collected in Beijing (China) and Mainz
523 (Germany). Consistently, a $C_2H_3O_4NS$ compound had the highest abundance among CHONS−
524 compounds in polluted Beijing aerosol samples. This agreement can be explained by the adjacent
525 locations of Beijing (39.99º N, 116.39º E) and Changchun (43.54º N, 125.13º E) and similar
526 residential heating patterns by coal combustion during wintertime. In conclusion, these results
527 further demonstrate that the precursors for CHONS− compounds in Shanghai− and Guangzhou−
528 are different from those in Changchun−, which is probably due to differences in anthropogenic
529 emissions.

530 **4 Conclusion**

531 The molecular composition of the organic fraction of PM$_{2.5}$ samples collected in three Chinese
532 megacities (Changchun, Shanghai and Guangzhou) was investigated using a UHPLC-Orbitrap
533 mass spectrometer. In total, 416−769 (ESI−) and 687−2943 (ESI+) organic compounds were
534 observed and separated into five subgroups: CHO, CHN, CHON, CHOS and CHONS. Specifically,
535 120 common formulas were detected in ESI− and 129 common formulas in ESI+ for all sample
536 locations, accounting for 57−71% and 30−75% in terms of peak abundance, respectively. Overall,
537 we found that urban OA in Changchun, Shanghai and Guangzhou shows a quite similar chemical
538 composition for organic compounds of high concentrations. The majority of these organic species
539 was assigned to mono-aromatic or poly-aromatic compounds, indicating that anthropogenic
540 emissions are the major source for urban OA in all three cities.

541 Despite the chemical similarity of the three sample locations for organic compounds in urban OA,
542 remarkable differences were found in chemical composition of the remaining particle constituents,
543 in particular for OA samples from Changchun. In general, a larger amount of polyaromatics was
544 observed for Changchun samples, most likely due to emissions from coal combustion during
545 wintertime residential heating period. Moreover, the peak abundance-weighted average DBE and
546 average Xc values of the total organic compounds in Changchun were found to be larger than those
547 for Shanghai and Guangzhou, showing that organic compounds in Changchun possess a higher
548 degree of unsaturation and aromaticity. For average H/C and O/C ratios a similar trend was
549 observed. While average H/C and O/C ratios detected in ESI− were found to be highest for
550 Guangzhou samples, relatively lower values were observed for Shanghai and Changchun samples,
551 indicating that OA collected in lower latitude regions of China experiences more intense
552 photochemical oxidation processes and/or are affect to a larger degree by biogenic sources.

553 **5 Limitations**

554 In this study, we used the peak abundance-weighted method to illustrate the difference in chemical
555 formulas assigned by Orbitrap mass spectrometry. This comparison was made based on the
556 assumption that the measured organic compounds have same peak abundance response in the mass
557 spectrometer. However, this assumption can bring some uncertainties because the ionization
558 efficiencies vary between different compounds (Schmidt et al., 2006; Leito et al., 2008; Perry et al.,
559 2008; Kruve et al., 2014). For example, the ionization efficiencies of nitrophenol species detected
560 in negative ESI mode can vary by a large degree depending on the position of the substituents at
561 the nitrobenzene ring (Schmidt et al., 2006; Kruve et al., 2014) and the ionization efficiencies of
562 carboxylic acids can also vary by several orders of magnitude depending on the structures (Kruve
563 et al., 2014). Nonetheless, it is a challenging analytical task to identify and quantify all compounds
564 in ambient OA due to the high chemical complexity of OA and the limits in authentic standards of
565 OA. Despite the inherent uncertainties, the peak abundance-weighted comparison of molecular
566 formulas provides an overview of the difference in chemical composition of OA in these three
567 representative Chinese cities. In particular, the chemical formulas assigned in this study can be
568 validated in future studies by authentic standards and the difference in ionization efficiencies can
569 be further evaluated.

570

571 **Author contributions.** RJH, TH and KW conducted the study design. LY, HN, JG and MW
572 collected the PM$_{2.5}$ filter samples. KW and YZ carried out the experimental work and data analysis.

573 KW wrote the manuscript. KW, TH, RJH, M. Brüggemann, YZ, JH, M. Bilde and MG interpreted
574 data and edited the manuscript. All authors commented on and discussed the manuscript.

575 **Competing interests.** The authors declare that they have no conflict of interest.

576 **Acknowledgements.** This study was supported by the National Natural Science Foundation of
577 China (NSFC, Grant No. 41925015, No. 91644219 and No. 41877408), the Chinese Academy of
578 Sciences (No. ZDBS-LY-DQC001), the National Key Research and Development Program of
579 China (No. 2017YFC0212701), and the German Research Foundation (Deutsche
580 Forschungsgemeinschaft, DFG) under Grant No. INST 247/664-1 FUGG. K. Wang and Y. Zhang
581 acknowledge the scholarship from Chinese Scholarship Council (CSC) and Max Plank Graduate
582 Center with Johannes Gutenberg University of Mainz (MPGC) and thanks Prof. Ulrich Pöschl, Dr.
583 Christopher J. Kampf and Dr. Yafang Cheng for their helpful suggestion on this study. K. Wang
584 also thanks Dr. Huanfeng Dong from Zhejiang University for the great support on the programming
585 of data process.

