1   Supplementary information for manuscript

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

5

6   C. Mohr[1], P. F. DeCarlo[1]*, M. F. Heringa[1], R. Chirico[1,**], J. G. Slowik[1], R. Richter[1],
7   C. Reche[2], A. Alastuey[2], X. Querol[2], R. Seco[3,***], J. Peñuelas[3], J. L. Jiménez[4,5], M. Crippa[1],
8   R. Zimmermann[6,7], U. Baltensperger[1], and A. S. H. Prévôt[1]

9

10   [1]Laboratory of Atmospheric Chemistry, Paul Scherrer Institut (PSI), Villigen, Switzerland

11   [2]Institute for Environmental Assessment and Water Research (IDAEA-CSIC), Barcelona, Spain

12   [3]Unitat d'Ecologia Global CREAF-CEAB-CSIC, Centre de Recerca Ecològica i Aplicacions

13   Forestals, Universitat Autònoma de Barcelona, Barcelona, Spain

14   [4]Department of Chemistry and Biochemistry, [5]Cooperative Institute for Research in the

15   Environmental Sciences (CIRES), University of Colorado, Boulder, USA

16   [6]Helmholtz Zentrum Munchen, Joint Mass Spectrometry Center, Institute of Ecological

17   Chemistry, Neuherberg, Germany

18   [7]University of Rostock, Rostock, Germany

19   * now at: Department of Civil, Architectural, and Environmental Engineering, Drexel University,

20   Philadelphia, USA

21   ** now at: Italian National Agency for New Technologies, Energy and Sustainable Economic

22   Development (ENEA), UTAPRAD-DIM, Frascati, Italy

23   ***now at: Atmospheric Chemistry Division, National Center for Atmospheric Research,
24   Boulder, USA

25

26

27

28   Correspondence to: A. S. H. Prévôt (andre.prevot@psi.ch)

29

1 Collection efficiency (CE)
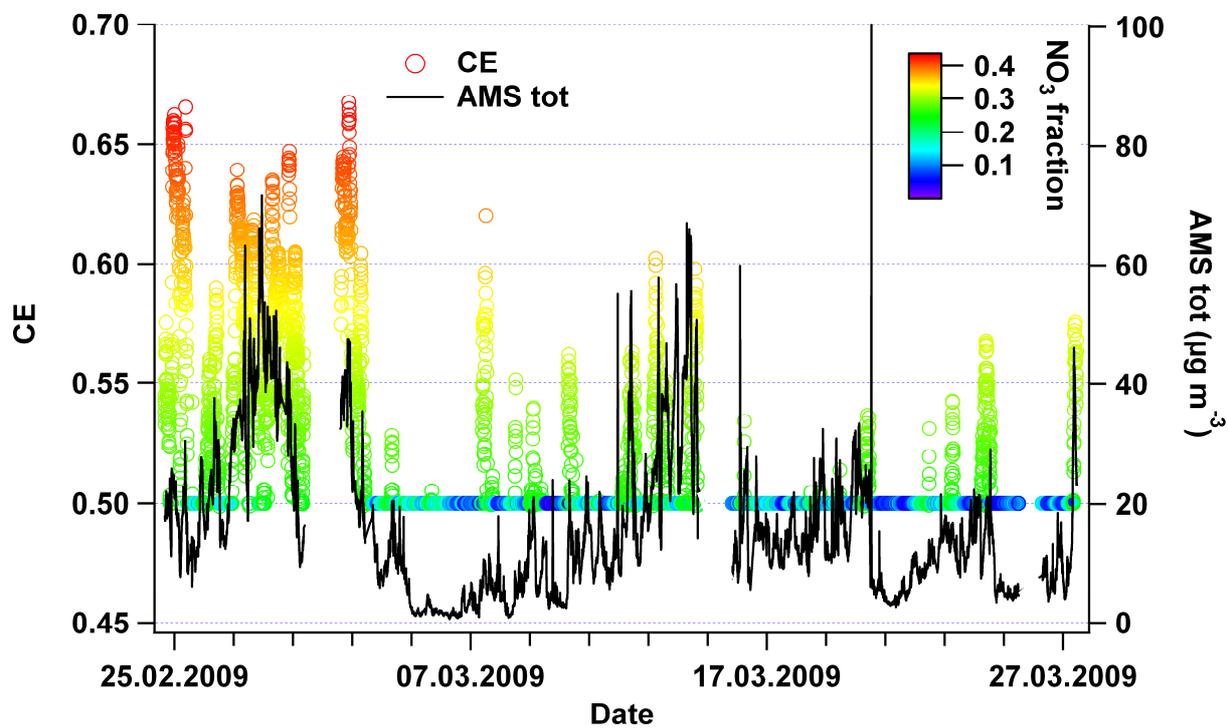
Figure S1: Time series of the collection efficiency (CE) used for the present dataset (left axis) and total concentration of species measured by AMS (right axis).

2   PM$_1$ time series

Least orthogonal distance fit
Coefficient values ± one standard deviation
intercept   =-3.60 ± 0.48
slope   =0.99 ± 0.02

AMS tot + BC (µg m^{-3})

Grimm PM$_1$ (µg m^{-3})
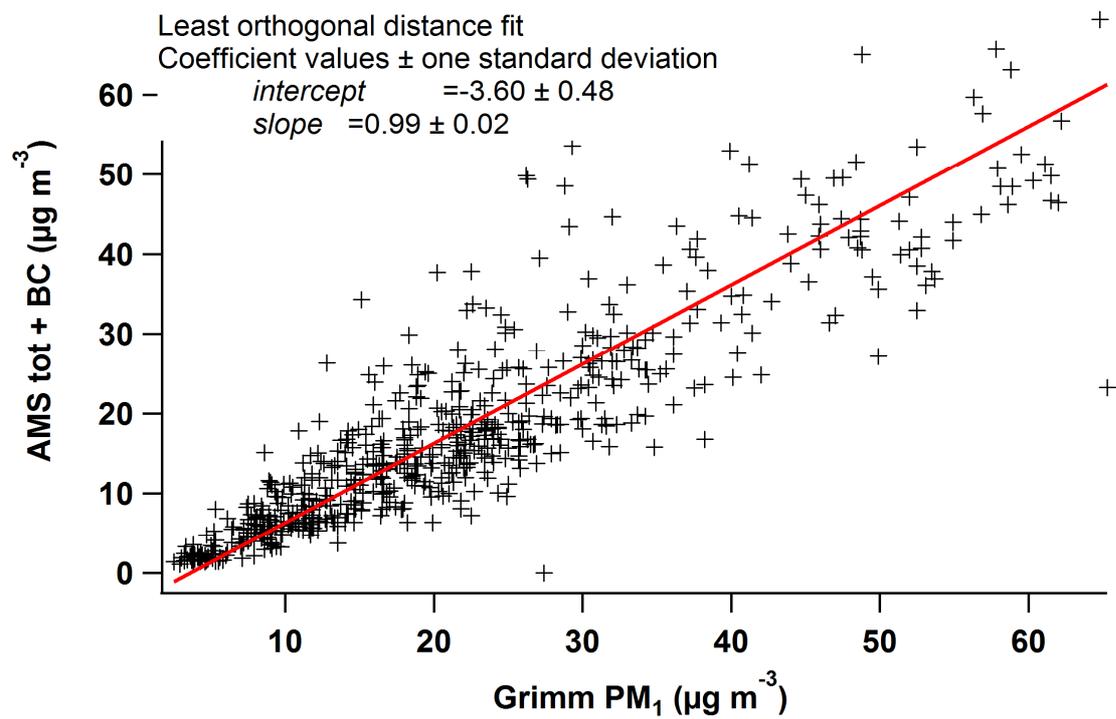
39   **Figure S2: Scatterplot of combined time series of total AMS species (HR) and BC (y-axis) and Grimm PM$_1$.**
40   **The data were fitted with a least orthogonal distance fit (red line).**

