the Creative Commons Attribution 3.0 License.

## Supplement of

## Organic aerosol source apportionment by offline-AMS over a full year in Marseille

## Carlo Bozzetti et al.

Correspondence to: André Stephan Henry Prévôt (andre.prevot@psi.ch) and Nicolas Marchand (nicolas.marchand@univ-amu.fr)

The copyright of individual parts of the supplement might differ from the CC BY 3.0 License.

1 Table S1. Supporting measurements. Batch 1 denotes the set of filters collected during the 2 yearly cycle from August 2011 to July 2012. Batch 2 indicated the set of filters collected 3 during February 2011.

| Analytical Method | Measured compounds / variable | Batch of filters |
| :---: | :---: | :---: |
| Tapered element oscillating microbalance equipped with a Filter Dynamic Measurement System | $\mathrm{PM}_{2.5}$ | 1 |
| IC (Jaffrezo et al., 1998) | $\mathrm{SO}_{4}{ }^{2-}, \mathrm{NO}_{3}{ }^{-}, \mathrm{Cl}^{-}, \mathrm{NH}_{4}{ }^{+}, \mathrm{Na}^{+}, \mathrm{K}^{+}, \mathrm{Ca}^{2+}$, <br> $\mathrm{Mg}^{2+}$, oxalate, malate, malonate and succinate | 1,2 |
| Thermal Optical Transmittance using Sunset Lab Analyzer (Birch and Cary, 1996) | $\mathrm{CO}_{3}{ }^{2-}$ | 1 |
| Thermal Optical Transmittance using Sunset Lab Analyzer (EUSAAR2, Cavalli et al., 2010) | EC/OC | 1,2 |
| Water extraction Thermal Decomposition ND-IR determination using TOC analyzer (description in Bozzetti et al., 2016a) | WSOC | 1 |
| Water extraction Thermal Decomposition <br> Chemilumenscence using TOC analyzer | Total nitrogen (TN) | 1 |
| (Chemical Derivatization) GC- MS <br> (El Haddad et al., 2009; Favez et | PAH: phenanthrene, anthracene, fluoranthene, acephenanthrene, pyrene, benzo[a]anthracene, | 1,2 |


| al., 2010) | chrysene/Triphenylene, benzo[b,k]fluoranthene, benzo[j]fluoranthene, benzo-e-pyrene, benzo-a-pyrene, indeno[1,2,3-cd]pyrene, dibenzo[a,h]anthracene, benzo - ghi perylene |
| :---: | :---: |
| (Chemical Derivatization) GC- <br> MS <br> (El Haddad et al., 2009; Favez et al., 2010) | Alkanes: octadecane (C18), nonadecane (C19), eicosane (C20), heneicosane (C21), docosane (C22), tricosane (C23), tetracosane (C24), pentacosane (C25), hexacosane (C26), heptacosane (C27), octacosane (C28), nonacosane (C29), triacontane (C30), untricontane (C31), dotriacontane (C32), tritriacontane (C33), tetratriacontane (C34), pentatriacontane (C35), hexatriacontane (C36) |
| (Chemical Derivatization) GC- <br> MS | Hopanes: $17 \alpha(\mathrm{H})-21 \beta(\mathrm{H})$ - norhopane (C29), $17 \alpha(\mathrm{H})-21 \beta(\mathrm{H})$ - hopane (C30), $17 \alpha(\mathrm{H})-21 \beta(\mathrm{H})-22 \mathrm{R}$ - homohopane |
| (El Haddad et al., 2009; Favez et al., 2010) | (C31), $17 \alpha(\mathrm{H})-21 \beta(\mathrm{H})-22 \mathrm{~S}-$ homohopane <br> (C31), $17 \alpha(\mathrm{H})-21 \beta(\mathrm{H})-22 \mathrm{~S}-$ <br> bishomohopane (C32), $17 \alpha(\mathrm{H})-21 \beta(\mathrm{H})-$ <br> 22R-bishomohopane (C32), 17 $\alpha(\mathrm{H})$ - <br> $21 \beta(\mathrm{H})$-22S-trishomohopane (C33), <br> $17 \alpha(\mathrm{H})-21 \beta(\mathrm{H})$-22R-trishomohopane <br> (C33) |
| (Chemical Derivatization) GC- <br> MS <br> (El Haddad et al., 2009; Favez et al., 2010) | Cellulose and lignin pyrolysis products: levoglucosan, vanilline, coniferaldehyde, syringaldehyde, acetosyringone, vanillic Acid, abietic Acid <br> Sterols: cholesterol, stigmasterol, $\beta$ - |

(Chemical Derivatization) GCMS
(El Haddad et al., 2009; Favez et al., 2010)

Fatty acids: stearic acid, oleic acid, linoleic Acid, palmitic Acid

Phthalate esters: di-ethyl phthalate, diisobutyl phthalate, dibutyl phthalate, benzyl butyl phthalate, bis (2-ethyl hexyl)
phthalate
Others: pinonic acid

$$
\mathrm{Al}, \mathrm{As}, \mathrm{Ba}, \mathrm{Ca}, \mathrm{Cd}, \mathrm{Ce}, \mathrm{Co}, \mathrm{Cr}, \mathrm{Cs}, \mathrm{Cu},
$$ $\mathrm{Fe}, \mathrm{K}, \mathrm{La}, \mathrm{Li}, \mathrm{Mg}, \mathrm{Mn}, \mathrm{Mo}, \mathrm{Na}, \mathrm{Ni}, \mathrm{Pb}$,

4-methyl-5-
UPLC-ESI-ToF-MS (Iinuma et al., 2010).
nitrocatechol
(major)/3-methyl-
5-nitrocatechol

3-methyl-4nitrocatechol
$\mathrm{Pd}, \mathrm{Pt}, \mathrm{Rb}, \mathrm{Sb}, \mathrm{Sc}, \mathrm{Se}, \mathrm{Sn}, \mathrm{Sr}, \mathrm{Ti}, \mathrm{Tl}, \mathrm{V}$, Zn, Zr
(Chauvel et al., 2010; El Haddad et al., 2011).