586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617

618
619
620
621
622
623
624 **References**
625
626 An, Z., Huang, R. J., Zhang, R., Tie, X., Li, G., Cao, J., Zhou, W., Shi, Z., Han, Y., Gu, Z., and Ji, Y.: Severe
627   haze in northern China: A synergy of anthropogenic emissions and atmospheric processes, Proc Natl Acad
628   Sci U S A, 116, 8657-8666, 10.1073/pnas.1900125116, 2019.
629 Brüggemann, M., Poulain, L., Held, A., Stelzer, T., Zuth, C., Richters, S., Mutzel, A., van Pinxteren, D.,
630   Iinuma, Y., Katkevica, S., Rabe, R., Herrmann, H., and Hoffmann, T.: Real-time detection of highly
631   oxidized organosulfates and BSOA marker compounds during the F-BEACh 2014 field study, Atmos.
632   Chem. Phys., 17, 1453-1469, 10.5194/acp-17-1453-2017, 2017.
633 Brüggemann, M., van Pinxteren, D., Wang, Y., Yu, J. Z., and Herrmann, H.: Quantification of known and
634   unknown terpenoid organosulfates in PM10 using untargeted LC–HRMS/MS: contrasting summertime
635   rural Germany and the North China Plain, Environmental Chemistry, -, https://doi.org/10.1071/EN19089,
636   2019.
637 Claeys, M., Graham, B., Vas, G., Wang, W., Vermeylen, R., Pashynska, V., Cafmeyer, J., Guyon, P., Andre,
638   M., Artaxo, P., and Maenhaut, W.: Formation of secondary organic aerosol through photooxidation of
639   isoprene, Science, 303, 1173-1175, 10.1126/science.1092805, 2004.
640 Daellenbach, K. R., Kourtchev, I., Vogel, A. L., Bruns, E. A., Jiang, J., Petäjä, T., Jaffrezo, J.-L., Aksoyoglu,
641   S., Kalberer, M., Baltensperger, U., El Haddad, I., and Prévôt, A. S. H.: Impact of anthropogenic and
642   biogenic sources on the seasonal variation in the molecular composition of urban organic aerosols: a field
643   and laboratory study using ultra-high-resolution mass spectrometry, Atmospheric Chemistry and Physics,
644   19, 5973-5991, 10.5194/acp-19-5973-2019, 2019.
645 Ding, X., Zhang, Y.-Q., He, Q.-F., Yu, Q.-Q., Shen, R.-Q., Zhang, Y., Zhang, Z., Lyu, S.-J., Hu, Q.-H., Wang,
646   Y.-S., Li, L.-F., Song, W., and Wang, X.-M.: Spatial and seasonal variations of secondary organic aerosol
647   from terpenoids over China, J. geophy. Res.-Atoms., 121, 14661-14678, doi:10.1002/2016JD025467,
648   2016.
649 Elzein, A., Dunmore, R. E., Ward, M. W., Hamilton, J. F., and Lewis, A. C.: Variability of polycyclic
650   aromatic hydrocarbons and their oxidative derivatives in wintertime Beijing, China, Atmospheric
651   Chemistry and Physics, 19, 8741-8758, 10.5194/acp-19-8741-2019, 2019.
652 Fleming, L. T., Lin, P., Laskin, A., Laskin, J., Weltman, R., Edwards, R. D., Arora, N. K., Yadav, A.,
653   Meinardi, S., Blake, D. R., Pillarisetti, A., Smith, K. R., and Nizkorodov, S. A.: Molecular composition
654   of particulate matter emissions from dung and brushwood burning household cookstoves in Haryana,
655   India, Atmos. Chem. Phys., 18, 2461-2480, 10.5194/acp-18-2461-2018, 2018.
656 Fu, P., Kawamura, K., Okuzawa, K., Aggarwal, S. G., Wang, G., Kanaya, Y., and Wang, Z.: Organic
657   molecular compositions and temporal variations of summertime mountain aerosols over Mt. Tai, North
658   China Plain, J. Geophy. Res., 113, 10.1029/2008jd009900, 2008.
659 Glasius, M., Hansen, A. M. K., Claeys, M., Henzing, J. S., Jedynska, A. D., Kasper-Giebl, A., Kistler, M.,
660   Kristensen, K., Martinsson, J., Maenhaut, W., Nøjgaard, J. K., Spindler, G., Stenström, K. E., Swietlicki,
661   E., Szidat, S., Simpson, D., and Yttri, K. E.: Composition and sources of carbonaceous aerosols in
662   Northern Europe during winter, Atmos. Environ., 173, 127-141, 10.1016/j.atmosenv.2017.11.005, 2018.
663 Hansen, A. M. K., Kristensen, K., Nguyen, Q. T., Zare, A., Cozzi, F., Noejgaard, J. K., Skov, H., Brandt, J.,
664   Christensen, J. H., Strom, J., Tunved, P., Krejci, R., and Glasius, M.: Organosulfates and organic acids in
665   Arctic aerosols: speciation, annual variation and concentration levels, Atmos. Chem. Phys., 14, 7807-
666   7823, https://doi.org/10.5194/acp-14-7807-2014, 2014.
667 Hansen, A. M. K., Hong, J., Raatikainen, T., Kristensen, K., Ylisirniö, A., Virtanen, A., Petäjä, T., Glasius,
668   M., and Prisle, N. L.: Hygroscopic properties and cloud condensation nuclei activation of limonene-
669   derived organosulfates and their mixtures with ammonium sulfate, Atmos. Chem. Phys., 15, 14071-14089,
670   https://doi.org/10.5194/acp-15-14071-2015, 2015.
671 He, X., Huang, X. H. H., Chow, K. S., Wang, Q., Zhang, T., Wu, D., and Yu, J. Z.: Abundance and Sources
672   of Phthalic Acids, Benzene-Tricarboxylic Acids, and Phenolic Acids in PM2.5 at Urban and Suburban

673   Sites in Southern China, ACS Earth and Space Chemistry, 2, 147-158,
674   10.1021/acsearthspacechem.7b00131, 2018.
675 Ho, C. S., Lam, C. W. K., Chan, M. H. M., Cheung, R. C. K., Law, L. K., Suen, M. W. M., and Tai, H. L.:
676   Electrospray ionisation mass spectrometry: principles and clinical application, Clin. Biochem. Rev., 24,
677   10, 2003.
678 Hoffmann, T., Huang, R. J., and Kalberer, M.: Atmospheric analytical chemistry, Anal. Chem., 83, 4649-
679   4664, 10.1021/ac2010718, 2011.
680 Huang, G., Liu, Y., Shao, M., Li, Y., Chen, Q., Zheng, Y., Wu, Z., Liu, Y., Wu, Y., Hu, M., Li, X., Lu, S.,
681   Wang, C., Liu, J., Zheng, M., and Zhu, T.: Potentially Important Contribution of Gas-Phase Oxidation of
682   Naphthalene and Methylnaphthalene to Secondary Organic Aerosol during Haze Events in Beijing,
683   Environ Sci Technol, 53, 1235-1244, 10.1021/acs.est.8b04523, 2019.
684 Huang, R. J., Zhang, Y., Bozzetti, C., Ho, K. F., Cao, J. J., Han, Y., Daellenbach, K. R., Slowik, J. G., Platt,
685   S. M., Canonaco, F., Zotter, P., Wolf, R., Pieber, S. M., Bruns, E. A., Crippa, M., Ciarelli, G., Piazzalunga,
686   A., Schwikowski, M., Abbaszade, G., Schnelle-Kreis, J., Zimmermann, R., An, Z., Szidat, S.,
687   Baltensperger, U., El Haddad, I., and Prevot, A. S.: High secondary aerosol contribution to particulate
688   pollution during haze events in China, Nature, 514, 218-222, 10.1038/nature13774, 2014.
689 Huang, R. J., Cao, J. J., and Worsnop, D.: Sources and Chemical Composition of Particulate Matter During
690   Haze Pollution Events in China, in: Air pollution in Eastern Asia: an integrated perspective, edited by
691   Bouarar, I., Wang, X. M., and Brasseur, G. P., Springer, Cham, Switzerland, 49-68, 2017.
692 Iinuma, Y., Müller, C., Berndt, T., Böge, O., Claeys, M., and Herrmann, H.: Evidence for the existence of
693   organosulfates from β-pinene ozonolysis in ambient secondary organic aerosol, Environ. Sci. Technol.,
694   41, 6678-6683, 10.1021/es070938t, 2007.
695 Jung, J., and Kawamura, K.: Enhanced concentrations of citric acid in spring aerosols collected at the Gosan
696   background site in East Asia, Atmos. Environ., 45, 5266-5272, 10.1016/j.atmosenv.2011.06.065, 2011.
697 Kautzman, K. E., Surratt, J. D., Chan, M. N., Chan, A. W., Hersey, S. P., Chhabra, P. S., Dalleska, N. F.,
698   Wennberg, P. O., Flagan, R. C., and Seinfeld, J. H.: Chemical composition of gas- and aerosol-phase
699   products from photooxidation of naphthalene, J. Phys. Chem. A, 114, 913-934, 10.1021/jp908530s, 2010.
700 Kind, T., and Fiehn, O.: Seven Golden Rules for heuristic filtering of molecular formulas obtained by
701   accurate mass spectrometry, BMC Bioinformatics, 8, 10.1186/1471-2105-8-105, 2007.
702 Kourtchev, I., O'Connor, I. P., Giorio, C., Fuller, S. J., Kristensen, K., Maenhaut, W., Wenger, J. C., Sodeau,
703   J. R., Glasius, M., and Kalberer, M.: Effects of anthropogenic emissions on the molecular composition of
704   urban organic aerosols: An ultrahigh resolution mass spectrometry study, Atmo. Environ., 89, 525-532,
705   10.1016/j.atmosenv.2014.02.051, 2014.
706 Kourtchev, I., Godoi, R. H. M., Connors, S., Levine, J. G., Archibald, A. T., Godoi, A. F. L., Paralovo, S. L.,
707   Barbosa, C. G. G., Souza, R. A. F., Manzi, A. O., Seco, R., Sjostedt, S., Park, J.-H., Guenther, A., Kim,
708   S., Smith, J., Martin, S. T., and Kalberer, M.: Molecular composition of organic aerosols in central
709   Amazonia: an ultra-high-resolution mass spectrometry study, Atmos. Chem. Phys., 16, 11899-11913,
710   https://doi.org/10.5194/acp-16-11899-2016, 2016.
711 Kruve, A., Kaupmees, K., Liigand, J., and Leito, I.: Negative electrospray ionization via deprotonation:
712   predicting the ionization efficiency, Anal Chem, 86, 4822-4830, 10.1021/ac404066v, 2014.
713 Laskin, A., Laskin, J., and Nizkorodov, S. A.: Chemistry of atmospheric brown carbon, Chem. Rev., 115,
714   4335-4382, 10.1021/cr5006167, 2015.
715 Laskin, J., Laskin, A., Roach, P. J., Slysz, G. W., Anderson, G. A., Nizkorodov, S. A., Bones, D. L., and
716   Nguyen, L. Q.: High-Resolution Desorption Electrospray Ionization Mass Spectrometry for Chemical
717   Characterization of Organic Aerosols, Anal. Chem., 82, 2048-2058, 10.1021/ac902801f, 2010.
718 Laskin, J., Laskin, A., and Nizkorodov, S. A.: Mass Spectrometry Analysis in Atmospheric Chemistry, Anal.
719   Chem., 90, 166-189, 10.1021/acs.analchem.7b04249, 2018.
720 Lee, A., Goldstein, A. H., Kroll, J. H., Ng, N. L., Varutbangkul, V., Flagan, R. C., and Seinfeld, J. H.: Gas-
721   phase products and secondary aerosol yields from the photooxidation of 16 different terpenes, J. Geophy.
722   Res., 111, 10.1029/2006jd007050, 2006.
723 Leito, I., Herodes, K., Huopolainen, M., Virro, K., Kunnapas, A., Kruve, A., and Tanner, R.: Towards the
724   electrospray ionization mass spectrometry ionization efficiency scale of organic compounds, Rapid
725   Commun Mass Spectrom, 22, 379-384, 10.1002/rcm.3371, 2008.
726 Li, Y. J., Huang, D. D., Cheung, H. Y., Lee, A. K. Y., and Chan, C. K.: Aqueous-phase photochemical
727   oxidation and direct photolysis of vanillin – a model compound of methoxy phenols from biomass burning,

