Supplement of Atmos. Chem. Phys., 21, 1775–1796, 2021
https://doi.org/10.5194/acp-21-1775-2021-supplement

Atmospheric Chemistry and Physics
Open Access
EGU

Supplement of

Characteristics, primary sources and secondary formation of water-soluble organic aerosols in downtown Beijing

Qing Yu et al.

Correspondence to: Jing Chen (jingchen@bnu.edu.cn)

S1 Estimation for the sampling artifacts of organic aerosols.

Sampling of organic carbon is accompanied by both positive and negative artifacts. The positive artifact is due to the adsorption of gaseous organics to the sampling filter, and the negative artifact is caused by the evaporation of collected particulate organic carbon. To eliminate the positive artifact, a denuder can be placed upstream the sample filter to remove the gaseous organics by diffusion to the adsorbent surface (Cheng et al., 2009). The use of a denuder in the sampling system has been reported in previous studies (Eatough et al., 1993, 1999; Mader et al., 2001; Matsumoto et al., 2003; Viana et al., 2006; Cheng et al., 2009, 2010, 2012; Kristensen et al., 2016). The use of a denuder may induce a larger negative artifact, however, as the removal of gaseous organics can enhance the evaporation of particulate OC. Thus a backup filter should also be included in the sampling system (Cheng et al., 2009). Besides, the flow rate passing through the denuder was very low in most studies (Matsumoto et al., 2003; Viana et al., 2006; Cheng et al., 2009, 2010, 2012; Kristensen et al., 2016). This might be due to the significantly decreased removal efficiency of the denuder as the air flow rate increased (Cui et al., 1998; Ding et al., 2002). To collect enough samples for the accurate measurement of trace organic species, the flow rate of 1.05 m^3 min^{-1} was chosen in this study. The air flow rate of about 1.05 m^3 min^{-1} has been frequently used in the field sampling of organic aerosols (Kawamura et al., 2013; Verma et al., 2012, 2015; Li et al., 2018; Ma et al., 2018; Huang et al., 2020). At this flow rate, a denuder with a high removal efficiency is hardly commercially available.

We estimated the sampling artifact of OC based on the literature results. Firstly, different OC fractions which have distinct volatility show different adsorption behavior. Besides, the adsorption behavior of the same OC fraction also vary with meteorological conditions. Cheng et al. (2015) compared the concentrations of different OC fractions (OC1, OC2, OC3, OC4) on bare quartz filters with those on denuded quartz filters in four seasons of Beijing, and the results are listed in Table S1. The contributions of different OC fractions measured in this study are also shown in Table S1.

In addition, the positive artifact of OC also depends on the sampling procedure. McDow (1986) systematically investigated the effect of sampling procedure on the OC measurement. The adsorption of organic vapors on the bare quartz filters ($C_{positive\ artifact}$) was a function of the sampling duration (t) multiplied by the face velocity (v) as follows:

$$C_{positive\ artifact} = \sum_i \rho_i \frac{1 - e^{\varepsilon_i v t}}{\varepsilon_i v t} \quad (1)$$

where the face velocity (v, cm·s^{-1}) is the ratio of the flow rate (cm^3·s^{-1}) to the sampling area of the filter (cm^2), ρ_i is the concentration of adsorptive vapor i (g·cm^{-3}), and ε_i is a constant which can be defined as:

$$\varepsilon_i = \frac{1}{l}[\beta f(t) t_0 \exp\left(\frac{Q_A}{RT}\right) + 1] \quad (2)$$

where l is the effective filter thickness. The average thickness of the quartz filter used in this study was 463 μm. The other parameters are all constants.

Therefore, it can be calculated that $\varepsilon_i > 1/l > 20$ cm^{-1}, and $1\text{-}e^{-\varepsilon v t} \approx 1$. Hence, the positive artifact ($C_{positive\ artifact}$) is inversely proportional to the product of the sampling duration and the face velocity ($v \times t$). The face velocity of Cheng et al. (2015) was 9.8 cm·s^{-1}, while that in our study was 47.3 cm·s^{-1}. The sampling duration of Cheng et al. (2015) was 24 h, while that in our study was 12 h. That is to say, the positive artifact of Cheng et al. (2015) was about 2.4 times higher than that in our study.

Based on the literature results and taking into account the above factors (seasons, OC fractions, sampling procedure), the contribution of positive artifact to the measured OC was estimated to be 2.3 %, 1.4 %, 9.9 %, and 2.2 % during the sampling periods in winter, spring, summer and autumn respectively in this study, which is roughly acceptable.

To further estimate the impact of gas-particle partitioning and potential reactions occurring on filters, we overlapped two quartz filters and took samples at a flow rate of 1.05 m^3·min^{-1} for a duration of 12 h. The organic tracers selected in this study were measured. The organic tracers on the backup filters typically originate from three sources: (1) adsorption of the gas-phase organic species; (2) adsorption of the semi-volatile species evaporated from the front filter; (3) secondary formation from the adsorbed organic vapors on the backup filter. Except for *cis*-pinonic acid, the tracer concentrations on the backup filter were all less than 5 % of those on the front filters, while the concentration of *cis*-pinonic acid on the backup filter was 21.6 % of that on the front filter. This result suggested that the sampling procedure in this study might bring some uncertainties for the measurement of *cis*-pinonic acid, and the sampling artifact was not significant for the other organic tracers.

S2 Detailed information for the chemical analysis.

S2-1 Chemical analysis of water-soluble ions and water-soluble organic carbon (WSOC)

To analyze the concentrations of water-soluble ions and water-soluble organic carbon (WSOC), a punch of each sampled filter was cut into pieces and extracted with 40 mL ultrapure water (>18.2 MΩ) for 30 min, then passed through a 0.45 μm PTFE filter. Five cations (Na^+, NH_4^+, K^+, Mg^{2+}, Ca^{2+}) and four anions (Cl^-, NO_3^-, SO_4^{2-}, $C_2O_4^{2-}$) were measured using the ion chromatography (Dionex 600), with the methanesulfonic acid (MSA) solution as cationic eluent and the potassium hydroxide (KOH) solution as anionic eluent. The concentration of WSOC was measured by a TOC analyzer (Shimazdu TOC-L CPN). The standard solution of total carbon (TC) was prepared by potassium acid phthalate ($C_8H_5KO_4$), and that of inorganic carbon (IC) was made by sodium carbonate (Na_2CO_3) and sodium bicarbonate ($NaHCO_3$). Total organic carbon (TOC) was calculated as total carbon minus inorganic carbon.

S2-2 The parameter settings of GC/MS/MS for analyzing organic tracers

The derivatives were immediately analyzed by a Shimadzu TQ8040 gas chromatography triple quadrupole mass spectrometry (GC/MS/MS). A JA-5MS capillary column (30 m × 0.25 mm i.d., film thickness 0.25 μm) was used as the GC column and helium was used as the carrier gas (1.0 mL min^{-1}). The injector was set splitless at a temperature of 290 °C. The programmed oven temperature increased from 70 °C to 150 °C at 2 °C min^{-1}, then to 200 °C at 5 °C min^{-1}, then to 300 °C at 25 °C min^{-1}, and stayed at 300 °C for 6 min. The MS was operated in EI mode at 70 eV with a scan range of 50-650 amu.

S3 The detailed procedures of the PMF source apportionment

1. Uncertainties of the input data

According to the User Guide of PMF5.0 (Norris et al., 2014), the uncertainties of the target species can be calculated as follow:

$$\mathrm{Unc} = 5/6 \times \mathrm{MDL} \quad (\mathrm{c} \le \mathrm{MDL}) \tag{3}$$

$$\mathrm{Unc} = \sqrt{(\mathrm{P} \times \mathrm{c})^2 + (0.5 \times \mathrm{MDL})^2} \quad (\mathrm{c} > \mathrm{MDL}) \tag{4}$$

where *Unc* is the data uncertainty, c is the concentration of the target species, MDL is the method detection limit, and P is the error fraction. Since the User Guide did not give the calculation method for the error fraction (P), we estimated the P values referring to the measured relative standard deviations (RSD) of the target species. The RSD values were calculated by measuring six identical portions of an ambient sample. P was set as 10 % when RSD < 10 %, and 15 % or 20 % when RSD > 10 %.

2. Selection of base solutions

The chemical components input into the PMF model were selected based on our understanding of the possible WSOC sources (Norris et al., 2014). Interpretability was usually considered to be the most important factor for selecting the optimum PMF solution (Shrivastava et al., 2007; Huang et al., 2014). The interpretable solutions are those which group tracers from different sources into distinct factors, while those grouping tracers from multiple sources into the same factor, distributing tracers for one source across multiple factors, or including factors with no distinct grouping of species are judged less interpretable (Shrivastava et al., 2007; Sowlat et al., 2016). In some previous literature, the optimal solution was defined as that with the maximum number of factors which had distinctive groupings of species, and explained at least 90 % of the total variable (Shrivastava et al., 2007). In this study, PMF was run repeatedly by changing the number of factors and the start seed numbers. The base solution was selected based on: (1) the interpretability of the derived factor profiles and the temporal variations of source contributions; (2) the reconstruction of the total variable and R^2 of input organic tracers (R^2>0.90); (3) the scaled residuals of the input species.