2 Figure S 1 . Scatter plot of $\mathrm{CO}_{3}{ }^{2-}$ measurements (Karanasiou et al., 2011) vs. $\mathrm{CO}_{3}{ }^{2-}$ estimates from the IC ion balance. The $\mathrm{CO}_{3}{ }^{2-}$ molar concentration from ion balance was estimated as 4 the difference between the equivalents of cations $\left(\mathrm{Ca}^{2+}, \mathrm{K}^{+}, \mathrm{NH}_{4}^{+}, \mathrm{Na}^{+}, \mathrm{Mg}^{2+}\right)$ and anions 5 $\left(\mathrm{NO}_{3}{ }^{-}, \mathrm{SO}_{4}{ }^{2-}, \mathrm{Cl}^{-}\right)$.
$\mathrm{HCO}_{3}{ }^{-}$correction of offline-AMS spectra

As mentioned in the manuscript, the measured pH of the filter extract never exceeded 8 , indicating the absence of $\mathrm{CO}_{3}{ }^{2-}$ in solution, and that we can assume that water-extracted $\mathrm{CO}_{3}{ }^{2-}$ is present as $\mathrm{HCO}_{3}{ }^{-}$. Considering all the measured $\mathrm{HCO}_{3}{ }^{-}$as deriving from $\mathrm{Ca}\left(\mathrm{HCO}_{3}\right)_{2}$ or from $\mathrm{NaHCO}_{3}$, none of the liquid extracts exceeded the $\mathrm{Ca}\left(\mathrm{HCO}_{3}\right)_{2}$ or $\mathrm{NaHCO}_{3}$ saturation concentrations at $20^{\circ} \mathrm{C}$. Even considering all the measured $\mathrm{CO}_{3}{ }^{2-}$ to be in the $\mathrm{CaCO}_{3}$ form (however we can exclude this assumption as the pH of the liquid extracts was always $<8$ ), only one filter extract showed a $\mathrm{CO}_{3}{ }^{2-}$ concentration exceeding the $\mathrm{CaCO}_{3}$ saturation concentration. Therefore we can assume all the estimated $\mathrm{CO}_{3}{ }^{2-}$ (from IC ion balance) to be solubilized and in the $\mathrm{HCO}_{3}{ }^{-}$form. This would be our best estimate of the $\mathrm{HCO}_{3}{ }^{-}$watersoluble concentration. In the following we also assess the sensitivity of the source apportionment results on the $\mathrm{HCO}_{3}{ }^{-}$correction of the PMF input matrices (described hereafter) by performing a source apportionment without $\mathrm{HCO}_{3}{ }^{-}$corrections.

The $\mathrm{HCO}_{3}^{-}$correction was implemented by estimating the $\mathrm{HCO}_{3}^{-}$relative ionization efficiency (RIE) with respect to $\mathrm{NO}_{3}^{-}$. We measured nebulized and size-selected $\mathrm{NaHCO}_{3}$ particles ( 400 nm mobility diameter, using a differential mobility analyzer, DMA) in the AMS. From the particle-ToF signal (pToF, Fig. S2) of the AMS we determined a $\mathrm{NaHCO}_{3}$ Jayne shape factor $S$ of $0.9 \pm 0.1$ (Jayne et al. 2000, variability from multiple $\mathrm{NaHCO}_{3}$ injections), defined as:

$$
\begin{equation*}
S=\frac{d_{v a}}{d_{m}} \cdot \frac{\rho_{0}}{\rho_{m}} \tag{S1}
\end{equation*}
$$

Here $d_{v a}$ denotes the aerodynamic diameter under vacuum measured by the AMS and $d_{m}$ the DMA mobility diameter (DeCarlo et al., 2004). $\rho_{0}$ is the standard density of $1 \mathrm{~g} \mathrm{~cm}^{-3}$ and $\rho_{m}$ represents the $\mathrm{NaHCO}_{3}$ density $\left(2.2 \mathrm{~g} \mathrm{~cm}^{-3}\right)$. Perfectly spherical particles are characterized by an $S$ value of 1 , in our case we observed $S$ values not significantly different from 1 (within our uncertainties) for standard $\mathrm{NaHCO}_{3}$ injections, indicating that there is no reason to consider non-spherical particles. Note that previous DMA and pToF calibrations were conducted using polystyrene (PSL) spherical particles with known diameters.


Figure $\mathrm{S} 2 . \mathrm{NaHCO}_{3} \mathrm{pToF}$ signal.
Data from $\mathrm{NaHCO}_{3}$ nebulization were collected using the single particle (brute force single particle (BFSP), Drewnik et al., 2004) AMS operating mode, tracing the $\mathrm{HCO}_{3}{ }^{-}$signal at $\mathrm{m} / \mathrm{z}$ 44. By using the $\mathrm{NaHCO}_{3}$ particle density, and the newly determined Jayne shape factor $(0.9 \pm 0.1)$ we calculated the number of $\mathrm{HCO}_{3}{ }^{-}$ions per particle. The $\mathrm{HCO}_{3}{ }^{-}$ionization efficiency $\left(\mathrm{IE}_{\mathrm{HCO}_{3}^{-}}\right)$was calculated by dividing the number of $\mathrm{CO}_{2}{ }^{+}$ions detected per particle by the number of $\mathrm{NaHCO}_{3}$ molecules per particle $\left(I P_{\mathrm{HCO}_{3}^{-}}\right)$) and by the $\mathrm{m} / \mathrm{z} 44$ fractional contribution $(f 44)$ to the $\mathrm{HCO}_{3}{ }^{-}$spectrum $(f 44=0.44)$ in order to account for the contribution of other fragments to the $\mathrm{HCO}_{3}{ }^{-}$spectrum. The $\mathrm{HCO}_{3}{ }^{-}$relative IE ( $\mathrm{RIE}_{\mathrm{HCO}_{3}}$ ) in comparison to $\mathrm{NO}_{3}{ }^{-}$was determined as follows:

$$
\begin{equation*}
R I E_{H C O_{3}^{-}}=\frac{I E_{\frac{m}{z}} 44, \mathrm{HCO}}{3}-\frac{\text { ionization cross section }\left(\mathrm{NO}_{3}^{-}\right)}{I E_{\mathrm{NO}_{3}^{-}}} \text {ionization cross section }\left(\mathrm{HCO}_{3}^{-}\right) \tag{S2}
\end{equation*}
$$

Here the molecular weights of $\mathrm{NO}_{3}{ }^{-}$and $\mathrm{HCO}_{3}{ }^{-}$were used as proxies for the corresponding ionization cross sections. $R I E_{\mathrm{HCO}_{3}^{-}}$was determined to be $1.4 \pm 0.2$ ( 0.2 is the variability from multiple $\mathrm{HCO}_{3}^{-}$nebulizations ( $n=3$ ) which includes the ion per particle counting uncertainty), which is not statistically different from the standard RIE assumed for organics ( $R I E_{\text {org }}=1.4$ ).