728   Atmospheric Chemistry and Physics, 14, 2871-2885, 10.5194/acp-14-2871-2014, 2014.

729   Lim, Y. B., Tan, Y., Perri, M. J., Seitzinger, S. P., and Turpin, B. J.: Aqueous chemistry and its role in
730   secondary organic aerosol (SOA) formation, Atmos. Chem. Phys., 10, 10521-10539, 10.5194/acp-10-
731   10521-2010, 2010.

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

735   Lin, P., Yu, J. Z., Engling, G., and Kalberer, M.: Organosulfates in humic-like substance fraction isolated
736   from aerosols at seven locations in East Asia: a study by ultra-high-resolution mass spectrometry, Environ.
737   Sci. Technol., 46, 13118-13127, 10.1021/es303570v, 2012b.

738   Lin, P., Laskin, J., Nizkorodov, S. A., and Laskin, A.: Revealing Brown Carbon Chromophores Produced in
739   Reactions of Methylglyoxal with Ammonium Sulfate, Environ. Sci. Technol., 49, 14257-14266,
740   10.1021/acs.est.5b03608, 2015.

741   McNeill, V. F., Woo, J. L., Kim, D. D., Schwier, A. N., Wannell, N. J., Sumner, A. J., and Barakat, J. M.:
742   Aqueous-phase secondary organic aerosol and organosulfate formation in atmospheric aerosols: a
743   modeling study, Environ Sci Technol, 46, 8075-8081, 10.1021/es3002986, 2012.

744   Ning, C., Gao, Y., Zhang, H., Yu, H., Wang, L., Geng, N., Cao, R., and Chen, J.: Molecular characterization
745   of dissolved organic matters in winter atmospheric fine particulate matters (PM2.5) from a coastal city of
746   northeast China, Sci Total Environ, 689, 312-321, 10.1016/j.scitotenv.2019.06.418, 2019.

747   Nizkorodov, S. A., Laskin, J., and Laskin, A.: Molecular chemistry of organic aerosols through the
748   application of high resolution mass spectrometry, Phys. Chem. Chem. Phys., 13, 3612-3629,
749   10.1039/c0cp02032j, 2011.

750   Noziere, B., Kalberer, M., Claeys, M., Allan, J., D'Anna, B., Decesari, S., Finessi, E., Glasius, M., Grgic, I.,
751   Hamilton, J. F., Hoffmann, T., Iinuma, Y., Jaoui, M., Kahnt, A., Kampf, C. J., Kourtchev, I., Maenhaut,
752   W., Marsden, N., Saarikoski, S., Schnelle-Kreis, J., Surratt, J. D., Szidat, S., Szmigielski, R., and
753   Wisthaler, A.: The molecular identification of organic compounds in the atmosphere: state of the art and
754   challenges, Chem. Rev., 115, 3919-3983, 10.1021/cr5003485, 2015.

755   Pereira, K. L., Hamilton, J. F., Rickard, A. R., Bloss, W. J., Alam, M. S., Camredon, M., Muñoz, A., Vázquez,
756   M., Borrás, E., and Ródenas, M.: Secondary organic aerosol formation and composition from the photo-
757   oxidation of methyl chavicol (estragole), Atmos. Chem. Phys., 14, 5349-5368, 10.5194/acp-14-5349-
758   2014, 2014.

759   Perry, R. H., Cooks, R. G., and Noll, R. J.: ORBITRAP MASS SPECTROMETRY: INSTRUMENTATION,
760   ION MOTION AND APPLICATIONS, Mass Spectrometry Reviews, 27, 661-699, 10.1002/mas.20186,
761   2008.

762   Rincón, A. G., Calvo, A. I., Dietzel, M., and Kalberer, M.: Seasonal differences of urban organic aerosol
763   composition - an ultra-high resolution mass spectrometry study, Environ. Chem., 9, 298,
764   10.1071/en12016, 2012.

765   Riva, M., Tomaz, S., Cui, T., Lin, Y.-H., Perraudin, E., Gold, A., Stone, E. A., Villenave, E., and Surratt, J.
766   D.: Evidence for an Unrecognized Secondary Anthropogenic Source of Organosulfates and Sulfonates:
767   Gas-Phase Oxidation of Polycyclic Aromatic Hydrocarbons in the Presence of Sulfate Aerosol, Environ.
768   Sci. Technol., 49, 6654-6664, 10.1021/acs.est.5b00836, 2015.

769   Schmidt, A. C., Herzschuh, R., Matysik, F. M., and Engewald, W.: Investigation of the ionisation and
770   fragmentation behaviour of different nitroaromatic compounds occurring as polar metabolites of
771   explosives using electrospray ionisation tandem mass spectrometry, Rapid Commun Mass Spectrom, 20,
772   2293-2302, 10.1002/rcm.2591, 2006.

773   Shi, Z., Vu, T., Kotthaus, S., Harrison, R. M., Grimmond, S., Yue, S., Zhu, T., Lee, J., Han, Y., Demuzere,
774   M., Dunmore, R. E., Ren, L., Liu, D., Wang, Y., Wild, O., Allan, J., Acton, W. J., Barlow, J., Barratt, B.,
775   Beddows, D., Bloss, W. J., Calzolai, G., Carruthers, D., Carslaw, D. C., Chan, Q., Chatzidiakou, L., Chen,
776   Y., Crilley, L., Coe, H., Dai, T., Doherty, R., Duan, F., Fu, P., Ge, B., Ge, M., Guan, D., Hamilton, J. F.,
777   He, K., Heal, M., Heard, D., Hewitt, C. N., Hollaway, M., Hu, M., Ji, D., Jiang, X., Jones, R., Kalberer,
778   M., Kelly, F. J., Kramer, L., Langford, B., Lin, C., Lewis, A. C., Li, J., Li, W., Liu, H., Liu, J., Loh, M.,
779   Lu, K., Lucarelli, F., Mann, G., McFiggans, G., Miller, M. R., Mills, G., Monk, P., Nemitz, E., amp, apos,
780   Connor, F., Ouyang, B., Palmer, P. I., Percival, C., Popoola, O., Reeves, C., Rickard, A. R., Shao, L., Shi,
781   G., Spracklen, D., Stevenson, D., Sun, Y., Sun, Z., Tao, S., Tong, S., Wang, Q., Wang, W., Wang, X.,
782   Wang, X., Wang, Z., Wei, L., Whalley, L., Wu, X., Wu, Z., Xie, P., Yang, F., Zhang, Q., Zhang, Y.,

Zhang, Y., and Zheng, M.: Introduction to the special issue "In-depth study of air pollution sources and processes within Beijing and its surrounding region (APHH-Beijing)", Atmospheric Chemistry and Physics, 19, 7519-7546, 10.5194/acp-19-7519-2019, 2019.

