

Supplement of Atmos. Chem. Phys., 20, 7693–7716, 2020  
<https://doi.org/10.5194/acp-20-7693-2020-supplement>  
© Author(s) 2020. This work is distributed under  
the Creative Commons Attribution 4.0 License.



*Supplement of*

## **Deconvolution of FIGAERO–CIMS thermal desorption profiles using positive matrix factorisation to identify chemical and physical processes during particle evaporation**

**Angela Buchholz et al.**

*Correspondence to:* Angela Buchholz ([angela.buchholz@uef.fi](mailto:angela.buchholz@uef.fi))

The copyright of individual parts of the supplement might differ from the CC BY 4.0 License.

## Supplement Information

### 1.1 Experimental setup of isothermal evaporation experiments

SOA was formed via combined photooxidation and ozonolysis of  $\alpha$ -pinene in a potential aerosol mass reactor (PAM, Aerodyne Research Inc., Kang et al., 2007; Lambe et al., 2011). The  $\alpha$ -pinene concentration (190 ppb), relative humidity (RH = 40%) and temperature (27 °C) in the PAM reactor were held constant throughout the experiments, but three different oxidative settings were chosen to create three types of  $\alpha$ -pinene SOA characterised by their oxidative age (Table S 2). The gas phase was monitored with a high-resolution time-of-flight proton transfer reaction mass spectrometer (PTR-MS, Ionicon, model 8000) while the particles were characterised with a scanning mobility particle sizer (SMPS, TSI Inc., model 3082+3775), a high resolution time-of-flight aerosol mass spectrometer (AMS, Aerodyne Research Inc., DeCarlo et al., 2006), and a filter inlet for gases and aerosols sampling unit (FIGAERO, Aerodyne Research Inc., Lopez-Hilfiker et al., 2014) in combination with a chemical ionisation mass spectrometer (CIMS, Aerodyne Research Inc., Lee et al., 2014) using iodide as reagent ion.

Two types of particles samples were collected with the FIGAERO-CIMS: “fresh” particles (labelled  $t_{\text{evap}} = 0.25\text{h}$ ) were collected directly after size selection with a nano differential mobility analyser (NanoDMA, TSI Inc., model 3085, 80 nm electro mobility size) and “RTC” particles (labelled  $t_{\text{evap}} = 4\text{h}$ ) which were left to evaporate at  $\sim 20$  °C for 3 - 4 h in a residence time chamber (RTC) prior to collection on the FIGAERO filter. The isothermal evaporation method has been described previously in detail (Buchholz et al., 2019; Yli-Juuti et al., 2017). The NanoDMA was operated with an open loop sheath flow (dry: 10 L min<sup>-1</sup>, wet: 8 L min<sup>-1</sup>) which together with the extremely short residence time inside the NanoDMA ( $\leq 0.3$  s) limited the diffusion of gaseous compounds into the selected sample flow (1 L min<sup>-1</sup>). This created a sudden shift in the gas-particle equilibrium which initiates the particle evaporation at the NanoDMA outlet. For evaporation experiments, this monodisperse particle sample was filled for 75 min into a 100 L stainless steel RTC, that was then closed off. After 3 – 4 h of isothermal evaporation, the RTC was reopened and the remaining particles sampled onto the FIGAERO filter. In addition, filter blank measurements were conducted on each experiment day to quantify the instrument background.

Generally, a mass loading of several 10s of ng is needed on the FIGAERO filter to achieve a sufficiently high signal-to-noise ratio in CIMS. In this study, 20 or 30 min of collection time were necessary to accumulate enough particular mass on the FIGAERO filter. The mass loadings in the RTC at the beginning of the evaporation experiment and on the FIGAERO filter estimated from AMS measurements are given in Table S 1. It has to be noted that in this specific setup, particles already collected on the FIGAERO filter will continue to evaporate during the remaining collection time as no

new gas-particle equilibrium can be reached after the dilution in the NanoDMA. Thus, the “fresh” particles on the FIGAERO filter actually represent particles with an average evaporation time of 15 min and may have lost some of the very volatile compounds before the thermal desorption begins. In Principle, particles sampled after hours of evaporation in the RTC exhibit the same behaviour. But as the majority of particle evaporation occurs within the first 1 h, the additional residence time on the FIGAERO filter does not contribute significantly to the overall evaporation (Buchholz et al., 2019).