As presented in Figure S1, the 7-factor solution separated cholesterol (the tracer for cooking) into multiple sources. It was difficult to explain why cholesterol appeared in the factor profiles of biomass burning, dust and fresh biogenic SOC. Besides, this solution led to poor fits for cholesterol

(R^2 = 0.28) and *cis*-pinonic acid (R^2 = 0.32), which were the key tracers selected in this study. Therefore, the 7-factor solution was not selected. As shown in Figure S2, the 8-factor solution also distributed cholesterol into multiple factors. This solution also resulted in a poor fit (R^2 = 0.28) for cholesterol. Therefore, the 8-factor solution was not chosen in this study. As shown in Figures S5-7, the solutions with 4 to 6 factors all showed poor interpretability for the derived factor profiles and poor fits for the key organic tracers. The 10-factor solution involved a factor without any tracer of high loading to indicate a specific source, thus could not be explained. By comparing the results with different factor numbers, the solution with 9 factors (Figure S3) was thought to be the most interpretable one.

3. Diagnostics for the base model run

The selected 9-factor solution was converged, and Q(Robust) was similar to Q(True). As shown in Figure S4, most of the input species showed normally distributed residuals between -2.0 and +2.0, indicating that these species were well modeled. The R^2 of WSOC and HULIS were 0.94 and 0.93, respectively, and the R^2 for all the organic tracers were higher than 0.96, again suggesting that these species were well modeled.

4. Error estimation

The selected base solution was subjected to displacement (DISP) and bootstrap (BS) tests for error estimation. For the DISP test, the percent change in Q (%dQ) was less than 0.1 %, indicating that this solution was the global minimum (Paatero et al., 2014). No factor swapped for any value of dQmax, indicating little rotational ambiguity in this solution (Paatero et al., 2014). For the BS test, the factor of "cooking" was mapped 79 % of the runs, the factor of "other primary combustion sources" was mapped 69 % of the runs, while other factors were mapped more than 91 % of the runs. The BS results indicated some uncertainties for the factors of cooking and other primary combustion sources, while the other factors were relatively stable.

Brown et al. (2015) indicated that the unstable PMF solution might be due to too many factors involved. To investigate the effect of factor number on the stability of solutions, the BS results for solutions with different factor numbers were compared and shown in Table S3. As shown in Table S3, reducing the number of factors did not significantly increase the successful rates of BS mapping, but decreased the interpretability of the derived factor profiles. As recommended by the previous

studies (Norris et al., 2014; Paatero et al., 2014), some constraints can be defined based on the priori information of the sources to reduce the variability of the solution.

5. Constrained model run

Bozzetti et al. (2017) exploited the markers' source specificity to set constraints for the profiles, so as to solve the problem of large mixtures of PMF factors associated with contributions of markers from different sources. They treated the contribution of unrelated source-specific markers as zero for each source, while non-source-specific variables were freely apportioned by the PMF algorithm. In addition, they set constraints for primary markers and combustion-related markers that can be seen as negligible in the secondary factors.

In the constrained model run, we set the constraints similar to those of Bozzetti et al. (2017), with a slight difference that we set the constraints by "soft pulling" so as to obtain a stable solution with a minimal change in the Q-value (dQ). The constraints were set as follows: (1) Levoglucosan was pulled up maximally with a limit of 0.25 % dQ for the factor of "primary biomass burning"; (2) Cholesterol was pulled up maximally with a limit of 0.50 % dQ for the factor of "cooking"; (3) Sulfate, *cis*-pinonic acid and 2-methylerythritol were pulled down maximally with limits of 0.25 % dQ for the factor of "other primary combustion sources"; (4) Phthalic acid was pulled up maximally with a limit of 0.25 % dQ for the factor of "aromatic SOA". The dQ(Robust) for all the constraints were 0.93 % in the final constrained model run, which was acceptable (below 1 %) as recommended by the PMF user guide (Norris et al., 2014). As shown in Table S4, all the factors were mapped more than 94 % of the runs, suggesting that this solution was stable. Thus the constrained 9-factor solution was chosen as the final solution.

6. Factor identification

The source profiles of the final solution are shown in Figure 4. Factor 1 showed high levels of levoglucosan and EC, thus was interpreted as the direct emissions from biomass burning. Factor 2 exhibited a high level of cholesterol, thus was regarded as cooking. Factor 3 showed a large fraction of EC that could not be explained by the direct emissions from biomass burning, suggesting that it was the direct emissions from other combustion sources, such as coal combustion, traffic emissions and waste burning. Factor 4 was featured by high loadings of Mg^{2+} and Ca^{2+}, thus was considered as dust. No significant EC but high fractions of 4-methyl-5-nitrocatechol and phthalic acid were found

in Factor 5 and Factor 6, respectively, which were regarded as SOC from biomass burning (biomass burning SOC) and aromatic precursors (aromatic SOC), respectively. Factor 7 exhibited a high level of *cis*-pinonic acid, thus was explained as fresh biogenic SOC. Factor 8 was characterized by high fractions of 2-methylerythritol and 3-hydroxyglutaric acid, which are the end products from isoprene and monoterpenes respectively, thus was identified as aged biogenic SOC. Note that *cis*-pinonic acid and 3-hydroxyglutaric acid were not grouped into the same factor though they are both SOA tracers of monoterpenes, owing to their different oxidation degree as discussed above. Factor 9 covered the secondary components (such as SO_4^{2-}, NO_3^-, NH_4^+ and $C_2O_4^{2-}$) that can not be well explained by the identified sources above, thus was considered to be SOC from other sources.

6-1 The reason why we separated the primary and secondary biomass burning SOA.

Levoglucosan correlated strongly with 4-methyl-5-nitrocatechol ($r=0.87$, $p<0.01$) in this study. It seems that levoglucosan and 4-methyl-5-nitrocatechol should be distributed into the same factor. Nevertheless, in fact, even though we reduced the factor number from nine to five, levoglucosan and 4-methyl-5-nitrocatechol could not be merged into one factor (Figure S1, 2, 3, 5, 6). When the factor number decreased to four, levoglucosan and 4-methyl-5-nitrocatechol were merged into one factor (Figure S8). However, this solution was less interpretable, and resulted in poorer fits for most of the input species (cholesterol: $R^2=0.17$; *cis*-pinonic acid: $R^2=0.25$; Ca^{2+}: $R^2=0.67$; Mg^{2+}: $R^2=0.73$; NO_3^-: $R^2=0.75$; etc). Furthermore, the slope of the fitting equation for the observed and predicted values of 4-methyl-5-nitrocatechol was even only 0.31, that is, the high values of 4-methyl-5-nitrocatechol in winter were not reproduced by the 4-factor solution. Hence, the 4-factor solution was also excluded in this study.

It was indeed interesting that levoglucosan and 4-methyl-5-nitrocatechol were not distributed in the same factor, though they showed strong correlation with each other. We attempted to explain this phenomenon as follows. The ratio of 4-methyl-5-nitrocatechol to levoglucosan showed significantly higher values ($p<0.01$) in winter (0.071 ± 0.029) than in other seasons (0.010 ± 0.009), which implied different types of biomass burning sources (primary and secondary). If they were merged into one factor, the ratio of 4-methyl-5-nitrocatechol to levoglucosan was regarded to be constant throughout the year, which was not the truth. According to the uncertainty estimation method for the input species (Equation 2), the data with lower concentrations usually have lower uncertainties, thus may

have a larger impact on the Q value. Taking the 4-factor solution (Figure S7) as an example, when these two tracers were merged into one factor, to minimize the Q value, the algorithm in the PMF model tended to assign a low value for the ratio of 4-methyl-5-nitrocatechol to levoglucosan in the factor profile of biomass burning (i.e. 0.024 in Factor 4). As shown in Figure S8, this ratio (orange line) was closer to the regression slope in other seasons (0.017), but much lower than that in winter (0.096). As a consequence, the high concentration of 4-methyl-5-nitrocatechol in winter could not be reproduced at all by such PMF solution. In conclusion, the solution which merged these two tracers into the same factor might bring about large uncertainties, and fail to reproduce the peak values of 4-methyl-5-nitrocatechol over the study period in winter.