41 3 PMF

42 3.1 UMR solution

43
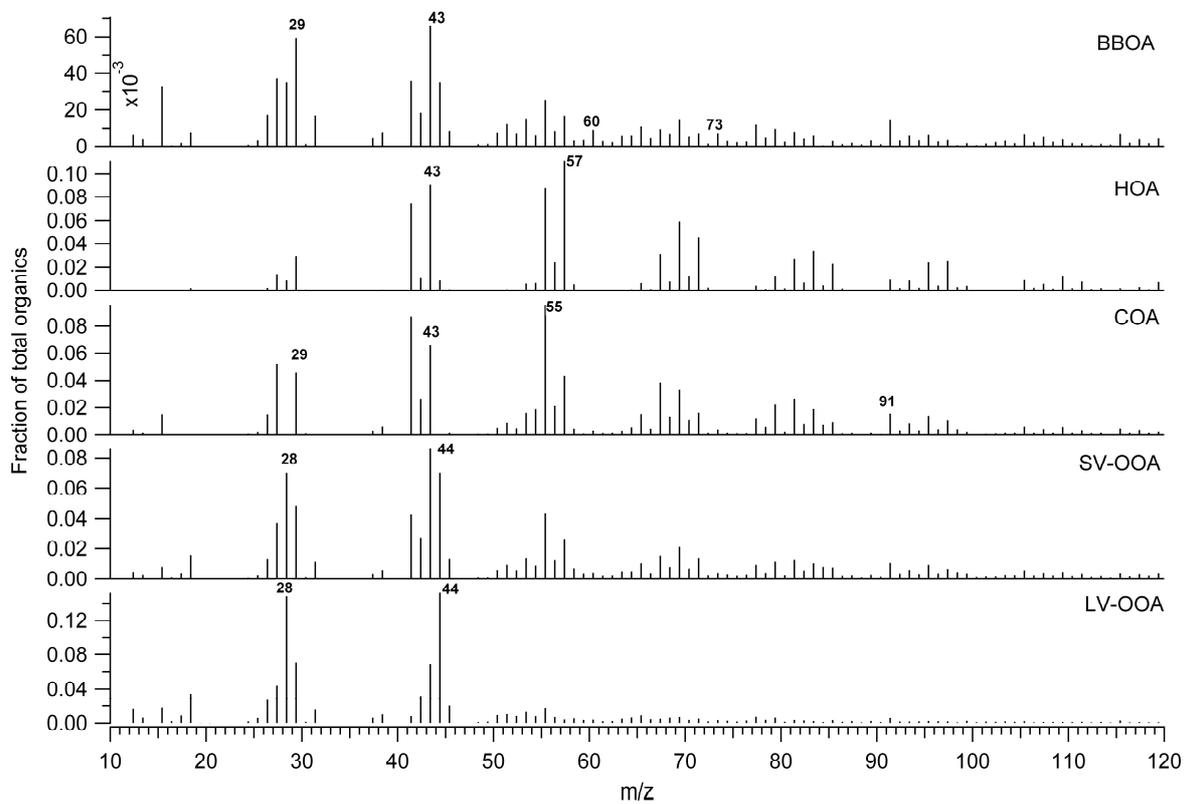
**Figure S3: Mass spectra of the UMR 5-factor-PMF solution.**
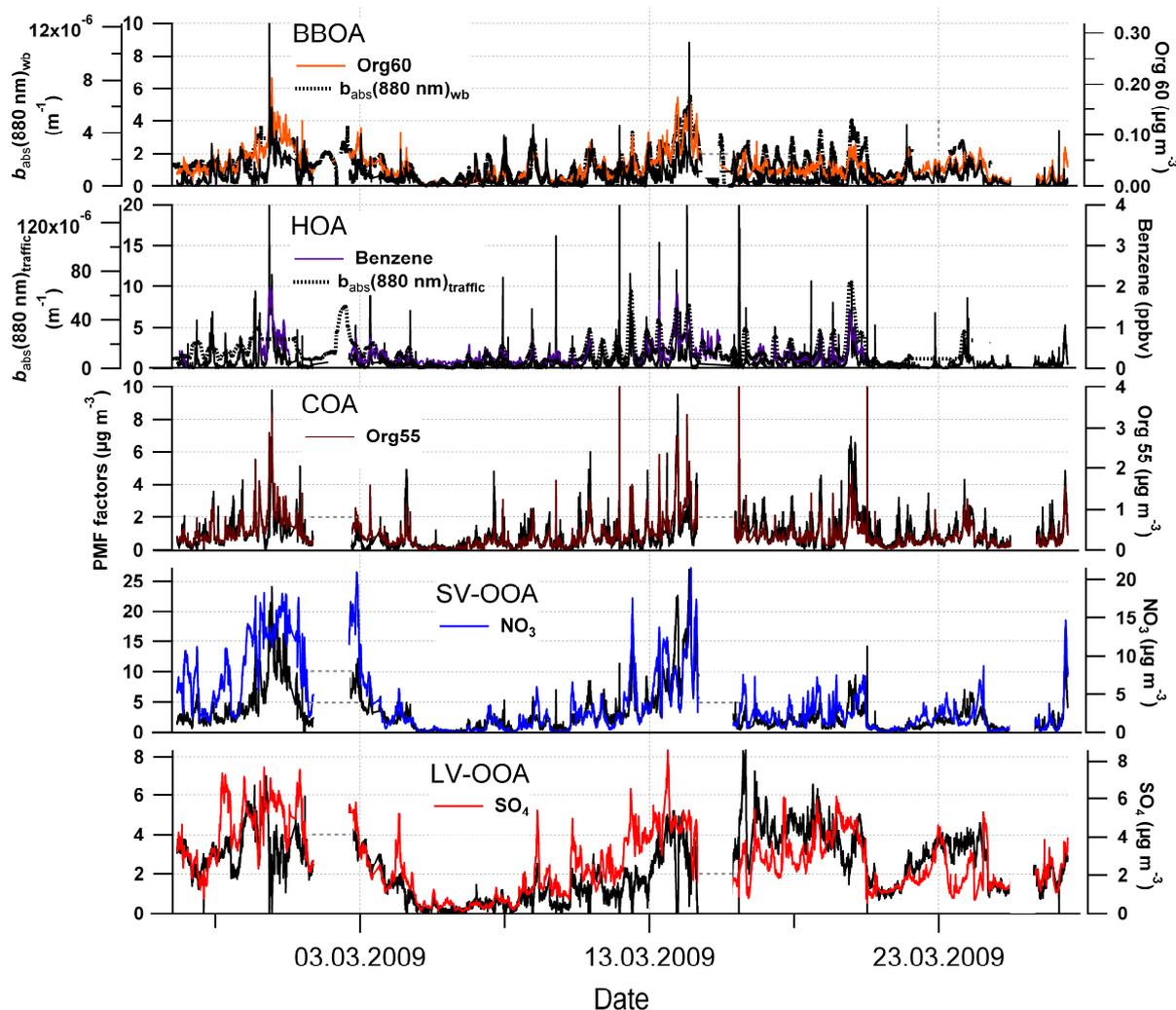
Figure S 4: Time series of the UMR 5-factor-PMF solution and ancillary data.

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

56 resolution and, related to that, the signal on an individual ion basis of the HR data matrix adds

57 additional information to the HR data matrix and thus allows for a better quantification of