## 1.2 PLError calculation

The error for mass spectra data consists of two parts: the counting statistic error and instrument noise (electronic background noise). Assuming a Poisson type distribution of the counting error, the total error  $S_{ij}$  for each ion  $i$  at time  $j$  can be described as (same as Eq. 7 in the main text):

$$S_{ij} = a \cdot \sqrt{\frac{X_{ij}}{t_s}} + \sigma_{noise,i} \quad (S1)$$

with  $X_{ij}$  signal intensity of the ion  $i$ ,  $t_s$  sampling (averaging) interval in  $s$ ,  $\sigma_{noise,i}$  the electronic noise for ion  $i$ , and the empirically derived parameters  $a$ . Yan et al. (2016) suggested a method to derive  $a$  from a data set of CIMS gas-phase measurements utilising that real changes were much slower than the noise in the data for most parts of their data set. The problem with FIGAERO data is that the changes in signal due to temperature are fast, and it is difficult to separate the noise/error from these changes. However, most ion traces reach a steady state in the last minutes of the measurements as most of the material has evaporated already, and background signals are stable or changing slowly. Additionally, the temperature ramp reaches a plateau (“soak” period). Thus, we modified the method from Yan et al. (2016). The last 20 data points of each thermogram (400 sec,  $T_{desorp} = \sim 195 - 205$  °C) are considered to be in steady state. To account for still persisting trends in the data, a linear fit (LineFit) is performed to these points for each ion trace. The residual between the data and this fit,  $res_{ij}$ , is calculated for each ion (20 values per ion):

$$res_{ij} = data_{ij} - LineFit_{ij} \quad (S2)$$

Yan et al. (2016) showed that the analytical uncertainty in gas-phase CIMS measurements is independent of the  $m/z$  range of the instrument and the specific ion. Thus, the values of all ions can be combined for further analysis.  $\sigma_{noise}$  is calculated as the median of the standard deviation of  $res_{ij}$  for one ion (Figure S 8).

$$\sigma_{noise} = median(stdev(res_{ij})) \quad (S3)$$

These  $\sigma_{noise}$  values are used in the CNerror scheme (Eq. 8 in main manuscript). Note that the values are higher for ions which have a higher plateau value, i.e. which show higher desorption during the end of the soak period. As this is an indication for a stronger impact of background compounds from the instrument/filter, it seemed justified to apply this

higher error value for the whole thermogram of these ions, decreasing their weight in the PMF optimisation algorithm. To account for the minimum error needed in the PMF algorithm,  $\sigma_{noise}$  values smaller than the median of all  $\sigma_{noise}$  values ( $\sigma_{min} = 0.013$ ) were set to the  $\sigma_{min}$  value.

5 Similar to the method in Yan et al. (2016), the  $res_{ij}$  values are grouped by signal strength ( $X_{ij}$ ) into 12 classes. The intervals were chosen to have at least 30 values in each class and the lowest concentration class starts at  $\sim 1 \text{ ct s}^{-1}$  (i.e.  $X_{ij} > 60\%$  of max signal, 50%-60%, 40%-50%, 30%-40%, 25%-30%, 20%-25%, 15%-20%, 10%-15%, 5%-10%, 2%-5%, 1%-2%, 0-1%). Note, that this is the average signal strength for the last 400 sec of the thermogram. The peak values during desorption may be much higher. A histogram is calculated for each class and the data are fitted with a gaussian function (examples in Figure S 9b and c). The half-width of that function ( $\sigma_{fit}$ : “width” in plots) is considered to be the overall  
10 measurement error  $S_{ij}$ .

To gain the factor  $a$  for equation 1,  $\sigma_{fit}$  is plotted against the average signal intensity of each class ( $I_{class}$ ) and a power law is fitted:

$$S_{ij} = \sigma_{fit} = a' \cdot (I_{class})^c + b \quad S4$$

15 With  $a' = a/\sqrt{(t_s)}$ ,  $b = \sigma_{noise}$ , and  $c$  should be 0.5 (i.e. the square root for the case where the error is following a Poisson distribution). The 95<sup>th</sup> percentile of  $\sigma_{fit}$  is used as “one-sigma-standard-deviation” to weight the data points. Figure S 9 shows the combined  $S_{ij}$  data set for all thermograms for all samples (4\*12 points per SOA type). The parameters yielded by freely fitting Eq. S4 (black line in Figure S 9) and fitting with  $c = 0.5$  (grey line) are given in Table S3. Note that the value for  $c$  is  $>0.5$  in the unconstrained case. This may be caused by some random error on the data or the fact that the error does not follow a Poisson style distribution. Having very few points with good quality above a signal strength of  
20  $10 \text{ ct s}^{-1}$  may also contribute to not very well constraining the fitted function. To avoid assumptions about the exact shape of the error distribution, we decided to use the parameters of the unconstrained fit in Eq. S4 to calculate the error values labelled PLerror. In addition, we apply the same minimum error criterium as in the CNerror case.