Water-soluble mass spectra were corrected as described hereafter. Inputs for this correction are:

- measured water-soluble normalized AMS spectra $(\overrightarrow{A M S})_{\text {norm,i }}$ and corresponding OM:OC ratios $\left(\frac{O M}{O C}\right)^{\prime}$,
- $\mathrm{WSOC}_{i}$ measurements (TOC analyzer),
- $\mathrm{HCO}_{3}{ }_{i}^{-}$estimates from IC ion balance,
- $\mathrm{HCO}_{3}{ }^{-}$normalized AMS spectrum $\overrightarrow{\mathrm{HCO}_{3}^{-}}$norm as measured from $\mathrm{NaHCO}_{3}$ solution nebulization, and corresponding $\left(\frac{\mathrm{HCO}_{3}^{-}}{\mathrm{C}_{\mathrm{HCO}}^{3}}\right)=4.01$ ratio determined from the $\mathrm{HCO}_{3}^{-}$ AMS spectrum.

No correction for gaseous $\mathrm{CO}_{2}$ was applied to the $\overrightarrow{\mathrm{HCO}_{3}}$ norm spectrum as the $\mathrm{CO}_{2}$ and $\mathrm{HCO}_{3}{ }^{-}$ fragmentation is supposed to be the same due to the $\mathrm{HCO}_{3}^{-}$thermal decomposition into $\mathrm{CO}_{2}$ and $\mathrm{H}_{2} \mathrm{O}$ onto the vaporizer. Moreover, the fragments deriving from the water fragmentation $\left(\mathrm{O}^{+}, \mathrm{OH}^{+}\right.$, and $\left.\mathrm{H}_{2} \mathrm{O}^{+}\right)$do not introduce differences into the $\mathrm{CO}_{2}$ and $\mathrm{HCO}_{3}{ }^{-}$spectra because their intensities were estimated from the $\mathrm{CO}_{2}{ }^{+}$fragment according to the standard AMS fragmentation table (Aiken et al., 2008).

For a generic filter sample $i$, the measured (OM/OC) ${ }_{i}{ }^{\prime}$ ratio represents a linear combination of (OM/OC) $)_{i}$ contributions deriving from $\mathrm{HCO}_{3}{ }_{i}$ and from $\mathrm{WSOM}_{i}$. Considering the statistically not different RIEs of organics and $\mathrm{HCO}_{3}{ }^{-}$, we can assume the organics and $\mathrm{HCO}_{3}{ }^{-}$AMS response to be not different. In the same way, considering internally mixed particles from filter extracts nebulization, we assumed equal CE for both WSOM and $\mathrm{HCO}_{3}{ }^{-}$.

$$
\begin{equation*}
\left(\frac{O M}{O C}\right)_{i}^{\prime}=\left(\frac{W S O M+H C O_{3}^{-}}{W S O C+C_{H C O_{3}^{-}}}\right)_{i} \tag{S3}
\end{equation*}
$$

where $C_{\mathrm{HCO}_{3}^{-}}$represents the C concentration deriving from $\mathrm{HCO}_{3}{ }^{-}$as measured by the AMS, calculated from the $\mathrm{HCO}_{3}^{-}{ }_{i}^{-}$absolute concentrations (from IC ion balance) divided by the $\left(\frac{\mathrm{HCO}_{3}^{-}}{\mathrm{C}_{\mathrm{HCO}}^{3}}{ }^{-}\right)$ratio determined from the AMS $\mathrm{HCO}_{3}^{-}$spectrum. Similarly to $\left(\frac{O M}{O C}\right)_{i}^{\prime}$, also the AMS mass spectral fingerprint can be considered as the sum of $\mathrm{WSOM}_{i}$ and $\mathrm{HCO}_{3}{ }^{-}$, therefore the normalized blank-subtracted AMS spectra $(\overrightarrow{A M S})_{\text {norm }, i}$ and corresponding errors $\sigma(\overrightarrow{A M S})_{n o r m, i}$ were rescaled to the sum of $\mathrm{WSOM}_{i}$ and $\mathrm{HCO}_{3 i}^{-}$calculated as the sum of $\mathrm{WSOC}_{i}$ (from TOC analyzer) and $C_{\mathrm{HCO}_{3}^{-,},}$(ion balance) multiplied by $\left(\frac{O M}{O C}\right)^{\prime}{ }_{i}$

$$
\begin{equation*}
(\overrightarrow{A M S})_{i}=\left(\mathrm{WSOC}_{i}+C_{H C O_{3}^{-}, i}\right) \cdot\left(\frac{O M}{O C}\right)_{i}^{\prime} \cdot(\overrightarrow{A M S})_{n o r m, i} \tag{S4}
\end{equation*}
$$

By dividing numerator and denominator of $\frac{W S O M}{W S O C+C_{\mathrm{HCO}_{3}^{-}}}$by WSOC and by dividing numerator and denominator of $\frac{\mathrm{HCO}_{3}^{-}}{W \mathrm{SOC}+\mathrm{CHO}_{3}^{-}}$by $\mathrm{C}_{\mathrm{HCO}_{3}^{-}}$, we can express Eq. S 3 as:

$$
\begin{equation*}
\left(\frac{O M}{O C}\right)_{i}^{\prime}=\frac{\left(\frac{O M}{O C}\right)_{W S O M, i}}{1+\frac{C_{H C O}^{-}}{W S O C}}+\frac{\left(\frac{\mathrm{HCO}_{3}^{-}}{C_{H C O}^{-}}\right)}{1+\frac{W S O C}{C_{H C O}^{-}}} \tag{S5}
\end{equation*}
$$

From Eq. S5 we can derive $\left(\frac{O M}{O C}\right)_{W S O M, i}$. The time dependent WSOM concentration was therefore calculated as $\left(\frac{O M}{O C}\right)_{W S O M} \cdot \mathrm{WSOC}$.

For a generic filter sample $i$, the $\mathrm{HCO}_{3}{ }^{-} \mathrm{AMS}$ signature $\left(\overrightarrow{\mathrm{HCO}_{3}^{-}}\right)_{i}$ can be determined as:

$$
\begin{equation*}
\left(\overrightarrow{\mathrm{HCO}_{3}^{-}}\right)_{i}=C_{\mathrm{HCO}_{3}^{-}} \cdot\left(\frac{\mathrm{HCO}_{3}^{-}}{\mathrm{C}_{\mathrm{HCO}_{3}^{-}}}\right) \cdot \overrightarrow{\mathrm{HCO}_{3}} \tag{S6}
\end{equation*}
$$

where $\overrightarrow{\mathrm{HCO}_{3}^{-}}$norm represents the normalized $\mathrm{HCO}_{3}{ }^{-}$AMS spectrum derived from standard injection. To derive the AMS signal purely generated by WSOM $(\overrightarrow{W S O M})_{i}$, we subtracted $\left(\overrightarrow{\mathrm{HCO}_{3}^{-}}\right)_{i}$ (calculated as in Eq. S6) from $(\overrightarrow{A M S})_{i}$ (calculated as in Eq. S4).

$$
\begin{equation*}
(\overrightarrow{W S O M})_{i}=(\overrightarrow{A M S})_{i}-\left(\overrightarrow{\mathrm{HCO}_{3}^{-}}\right)_{i} \tag{S7}
\end{equation*}
$$