Fresh biomass burning emissions show a high fraction of anhydrosugar, such as levoglucosan. The relative intensity of anhydrosugar decreased due to the degradation or oxidation reactions (Gilardon et al., 2016; Sengupta et al., 2020). Chamber studies indicated that substantial amounts of nitrogen-containing organic compounds, such as nitrophenols and methyl-nitrocatechols, were generated during aging (Bertrand et al., 2018; Hartikainen et al., 2018). 4-Methyl-5-nitrocatechol was recommended to be the secondary tracer for aged biomass burning SOA (Bertrand et al., 2018). In the previous studies (Gilardon et al., 2016; Zhou et al., 2017; Srivastava et al., 2018), the biomass burning source was also separated into primary and secondary fractions with the PMF model. In Gilardon et al. (2016), the factor profile of primary biomass burning was featured by a high loading of anhydrosugar (signal at m/z=60) but a low level of the aged OA (signal at m/z=44), while the factor profile of biomass burning SOA was featured by a high fraction of the aged OA (signal at m/z=44) but a low fraction of anhydrosugar (signal at m/z=60) (Gilardon et al., 2016). Srivastava et al. (2018) also used levoglucosan and 4-methyl-5-nitrocatechol as tracers to differentiate the primary biomass burning and biomass burning SOA using the PMF model (92 samples, which was less than that in our study). In this study, as shown in Figure 4, Factor 1 had high fractions of levoglucosan and EC, but a low fraction of 4-methyl-5-nitrocatechol, thus was considered as the direct emission from biomass burning. The concentration ratio of levoglucosan to WSOC in this factor was 0.085 $\mu g \cdot \mu g^{-1}$, similar to that measured in the primary combustion of crop straws (0.097 $\mu g \cdot \mu g^{-1}$), wood (0.081 $\mu g \cdot \mu g^{-1}$) and leaves (0.095 $\mu g \cdot \mu g^{-1}$) in North China (Yan et al., 2018). Factor 5 showed a high level of 4-methyl-5-nitrocatechol, but low loadings of EC and levoglucosan, thus was identified as biomass burning SOA.

6-2 The interpretation for Factor 7, Factor 8, and Factor 9.

As presented in Figures S1-3 and Figures S5-7, even if we reduced the factor number from nine to four, 2-methylerythritol, 3-hydroxyglutaric acid and *cis*-pinonic acid could not be merged into the same factor. Large fractions of 3-hydroxyglutaric acid and 2-methylerythritol were usually grouped into one factor, since they strongly correlated with each other ($r=0.94$, $p<0.01$). *Cis*-pinonic acid could not be distributed in this factor since it correlated less strongly with 2-methylerythritol ($r=0.51$, $p<0.01$) and 3-hydroxyglutaric acid ($r=0.58$, $p<0.01$). As stated in Section 3.2, *cis*-pinonic acid is a lower-generation oxidative product from monoterpenes, while 2-methylerythritol and 3-hydroxyglutaric acid are more aged products from isoprene and monoterpenes, respectively (Kourtchev et al., 2009). Hence, Factor 7 with a high level of cis-pinonic acid was interpreted as the fresh biogenic SOC, and Factor 8 with high loadings of 2-methylerythritol and 3-hydroxyglutaric acid was interpreted as aged biogenic SOC. As shown in Figure 5, the seasonal variation of their source contributions also supported this interpretation.

Since the major fraction of 3-hydroxyglutaric acid was distributed in Factor 8, it was not proper to interpret Factor 9 as monoterpene SOC. In fact, as shown in Figure S1-3 and Figure S5-7, a minor fraction of 3-hydroxyglutaric acid was always distributed in factors other than the biogenic SOC. It was more interpretable when this minor fraction of 3-hydroxyglutaric acid was distributed in the same factor together with SO_4^{2-}, NO_3^-, NH_4^+ and $C_2O_4^{2-}$. In this case, Factor 9 of the selected 9-factor solution could be interpreted as a mixed secondary source and explain the secondary species that were not well fitted by other identified secondary sources. Similar factor profile has also been resolved in the literature, and was usually interpreted as the "inorganic-rich SOA" (Huang et al., 2014).

6-3 The interpretation for Factor 3 (Other primary combustion sources).

As shown in Figure 4, for the constrained 9-factor solution, Factor 3 showed a significant level of EC that could not be explained by direct emissions of biomass burning, implying that it could be associated with the primary emissions from other combustion sources, such as coal combustion, traffic emissions, and waste burning, etc. Indeed, a minor fraction of SO_4^{2-} (20.8 %), NH_4^+ (19.3 %) and phthalic acid (20.0 %) were also distributed in Factor 3. However, in fact, previous studies have indicated that SO_4^{2-}, NO_3^-, NH_4^+ and phthalic acid can also be directly emitted from coal combustion

(Zhang et al., 2008; Dai et al., 2019) and traffic emissions (Al-Naiema and Stone, 2017; Hao et al., 2019). Dai et al. (2019) suggested that primary SO_4^{2-} accounted for 38.9 % and 16.9 % to the total SO_4^{2-} in $PM_{2.5}$ in the heating and non-heating seasons respectively. Accordingly, such distribution of SO_4^{2-} (20.8 %) in Factor 3 was acceptable. Similar loadings of SO_4^{2-} and NH_4^+ were also found in the factor profile of coal combustion in the previous source apportionment study (Huang et al., 2014). To sum up, it was reasonable that a minor fraction of SO_4^{2-}, NH_4^+ and phthalic acid presented in the factor profile of Factor 3, i.e., other primary combustion sources.

Table S1 The ratio of the OC concentrations on the bare quartz filters to those on the denuded quartz filters in Cheng et al. (2015), as well as the contribution of different OC fractions measured in this study.

	The ratio of OC on bare quartz filters to denuded quartz filters (Cheng et al., 2015)				The contribution of different OC fractions measured in this study			
	OC1	OC2	OC3	OC4	OC1	OC2	OC3	OC4
Winter	1.27	1.03	1.02	1.05	10.8 %	19.6 %	24.7 %	44.8 %
Spring	2.05	1.05	1.00	1.01	3.9 %	27.2 %	43.1 %	25.7 %
Summer	2.45	1.60	1.17	1.08	4.4 %	37.6 %	36.0 %	22.0 %
Autumn	2.08	1.05	0.99	1.01	7.9 %	26.5 %	40.2 %	25.3 %

Table S2 The detailed information of the authentic standards used in this study.

Authentic standard	Molecular formula	CAS number	Company	Purity
Levoglucosan	$C_6H_{10}O_5$	498-07-7	Sigma-Aldrich	99%
Cholesterol	$C_{27}H_{46}O$	57-88-5	Sigma-Aldrich	93%
Phthalic acid	$C_8H_6O_4$	88-99-3	Sigma-Aldrich	99%
4-Methyl-5-nitrocatechol	$C_7H_7NO_4$	68906-21-8	Toronto Research Chemicals	98%
2-Methylerythritol	$C_5H_{12}O_4$	58698-37-6	Sigma-Aldrich	90%
3-Hydroxyglutaric acid	$C_5H_8O_5$	638-18-6	Sigma-Aldrich	95%
cis-Pinonic acid	$C_{10}H_{16}O_3$	61826-55-9	Sigma-Aldrich	98%

Table S3 The successful rates of BS mapping for the solutions with different numbers of factors. The values no more than 85 % were shown in bold.

BS mapping	3-factor solution	4-factor solution	5-factor solution	6-factor solution	7-factor solution	8-factor solution	9-factor solution
Factor 1	93 %	93 %	95 %	95 %	**82 %**	99 %	99 %
Factor 2	**83 %**	**85 %**	99 %	98 %	88 %	95 %	**79 %**
Factor 3	100 %	95 %	**78 %**	99 %	**74 %**	100 %	100 %
Factor 4		87 %	**74 %**	**75 %**	**79 %**	98 %	**69 %**
Factor 5			94 %	91 %	92 %	92 %	96 %
Factor 6				**81 %**	99 %	**67 %**	96 %
Factor 7					98 %	100 %	100 %
Factor 8						98 %	91 %
Factor 9							99 %

Table S4 The successful rates of BS mapping for the constrained 9-factor solution.

BS Mapping:	Base Factor 1	Base Factor 2	Base Factor 3	Base Factor 4	Base Factor 5	Base Factor 6	Base Factor 7	Base Factor 8	Base Factor 9	Unmapped
Boot Factor 1	97 %	0	0	0	3 %	0	0	0	0	0
Boot Factor 2	1 %	95 %	0	0	0	0	0	1 %	3 %	0
Boot Factor 3	0	0	100 %	0	0	0	0	0	0	0
Boot Factor 4	4 %	0	0	94 %	0	1 %	0	1 %	0	0
Boot Factor 5	3 %	0	0	0	97 %	0	0	0	0	0
Boot Factor 6	0	0	0	1 %	0	99 %	0	0	0	0
Boot Factor 7	0	0	0	0	0	0	100 %	0	0	0
Boot Factor 8	0	0	0	0	0	1 %	0	97 %	2 %	0
Boot Factor 9	0	0	0	0	0	0	0	0	100 %	0

Table S5 WSOC, OC concentrations and WSOC/OC ratios in $PM_{2.5}$ Beijing in recent years.