58 primary and secondary OA.
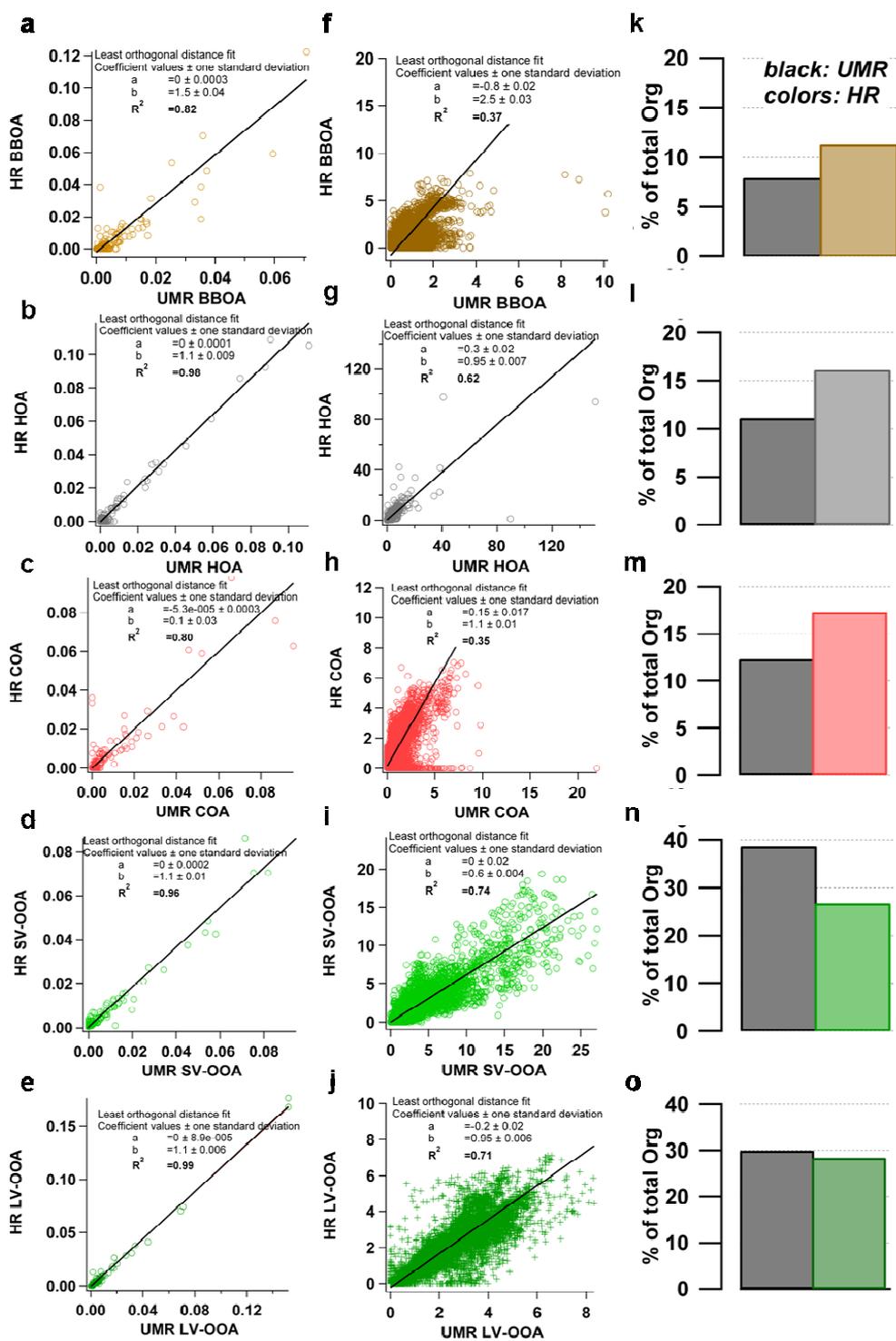
Figure S5: Scatter plots of UMR and HR PMF spectra (a-e), time series (f-j) and a comparison of the mass attributed to each factor relative to OA (k-o).

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 σ_{ij} of the measured input values x_{ij} are used to determine the weights of the residuals e_{ij}.

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

If the model is appropriate and the data uncertainties estimations are accurate, $(e_{ij} / \sigma_{ij})^2$ is ~1 and the expected Q ($Qexpected$) = $mn-p(m+n) \approx mn$, the degrees of freedom of the fitted data. The Q-value is one mathematical criterion for the quality of the PMF solution: $Q/Qexpected>>1$ indicates an underestimation, $Q/Qexpected <<1$ 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 *fpeak*. *Fpeak* > 0 tries to impose rotations on the emerging solutions using positive coefficients r in **T**, *fpeak* < 0 vice versa. *Fpeak* = 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.

91   For the PMF solution presented in the manuscript, the 6-factor solution was chosen and the two

92   factors assigned to SV-OOA (black and purple) regrouped to one SV-OOA, using the sum for

93   the time series and the loadings-weighted average of the spectra.

94   Figure S10 presents the explained variance of the organics as a function of *fpeak* for the chosen

95   6-factor solution. *fpeak* was chosen to b -0.7 based on correlations of the corresponding factors

96   with reference spectra (Ng et al., 2011).

97   A boxplot of the scaled residuals (boxes are +/- 25% of points) per *m/z* is shown in Fig. S11,

98   time series of the residuals and *Q/Qexpected* are shown in Fig. S12. On 16 March 2009, a power

99   failure led to a breakdown of the instrument and a subsequent pumping down effect (Fig. S12).

100   Downweighting this period in the input for PMF did not alter the solution.

101   The solution space for the chosen $p = 6$ (central rotation) was explored by running PMF with 50

102   random initial values (*SEED*) at iteration start (Figs. S13 – 14). Roughly three solution groups

103   can be identified (numbers in Fig. S14). Groups 1 and 2 feature a factor spectrum predominantly

104   consisting of *m/z* 43 and two spectra that are basically identical. The spectrum with BBOA-like

105   features shows no contributions at *m/z* 44, which is inconsistent with previous studies. For group

106   3, all spectra not assigned to OOA show very high similarities. The solution with a central

107   rotation (*fpeak* = 0) was thus discarded regardless of *SEED* values. Similar information was also

108   published in the supplementary information of Mohr et al. (2011).
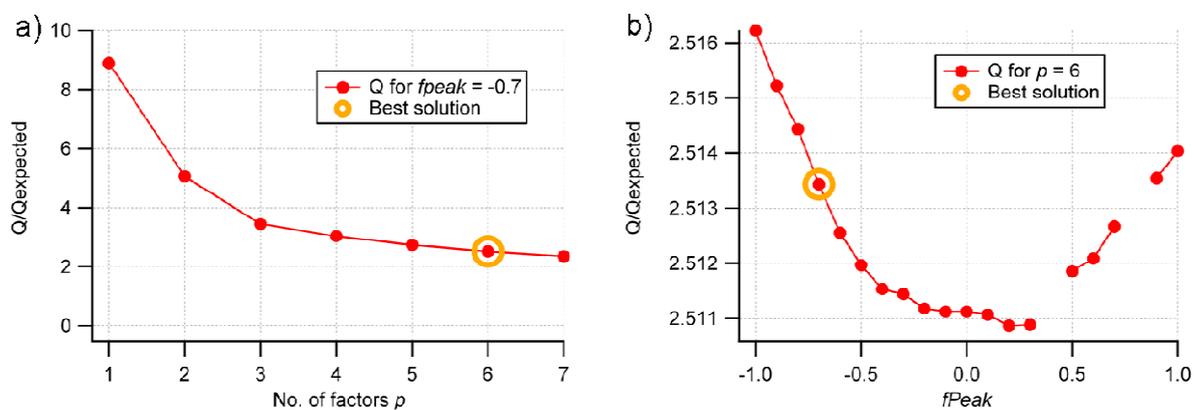
110 **Figure S 6: *Q/Qexpected* versus the number of factors *p* (a) or *fpeak* (b). The orange circle denotes the chosen**