The $\mathrm{CO}_{3}{ }^{-}$concentration uncertainty $\left(\sigma_{\mathrm{CO}_{3}^{-}}\right)$was estimated by propagating the error for all the ions (measured by IC) used to estimate the $\mathrm{CO}_{3}{ }^{-}$concentration from the ion balance $\left(\mathrm{Ca}^{2+}\right.$, $\mathrm{Mg}^{2+}, \mathrm{K}^{+}, \mathrm{Na}^{+}, \mathrm{NH}_{4}^{+}, \mathrm{Cl}^{-}, \mathrm{NO}_{3}{ }^{-}, \mathrm{SO}_{4}{ }^{2-}$ ). For a generic ion $s$, the errors $\left(\sigma_{i, s}\right)$ were estimated by propagating the detection limits $(\mathrm{DL})_{s}$ and the relative repeatability $(R R)_{s}$ multiplied by the ion concentration according to Eq. (S8) (Rocke and Lorenzato, 1995):

$$
\begin{equation*}
\sigma_{i, s}=\sqrt{D L_{s}^{2}+\left(x_{i} \cdot R R\right)_{s}^{2}} \tag{S8}
\end{equation*}
$$

We assumed for each ion an uncertainty deriving from IC detection limit. On average this $\mathrm{CO}_{3}{ }^{2-}$ uncertainty was $28 \%$. The uncertainty associated to $\left(\overrightarrow{\mathrm{HCO}_{3}}\right)_{i}$ was instead estimated as:

$$
\begin{equation*}
\sigma\left(\overrightarrow{\mathrm{HCO}_{3}^{-}}\right)_{i}=\sigma_{\mathrm{CO}_{3}^{-}} \cdot \frac{A W_{C}}{F W_{C O_{3}^{2-}}} \cdot\left(\frac{\mathrm{HCO}_{3}^{-}}{\mathrm{C}_{\mathrm{HCO}_{3}^{-}}}\right) \cdot \overrightarrow{\mathrm{HCO}_{3}^{-}} \text {norm } \tag{S9}
\end{equation*}
$$

where $A W_{C}$ is the carbon atomic weight and $\mathrm{FW}_{\mathrm{HCO}_{3}^{-}}$represents the $\mathrm{CO}_{3}{ }^{2-}$ molecular weight. The final $\mathrm{WSOM}_{i}$ mass spectral uncertainty $\sigma(\overrightarrow{W S O M})_{i}$ error was estimated by summing
under quadrature $\left(\sigma\left(\overrightarrow{\mathrm{COO}_{3}}\right)_{i}\right)$ and the error associated to the total AMS signal rescaled for the sum of WSOM and $\mathrm{HCO}_{3}{ }^{-}$.
$\sigma(\overrightarrow{W S O M})_{i}=\sqrt{\left(\sigma\left(\overrightarrow{\mathrm{HCO}_{3}^{-}}\right)_{i}\right)^{2}+\left(\sigma(\overrightarrow{\mathrm{AMS}})_{i}\right)^{2}}$

## Influence of the $\mathrm{HCO}_{3}{ }^{-}$correction on the offline-AMS source apportionment results

As we have mentioned in the main text, Section 2.4, the offline AMS measurements of the WSOC are influenced by the presence of inorganic carbonates. This influence has been corrected using the carbonate mass estimated from the ion balance obtained by IC measurements. If this correction is not applied, PMF separates an additional factor with a highly oxidized fingerprint similar to inorganic carbonate and whose time series strongly correlates with that of $\mathrm{Ca}^{2+}$. In the following we compare the carbonate mass estimated from IC measurements and from PMF, and assess the influence of the correction applied on the estimation of the different factors.
We performed a source apportionment on the non $\mathrm{HCO}_{3}{ }^{-}$corrected input matrices. Input matrices were scaled to $\mathrm{WSOM}_{i}$ concentration calculated as $\mathrm{WSOC}_{i}$ multiplied by $\left(\frac{O M}{O C}\right)^{\prime}$. We explored a 6 -factor solution where the additional separated factor was attributed to inorganic dust. For a generic water soluble K factor (WSKOA) and a generic time element $i$, the corresponding water soluble OC concentrations (WSKOC) $)_{i}$ were multiplied by a factor (1/(1$f_{\text {inorganic }}$ dust) $)_{i}$, where $f_{\text {inorganic }}$ dust represents the relative contribution of the water soluble inorganic dust factor as obtained from PMF. The application of the $\left(1 /\left(1-f_{\text {inorganic }} \text { dust }\right)\right)_{i}$ factor enables rescaling the sum of the water soluble OC concentrations of the five organic factors (WSHOC, WSBBOC, WSINDOC, WSCOC, WSOOC) to the measured WSOC concentration, under the assumption of no organic contributions to the inorganic dust factor. The accuracy of this assumption is discussed in the following.
In total 240 PMF runs were performed. PMF solutions were retained according to the acceptance criteria 1-6 listed in Section 2.4. We found that primary sources (WSHOC, WSCOC, WSBBOC, and WSINDOC) showed statistically not different contributions with the offline-AMS source apportionment conducted on $\mathrm{HCO}_{3}{ }^{-}$-corrected spectra. However, for some of the retained PMF solutions, the inorganic dust-related factor tended to strongly mix with WSOOC because the two factors were characterized by the highest $f \mathrm{CO}_{2}{ }^{+}$(in this case none of the factors showed a strong correlation with $\mathrm{HCO}_{3}{ }^{-}$concentrations). In order to retain
the solutions that best resolved the inorganic dust from the WSOOC factor we introduced two further acceptance criteria:

1) Significantly positive $R$ between inorganic dust and $\mathrm{HCO}_{3}{ }^{-}$.
2) Inorganic dust correlation with $\mathrm{HCO}_{3}^{-}(R)$ significantly higher than the correlation between WSOOC and $\mathrm{HCO}_{3}{ }^{-}$.

Consequently, half of the solutions retained according to criteria 1-6 were discarded.


Figure S3. Source apportionment results obtained with and without $\mathrm{HCO}_{3}{ }^{-}$correction.

## Residual analysis

WSHOC, WSBBOC, WSOOC, and WSINDOC showed not statistically different concentrations ( $99 \%$ confidence interval) with and without the $\mathrm{HCO}_{3}{ }^{-}$correction. The WSCOC factor instead revealed statistically different concentrations within a $99 \%$ confidence interval, but only for $12 \%$ of the data points. Overall the PMF estimate of the C from inorganic dust was higher than the C estimate from $\mathrm{HCO}_{3}{ }^{-}$derived from the IC ion balance, and was more uncertain. This can be explained by an imperfect separation from other factors (especially WSOOC). The difficult separation between WSOOC and inorganic dust hampered an accurate post-PMF $\mathrm{HCO}_{3}{ }^{-}$correction. Therefore we opted to show in the manuscript source apportionment results obtained performing a pre-PMF $\mathrm{HCO}_{3}{ }^{-}$correction of the OA input matrices, but we note that this correction while uncertain does not have a significant effect on factor retrieval.