Song, J., Li, M., Jiang, B., Wei, S., Fan, X., and Peng, P.: Molecular Characterization of Water-Soluble Humic like Substances in Smoke Particles Emitted from Combustion of Biomass Materials and Coal Using Ultrahigh-Resolution Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry, Environ. Sci. Technol., 52, 2575-2585, 10.1021/acs.est.7b06126, 2018.

Sun, Y., Jiang, Q., Zhang, Z., Fu, P., Li, J., Yang, T., and Yin, Y.: Investigation of the sources and evolution processes of severe haze pollution in Beijing in January 2013, J. Geophy. Res.-Atmos., 119, 4380-4389, 10.1002/, 2014.

Surratt, J. D., Gomez-Gonzalez, Y., Chan, A. W., Vermeylen, R., Shahgholl, M., Kleindienst, T. E., Jaoui, M., Maenhaut, W., Claeys, M., Flagan, R. C., and Seinfeld, J. H.: Evidence for Organosulfate in Secondary Organic Aerosol, Environ. Sci. Technol., 41, 517-527, 10.1021/es062081q, 2007.

Surratt, J. D., Gómez-González, Y., Chan, A. W., Vermeylen, R., Shahgholi, M., Kleindienst, T. E., Edney, E. O., Offenberg, J. H., Lewandowski, M., Jaoui, M., Maenhaut, W., Claeys, M., Flagan, R. C., and Seinfeld, J. H.: Organosulfate Formation in Biogenic Secondary Organic Aerosol, J. Phys. Chem. A, 112, 8345-8378, 2008.

Tao, S., Lu, X., Levac, N., Bateman, A. P., Nguyen, T. B., Bones, D. L., Nizkorodov, S. A., Laskin, J., Laskin, A., and Yang, X.: Molecular Characterization of Organosulfates in Organic Aerosols from Shanghai and Los Angeles Urban Areas by Nanospray-Desorption Electrospray Ionization High-Resolution Mass Spectrometry, Environ. Sci. Technol., 48, 10993-11001, 10.1021/es5024674, 2014.

Tong, H., 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, 10.1039/c5fd00206k, 2016.

Tong, H., Zhang, Y., Filippi, A., Wang, T., Li, C., Liu, F., Leppla, D., Kourtchev, I., Wang, K., Keskinen, H. M., Levula, J. T., Arangio, A. M., Shen, F., Ditas, F., Martin, S. T., Artaxo, P., Godoi, R. H. M., Yamamoto, C. I., de Souza, R. A. F., Huang, R. J., Berkemeier, T., Wang, Y., Su, H., Cheng, Y., Pope, F. D., Fu, P., Yao, M., Pohlker, C., Petaja, T., Kulmala, M., Andreae, M. O., Shiraiwa, M., Poschl, U., Hoffmann, T., and Kalberer, M.: Radical Formation by Fine Particulate Matter Associated with Highly Oxygenated Molecules, Environ Sci Technol, 53, 12506-12518, 10.1021/acs.est.9b05149, 2019.

Tu, P., Hall, W. A. t., and Johnston, M. V.: Characterization of Highly Oxidized Molecules in Fresh and Aged Biogenic Secondary Organic Aerosol, Anal. Chem., 88, 4495-4501, 10.1021/acs.analchem.6b00378, 2016.

Wang, G., Kawamura, K., Umemoto, N., Xie, M., Hu, S., and Wang, Z.: Water-soluble organic compounds in PM2.5and size-segregated aerosols over Mount Tai in North China Plain, J. Geophy. Res., 114, 10.1029/2008jd011390, 2009.

Wang, K., Zhang, Y., Huang, R.-J., Cao, J., and Hoffmann, T.: UHPLC-Orbitrap mass spectrometric characterization of organic aerosol from a central European city (Mainz, Germany) and a Chinese megacity (Beijing), Atmos. Environ., 189, 22-29, 10.1016/j.atmosenv.2018.06.036, 2018.

Wang, K., Zhang, Y., Huang, R.-J., Wang, M., Ni, H., Kampf, C. J., Cheng, Y., Bilde, M., Glasius, M., and Hoffmann, T.: Molecular characterization and source identification of atmospheric particulate organosulfates using ultrahigh resolution mass spectrometry, Environ. Sci. Technol., 10.1021/acs.est.9b02628, 2019a.

Wang, M., Huang, R.-J., Cao, J., Dai, W., Zhou, J., Lin, C., Ni, H., Duan, J., Wang, T., Chen, Y., Li, Y., Chen, Q., Haddad, I. E., and Hoffmann, T.: Determination of n-alkanes, PAHs and hopanes in atmospheric aerosol: evaluation and comparison of thermal desorption GC-MS and solvent extraction GC-MS approaches, Atmos. Mea. Tech. Discuss., 1-21, 10.5194/amt-2019-4, 2019b.

Wang, X. K., Rossignol, S., Ma, Y., Yao, L., Wang, M. Y., Chen, J. M., George, C., and Wang, L.: Molecular characterization of atmospheric particulate organosulfates in three megacities at the middle and lower reaches of the Yangtze River, Atmos. Chem. Phys., 16, 2285-2298, https://doi.org/10.5194/acp-16-2285-2016, 2016.

Wang, X. K., Hayeck, N., Brüggemann, M., Yao, L., Chen, H. F., Zhang, C., Emmelin, C., Chen, J. M., George, C., and Wang, L.: Chemical characterization of organic aerosol in: A study by Ultrahigh-Performance Liquid Chromatography Coupled with Orbitrap Mass Spectrometry, J. Geophy. Res.-Atmos., 122, 703-722, https://doi.org/10.1002/2017JD026930, 2017.

838 Xu, W., Sun, Y., Wang, Q., Zhao, J., Wang, J., Ge, X., Xie, C., Zhou, W., Du, W., Li, J., Fu, P., Wang, Z.,
839   Worsnop, D. R., and Coe, H.: Changes in Aerosol Chemistry From 2014 to 2016 in Winter in Beijing:
840   Insights From High-Resolution Aerosol Mass Spectrometry, J. Geophy. Res.-Atmos., 124, 1132-1147,
841   10.1029/2018jd029245, 2019.
842 Yassine, M. M., Harir, M., Dabek-Zlotorzynska, E., and Schmitt-Kopplin, P.: Structural characterization of
843   organic aerosol using Fourier transform ion cyclotron resonance mass spectrometry: aromaticity
844   equivalent approach, Rapid Commun. Mass Spectrom., 28, 2445-2454, 10.1002/rcm.7038, 2014.
845 Zhang, P.: Revitalizing old industrial base of Northeast China: Process, policy and challenge, Chin. Geogra.
846   Sci., 18, 109-118, 10.1007/s11769-008-0109-2, 2008.
847 Zielinski, A. T., Kourtchev, I., Bortolini, C., Fuller, S. J., Giorio, C., Popoola, O. A. M., Bogialli, S., Tapparo,
848   A., Jones, R. L., and Kalberer, M.: A new processing scheme for ultra-high resolution direct infusion
849   mass spectrometry data, Atmos. Environ., 178, 129-139, 10.1016/j.atmosenv.2018.01.034, 2018.