111 **UMR solution.**
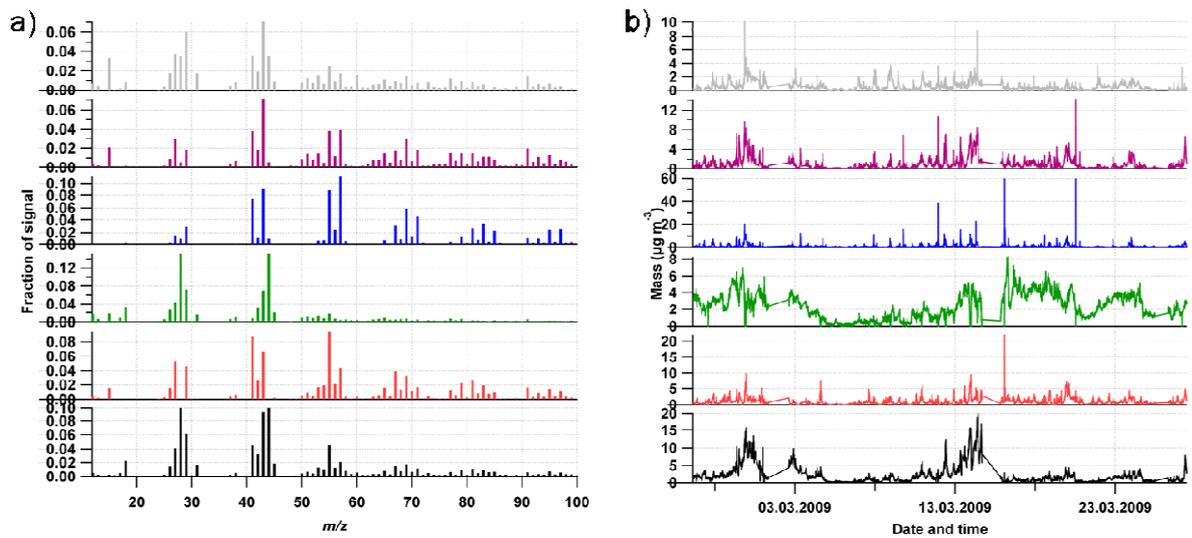
Figure S7: 6-factor UMR solution chosen, mass spectra (a) and time series (b). The black and the purple factor (SV-OOA 1 and 2) were regrouped to SV-OOA.
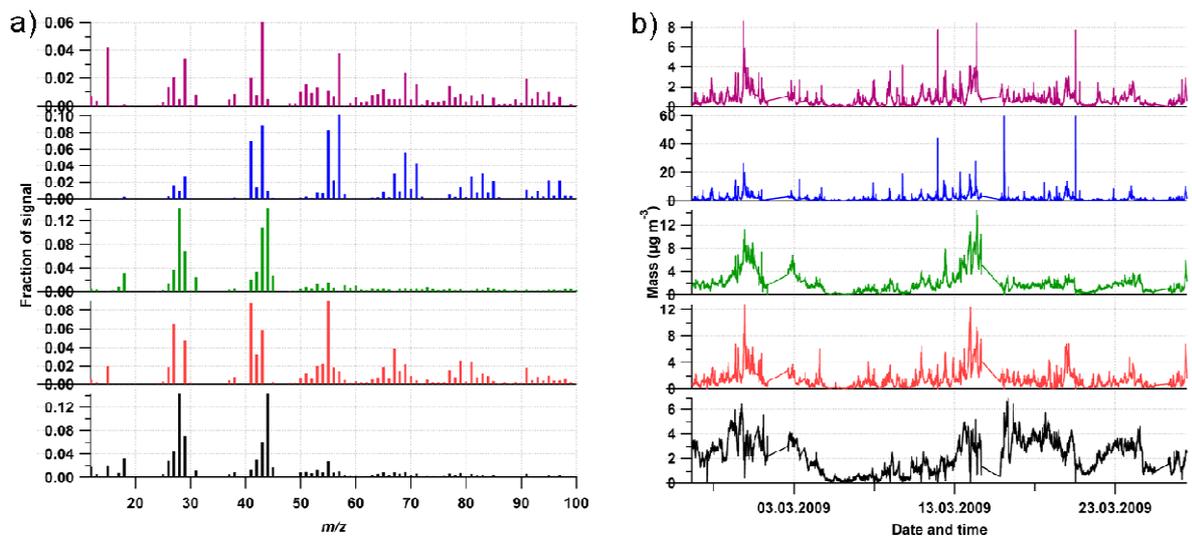

Figure S8: 5-factor UMR solution, mass spectra (a) and time series (b).
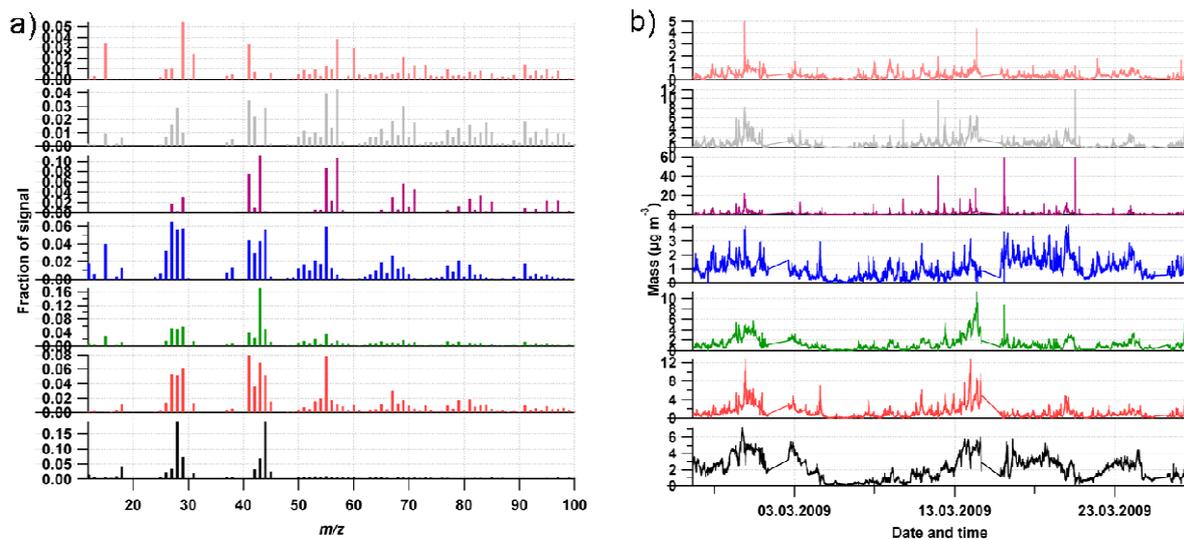
118

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

120
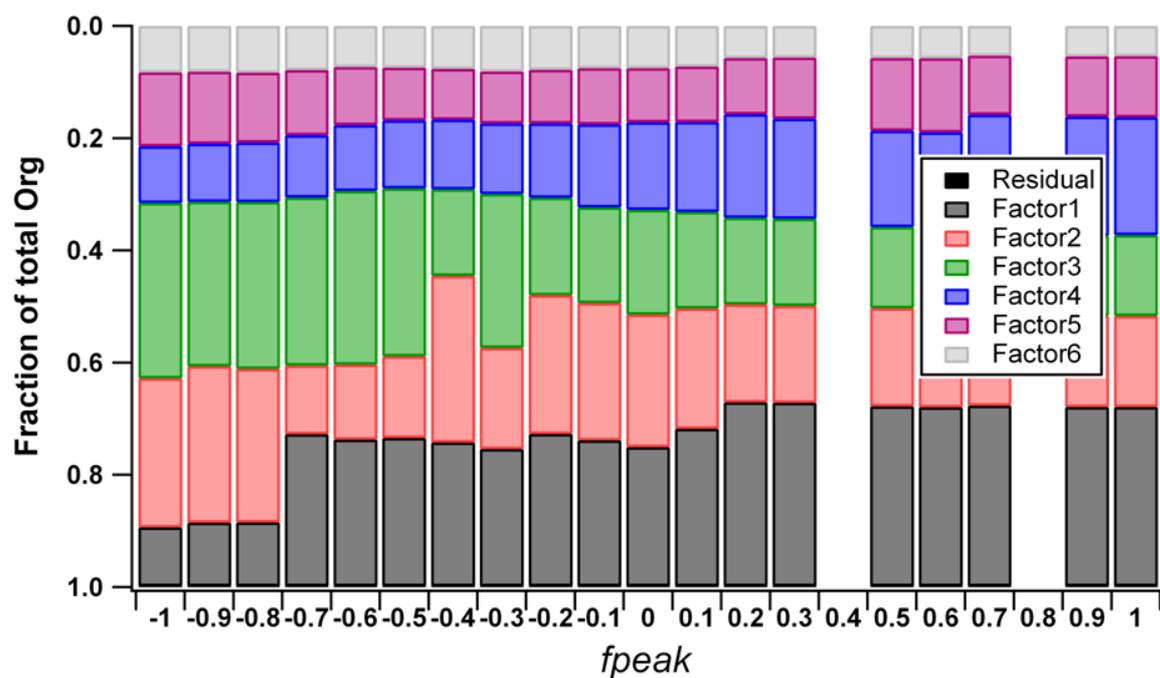
121

122   **Figure S10: Variance explained by PMF due to the 6-factor UMR solution as a function of *fpeak*. For the**

123   **solution presented, *fpeak* =-0.7.**

124

125

126

127
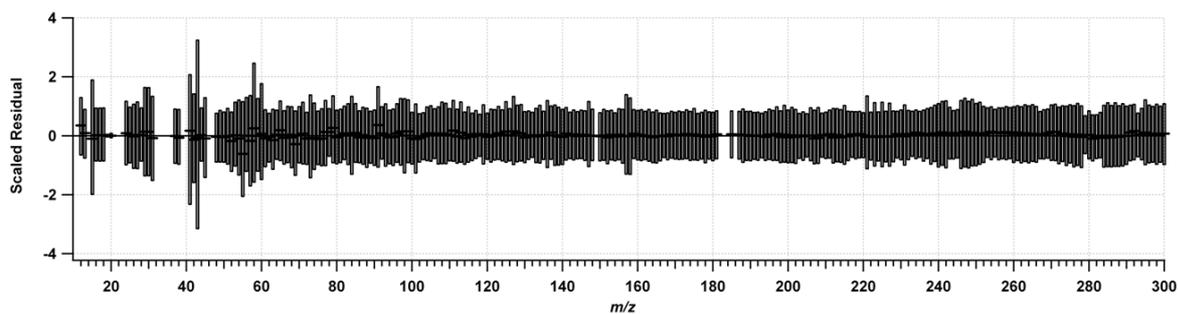
Figure S11: Median black strokes) and lower/upper quartiles (boxes) of the scaled residuals per *m/z*.