Sampling period	WSOC (μg m^{-3})	OC (μg m^{-3})	WSOC/OC	OC method*	Reference
2017 Winter	11.71 ± 13.92	20.56 ± 21.89	0.52 ± 0.08	IMPROVE (TOR)	This study
2017 Spring	4.40 ± 2.34	8.70 ± 3.12	0.49 ± 0.12		
2017 Summer	4.66 ± 2.46	7.75 ± 3.91	0.60 ± 0.11		
2017 Autumn	4.77 ± 2.83	9.71 ± 3.69	0.46 ± 0.12		
2016 Spring	3.8 ± 3.8[a]	7.9 ± 7.4	0.48	Optical transmittance	Yang et al. (2019)
2016 Summer	2.3 ± 0.9[a]	3.4 ± 1.0	0.68		
2016 Autumn	4.8 ± 4.1[a]	7.9 ± 7.6	0.61		
2016 Winter	10.0 ± 10.1[a]	20.7 ± 16.2	0.49		
2013 Autumn	Not mentioned	11.4	0.70 ± 0.27	IMPROVE (TOR)	Zhao et al. (2018)
2013 Winter		19.4	0.49 ± 0.11		
2014 Spring		8.53	0.56 ± 0.07		
2014 Summer		5.29	0.58 ± 0.10		
2013-2014 Winter	12.8	29.1	0.46	IMPROVE-A	Huang et al. (2020)
2013 Winter	10.8 ± 3.1	32.9 ± 16.8	0.39 ± 0.16	IMPROVE-A (TOT)	Yan et al. (2015)
2013 Summer	6.4 ± 2.2	9.7 ± 2.9	0.66 ± 0.06		
2012 Summer	4.4 ± 3.6	8.5 ± 5.2	0.52	IMPROVE-A (TOR)	Li et al. (2019a)
2012 Autumn	5.2 ± 4.0	10.3 ± 7.4	0.50		
2013 Winter	10.3 ± 9.8	28.9 ± 22.0	0.36		
2013 Spring	5.9 ± 4.9	14.6 ± 10.8	0.40		
2011-2012 Winter	Not mentioned	Not mentioned	0.36 ± 0.05[b] 0.44 ± 0.05[c] 0.47 ± 0.05[d]	IMPROVE-A (TOT)	Cheng et al. (2015)
2011 Summer	4.48	13.55	0.33	IMPROVE (TOR)	Xiang et al. (2017)
2011 Autumn	5.82	25.42	0.25		
2011 Winter	5.53	28.16	0.20		
2012 Spring	3.90	16.57	0.27		
2012 Summer	5.81	16.54	0.34		
2011 Summer	7.8 ± 4.4	12.0 ± 6.3	0.65	IMPROVE-A (TOT)	Cheng et al. (2013)
2011-2012 Winter	11.2 ± 8.2	24.6 ± 17.1	0.46		
2010 Fall	8.6 ± 6.4	20.4 ± 15.4	0.42	IMPROVE (TOT)	Du et al. (2014)
2010 Winter	8.0 ± 6.7	20.6 ± 16.1	0.39		
2011 Spring	4.7 ± 3.1	10.2 ± 6.8	0.46		
2011 Summer	6.7 ± 4.4	10.7 ± 6.2	0.61		
2011 Fall	8.6 ± 6.1	19.7 ± 15.4	0.44		
2009 Spring	6.7 ± 1.8	13.7 ± 4.4	0.49	Not mentioned	Tao et al.(2016)
2009 Summer	3.2 ± 1.1	11.1 ± 1.8	0.29		
2009 Autumn	7.7 ± 5.0	17.8 ± 5.6	0.43		

2010 Winter	7.7 ± 3.6	24.9 ± 15.6	0.31		
2009 Winter	7.28	27.7 ± 15.4[e]	0.26	IMPROVE-A (TOR)	Cheng et al. (2011)
		30.9 ± 16.3[f]	0.24		
		32.6 ± 18.6[e]	0.22	IMPROVE-A (TOT)	
		36.1 ± 19.5[f]	0.20		
2009 Summer	3.36	7.2 ± 2.4[e]	0.48	IMPROVE-A (TOR)	
		9.4 ± 2.7[f]	0.36		
		8.8 ± 3.3[e]	0.38	IMPROVE-A (TOT)	
		11.4 ± 3.6[f]	0.30		

[*] The thermal-optical reflectance (TOR) method and thermal-optical transmittance (TOT) method are two different charring correction methods to determine the split of OC and EC. The transmittance-defined EC is the carbon measured after the filter transmittance returns to its initial value in the He/O_2 atmosphere, whereas the reflectance-defined EC is the carbon measured after the filter reflectance returns to its initial value (Cheng et al., 2011).

[a] In Yang et al. (2019), the concentrations of WSOC were measured by UV/VIS absorption (at wavelengths of about 250 nm)

[b,c,d] In Cheng et al. (2015), "b" refers to the constructed $PM_{2.5}$ below 30 μg m^{-3}, "c" between 30 μg m^{-3} and 90 μg m^{-3}, and "d" above 90 μg m^{-3}.

[e,f] In Cheng et al. (2011), "e" was measured using the denuded quartz filter and "f" was measured using the un-denuded (bare) quartz filter.

Table S6 The mean values and standard deviations (SD) of planetary boundary layer (PBL) heights and wind speeds in Beijing during the study periods in four seasons.

		Winter		Spring		Summer		Autumn	
		Day	**Night**	**Day**	**Night**	**Day**	**Night**	**Day**	**Night**
PBL (m)	**Mean**	309.7	283.2	1149.4	303.9	871.6	210.5	423.2	90.9
	SD	236.7	259.1	841.1	363.2	405.6	164.3	278.9	90.4
WS (m s^{-1})	**Mean**	1.14	0.83	1.75	0.83	1.03	0.63	0.72	0.39
	SD	0.79	0.61	0.71	0.75	0.68	0.44	0.62	0.25

Table S7 Spearman correlation coefficients between SOA tracers as well as the WSOC/OC ratio and meteorological parameters, O_3, aerosol acidity, and aerosol liquid water content (LWC) in Beijing during the sampling periods in four seasons.

Compounds	Sampling periods	T	RH	WS	SR	O_3	Acidity	LWC
WSOC/OC	Whole	0.296**	0.290**	-0.142	0.021	0.152	0.444**	0.387**
	January	0.495**	0.550**	-0.298	-0.508	-0.267	0.783**	0.684**
	April	-0.149	0.640**	-0.317	-0.762**	-0.424*	0.580**	0.680**
	July	0.273	-0.066	-0.053	-0.082	0.081	0.499**	0.063
	October	0.469**	0.365*	-0.074	-0.068	-0.073	0.803**	0.619**
4-Methyl-5-nitrocatechol	Whole	-0.780**	0.051	-0.229*	-0.691**	-0.841**	-0.247*	0.227*
	January	0.201	0.688**	-0.712**	-0.618*	-0.766**	0.730**	0.875**
	April	-0.516**	0.880**	-0.553**	-0.932**	-0.685**	0.482**	0.816**
	July	-	-	-	-	-	-	-
	October	-0.259	0.211	-0.281	-0.044	-0.504**	-0.037	0.186
Phthalic acid	Whole	0.048	0.212*	-0.198*	-0.137	-0.146	0.181	0.452**
	January	0.389*	0.657**	-0.621**	-0.648*	-0.645**	0.815**	0.869**
	April	-0.158	0.681**	-0.300	-0.865**	-0.486**	0.660**	0.768**
	July	0.714**	-0.597*	0.308	0.186	0.653**	0.547**	-0.321
	October	0.549**	0.325	-0.082	-0.103	-0.013	0.803**	0.579**
2-Methylerythritol	Whole	0.595**	0.545**	-0.343**	0.211	0.333**	0.786**	0.562**
	January	0.304	0.686**	-0.624**	-0.530	-0.679**	0.789**	0.868**
	April	-0.079	0.563**	-0.440*	-0.524*	-0.535**	0.694**	0.545**
	July	0.657**	-0.399*	0.249	0.343	0.563**	0.759**	-0.074
	October	0.255	0.371*	-0.243	0.109	-0.253	0.627**	0.520**
3-Hydroxyglutaric acid	Whole	0.626**	0.534**	-0.299**	0.212	0.372**	0.822**	0.576**
	January	0.495**	0.649**	-0.640**	-0.653*	-0.615**	0.893**	0.844**
	April	-0.146	0.672**	-0.338	-0.715**	-0.429*	0.699**	0.668**
	July	0.533**	-0.293	0.059	0.186	0.536**	0.718**	0.108
	October	0.482**	0.340	-0.002	0.018	0.012	0.832**	0.613**
cis-Pinonic acid	Whole	0.591**	0.032	-0.092	0.178	0.348**	0.586**	0.111
	January	0.368*	0.577**	-0.650**	-0.473	-0.660**	0.728**	0.790**
	April	0.263	0.232	-0.128	-0.785**	-0.141	0.675**	0.390*
	July	-0.007	-0.204	-0.149	0.236	-0.139	-0.314	-0.336
	October	0.778**	-0.426*	0.306	0.344	0.537**	0.714**	-0.161

T: temperature; RH: relative humidity; WS: wind speed; SR: solar radiation; LWC: liquid water content.

