## 1 Supplementary Material

### 1.1 Weighting $m / z 29$ for PMF analysis

Since it was not clear whether the signal at $\mathrm{m} / \mathrm{z} 29$ was partly an artefact due to a leakage in the vacuum chamber, this mass to charge ratio was downweighted. Given the case of a leakage, more air molecules would be available to produce a signal predominantly at $\mathrm{m} / \mathrm{z} 28$ $\left(\mathrm{N}_{2}{ }^{+}\right)$. This would result in an elevated baseline influencing the neighbor $\mathrm{m} / \mathrm{z} 29$ which in turn would be overestimated.

When not downweighting $m / z 29$ a six factor PMF solution seemed to best meet the criteria of a minimum $Q / Q_{\text {exp. }}$. However, one factor mainly consisting of $m / z 29(62.3 \%)$ and $m / z 44(9.5$ $\%$ ) was derived. This factor, based on only two mass to charge ratios and based to more than $50 \%$ on $\mathrm{m} / \mathrm{z} 29$ was not trusted, due to steps in the signal intensity occurring in the time series (see Fig. S1). To test whether this factor might be an artefact, $m / z 29$ was downweighted by a factor 100 and 1000. The factor profile disappeared when downweighting by either factor. Table S1 shows the relative differences between downweighting by a factor 100 or 1000 in the time series and mass spectra for the five factor solution at fpeak $=0$ and seed $=0$. For the final analysis $m / z 29$ was downweighted by a factor of 1000 .

Fig. S1: Time series and mass spectrum of the $\mathrm{m} / \mathrm{z} 29$ dominated factor


20 \begin{tabular}{lll}
\multicolumn{2}{l}{ Table S1: Relative difference between downweighting $\mathrm{m} / \mathrm{z} 29$ by a factor 100} <br>

\cline { 2 - 3 } \& Factor \& | relative difference $100 / 1000$ |
| :--- |
| in mass spectra (\%) | <br>


\hline MSA-OA \& 2.15 \& | relative difference 100/1000 in |
| :--- |
| time series (\%) | <br>

AA \& 1.68 \& 0.36 <br>
M-OOA \& 1.04 \& 0.71 <br>
Sea Salt \& 4.06 \& 1.30 <br>
HOA \& 3.02 \& 2.59 <br>
\hline
\end{tabular}

### 1.2 Key diagnostic plots for the PMF 5 factor solution

The change in slope of the $Q / Q_{\text {exp }}$ vs. the number of PMF solutions curve between factors 4 and 5 (see Fig. S 2 a ) indicates that at least 5 factors should be considered. The $Q$ value or "PMF quality of fit parameter" (Zhang et al., 2011) refers to the sum of weighed square residuals. Five factors were chosen, because factor splitting was identified in the solutions using 6 or more factors (see Fig. S3). For example, factor 1 (count upwards from bottom) in the 5 -factor solution is split into factors 1 and 2 in the 6 -factor solution, as the time series clearly show (cf. Figs. S3c and S3e). Also, the correlation between the two mass spectra is very high at $99 \%$ (not shown). In the 5 -factor solution, factor 1 contributes $23.5 \%$ of the mass, while factors 1 and 2 contribute 17.8 and $12.2 \%$ in the 6 factor solution. The 7 -factor solution yields factors that do not represent meaningful OA aerosol mass spectra, where e.g., $\mathrm{m} / \mathrm{z} 28$ and 44 make up $50 \%$ of the mass. Choosing only 4 factors results in factors 3 and 5 from the five factor solution being merged. Factor 5 in the 5 -factor solution was identified as hydrocarbon-like OA (HOA), based on the high correlation $\left(\mathrm{R}^{2}=0.92\right)$ with the reference mass spectrum (A-DEC-Q-015) of the AMS UMR database, based on Ng et al. (2011). When only 4 factors are chosen, the correlation to the HOA reference spectrum decreases to $\mathrm{R}^{2}=0.76$. The additional profile in the 5 -factor solution, which was identified as marine oxygenated OA (M-OOA), has a low correlation with the HOA reference spectrum $\left(R^{2}=\right.$ 0.10 ), so that the separation of the factors is meaningful. Also, the time series of factors MOOA and HOA have a correlation coefficient of $\mathrm{R}^{2}=0.12$ (see Fig. S2 d), so that a 5 -factor solution is plausible. All time series' correlation coefficients are in the range between -0.25 and $0.34\left(R^{2}\right.$ between 0.06 and 0.12 ), while the mass spectrum $R$ ranges between 0.14 and 0.59 ( $\mathrm{R}^{2}$ between 0.02 and 0.35 ). The relatively high correlations in the mass spectra between factors 1 and 2 as well as 2 and 3 (see Fig. S2d) are based on the relatively high contributions of $m / z 28,29$ and 44 to the signals. Other marker fragments, however, do not coincide.