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

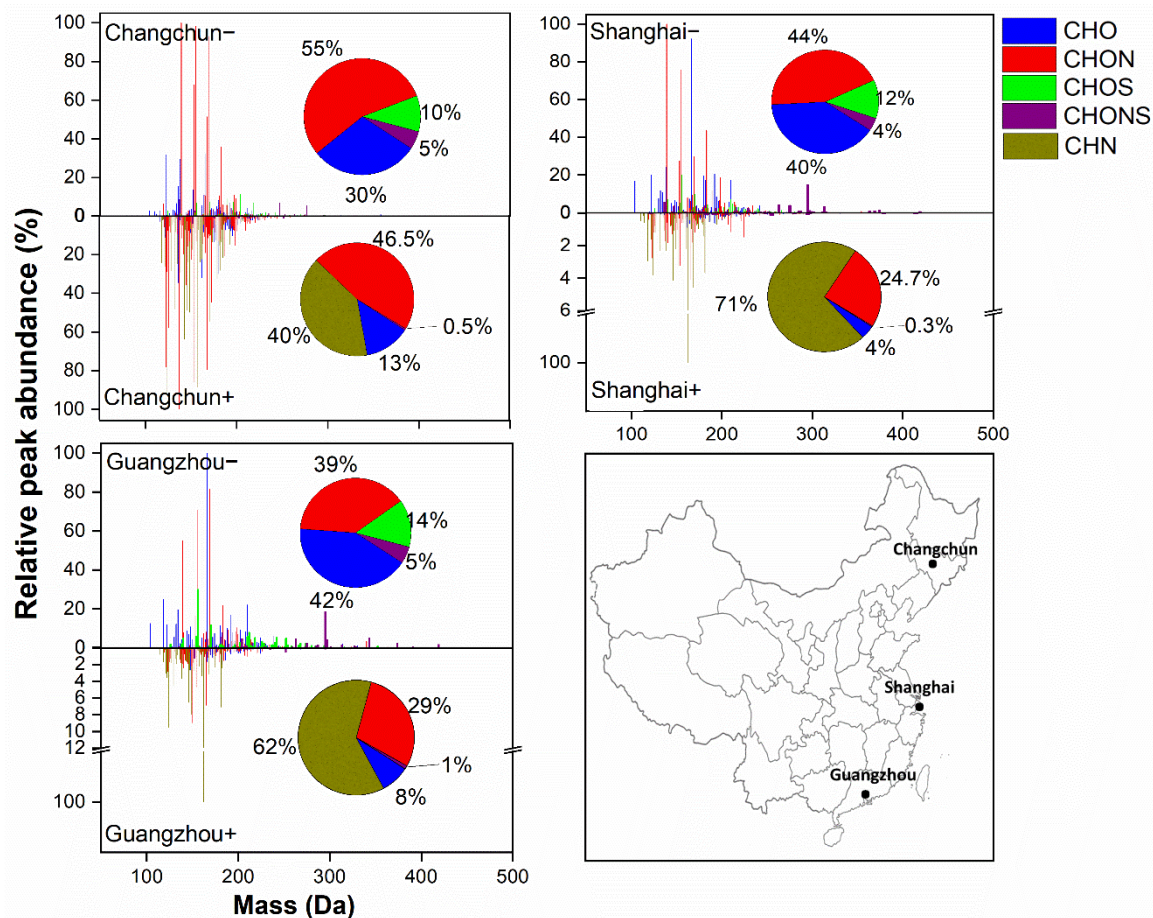
875

876 Table 1. Number of organic compounds and molecular formulas in each subgroup and the peak
877 abundance-weighted average values of molecular mass (MM_{avg}), elemental ratios, double bond
878 equivalent (DBE), aromaticity equivalent (Xc) and isomer number fraction (meaning the
879 percentage of formula numbers that have isomers among all assigned formulas) for detected
880 organic compounds in ESI− and ESI+ in the three Chinese cities.

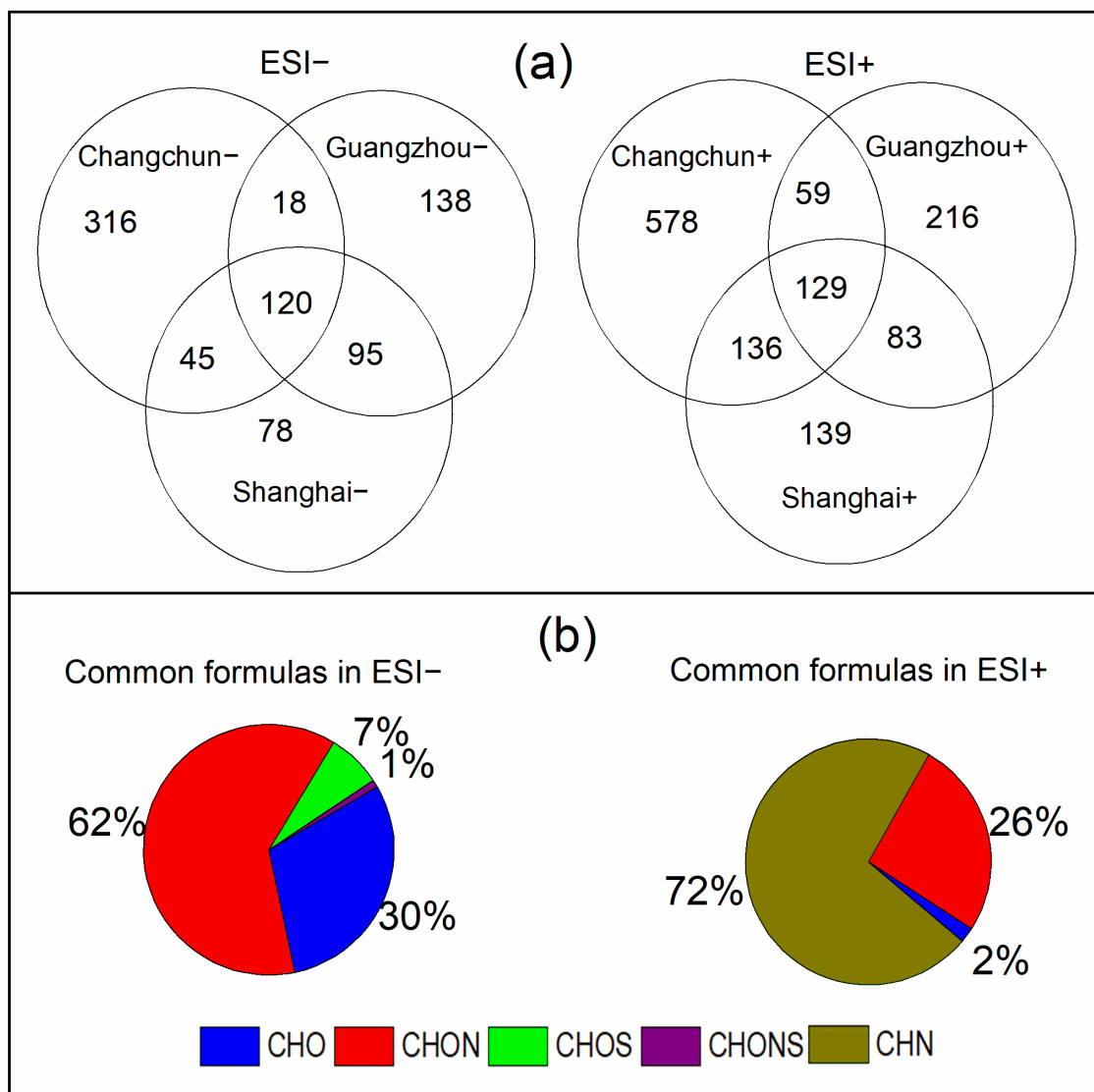
Sample ID	Subgroup	Number of compounds[*]	Relative abundance (%)	MM_{avg}	H/C	O/C[**]	DBE	Xc	Isomer number fraction (%)
Changchun−	total	769(415)	100	169	1.03	0.58	5.02	2.13	34
	CHO−	346(136)	30	162	0.96	0.41	5.65	2.28	52
	CHON−	180(96)	55	163	0.94	0.51	5.24	2.44	36
	CHOS−	155(105)	10	198	1.56	1.17(0.52)	2.55	0.50	28
	CHONS−	88(78)	5	214	1.35	1.07(-1.4)	3.75	1.06	8
Shanghai−	total	416(272)	100	176	1.05	0.69	4.99	1.92	31
	CHO−	164(90)	40	171	0.97	0.59	5.37	1.94	41
	CHON−	135(89)	44	169	0.86	0.56	5.67	2.47	37
	CHOS−	75(62)	12	190	1.85	1.41(0.61)	1.79	0.34	15
	CHONS−	42(31)	4	266	1.56	1.00(0.11)	3.30	0.44	13
Guangzhou−	total	488(304)	100	183	1.14	0.74	4.55	1.65	34
	CHO−	196(110)	42	172	1.10	0.65	4.68	1.57	44
	CHON−	161(98)	39	173	0.89	0.58	5.56	2.41	35
	CHOS−	86(67)	14	201	1.85	1.48(0.71)	1.71	0.21	21
	CHONS−	45(29)	5	293	1.56	0.82(0.06)	3.45	0.43	28
Changchun+	total	2943(679)	100	160	1.21	0.13	5.58	2.36	56
	CHO+	609(162)	13	174	0.94	0.28	6.55	2.22	50
	CHN+	696(126)	40	154	1.22	0.00	5.84	2.60	77
	CHON+	1594(352)	46.5	161	1.27	0.19	5.11	2.22	55
	CHONS+	44(39)	0.5	196	1.91	0.70	2.64	0.09	13
Shanghai+	total	704(383)	100	162	1.37	0.09	4.91	2.32	32
	CHO+	87(67)	4	184	1.13	0.43	5.46	1.46	19
	CHN+	253(84)	71	159	1.38	0.00	5.08	2.55	54
	CHON+	350(218)	24.7	167	1.40	0.27	4.34	1.81	30
	CHONS+	14(14)	0.3	241	1.17	0.61	5.32	0.91	0
Guangzhou+	total	687(412)	100	161	1.41	0.17	4.58	2.07	30
	CHO+	125(87)	8	185	1.12	0.42	5.19	1.20	26
	CHN+	205(78)	62	156	1.42	0.00	4.80	2.47	54
	CHON+	336(227)	29	165	1.47	0.45	4.00	1.51	26