129

130
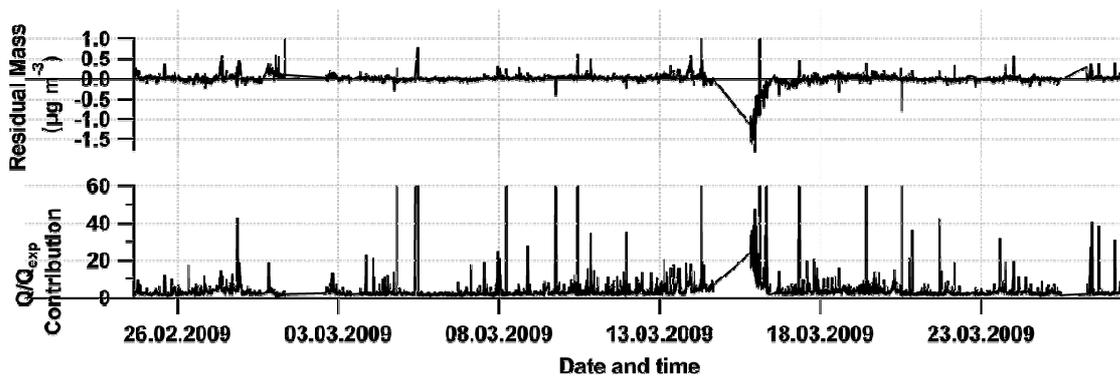
Figure S12: Time series of scaled residuals (top panel) and *Q/Qexpected* (lower panel).

132

133
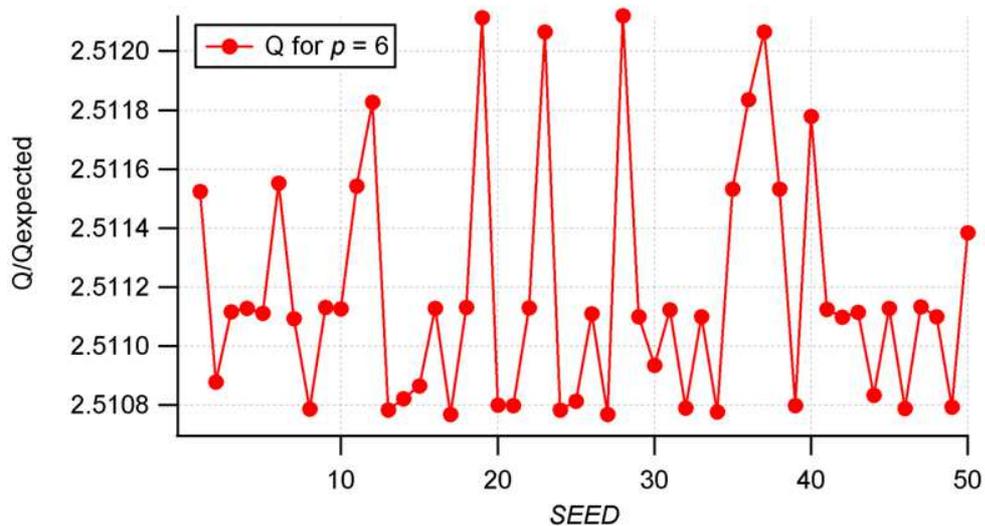
Figure S13: *Q/Qexpected* as a function of different *SEED* values.

135

12

136
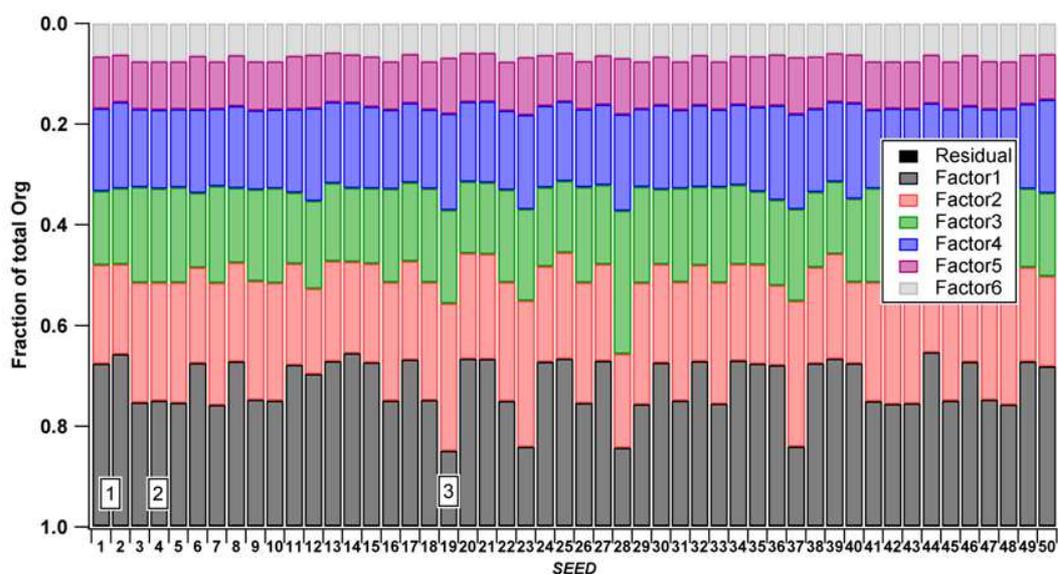
Figure S14: Variance explained by PMF due to the 6-factor UMR solution as a function of *SEED*. The

138 **numbers 1, 2 and 3 denote the three solution groups identified (see text).**

139

3.4   HR solution criteria

141

142
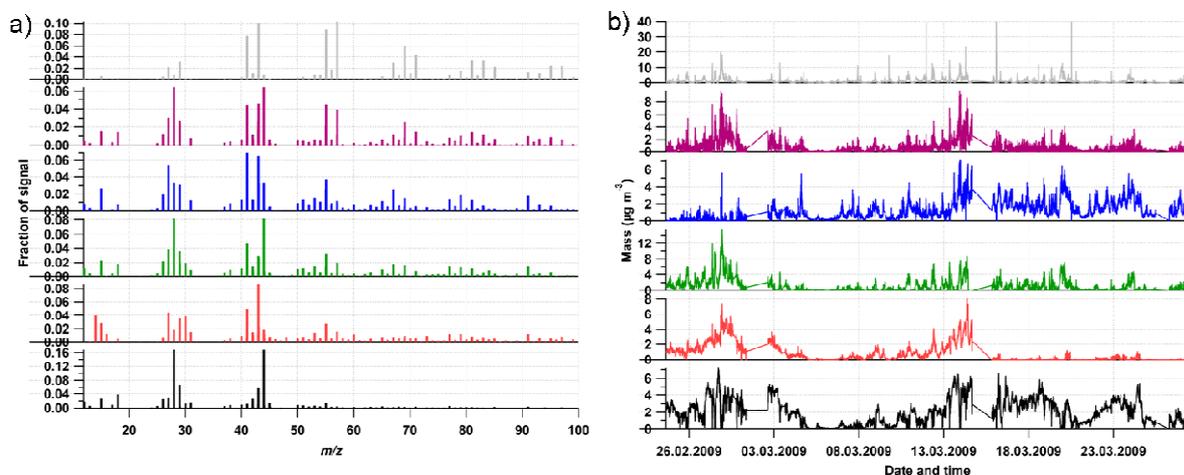
Figure S15: Chosen 6-factor solution of the HR dataset, mass spectra (a) and time series (b).