Level of significance: *: $p<0.05$; **: $p<0.01$.

Table S8 Correlations between the ratios of SOA tracers to OC and RH as well as LWC.

Compounds	Sampling periods	RH	LWC
4-Methyl-5-nitrocatechol/OC	Whole	0.192	0.248*
	January	0.733**	0.819**
	April	0.856**	0.750**
	July	-	-
	October	0.220	0.105
Phthalic acid/OC	Whole	0.105	0.220*
	January	0.380*	0.666**
	April	0.536**	0.599**
	July	-0.272	-0.156
	October	0.265	0.499**
2-Methylerythritol/OC	Whole	0.506**	0.385**
	January	0.593**	0.690**
	April	0.244	0.174
	July	-0.079	0.081
	October	0.399*	0.438**
3-Hydroxyglutaric acid/OC	Whole	0.439**	0.323**
	January	-0.119	-0.168
	April	0.396*	0.366*
	July	0.136	0.356
	October	0.327	0.590**
cis-Pinonic acid/OC	Whole	-0.076	-0.208*
	January	-0.668**	-0.788**
	April	-0.214	-0.126
	July	0.292	-0.016
	October	-0.547**	-0.389*

Level of significance: *: $p<0.05$; **: $p<0.01$.

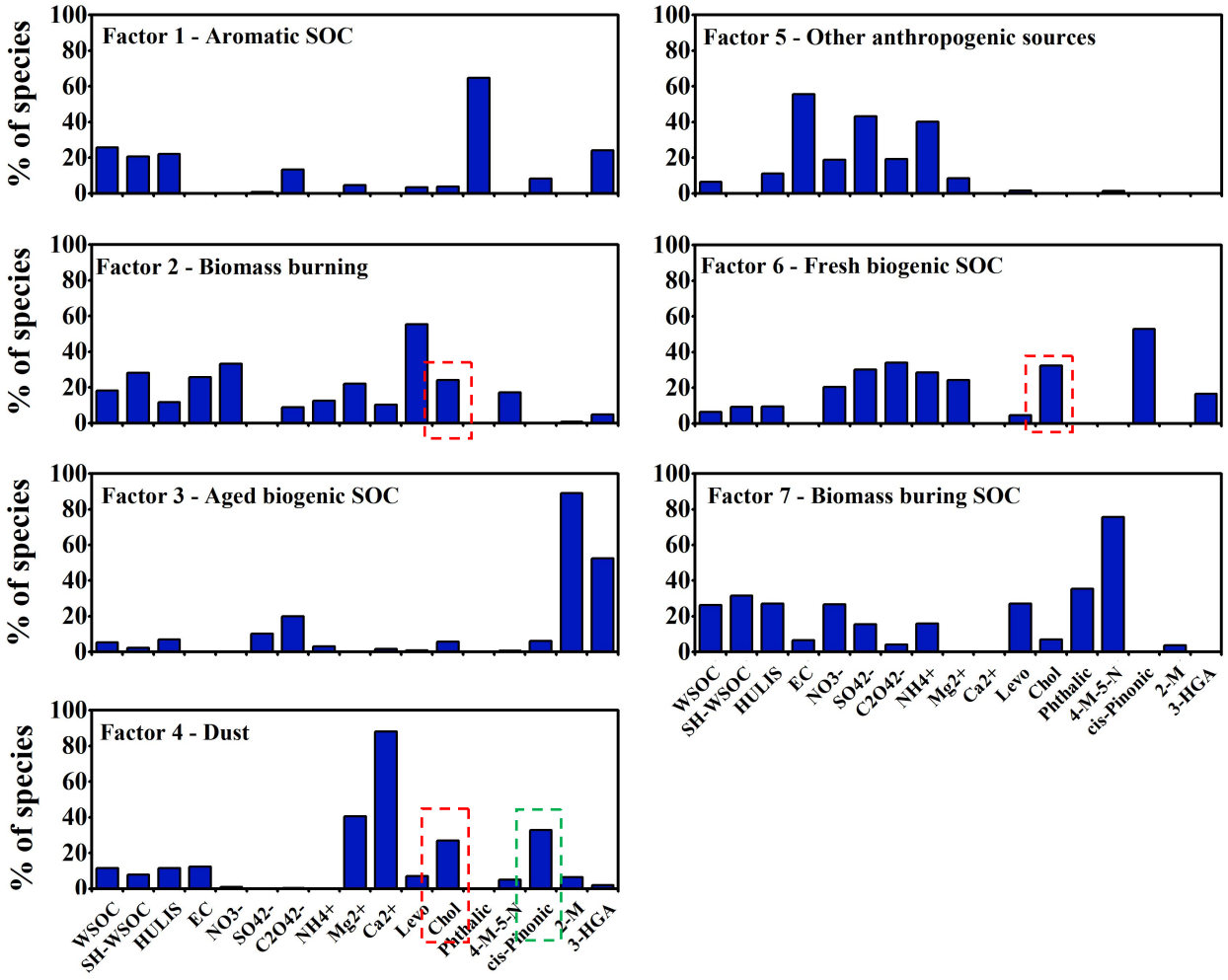

Figure S1. A 7-factor solution resolved by the PMF model.

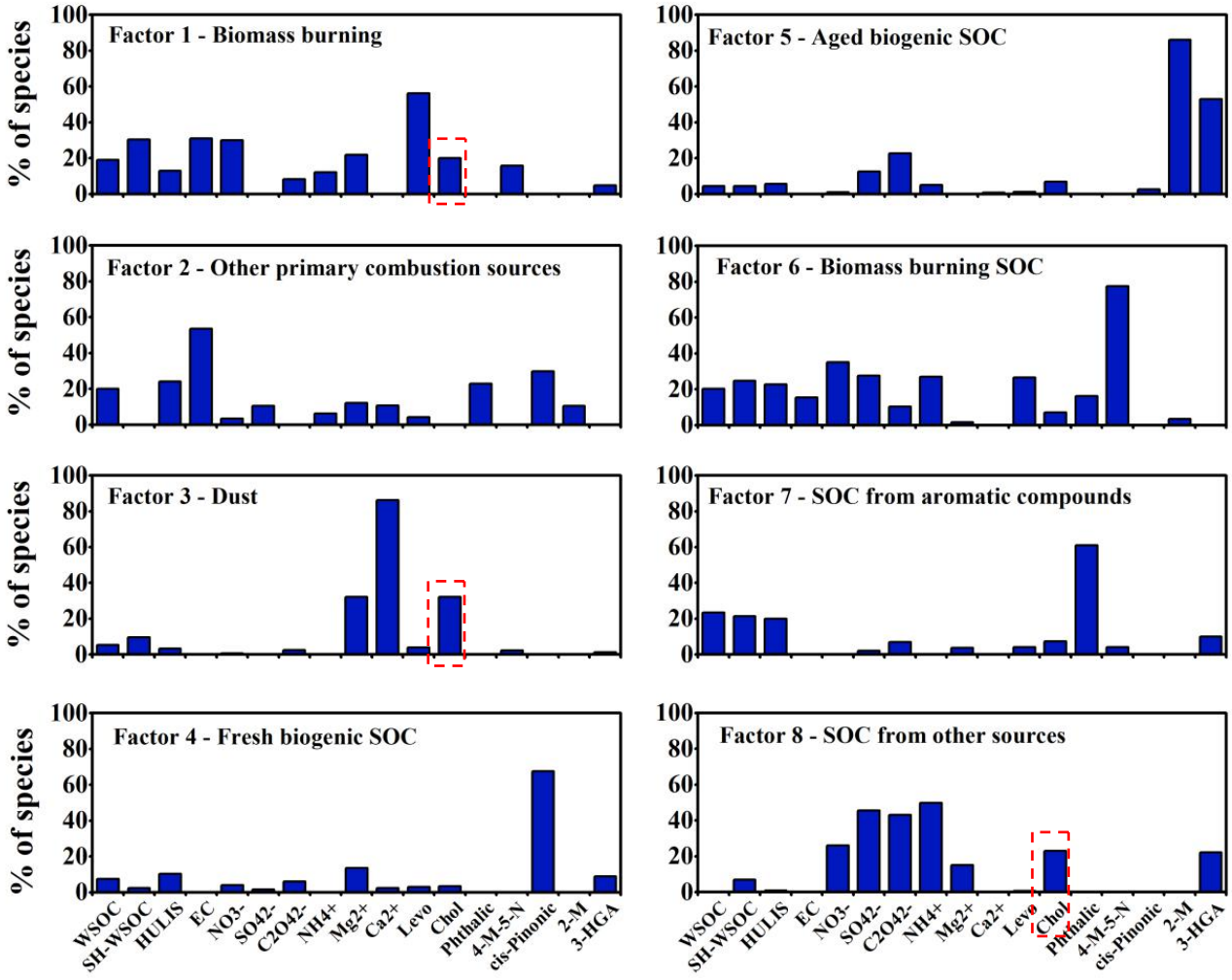

Figure S2. An 8-factor solution resolved by the PMF model.