| | CHONS+ | 21(20) | 1 | 209 | 1.84 | 0.71 | 3.05 | 0.31 | 5 |

881 *The values in brackets indicate the number of unique molecular formulas. **The values in brackets indicate the
882 (O−3S)/C and (O−3S−2N)/C ratios for CHOS and CHONS compounds, respectively, detected in ESI− mode

883

884

885 Figure 1. Mass spectra of detected organic compounds reconstructed from extracted ion
886 chromatograms in ESI− and ESI+. The horizontal axis refers to the molecular mass (Da) of the
887 identified species. The vertical axis refers to the relative peak abundance of each individual
888 compound to the compound with the greatest peak abundance. The pie charts show the percentage
889 of each organic compound subgroup (i.e. CHO, CHON, CHOS, CHONS and CHN) in each sample
890 in terms of peak abundance. The map in the lower right corner shows the locations of these three
891 megacities in China.

892

893

894   Figure 2. (a) Venn diagrams showing the number distribution of all molecular formulas detected in

895   ESI− and ESI+ for all sample locations. The overlapping molecular formulas refer to the

896   compounds detected in each city with the same molecular formulas and with the same retention

897   times (retention time difference ≤ 0.1 min). (b) Peak abundance contribution of each elemental

898   formula category to the total common formulas.

899

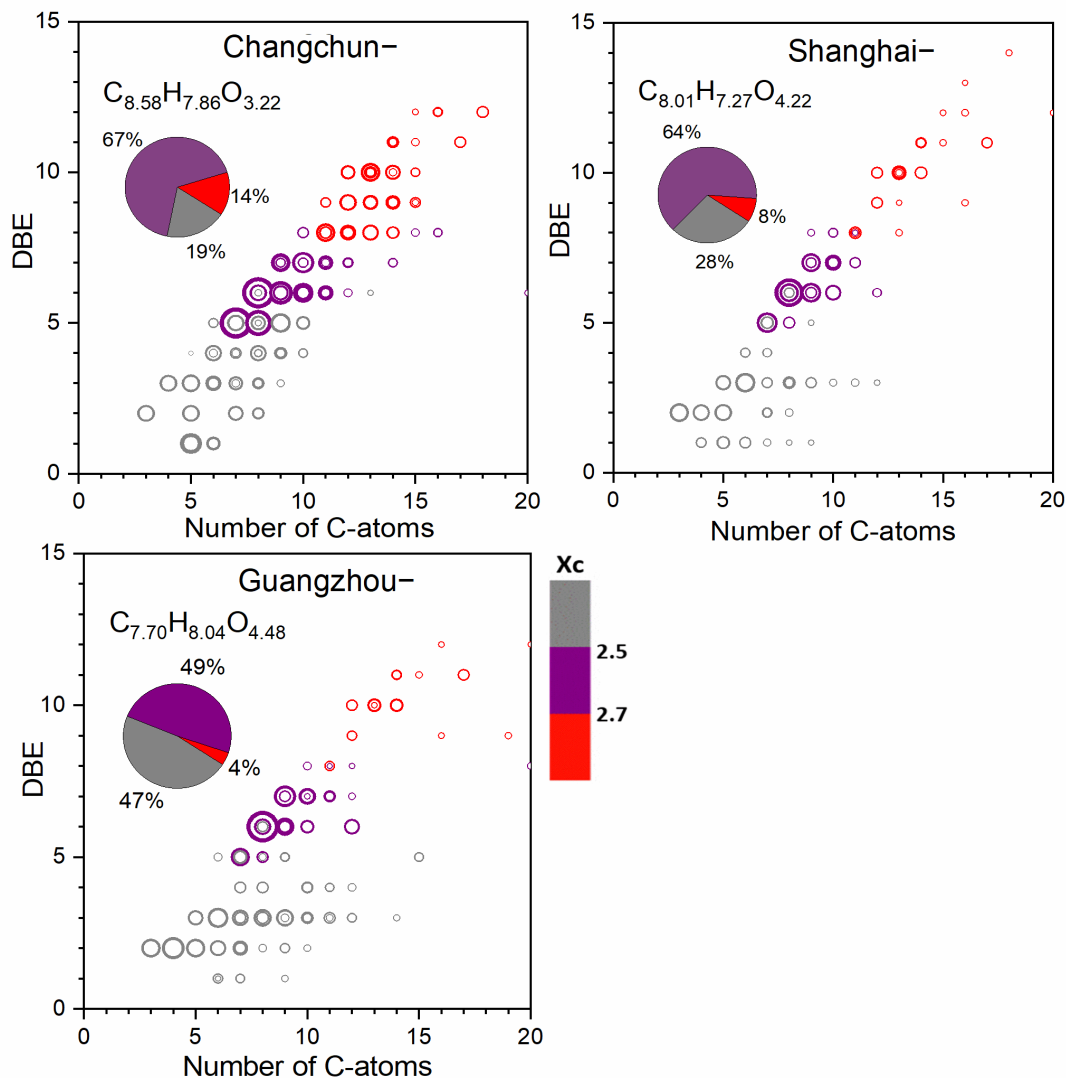



Figure 3. Double bond equivalent (DBE) versus carbon number for all CHO− compounds for all sample locations. The molecular formula represents the abundance-weighted average CHO− formula and the area of the circles is proportional to the fourth root of the peak abundance of an individual compound (a diagram with circle areas related to the absolute peak abundances is presented in Fig. S2). The color bar denotes the aromaticity equivalent (gray with $Xc < 2.50$, purple with $2.50 \leq Xc < 2.70$ and red with $Xc \geq 2.70$). The pie charts show the percentage of each Xc category (i.e., gray color-coded compounds, purple color-coded compounds and red color-coded compounds) in each sample in terms of peak abundance.