144

145
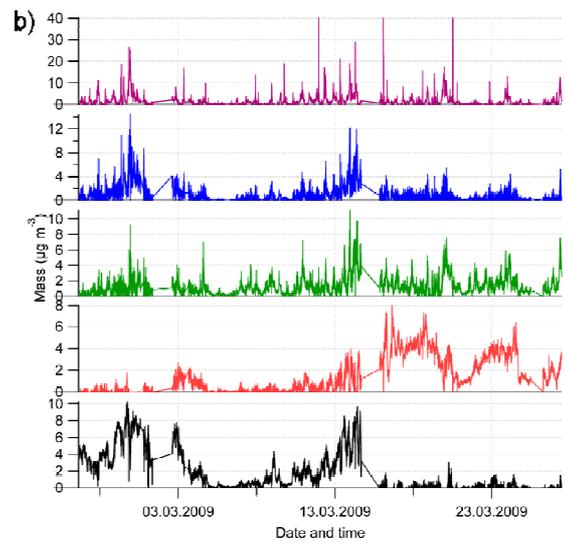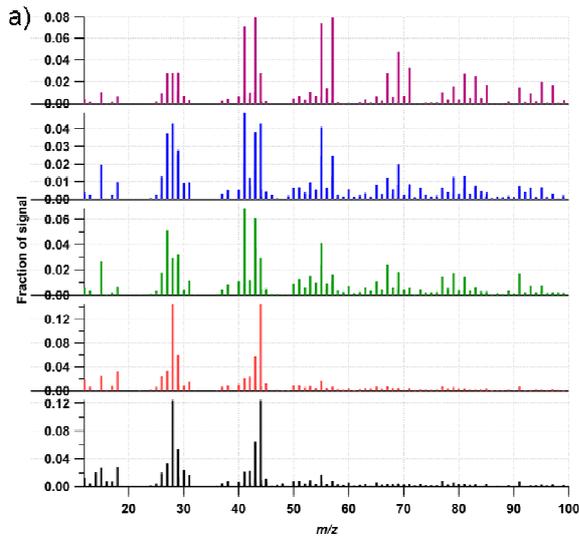
Figure S16: 5-factor solution for the HR dataset, mass spectra (a) and time series (b).

147

148
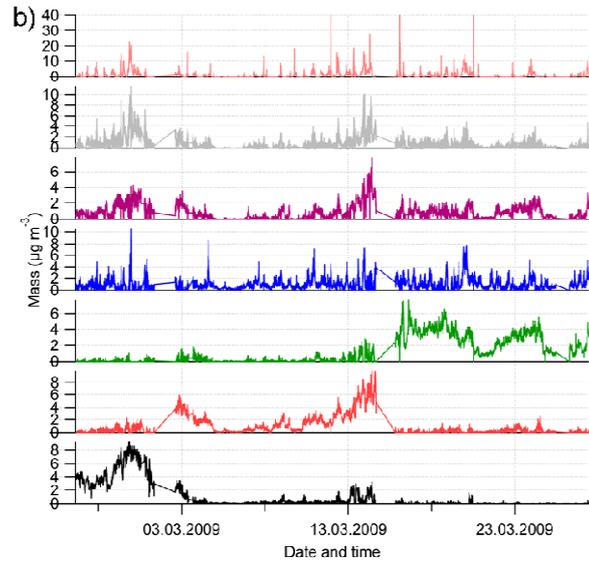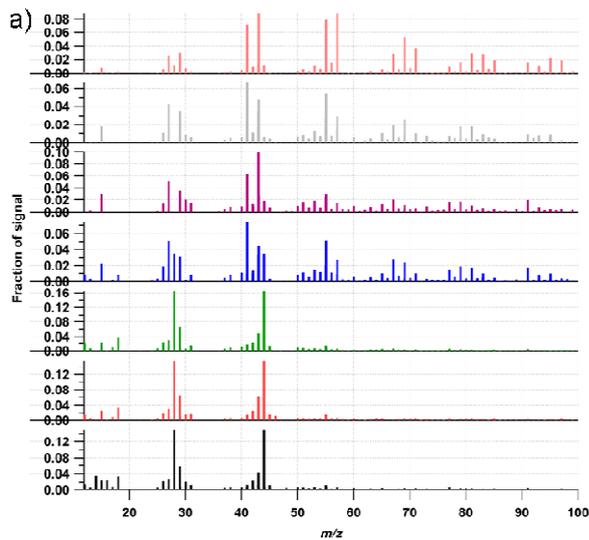
Figure S17: 7-factor solution for the HR dataset, mass spectra (a) and time series (b).

150
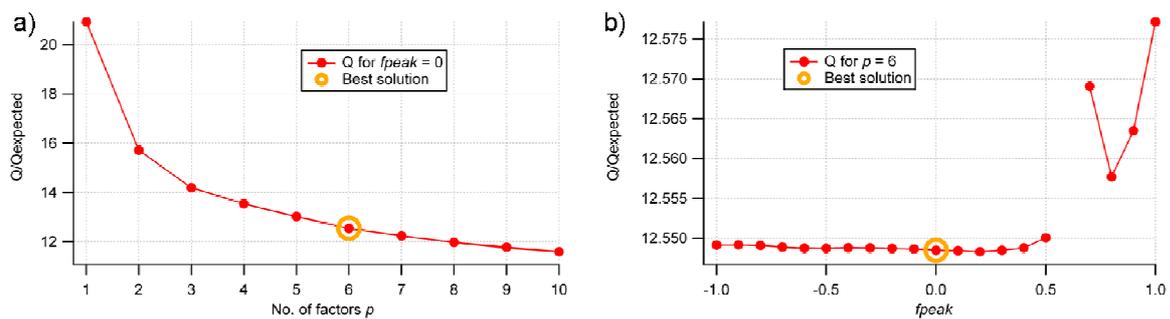
Figure S18: *Q/Qexpected* versus the number of factors *p* (a) or *fpeak* (b), HR PMF. For *fpeak < -1*, *Q/Qexpected* starts to increase again (not shown). The orange circle denotes the chosen solution.
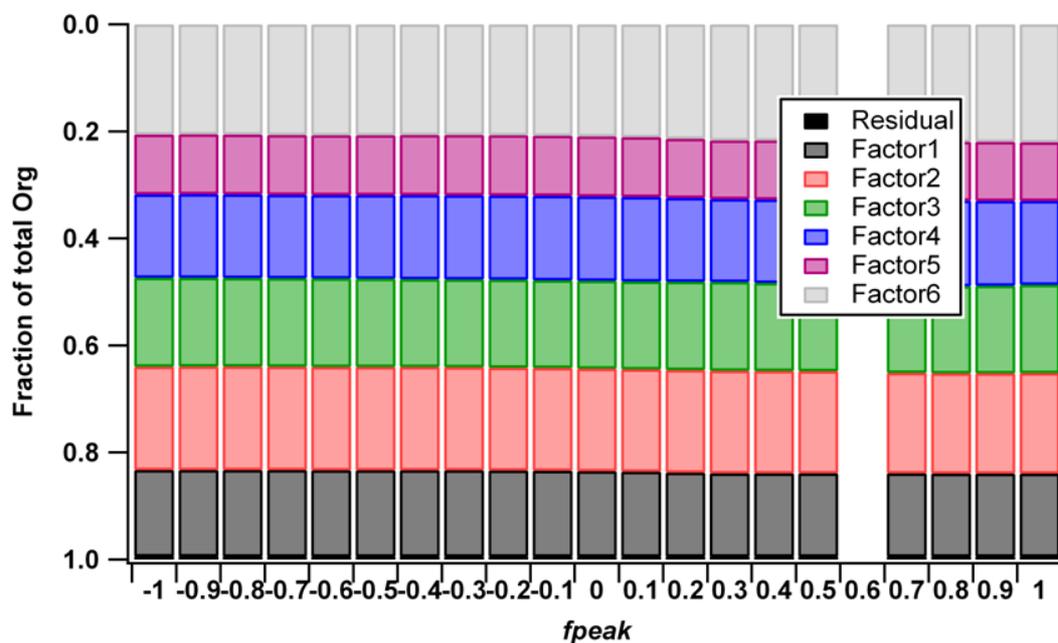
**Figure S19: Variance explained by PMF due to the 6-factor HR solution as a function of *fpeak*. For the solution presented,** *fpeak* **=0.**

159
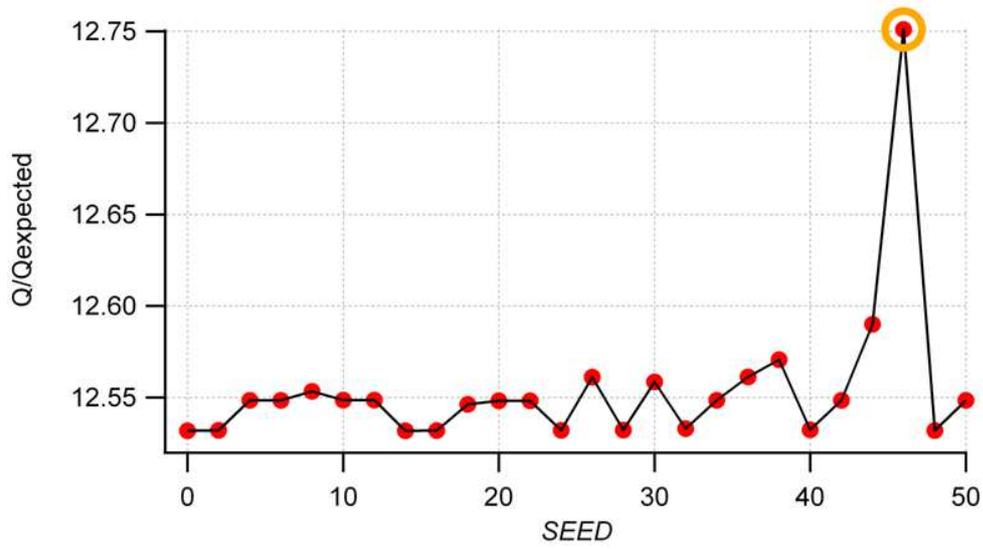
Figure S20: *Q/Qexpected* **versus** *SEED* **for the HR solution.**