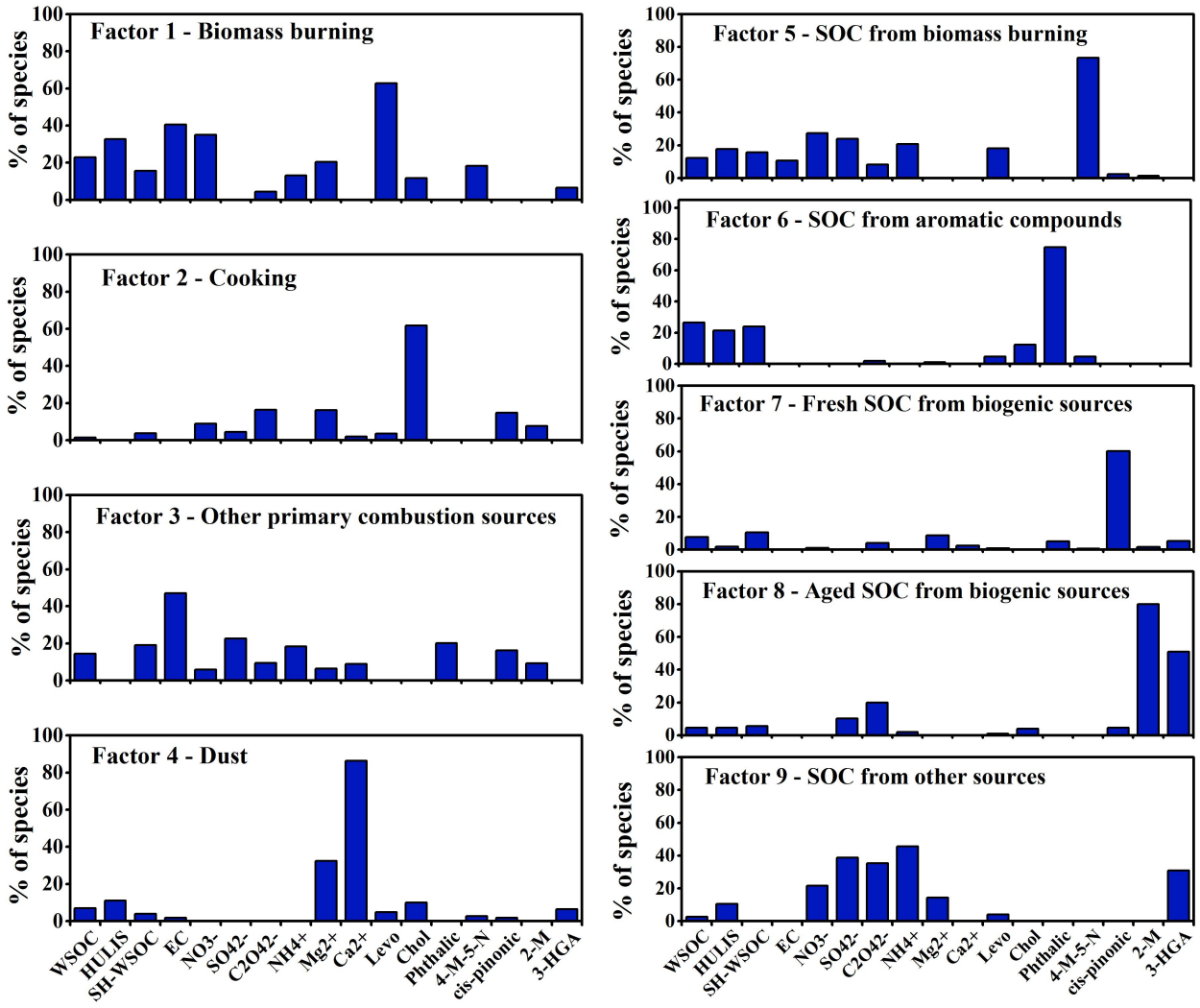

Figure S3. A 9-factor solution resolved by the PMF model.

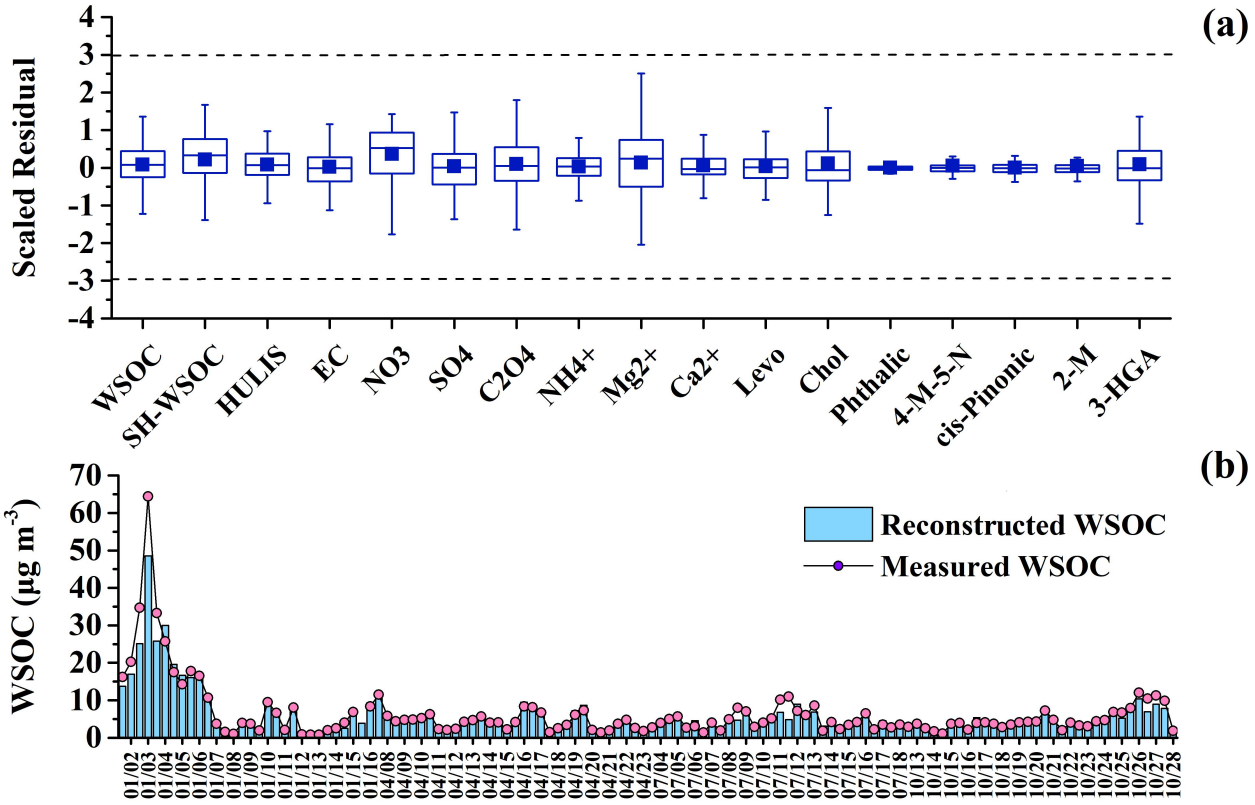

Figure S4. (a) The box plots showing the distributions of the scaled residuals for each species; (b) The time series of the measured WSOC and the reconstructed WSOC based on the 9-factor solution.

