Supplementary information for manuscript

Identification and quantification of organic aerosol from cooking and other sources in Barcelona using aerosol mass spectrometer data

C. Mohr¹, P. F. DeCarlo¹*, M. F. Heringa¹, R. Chirico¹**, J. G. Slowik¹, R. Richter¹, C. Reche², A. Alastuey², X. Querol², R. Seco³**, J. Peñuelas³, J. L. Jiménez⁴, M. Crippa¹, R. Zimmermann⁶, U. Baltensperger¹, and A. S. H. Prévôt¹

¹Laboratory of Atmospheric Chemistry, Paul Scherrer Institut (PSI), Villigen, Switzerland
²Institute for Environmental Assessment and Water Research (IDAEA-CSIC), Barcelona, Spain
³Unitat d’Ecologia Global CREAF-CEAB-CSIC, Centre de Recerca Ecològica i Aplicacions
Forestals, Universitat Autònoma de Barcelona, Barcelona, Spain
⁴Department of Chemistry and Biochemistry, Cooperative Institute for Research in the
Environmental Sciences (CIRES), University of Colorado, Boulder, USA
⁵Helmholtz Zentrum Munchen, Joint Mass Spectrometry Center, Institute of Ecological
Chemistry, Neuherberg, Germany
⁶University of Rostock, Rostock, Germany
* now at: Department of Civil, Architectural, and Environmental Engineering, Drexel University,
Philadelphia, USA
** now at: Italian National Agency for New Technologies, Energy and Sustainable Economic
Development (ENEA), UTAPRAD-DIM, Frascati, Italy
***now at: Atmospheric Chemistry Division, National Center for Atmospheric Research,
Boulder, USA

Correspondence to: A. S. H. Prévôt (andre.prevot@psi.ch)
Figure S1: Time series of the collection efficiency (CE) used for the present dataset (left axis) and total concentration of species measured by AMS (right axis).
Figure S2: Scatterplot of combined time series of total AMS species (HR) and BC (y-axis) and Grimm PM$_1$.

The data were fitted with a least orthogonal distance fit (red line).
3 PMF

3.1 UMR solution

Figure S3: Mass spectra of the UMR 5-factor-PMF solution.
3.2 Comparison of UMR and HR PMF solution

The $R^2$ of the correlation of the mass spectra of the UMR and HR PMF solution range from 0.80 (COA) to 0.99 (LV-OOA), confirming their similarities. Bigger differences can be seen in the time series of the corresponding factors. The COA time series show discrepancies in the total mass especially in the beginning of the campaign (until 02 March 2009), visualized in the data points with a much lower slope in Fig. S5 h. For the BBOA, the UMR time series features peaks not inherent to the HR time series. Concerning the mass attribution to each factor, HR generally assigns more mass to the primary OA factors and less to the OOA factors. Here the higher
resolution and, related to that, the signal on an individual ion basis of the HR data matrix adds additional information to the HR data matrix and thus allows for a better quantification of primary and secondary OA.
Figure S5: Scatter plots of UMR and HR PMF spectra (a-e), time series (f-j) and a comparison of the mass attributed to each factor relative to OA (k-o).
3.3 UMR solution criteria

The PMF solution for a chosen number of factors $p$ is a weighted iterative least squares fit minimizing $Q$ as in Eq. (1), with $m$ and $n$ denoting the rows and columns of the input matrices, respectively. The known standard deviations $\sigma_{ij}$ of the measured input values $x_{ij}$ are used to determine the weights of the residuals $e_{ij}$.

$$Q = \sum_{i=1}^{m} \sum_{j=1}^{n} \left( \frac{e_{ij}}{\sigma_{ij}} \right)^2$$

If the model is appropriate and the data uncertainties estimations are accurate, $(e_{ij}/\sigma_{ij})^2$ is ~1 and the expected $Q$ ($Q_{expected}$) = $mn-p(m+n) \approx mn$, the degrees of freedom of the fitted data. The $Q$-value is one mathematical criterion for the quality of the PMF solution: $Q/Q_{expected} >> 1$ indicates an underestimation, $Q/Q_{expected} << 1$ an overestimation of errors in the input data (Paatero et al., 2002). The mathematically correct value of $p$ in PMF would be where the line changes the slope in the plot of a series of $p$ values versus their respective minimized $Q$ (Fig. S6 a). However, a PMF solution has to be feasible in an ambient context and thus does not necessarily correspond to the mathematically correct value of $p$.

Rotational ambiguity is a significant problem in the use of factor analysis (Paatero et al., 2002). PMF solutions are not unique since linear transformation (still conserving the non-negativity constraint) are possible ($GF = GTT^-1F$). The rotational freedom of the chosen solution can be explored through a non-zero valued user-specified rotational parameter $f_{peak}$. $F_{peak} > 0$ tries to impose rotations on the emerging solutions using positive coefficients $r$ in $T$, $f_{peak} < 0$ versa. $F_{peak} = 0$ produces the most central solution (Fig. S6 b).