908

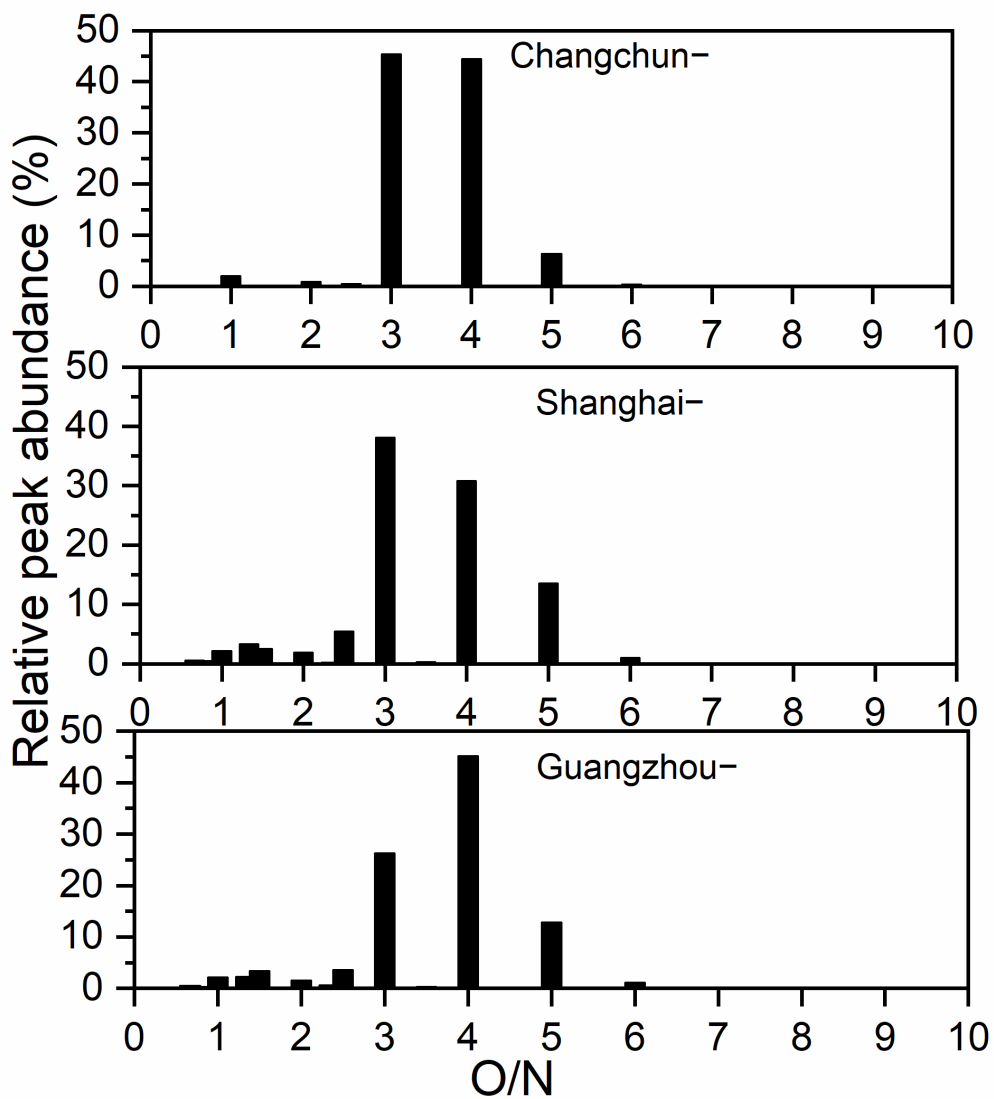
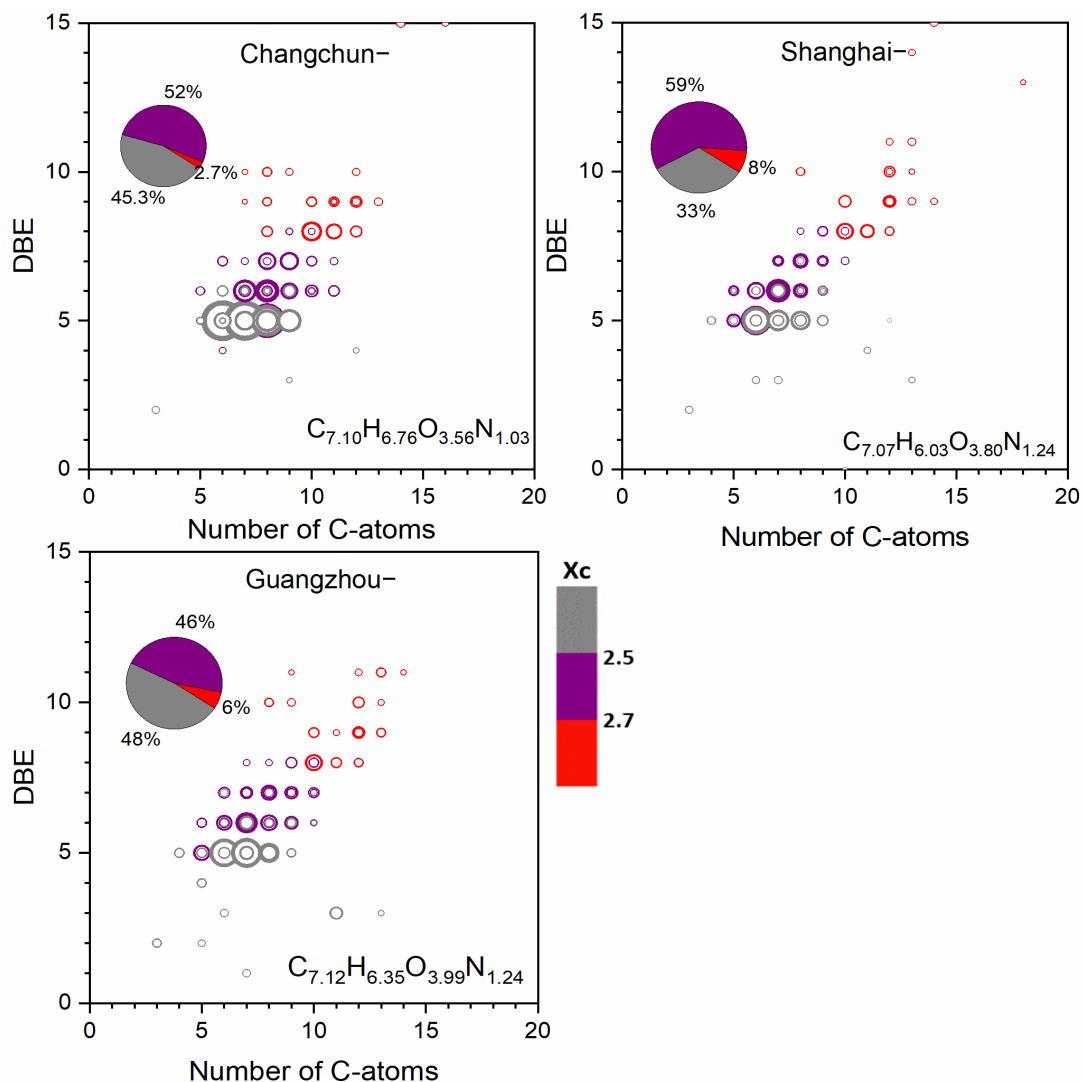



Figure 4. Classification of CHON− compounds into different subgroups according to O/N ratios in their formulas. The y-axis indicates the relative contribution of each specific O/N ratio subgroup to the sum of peak abundances of CHON− compounds.

912

913 Figure 5. Double bond equivalent (DBE) versus carbon number for all CHON− compounds for all

914 sample locations. The molecular formula represents the abundance-weighted average CHON-