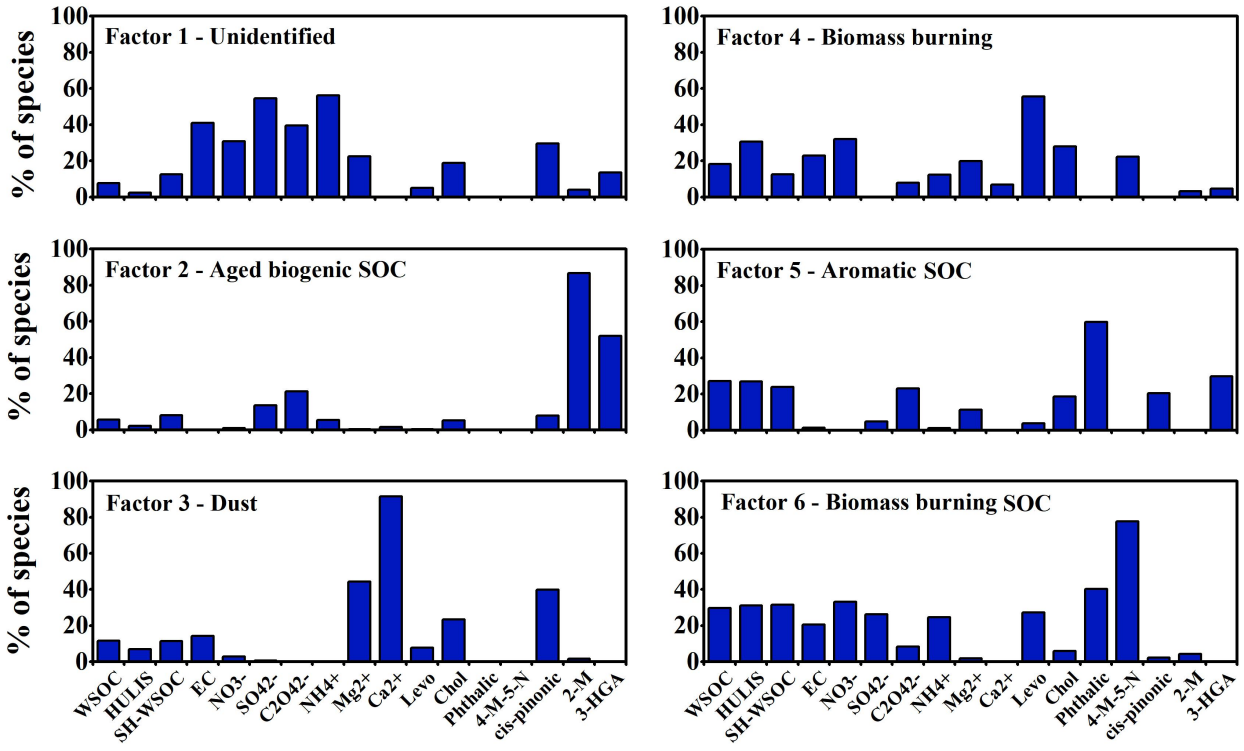

Figure S5. A 6-factor solution resolved by the PMF model.

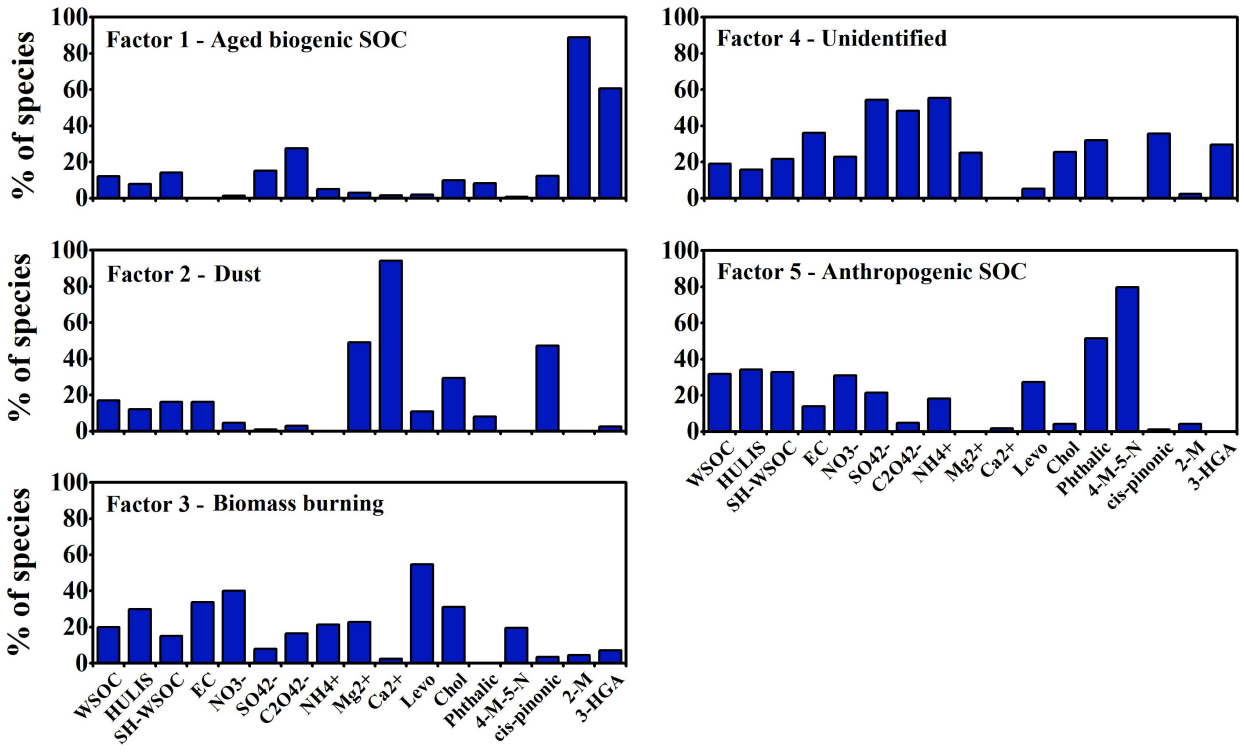

Figure S6. A 5-factor solution resolved by the PMF model.

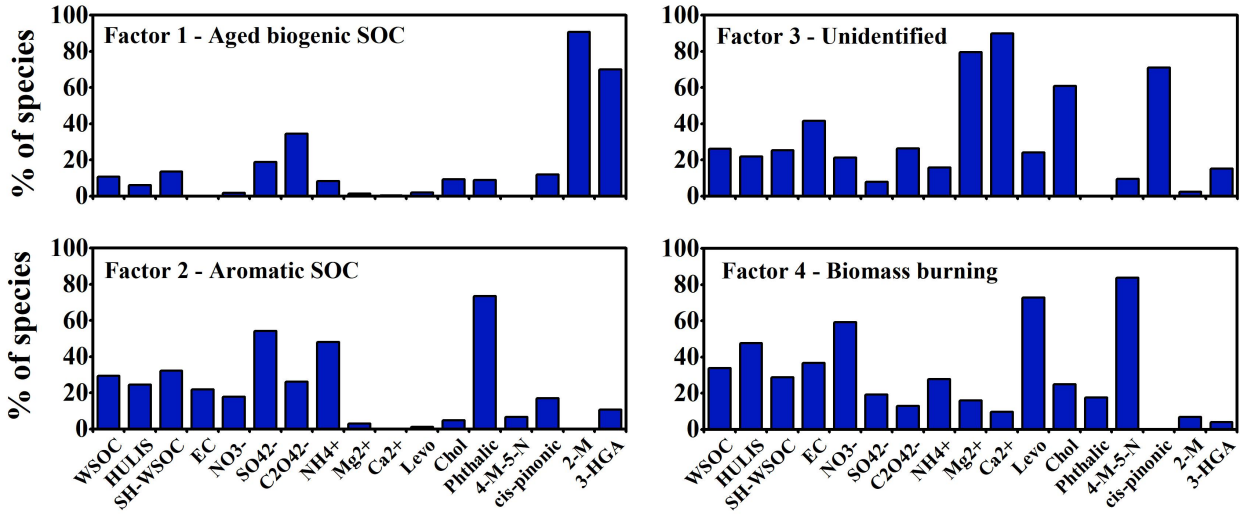

Figure S7. A 4-factor solution resolved by the PMF model.