The number of factors $p$ was chosen to be 6 for the UMR dataset (Fig. S7). In the solution with $p = 5$ (Fig. S8 a), the spectra of BBOA, HOA, and COA are less clearly separated (e. g. high signal at $m/z$ 57 in the top factor resembling BBOA, but very little signal at $m/z$ 57 in the red factor resembling COA). Figure S8 b shows the time series of the 5-factor solution – they are less clearly distinct than those of the 6-factor solution. The 7-factor solution (Fig. S9, a) features a factor consisting mostly of signal at $m/z$ 43 and a factor (orange) with single, isolated peaks inconsistent with regular ion series. The time series show a more similar evolution (Fig. S9 b), indicating a split of factors.
For the PMF solution presented in the manuscript, the 6-factor solution was chosen and the two factors assigned to SV-OOA (black and purple) regrouped to one SV-OOA, using the sum for the time series and the loadings-weighted average of the spectra.

Figure S10 presents the explained variance of the organics as a function of \(f_{\text{peak}}\) for the chosen 6-factor solution. \(f_{\text{peak}}\) was chosen to be -0.7 based on correlations of the corresponding factors with reference spectra (Ng et al., 2011). A boxplot of the scaled residuals (boxes are +/- 25% of points) per \(m/z\) is shown in Fig. S11, time series of the residuals and \(Q/Q_{\text{expected}}\) are shown in Fig. S12. On 16 March 2009, a power failure led to a breakdown of the instrument and a subsequent pumping down effect (Fig. S12). Downweighting this period in the input for PMF did not alter the solution.

The solution space for the chosen \(p = 6\) (central rotation) was explored by running PMF with 50 random initial values (\(SEED\)) at iteration start (Figs. S13 – 14). Roughly three solution groups can be identified (numbers in Fig. S14). Groups 1 and 2 feature a factor spectrum predominantly consisting of \(m/z\) 43 and two spectra that are basically identical. The spectrum with BBOA-like features shows no contributions at \(m/z\) 44, which is inconsistent with previous studies. For group 3, all spectra not assigned to OOA show very high similarities. The solution with a central rotation (\(f_{\text{peak}} = 0\)) was thus discarded regardless of \(SEED\) values. Similar information was also published in the supplementary information of Mohr et al. (2011).

Figure S 6: \(Q/Q_{\text{expected}}\) versus the number of factors \(p\) (a) or \(f_{\text{peak}}\) (b). The orange circle denotes the chosen UMR solution.
Figure S7: 6-factor UMR solution chosen, mass spectra (a) and time series (b). The black and the purple factor (SV-OOA 1 and 2) were regrouped to SV-OOA.

Figure S8: 5-factor UMR solution, mass spectra (a) and time series (b).
Figure S9: 7-factor UMR solution, mass spectra (a) and time series (b).

Figure S10: Variance explained by PMF due to the 6-factor UMR solution as a function of \( f_{\text{peak}} \). For the solution presented, \( f_{\text{peak}} = -0.7 \).
Figure S11: Median black strokes) and lower/upper quartiles (boxes) of the scaled residuals per m/z.

Figure S12: Time series of scaled residuals (top panel) and $Q/Q_{expected}$ (lower panel).

Figure S13: $Q/Q_{expected}$ as a function of different SEED values.
Figure S14: Variance explained by PMF due to the 6-factor UMR solution as a function of SEED. The numbers 1, 2 and 3 denote the three solution groups identified (see text).

3.4 HR solution criteria

Figure S15: Chosen 6-factor solution of the HR dataset, mass spectra (a) and time series (b).
Figure S16: 5-factor solution for the HR dataset, mass spectra (a) and time series (b).

Figure S17: 7-factor solution for the HR dataset, mass spectra (a) and time series (b).
Figure S18: $Q/Q_{\text{expected}}$ versus the number of factors $p$ (a) or $f_{\text{peak}}$ (b), HR PMF. For $f_{\text{peak}} < -1$, $Q/Q_{\text{expected}}$ starts to increase again (not shown). The orange circle denotes the chosen solution.

Figure S19: Variance explained by PMF due to the 6-factor HR solution as a function of $f_{\text{peak}}$. For the solution presented, $f_{\text{peak}} = 0$. 
Figure S20: $Q/Q_{\text{expected}}$ versus SEED for the HR solution.

Figure S21: Variance explained by PMF due to the 6-factor HR solution as a function of SEED.
Figure S22: Median (black strokes) and lower/upper quartiles (boxes) of the scaled residuals per m/z (HR solution).

Figure S23: Time series of scaled residuals (top panel) and $Q/Q_{expected}$ (lower panel) for the HR solution.
Figure S24: Scatter plot of the time series of $b_{abs}(880 \text{ nm})_{\text{traffic}}$ and HOA. The red line is the least orthogonal distance fit where the circle data points were removed.

Least orthogonal distance fit

$R^2$ all data: 0.17

$R^2$ circles removed from the fit: 0.52
Figure S25: $m/z$ 55/Org ($f_{55}$) plotted against $m/z$ 57/Org ($f_{57}$).

Figure S26. Signal at $m/z$ 55 in the HR spectra of meat cooking sources (a) and vehicle engine sources (b). In the engine exhaust spectra, the signal is almost entirely due to the reduced hydrocarbon ion $C_4H_7^+$, whereas in
the cooking spectra there is also substantial contribution from the oxygen-containing ion $\text{C}_3\text{H}_3\text{O}^+$. Reprinted from Mohr et al. (2009).

4 References