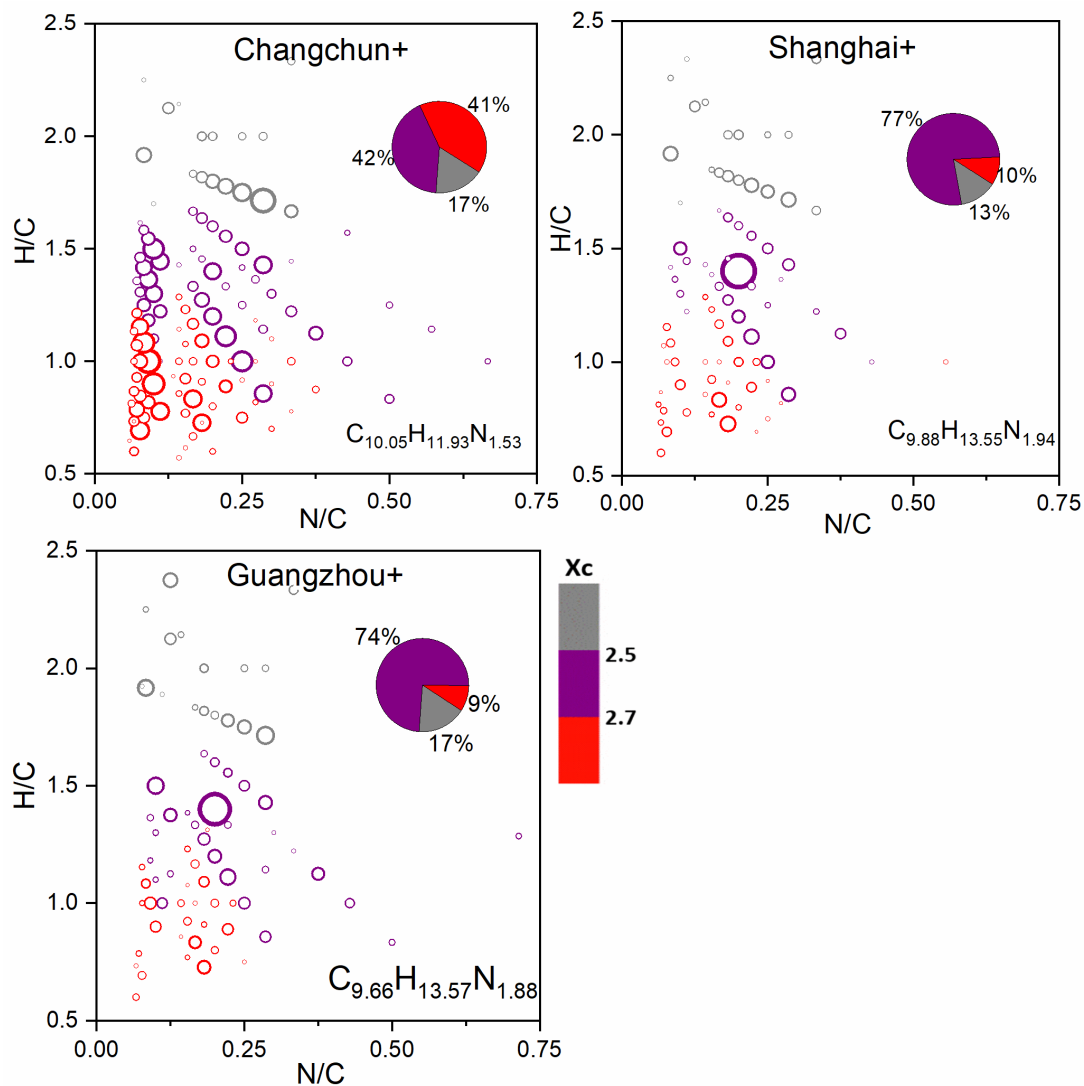
915 formula and the area of circles is proportional to the fourth root of the peak abundance of an

916 individual compound (a diagram with circle areas related to absolute peak abundances is presented

917 in Fig. S6). The color bar denotes the aromaticity equivalent (gray with $X_c < 2.50$, purple with 2.50

918 $\leq X_c < 2.70$ and red with $X_c \geq 2.70$). The pie charts show the percentage of each X_c category (i.e.,

919 gray color-coded compounds, purple color-coded compounds and red color-coded compounds) in

920 each sample in terms of peak abundance.

921

922 Figure 6. Van Krevelen diagrams for CHN+ compounds in Changchun, Shanghai and Guangzhou

923 samples. The area of circles is proportional to the fourth root of the peak abundance of an individual

924 compound (a diagram with circle areas related to absolute peak abundances is presented in Fig.

925 S10) and the color bar denotes the aromaticity equivalent (gray with $Xc < 2.50$, purple with $2.50 \leq$

926 $Xc < 2.70$ and red with $Xc \geq 2.70$). The pie charts show the percentage of each Xc category (i.e.,

927 gray color-coded compounds, purple color-coded compounds and red color-coded compounds) in

928 each sample in terms of peak abundance.

929

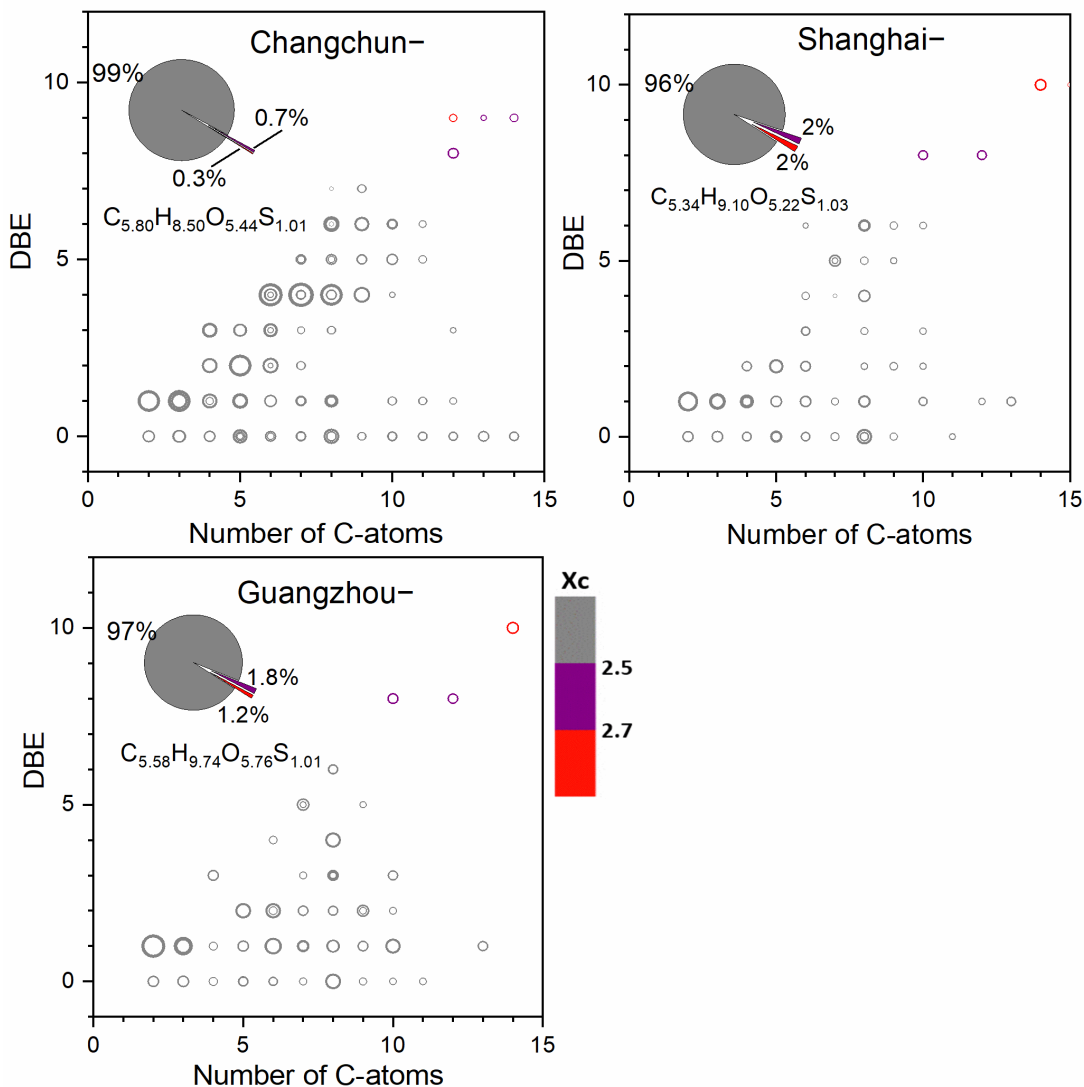



Figure 7. Double bond equivalent (DBE) versus carbon number for all CHOS− compounds for all sample locations. The molecular formula represents the abundance-weighted average CHOS− formula and the area of circles is proportional to the fourth root of the peak abundance of an individual compound (a diagram with circle areas related to absolute peak abundances is presented in Fig. S11). The color bar denotes the aromaticity equivalent (gray with Xc < 2.50, purple with $2.50 \leq Xc < 2.70$ and red with $Xc \geq 2.70$). The pie charts show the percentage of each Xc category (i.e., gray color-coded compounds, purple color-coded compounds and red color-coded compounds) in each sample in terms of peak abundance.

938

939

940