Figure S4. Online-AMS PMF $\Delta Q / Q_{\text {exp }}$ analysis. In this study, a large $Q / Q_{\text {exp }}$ decrease could be observed up to 4 factors. The 4 -factor solution enabled resolving BBOA which is mixed with OOA in the 3 -factor solution. Increasing further the number of factors provided only small additional contributions to the explained variability, resulting in a splitting of HOA in the 5factor solution, and BBOA in the 6 -factor solution; however the newly separated factors could not be attributed to specific different sources. In terms of residuals, a clear removal of the structure can be observed up to 4 factors.

Comparison between constrained and unconstrained PMF solution


Figure S5. Comparison of COA and HOA diurnal cycles from constrained and unconstrained PMF solutions.

## Best cluster selection



Figure S6. Number of cluster selection: $C^{\prime}$ values (Eq. 7) as a function of the number of clusters.

From the $a$ value sensitivity analysis, 121 solutions were obtained, the diurnal time series of which were clustered using a k mean clustering approach. The clusters were then filtered based on the cosine similarity of the HOA, COA, and BBOA average cluster mass spectra with the average mass spectra reported in the AMS literature for the same factors (Crippa et al., 2013b, Mohr et al., 2012 and 2009, Bruns et al., 2015, Docherty et al., 2011, Setyan et al., 2012, He et al., 2010). Given two vectors (in our case mass spectra) $A$ and $B$ with $n$ elements each, the cosine similarity is defined as:

$$
\begin{equation*}
\text { cosine similarity }=\frac{\sum_{i=1}^{n} A_{i} \cdot B_{i}}{\sqrt{\sum_{i=1}^{n}\left(A_{i}\right)^{2}} \cdot \sqrt{\sum_{i=1}^{n}\left(B_{i}\right)^{2}}} \tag{S11}
\end{equation*}
$$

Eq. S11 returns values between -1 and 1 , with 1 meaning collinearity, and 0 orthogonality. Because of the different HR fittings performed in different works, we considered only fragments in common with our HR fit for comparison. Within the common variables, we selected a subset of fragments characterized by small variability in comparison to the average value ( $S / N>2$ ) for at least one average mass spectrum among COA, HOA, and BBOA. Here $S$ denotes the average literature value for a certain fragment, and N represents its standard deviation. This selection was performed in order to choose the most stable and therefore certain fragments characterizing the reference spectra. Following this step, 95,92 , and $87 \%$ of the HOA, COA, and BBOA mass was retained for the average reference mass spectra,
respectively. For our dataset, depending on the cluster we selected 91-93\% of the COA mass, $84-86 \%$ for HOA, and $91-93 \%$ for BBOA. We explored the deviation of the excluded fragments from the literature values by checking whether their relative contributions to the factor mass spectra were within the literature range (maximum and minimum). For this comparison we calculated the average HOA, COA, BBOA, and OOA spectra for each of the 5 clusters. This average includes all the PMF factor profiles attributed to a specific cluster for the 100 random initiations of the $k$-mean algorithm. Overall, depending on the cluster, only 1$2 \%$ of the total HOA mass wasn't included within the literature range, $2-5 \%$ for COA, while for BBOA the value was smaller than $0.5 \%$. These diagnostics highlight the relevance in terms of mass of the fragments retained for the cosine similarity comparison.

For each source (BBOA, COA, HOA), individual literature spectra were compared with the corresponding average literature spectrum in order to estimate the minimum cosine similarity value characterizing the average profiles. This minimum value is used as a threshold above which a spectrum can be considered not different from literature profiles within $1 \sigma$. The obtained cosine similarities were $0.965 \pm 0.008$ for $\mathrm{HOA}, 0.96 \pm 0.05$ for COA, and $0.94 \pm 0.06$ for BBOA. Therefore the minimum cosine similarities to define a mass spectrum as not statistically different from the average reference spectra were 0.957 for HOA, 0.91 for COA, and 0.88 for BBOA.

We note that that reference HOA and COA spectra are less variable $(0.965 \pm 0.008$ and $0.96 \pm 0.05$ respectively) than $\operatorname{BBOA}(0.94 \pm 0.06)$. This is probably due to the different fuels and burning conditions characterizing the different BBOA ambient and chamber profiles.

Subsequently we checked whether the average HOA, COA, and BBOA reference profiles showed statistically different spectra with each other. This was tested by calculating the cosine similarity between the average HOA, COA, and BBOA literature profiles and all the aforementioned profiles reported in literature (Crippa et al., 2013b, Mohr et al., 2012 and 2009, Bruns et al., 2015, Docherty et al., 2011, Setyan et al., 2012, He et al., 2010) for factors of a different type (e.g. average HOA vs. all COA). Our results indicate that HOA, BBOA, and COA average literature mass spectra show statistically different profiles within $1 \sigma-$ average cosine similarities: HOA - BBOA: $0.5 \pm 0.1$; HOA - COA: $0.83 \pm 0.05$, COA - BBOA: $0.6 \pm 0.1$; COA - HOA: $0.83 \pm 0.08$, BBOA - COA: $0.6 \pm 0.1$; BBOA - HOA: $0.5 \pm 0.2$. This means that PMF factors can be identified based on the analysis approach we have adopted, as factor mass spectra are characteristic: i.e. differences between mass spectra pertaining to the
same factor are significantly smaller then differences between mass spectra related to different factors.

In order to select the best clusters we determined the cosine similarity of the average cluster mass spectra with the average reference profiles. A cluster was retained if the HOA, COA, and BBOA average cluster spectra were not statistically different from the corresponding average reference profiles. However the average cluster mass spectra are also characterized by an uncertainty deriving from the $k$-mean algorithm random initiation. To calculate this uncertainty, we generated 100 random cluster profiles by randomly varying the average cluster mass spectra within the corresponding standard deviation (calculated as the standard deviation of the cluster profiles obtained initiating the $k$-means algorithm 100 times) assuming a normal distribution of the error. Each randomly generated profile was compared with the average reference spectrum by calculating the corresponding cosine similarity. This provides the cosine similarity uncertainty of an average cluster spectrum with the literature average reference profile. From the comparison of average cluster mass spectra with reference spectra, we observed that HOA and BBOA showed statistically not different fingerprints with the corresponding average reference profiles (within $1 \sigma$ ) for all clusters, while COA showed a statistically different mass spectrum with respect to the average profile for cluster 5, which therefore was not retained for further analysis. For cluster 4, although COA average spectrum was statistically not different from the average COA reference profile within $1 \sigma$, the mass spectrum was not statistically different either from the HOA average mass spectrum within $1 \sigma$, suggesting a certain mixing of the two sources, therefore also cluster 4 was rejected (Table S2).