161

162
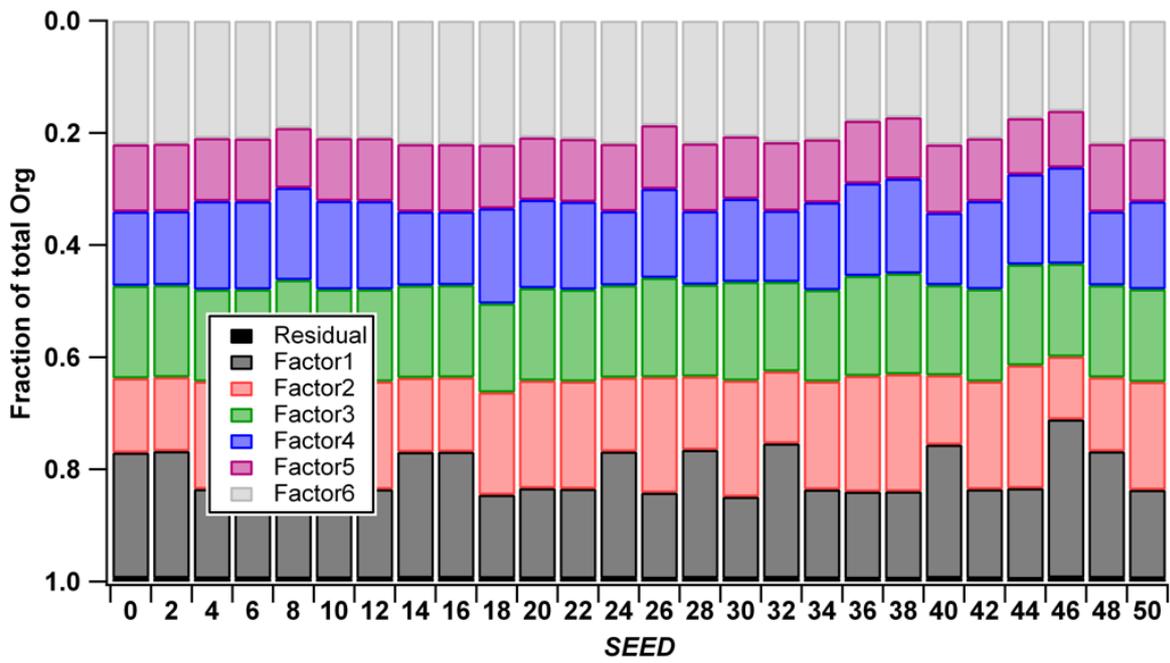
Figure S21: Variance explained by PMF due to the 6-factor HR solution as a function of *SEED*.

164
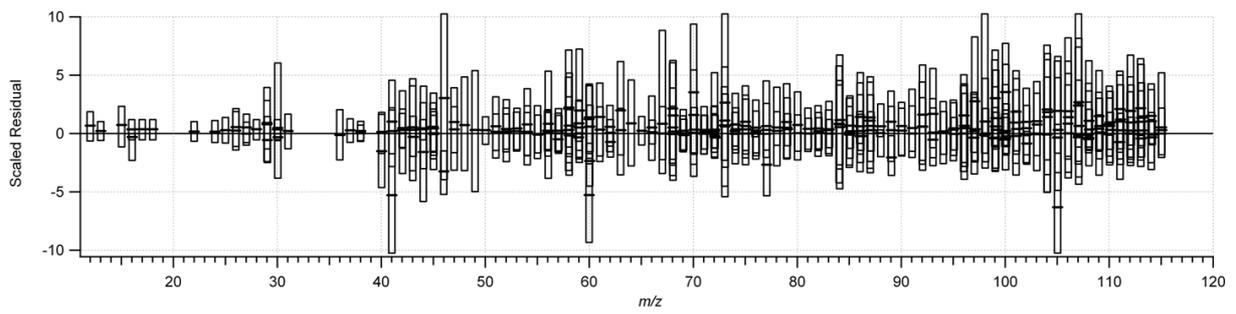
Figure S22: Median (black strokes) and lower/upper quartiles (boxes) of the scaled residuals per *m/z* (HR solution).
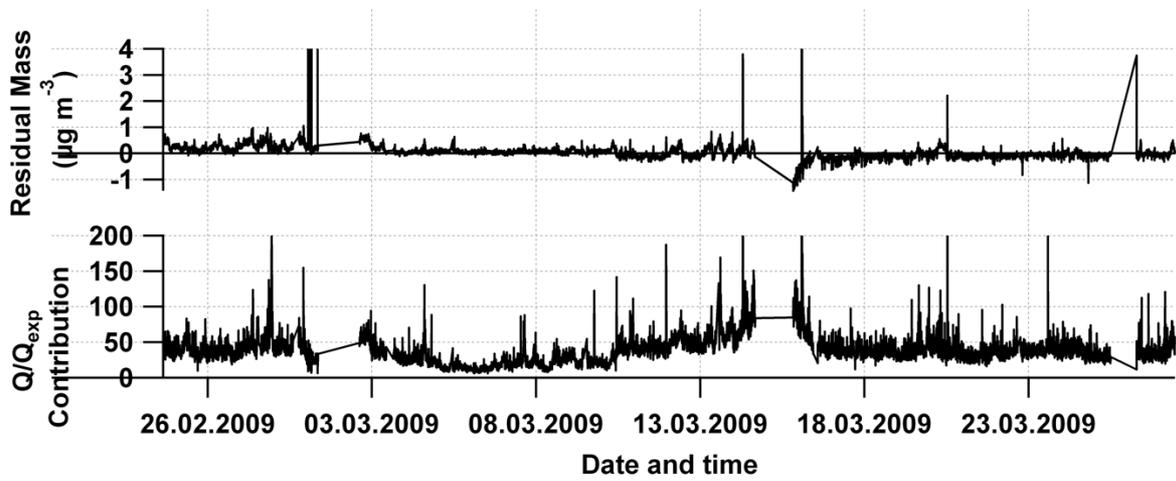