The chosen final fpeak is 0 , where $Q / Q_{\text {exp }}$ is at a minimum (see Fig. S2c). All other values of fpeak between -1 and 0.2 show only rotational ambiguity in the 5 -factor solution, while fpeak $>0.2$ introduces a factor consisting to $80 \%$ of $\mathrm{m} / \mathrm{z} 29$, which does not represent a meaningful chemical mass spectrum. Fig. S2 f) shows that this factor starts to dominate the contributions to the total reconstructed mass of OA . The $Q / Q_{\text {exp }}$ value $<1$ at fpeak $=0$ indicates that the error matrix was somewhat overestimated which is consistent with downweighting $\mathrm{m} / \mathrm{z} 29$ by
a factor of 1000 . Seed 0 has been chosen to represent the final 5 -factor solution, as $Q / Q_{\text {exp }}$ is very stable over the whole range of tested seeds. Also, the contribution of each factor to the total organic mass when varying seeds does not vary significantly (see Fig. S2b).

Fig. S2: a) $Q / Q_{\text {exp }}$ as a function of the number of factors $(P)$ used in the PMF analysis; the yellow circle denotes the best solution presented in this work $\mathbf{b}) Q / Q_{\text {exp }}$ as a function of seeds between 0 and 50 in steps of two, c) $Q / Q_{\text {exp }}$ as a function of fpeaks between -1 and 1 in steps of $0.2, \mathbf{d}$ ) Pearson's correlation coefficient for time series and mass spectra for the 5 factor solution, e) variation of factor contributions to total OA as a function of seeds, $\mathbf{f}$ ) variation of factor contributions to total OA as a function of fpeaks.


## c)


d)



Fig. S4: a) scaled residuals for each $\mathrm{m} / \mathrm{z}$, the horizontal bars indicate the median while the boxes represent the interquartile, b) comparison time series of the reconstructed OA (sum of the five factors) and the measured $\mathrm{OA}, \mathbf{c}$ ) sum of the residuals (measured - reconstructed) of the fit, d) $Q / Q_{\text {exp }}$ for each time step, and e) $Q / Q_{\text {exp }}$ for each mass to charge ratio




 bootstrapping method (all numbers in \%).

| Factor | seeds $_{M S}$ | seeds $_{T S}$ | fpeak $_{M S}$ | fpeak $_{T S}$ | Bootstrap $_{M S}$ | Bootstrap $_{T S}$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| MSA-OA | 1.61 | 0.52 | 75.9 | 29.1 | 2.51 | 2.37 |
| AA | 4.67 | 0.80 | 28.2 | 31.9 | 3.56 | 2.59 |
| M-OOA | 0.54 | 0.69 | 37.4 | 130.2 | 7.92 | 4.26 |
| Sea Salt | 1.15 | 0.40 | 29.1 | 15.3 | 16.1 | 7.26 |
| HOA | 0.19 | 0.90 | 16.5 | 14.5 | 13.5 | 9.19 |

Table S2: Relative standard deviations for each factor profile mass spectrum and time series based on the variations of seeds and fpeaks and based on the statistical variations through the