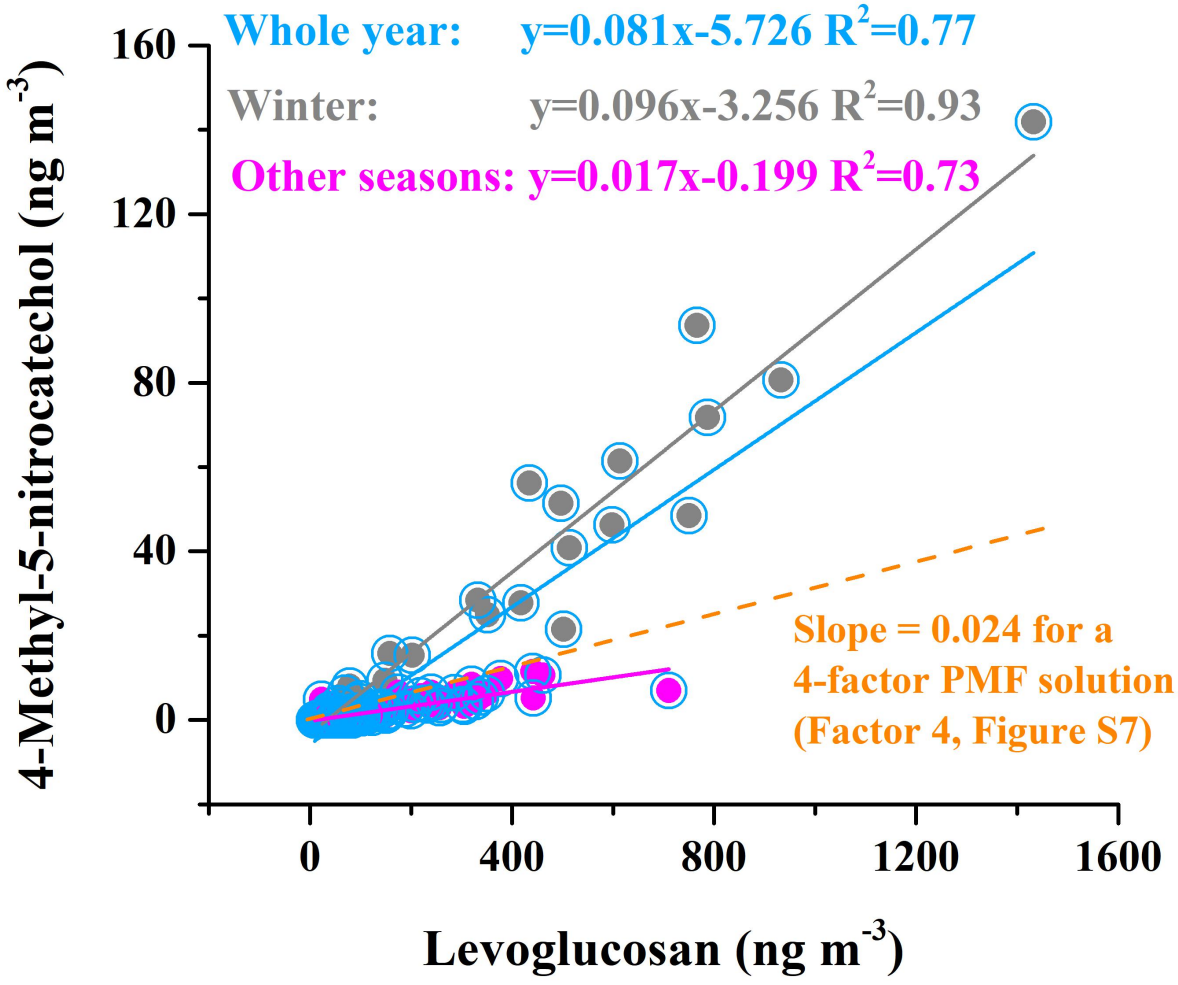

Figure S8 The relationship between levoglucosan and 4-methyl-5-nitrocatechol.

Reference

Brown, S. G., Eberly, S., Paatero, P., and Norris, G. A.: Methods for estimating uncertainty in PMF solutions: Examples with ambient air and water quality data and guidance on reporting PMF results, Sci. Total Environ., 518-519, 626-635, doi:10.1016/j.scitotenv.2015.01.022, 2015.

Cheng, Y., He, K., Duan, F., Zheng, M., Ma, Y., and Tan, J.: Positive sampling artifact of carbonaceous aerosols and its influence on the thermal-optical split of OC/EC, Atmos. Chem. Phys., 9, 7243-7256, 2009.

Cheng, Y., He, K., Duan, F., Zheng, M., Ma, Y., Tan, J., and Du, Z.: Improved measurement of carbonaceous aerosol: evaluation of the sampling artifacts and inter-comparison of the thermal-optical analysis methods, Atmos. Chem. Phys., 10, 8533-8548, doi:10.5194/acp-10-8533-2010, 2010.

Cheng, Y., He, K., Duan, F., Zheng, M., Du, Z., Ma, Y., and Tan, J.: Ambient organic carbon to elemental carbon ratios: Influences of the measurement methods and implications, Atmos. Environ., 45, 2060-2066, doi:10.1016/j.atmosenv.2011.01.064, 2011.

Cheng, Y., Duan, F., He, K., Du, Z., Zheng, M., and Ma, Y.: Sampling artifacts of organic and inorganic aerosol: Implications for the speciation measurement of particulate matter, Atmos. Environ., 55, 229-233, doi:10.1016/j.atmosenv.2012.03.032, 2012.

Cheng, Y., and He, K.: Uncertainties in observational data on organic aerosol: An annual perspective of sampling artifacts in Beijing, China, Environ. Pollut., 206, 113-121, doi:10.1016/j.envpol.2015.06.012, 2015.

Cui, W., Eatough, D. J., and Eatough, N. L.: Fine particulate organic material in the Los Angeles Basin-I: Assessment of the high-volume Brigham Young University Organic Sampling System, BIG BOSS, J. Air Waste Manage Assoc., 48, 1024-1037, doi: 10.1080/10473289.1998.10463760, 1998.

Dai, Q., Bi, X., Song, W., Li, T., Liu, B., Ding, J., Xu, J., Song, C., Yang, N., Schulze, B. C., Zhang, Y., Feng, Y., and Hopke, P. K.: Residential coal combustion as a source of primary sulfate in Xi'an, China, Atmos. Environ., 196, 66-76, doi: 10.1016/j.atmosenv.2018.10.002, 2019.

Ding, Y. M., Pang, Y. B., and Eatough, D. J.: High-volume diffusion denuder sampler for the routine monitoring of fine particulate matter: I. design and optimization of the PC-BOSS, Aerosol Sci. Tech., 36, 369–382, 2002.

Eatough, D. J., Wadsworth, A., Eatough, D. A., Crawford, J. W., Hansen, L. D., and Lewis, E. A.: A multiple-system, multichannel diffusion denuder sampler for the determination of fine particulate organic material in the atmosphere, Atmos. Environ., 27A, 1213–1219, 1993.

Eatough, D. J., Obeidi, F., Pang, Y. B., Ding, Y. M., Eatough, N. L., and Wilson, W. E.: Integrated and real-time diffusion denuder sampler for $PM_{2.5}$, Atmos. Environ., 33, 2835–2844, 1999.

Hao, Y., Gao, C., Deng, S., Yuan, M., Song, W., Lu, Z., and Qiu, Z.: Chemical characterisation of $PM_{2.5}$ emitted from motor vehicles powered by diesel, gasoline, natural gas and methanol fuel, Sci. Total Environ., 674, 128-139, doi: 10.1016/j.scitotenv.2019.03.410, 2019.

Kawamura, K., Tachibana, E., Okuzawa, E., Aggarwal, S. G., Kanaya, Y., and Wang, Z.: High abundances of water-soluble dicarboxylic acids, ketocarboxylic acids and α-dicarbonyls in the mountaintop aerosols over the North China Plain during wheat burning season, Atmos. Chem. Phys., 13, 8285-8302, doi: 10.5194/acp-13-8285-2013, 2013.

Kristensen, K., Bilde, M., Aalto, P. P., Petaja, T., and Glasius, M.: Denuder/filter sampling of organic acids and organosulfates at urban and boreal forest sites: Gas/particle distribution and possible sampling artifacts, Atmos. Environ., 130, 36-53, doi:10.1016/j.atmosenv.2015.10.046, 2016.

Mader, B. T., Flagan, R. C., and Seinfeld, J. H.: Sampling atmospheric carbonaceous aerosols using a particle trap impactor/denuder sampler, Environ. Sci. Technol., 35, 4857-4867, 2001.

Matsumoto, K., Hayano, T., and Uematsu, M: Positive artifact in the measurement of particulate carbonaceous

substances using an ambient carbon particulate monitor, Atmos. Environ., 37, 4713-4717, doi: 10.1016/j.atmosenv.2003.07.006, 2003.

McDow, S. R.: The effect of sampling procedures on organic aerosol measurement, Ph.D. Dissertation, Oregon Graduate Center, 1987.

Sengupta, D., Sengupta, V., Bhattarai, C., Watts, A. C., Moosmuller, H., and Khlystov, A. Y.: Polar semivolatile organic compounds in biomass-burning emissions and their chemical transformations during aging in an oxidation flow reactor, Atmos. Chem. Phys., 20, 8227-8250, doi:10.5194/acp-20-8227-2020, 2020.

Shrivastava, M. K., Subramanian, R., Rogge, W. F., and Robinsen, A. L.: Sources of organic aerosol: Positive matrix factorization of molecular marker data and comparison of results from different source apportionment models, Atmos. Environ., 41, 9353-9369, doi:10.1016/j.atmosenv.2007.09.016, 2007.

Sowlat, M. H., Hasheminassab, S., and Sioutas, C.: Source apportionment of ambient particle number concentrations in central Los Angeles using positive matrix factorization (PMF), Atmos. Chem. Phys., 16, 4849-4866, doi: 10.5194/acp-16-4849-2016, 2016.

Viana, M., Chi, X., Maenhaut, W., Cafmeyer, J., Querol, X., Alastuey, A., Mikuska, P., and Vecera, Z.: Influence of sampling artifacts on measured PM, OC, and EC levels in carbonaceous aerosols in an urban area, Aerosol Sci. Tech., 40, 107-117, 2006.

Zhang, Y., Schauer, J. J., Zhang, Y., Zeng, L., Wei, Y., Liu, Y., and Shao, M.: Characteristics of particulate carbon emissions from real-world Chinese coal combustion, Environ. Sci. Technol., 42, 5068-5073, doi:10.1021/es7022576, 2008.

Zhou, S., Collier. S., Jaffe, D. A., Briggs, N. L., Hee, J., Sedlacek III, A. J., Kleinman, L., Onasch, T. B., and Zhang, Q.: Regional influence of wildfires on aerosol chemistry in the western US and insights into atmospheric aging of biomass burning organic aerosol, Atmos. Chem. Phys., 17, 2477-2493, doi:10.5194/acp-17-2477-2017, 2017.