1 Table S2. Cosine similarity between COA, HOA, and BBOA average cluster spectra with the

|  | threshold | 0.957 | 0.91 | 0.898 |
| :---: | :---: | :---: | :---: | :---: |
|  |  | HOA <br> reference <br> spectrum | COA <br> reference <br> spectrum | BBOA <br> reference <br> spectrum |
| HOA | cluster1 | $0.99188(3)$ | $0.6931(1)$ | $0.4677(2)$ |
|  | cluster2 | $0.99532(2)$ | $0.69320(8)$ | $0.5237(1)$ |
|  | cluster3 | $0.99177(6)$ | $0.6752(2)$ | $0.5254(3)$ |
|  | cluster4 | $0.99025(9)$ | $0.6753(4)$ | $0.5448(5)$ |
|  | cluster5 | $0.99674(4)$ | $0.7079(2)$ | $0.5195(4)$ |
|  | cluster1 | $0.8373(1)$ | $0.99208(5)$ | $0.6453(4)$ |
|  | cluster2 | $0.85270(9)$ | $0.98997(2)$ | $0.62187(8)$ |
|  | cluster3 | $0.8861(2)$ | $0.98598(4)$ | $0.6045(2)$ |
|  | cluster4 | $0.9565(8)$ | $0.9648(3)$ | $0.6219(8)$ |
|  | cluster5 | $0.80(1)$ | $0.89(1)$ | $0.855(2)$ |
|  | cluster1 | $0.594(1)$ | $0.588(1)$ | $0.9687(5)$ |
|  | cluster2 | $0.7134(3)$ | $0.6874(1)$ | $0.94746(7)$ |
| BBOA | cluster3 | $0.6945(2)$ | $0.6739(2)$ | $0.95441(6)$ |
|  | cluster4 | $0.578(1)$ | $0.598(1)$ | $0.96752(5)$ |
|  | cluster5 | $0.713(2)$ | $0.682(2)$ | $0.943(1)$ | corresponding reference profiles from literature (average of the profiles reported in: Crippa et al., 2013b, Mohr et al., 2012 and 2009, Bruns et al., 2015, Docherty et al., 2011, Setyan et al., 2012, He et al., 2010). Threshold cosine similarity indicates the minimum cosine similarity value which defines a cluster profile as not statistically different from the reference profiles. Highlighted values indicate cluster profiles not statistically different from the reference profiles.



Figure S7. Average COA cluster spectra.
These results are also reflected by the high average COA diurnal pattern correlation with $\mathrm{NO}_{\mathrm{x}}$ (typical traffic tracer) for cluster $4(R=0.64)$, while for cluster 5 the average COA diurnal pattern correlates well with AMS-PAH (highly correlated with BBOA) and fragment $f \mathrm{C}_{2} \mathrm{H}_{4} \mathrm{O}_{2}{ }^{+}$diurnals ( $R=0.94$ and 0.98 respectively), suggesting a certain mixing with BBOA (Fig. S8). For clusters 1-3, the COA diurnal correlation with $\mathrm{NO}_{x}$ and AMS-PAH was smaller than for clusters 4-5, and smaller than the correlation of HOA with $\mathrm{NO}_{x}$ and BBOA with AMS-PAH indicating a good COA separation from HOA and $\mathrm{NO}_{\mathrm{x}}$ (Table S2).

Table S3. Correlation of COA cluster diurnals with $\mathrm{NO}_{\mathrm{x}}$, AMS-PAH, $m / z 60$, and suboptimal clusters. High correlations with $\mathrm{NO}_{x}$ suggest possible mixing between COA and HOA ; high correlations with AMS-PAH and $m / z 60$ are suggestive of possible mixings between BBOA and COA; high correlations with clusters 4 and 5 are probably indicative of uncertain attribution of the PMF runs between the optimal and suboptimal clusters.

COA diurnals

| Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | Pearson correlation <br> coefficient |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 0.84 | 0.52 | 0.48 | 0.72 | 0.94 | $\mathrm{AMS}-\mathrm{PAH}$ |
| 0.19 | 0.34 | 0.57 | 0.64 | 0.09 | $\mathrm{NO}_{\mathrm{x}}$ |
| 0.82 | 0.42 | 0.30 | 0.54 | 0.98 | $\mathrm{~m} / \mathrm{z} 60$ |
| 0.77 | 0.83 | 0.93 | 1 | 0.61 | cluster 4 COA diurnal |
| 0.91 | 0.56 | 0.42 | 0.61 | 1 | cluster 5 COA diurnal |

Because of the small variability and the relatively high correlation coefficients between HOA and $\mathrm{NO}_{x}$, and between BBOA and AMS-PAHs among solutions belonging to the retained
clusters ( $R$ ranging between 0.76-0.79 and 0.92-0.93 respectively, Fig. S10), we did not select additional acceptance criteria based on HOA and BBOA temporal trends.


| 120 | $-25 \times 10^{-3}$ |
| :---: | :---: |
| - 100 | -20 |
| - 80 入 ${ }_{\text {又 }}$ | -20 |
| $60 \stackrel{\text { E }}{ }$ | -15 |
| $40{ }^{\text {E }}$ | -10 |
|  | $3_{i}$ |
| 20 |  |
| 0 | $\square_{0}$ |



| $-m / z 60$ |
| :---: |
| $\mathrm{NO}_{\mathrm{x}}$ |
| AMS-PAH |
| Selected clusters: |
| Cluster 1 |
| Cluster 2 |
| Cluster 3 |
| Rejected clusters |
| Cluster 4 |
| Cluster 5 |

Figure S8. Top: diurnal cycles of $m / z 60$ (AMS tracer for BBOA), AMS-PAHs, and $\mathrm{NO}_{\mathrm{x}}$. Bottom: Average COA diurnal cycles for the different clusters.


Figure S9. a value sensitivity analysis: cluster analysis. a) Attribution of the PMF solutions to the clusters. b) Optimal PMF solutions. We initiated the cluster analysis 100 times; the selected solutions were those attributed to the optimal clusters (\#1-\#3) more than $95 \%$ of the cluster analyses. We note that although cluster 1 and 3 were chosen among the optimal clusters (clusters \#1-\#3), many of the PMF runs belonging to these clusters were attributed to suboptimal clusters (\#4-\#5) more than $5 \%$ of the time. Not surprisingly clusters 1 and 3 show better COA diurnal cycle correlations with the suboptimal clusters than cluster 2 (Table S3).