**Figure S23: Time series of scaled residuals (top panel) and *Q/Qexpected* (lower panel) for the HR solution.**
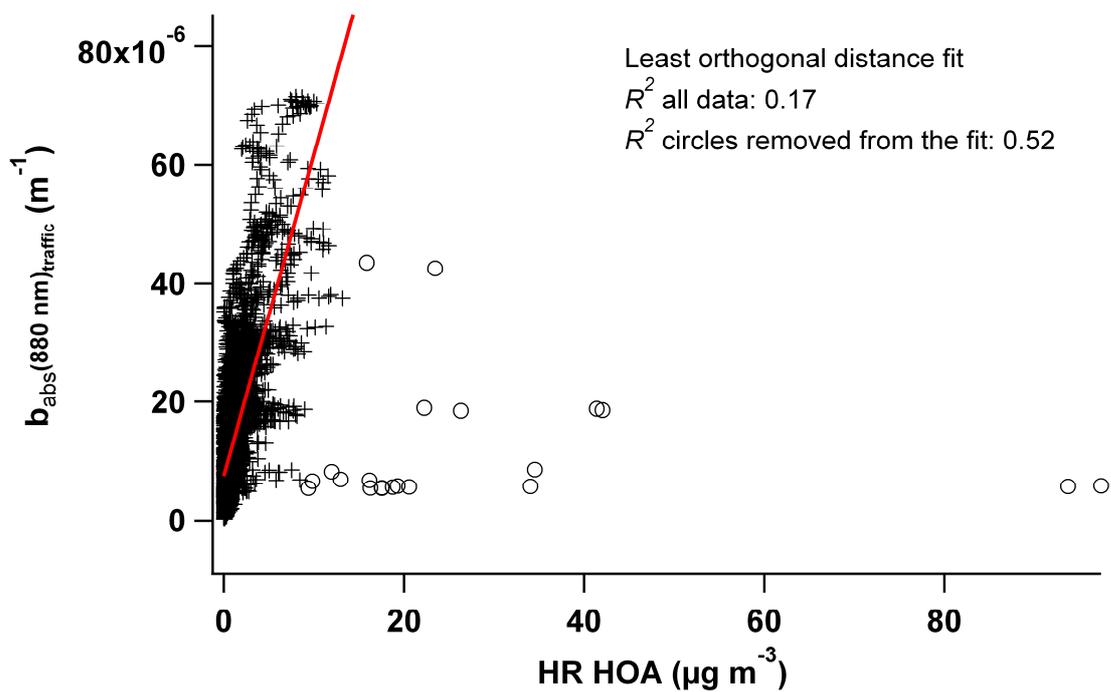
Figure S24: Scatter plot of the time series of $b_{abs}(880 \text{ nm})_{traffic}$ and HOA. The red line is the least orthogonal distance fit where the circle data points were removed.
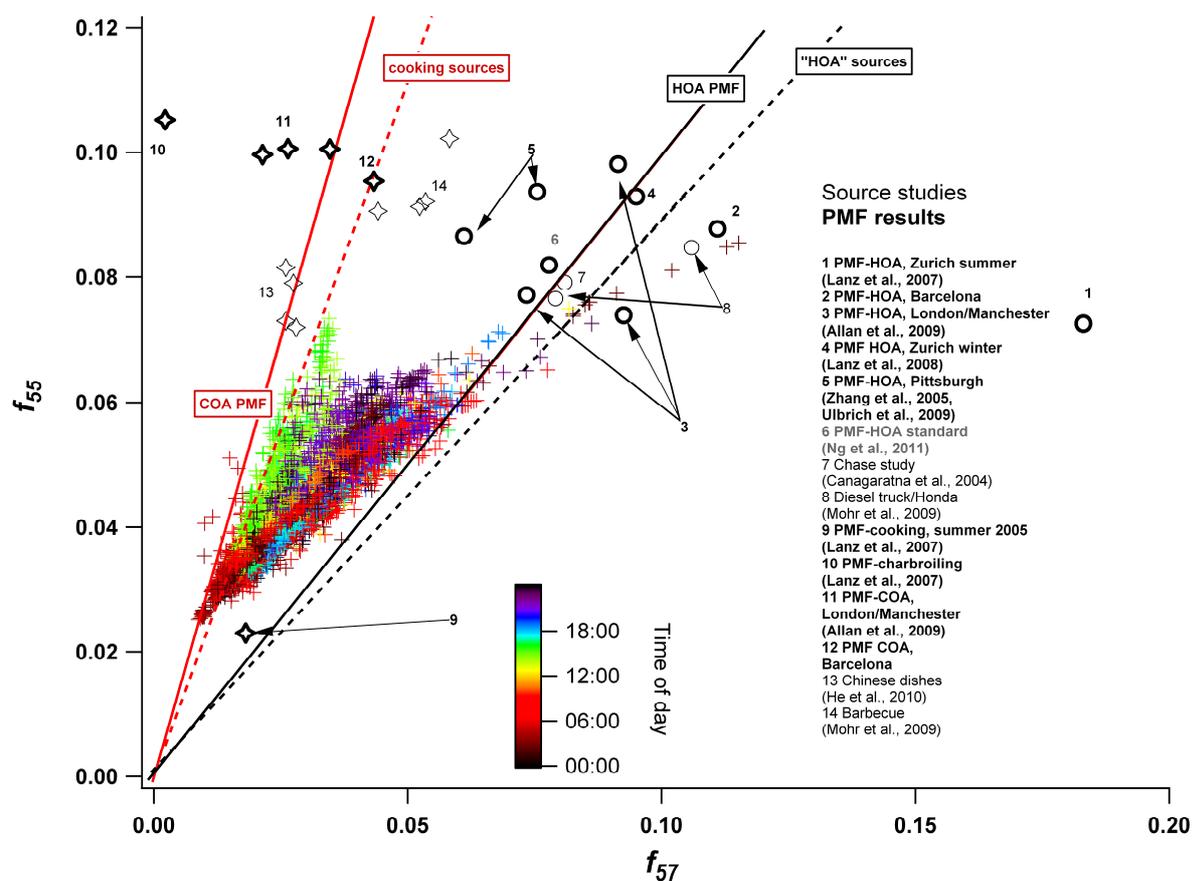
Figure S 25: *m/z* 55/Org (*f₅₅*) plotted against *m/z* 57/Org (*f₅₇*).
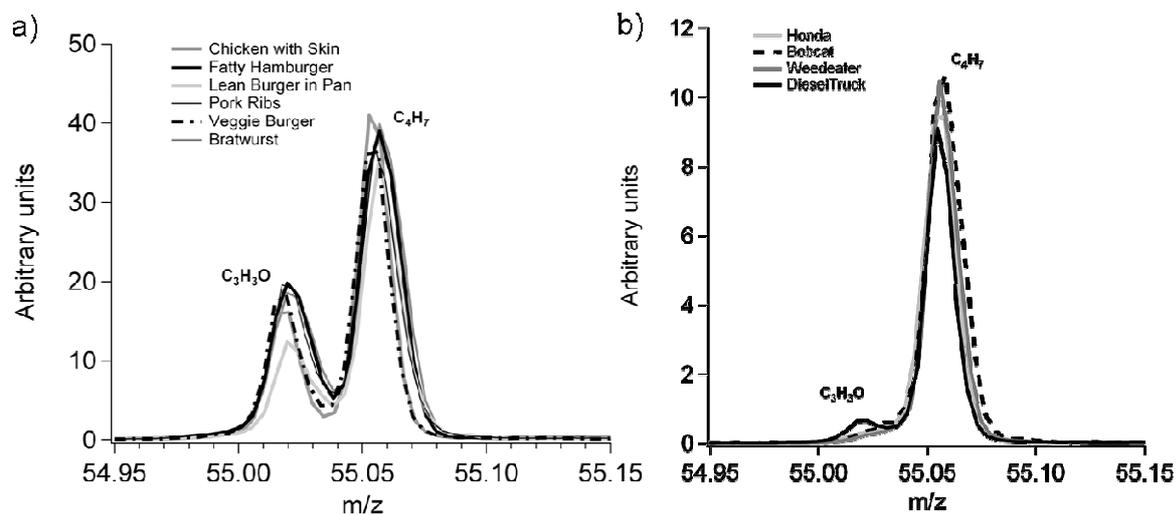
Figure S26. Signal at *m/z* 55 in the HR spectra of meat cooking sources (a) and vehicle engine sources (b). In the engine exhaust spectra, the signal is almost entirely due to the reduced hydrocarbon ion $C_4H_7^+$, whereas in

182 **the cooking spectra there is also substantial contribution from the oxygen-containing ion $C_3H_3O^+$. Reprinted**

183 **from Mohr et al. (2009).**

184

4 References

186 Mohr, C., Huffman, J. A., Cubison, M. J., Aiken, A. C., Docherty, K. S., Kimmel, J. R., Ulbrich,
187 I. M., Hannigan, M., and Jimenez, J. L.: Characterization of primary organic aerosol emissions
188 from meat cooking, trash burning, and motor vehicles with high-resolution aerosol mass
189 spectrometry and comparison with ambient and chamber observations, Environ. Sci. Technol.,
190 43, 2443-2449, 2009.

191 Mohr, C., Richter, R., DeCarlo, P. F., Prévôt, A. S. H., and Baltensperger, U.: Spatial variation of
192 chemical composition and sources of submicron aerosol in Zurich during wintertime using
193 mobile aerosol mass spectrometer data, Atmos. Chem. Phys., 11, 7465-7482, 2011.

194 Ng, N. L., Canagaratna, M. R., Jimenez, J. L., Zhang, Q., Ulbrich, I. M., and Worsnop, D. R.:
195 Real-time methods for estimating organic component mass concentrations from aerosol mass
196 spectrometer data, Environ. Sci. Technol., 45, 910-916, 2011.

197 Paatero, P., Hopke, P. K., Song, X. H., and Ramadan, Z.: Understanding and controlling rotations
198 in factor analytic models, Chemometr. Intell. Lab., 60, 253-264, 2002.
199
200