Figure S10. a value sensitivity analysis diagnostics: correlations of factor profiles and corresponding tracers as a function of COA and HOA $a$ values. The area surrounded by black lines corresponds to the retained PMF solutions.


2 Figure S11. Offline-AMS PMF $\Delta\left(Q / Q_{\text {exp }}\right)$ analysis. In this study, a large $Q / Q_{\text {exp }}$ decrease not be attributed to specific aerosol sources/processes.


7 Figure S12. Probability density functions of the OC residuals from $R_{Z}$ sensitivity analysis (Eq. 8 could be observed for solutions with up to 5 factors. Increasing the number of factors $Q / Q_{\text {exp }}$ leads to smaller increase in the explained variability. The newly separated OOA factors could 11).


2 Figure S13. Factor recoveries: probability density functions. Vertical sticks represent the recoveries determined by Daellenbach et al. (2016). Estimated recoveries: $R_{H O A, \text { med }}=0.11$ (first quartile 0.10 , third quartile 0.12 ); $R_{B B O A, \text { med }}=0.65$ (first quartile 0.63 , third quartile 0.69 ); $R_{\text {COA, med }}=0.53$ (first quartile 0.48 , third quartile 0.59 ); $R_{\text {OOA, med }}=0.89$ (first quartile 0.87 , third quartile 0.91 ).


Figure S14. Offline-AMS: INDOA vs. residuals concentrations (calculated according to Eq. 8) scatter plot. Residuals $>6 \mu \mathrm{~g} \mathrm{~m}{ }^{-3}$ represented $<2 \%$ of the points and were associated to 25 November, where the C bulk extraction efficiency (Bulk EE = WSOC:OC) was estimated at $22 \%$, representing an outlier in comparison to the median Bulk $\mathrm{EE}=0.61$ (first quartile $=$ 0.54 ; third quartile $=0.71$ ).


Figure S15. Online-AMS: AMS-PAHs rose plot. Color code denotes the hour of the day; marker size is proportional to the AMS-PAHs concentration. The distance from the center is proportional to the wind speed. El Haddad et al. (2013) revealed industrial-related emissions associated to wind direction from W-SW $\left(225^{\circ}-270^{\circ}\right)$. In this work (wintertime monitoring period) no significant increase of AMS-PAHs with wind direction from W-SW is observed.


Figure $\mathrm{S} 16 . \mathrm{C}_{2} \mathrm{H}_{4} \mathrm{O}_{2}{ }^{+}$vs. AMS-PAHs scatter plot. The good correlation indicates that AMS-
PAHs are mainly emitted by BBOA.
4


2 Figure S17. Online-AMS: BBOA rose plot. Color code denotes the hour of the day, marker from E-NE during nighttime.


7 Figure S18. Organic-N and OC time series.


Figure $\mathrm{S} 19 . \mathrm{C}_{2} \mathrm{H}_{3} \mathrm{O}^{+}{ }_{\text {OOA }}$ and $\mathrm{CO}_{2}{ }^{+}$OOA yearly cycle. The absence of a clear seasonality hampers the separation of two distinct OOA factors.

## Comparison of online-AMS and offline-AMS mass spectra

Offline-AMS source apportionment was conducted on water soluble OA AMS spectra. In this section we compare the online- and offline-AMS factor profiles obtained from PMF analysis. In general the measured water soluble AMS spectra show a higher $\mathrm{O}: \mathrm{C}$ ratio (winter median $=$ 0.51 ) than the online-AMS spectra (median $=0.44$ ). This is also indirectly confirmed by the high factor recoveries for the factors characterized by high $\mathrm{O}: \mathrm{C}$ ratios (e.g. OOA and BBOA), while lower recoveries were estimated for the factors associated with low O:C ratios (e.g. COA and HOA, Fig. S14), indicating that offline-AMS accesses the most oxidized (i.e. watersoluble) OA fraction, consistent with Daellenbach et al. (2016). Table S4 reports the cosine similarities between the offline-AMS PMF factor profiles and the average reference onlineAMS profiles (Crippa et al., 2013b; Mohr et al., 2012; 2009, Bruns et al., 2015; Docherty et al., 2011; Setyan et al., 2012; He et al., 2010).

Table S4. Cosine similarities between offline-AMS PMF factors and average online-AMS reference profiles (Crippa et al., 2013b; Mohr et al., 2012; 2009; Bruns et al., 2015; Docherty et al., 2011; Setyan et al., 2012; He et al., 2010). The threshold cosine similarity indicates the minimum cosine similarity value which defines a mass spectrum as not statistically different from the average online-AMS reference profiles. Values highlighted in green indicate offlineAMS factor profiles not statistically different from online-AMS reference spectra. Cosine similarities were calculated for all the retained offline-AMS PMF spectra. The values reported in the table represent the average cosine similarity; the corresponding uncertainty is reported in parentheses.

| Threshold <br> cosine <br> similarity |  |  | Offline-AMS factor profiles |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | HOA | COA | BBOA |
| 0.957 | Online-AMS | HOA | $0.87(1)$ | $0.87(1)$ | $0.32(1)$ |
| 0.91 | average | COA | $0.931(3)$ | $0.933(9)$ | $0.46(2)$ |
| 0.89 | reference | spectra | BBOA | $0.479(7)$ | $0.48(1)$ |

As expected from the relatively high recovery (median 0.65), the water soluble fraction of BBOA shows a mass spectral fingerprint not statistically different from the BBOA onlineAMS reference spectra, although more oxidized ( $\mathrm{O}: \mathrm{C}=0.54$ for offline-AMS, and 0.35 for online-AMS). Despite the different degree of oxidation, the two spectra were not considered as different within our uncertainty due to the large variability of the BBOA AMS spectra reported in literature. As already mentioned, this variability mostly derives from the different burning conditions and fuels. The offline- and online-AMS spectra for COA are not statistically different from each other either. By contrast HOA, which has the lowest recoveries among the separated offline-AMS factors, shows a fingerprint that is statistically different from the HOA online-AMS reference spectrum, moreover the offline-AMS HOA fingerprint is statistically not different from the COA online-AMS reference spectrum. This is due to the higher HOA $a$ values associated with the accepted solutions $(0.5,0.9,1)$ in comparison to the COA $a$ values ( $0.2,0.3,0.4,0.8$ ), which therefore enabled a larger variation of the HOA mass spectrum. The online-AMS HOA literature spectra are characterized by similar values of saturated and unsaturated hydrocarbon fragments, while the online-AMS COA reference spectra show higher values for saturated hydrocarbon fragments in comparison to unsaturated ones (Mohr et al., 2009). The lower water solubility of saturated
hydrocarbons in comparison to unsaturated hydrocarbons (Daellenbach et al. 2016) therefore increases the resemblance of HOA and COA water-soluble spectra.

## Comparison of AMS-PAHs and GC-MS PAHs

AMS-PAHs concentrations were found on average $19 \%$ higher than the sum of GC-MS quantified PAHs, and showed a significant correlation ( $R=0.68$ ). This relatively small concentration discrepancy can originate from different causes:
a) GC-MS quantified PAHs do not represent the total PAHs mass.
b) AMS-PAHs RIE could differ from the average organic RIE (1.4) assumed in this work.
c) PAHs might be formed on the AMS vaporizer surface from the pyrolysis of refractory organic compounds.

## Factor solubility sensitivity analysis.

Daellenbach et al. (2016) stated that the calculated factor recoveries are consistent with the water solubility of these fractions, with HOA being barely water soluble ( $\sim 13 \%$ ), BBOA moderately water soluble ( $65 \%$ ) and OOA almost entirely water soluble ( $90 \%$ ). These factor recoveries have been revaluated in this study (see "Offline-AMS source apportionment optimization" section) and are consistent with the estimates of Daellenbach et al. (2016) (Fig. S13), based on collocated ACSM and offline-AMS measurements at another site. In addition, here the recovery for industrial OA has been assessed to be similar to that of BBOA (69\%).

We also assessed the relationship between the factor recoveries and the water solubility of the compounds therein. We assumed each of the components to comprise a single average surrogate in equilibrium between the aqueous solution and an ideal solution of water insoluble organic species. The adaptation of the equations in Psichoudaki and Pandis (2013) to our case yields the following expression of the surrogate extracted fraction/recovery. We will refer to this fraction as $R_{k}^{\prime}$, to discriminate between measured and calculated recoveries.
$R_{k}^{\prime}=\frac{m_{k}+V_{W} S_{k}+m_{O A}\left(1-f_{W S O C}\right)-\left(\left(m_{k}+V_{W} S_{k}+m_{O A}\left(1-f_{W S O C}\right)\right)^{2}-4 m_{k} V_{W} S_{k}\right)^{1 / 2}}{2 m_{k}}$

Here, $m_{k}$ and $m_{O A}$ are the total mass of a factor $k$ and of the organic aerosol in the extracted sample, respectively. $V_{W}$ is the volume of water used for extraction, $f_{\text {WSOC }}$ the fraction of water soluble organics and $S_{k}$ the water solubility of the average surrogate compound representative of the bulk composition of the component $k$. This formulation should provide the highest estimate of $S_{k}$ compared for example to considering the extraction of a single component $k$ to be independent of the presence of the other organics ( $k$ forms its own phase). Using this formulation, we estimate the recoveries obtained under our conditions to be consistent with $S_{k}$ values of $10^{-3} \mathrm{~g} \mathrm{~L}^{-1}, 10^{-2} \mathrm{~g} \mathrm{~L}^{-1}$ and $10^{-1} \mathrm{~g} \mathrm{~L}^{-1}$, for HOA, BBOA/COA/INDOA, and OOA, respectively (Fig. S20). We also assessed the sensitivity of $R_{k}^{\prime}$ towards the bulk aerosol composition, by varying $f_{W S O C}$, the total organic aerosol concentration and the contribution of the factor of interest within the observed ranges. This sensitivity analysis suggests that for a similar solubility, the variability in the extraction conditions may influence the recoveries by 10 percentage points on average (see the upper and lower curves in Figure S20). These variations are relatively small, within our confidence interval of the determined recovery parameters. We note that the extraction procedure adopted here favors the compounds' partitioning into the aqueous phase, given the high extraction volume compared to the sampled air volume in the extracted filter fraction: $\sim 0.5 \mathrm{~cm}^{3} \mathrm{~m}^{-3} v s .0 .1 \mathrm{~cm}^{3} \mathrm{~m}^{-3}$ in the other studies (Psichoudaki and Pandis, 2013 and references therein). Under these conditions, all typical functionalized compounds would be extracted (Cappelli et al., 2013; Meylan and Howard, 1994a,b; Meylan et al., 1996)

We also note that the model used here is rather simplistic and the different components are expected to comprise a suite of compounds with a wide range of water solubility. This can be expressed in a solubility basis set by analogy to the volatility basis set (VBS). This simplification implies on the one hand that the solubility values provided here are only weighted average values for the solubility of different compounds contained in these components. On the other hand, the model provided here would significantly over-predict the sensitivity of the recoveries to the extraction procedure adopted (filter loading, bulk OA solubility and extraction volume). Again by analogy to the VBS, most of the compounds contained in one component may be either fully water soluble or fully insoluble under most of the extraction conditions, and only a minor fraction of semi-soluble compounds would be sensitive to the extraction procedure. Still, we note that the data we present here cannot be directly extrapolated to other studies and establishing a solubility basis set for the different
components would require significantly varying the extraction conditions of the different samples followed by an assessment of the recovery, which is beyond the scope of this study.


Figure S20. Sensitivity of the calculated factor recoveries $R^{\prime}{ }_{k}$ to the factor solubility $S_{K}$ $\left(\mathrm{g} \mathrm{L}^{-1}\right)$. Vertical lines define the factor solubility calculated from the median factor recoveries determined in this work.

References (not already included in the main text):
Cappelli, C. I., Manganelli, S., Lombardo, A., Gissi, A., and Benfenati, E.: Validation of quantitative structure-activity relationship models to predict water-solubility of organic compounds, Sci Tot. Environ., 463-464, 781-789, 2013.

Jayne, J. T., Leard, D. C., Zhang, X. F., Davidovits, P., Smith, K. A., Kolb, C. E., and Worsnop, D. R.: Development of an aerosol mass spectrometer for size and composition analysis of submicron particles, Aerosol Sci. Tech., 33, 49-70, 2000.

Meylan W. M., and Howard, P. H.: Upgrade of PCGEMS water solubility estimation method (draft). U.S. Environmental Protection Agency, Office of pollution Prevention and Toxics, Washinghton DC, 1994a.

Meylan W. M., and Howard, P. H.: Validation of water solubility estimation methods using $\mathrm{K}_{\mathrm{Ow}}$ for application in PCGEMS \& EPI. Final report. U.S. Environmental Protection Agency, Office of pollution Prevention and Toxics, Washinghton DC, 1994b.

Meylan W. M., Howard, P. H., and Boethling, R. S.: Improved method for estimating water solubility from octanol/water partition coefficient. Environ. Toxicol. Chem. 15, 100-106, doi:10.1002/etc.5620150205, 1996.

Rocke, D. M., and Lorenzato, S.: A two-component model for measurement error in analytical chemistry, Technometrics, 37, 176-184, 1995.