As a test case not shown in the manuscript, PMF analysis was also conducted with error values calculated according to Eq. S1 with  $\sigma_{noise} = 0.013$  and  $a = 1$  or  $a = 1.28$  (value typically used for AMS data). The results were very similar to the  
25 ones with the PLerror scheme.

### 1.3 Drivers controlling the grouping of compounds into PMF factors

When analysing a FIGAERO-CIMS thermal desorption data set containing multiple samples with PMF, there are two driving forces for the grouping of compounds into factors: their volatility and their “source” in the atmosphere/chamber/OFR (biomass burning, oxidation of different precursors, day-time/night-time chemistry, etc.). To

investigate the competition between these two drivers, we create simple artificial data sets (Figure S 10). We combine 4 compounds A, B, C, and D (with nominal ion masses of 1, 2, 3, and 4) to form 4 SOA types (SOA1, SOA2, SOA3, SOA4). A, B, and C have the same volatility. A has the same thermogram in each SOA type.

In scenario X, we investigate as an extreme case the combination of SOA1 (containing A, B, and D) and SOA2 (containing A, C, and D). This scenario can be interpreted as one source/process creating A and B at the same time for SOA1, but than a different pathway created A together with C in the case of SOA2. The 3-factor solution PMF result for this scenario (SOA1 and SOA2) is shown in Figure S 11. The common compound A has exactly the same thermogram behaviour in both samples. But as it once correlates with B (in SOA1) and once with C (SOA2), A is explained with Factor 1 (black) for the SOA1 sample and with Factor 2 (red) for the SOA 2 sample. Even increasing the number of factors does not create a “common” factor which contains only A. For this scenario, the different “source” for A (which lead to different compounds correlating with it) dominates the factor identification and not the fact that A, B, and C have the same volatility. Note that compound D which also does not change between SOA1 and SOA2 is separated into its own factor. Here the difference in volatility (from A, B, and C) is the driving force for the factor grouping.

In scenario Z, SOA3 and SOA4 each contain all 4 compounds. The concentration of A is the same in both types while the contribution of C is higher in SOA4 than in SOA3 and that of B is lower. The 3-factor solution for this scenario is depicted in Figure S 12. Again, we find two factors explaining the behaviour of the three compounds with the same volatility. But now factor 2 (red, dominated by A and C) has a considerable contribution to both SOA types. Figure S 13 shows how the thermograms of A are explained by the two factors. Similar to scenario X, we can interpret the factor grouping by changes in the processes/sources producing the compound A, B, and C. Factor 1 (black) stands for process 1 creating the majority of C and some A. Process 2 creates mostly A and B and is explained by factor 2. Again, factor 3 (containing D) is differentiated by the different volatility of D. Integrating the factor thermograms shows that the more volatile fraction of SOA4 is formed by process 2 while that the fraction of SOA3 with the same volatility is formed by both processes 1 and 2 (Figure S 14).

This Scenario Z is very similar to comparing, e.g., samples of low- and medium-O : C SOA. Many ions occur in both SOA types, but the ratios between them change. Many products of similar volatility can be formed via different pathways in the oxidation of  $\alpha$ -pinene, but depending on the reaction path, they will correlate with different other products. The change in the oxidation field (increase of  $[O_3]$  and  $[OH]$ ) probably affected the  $HO_2/RO_2$  chemistry which has a strong influence on, e.g., highly oxygenated material (HOM) and dimer formation, i.e. we changed the ratio between reaction paths. Note that compounds of significantly different volatility will be grouped into different factors even if they were produced by the same process/pathway (see behaviour of compound D). When investigating ambient aerosol samples, this behaviour may impact the identification of SOA sources and/or types. One SOA source/type that would be identified

as a factor in a PMF analysis of FIGAERO-CIMS data integrated over each thermogram scan, may be split into multiple factors when analysing the thermal desorption data with PMF. This needs to be investigated in future studies comparing results for the two methods with ambient data sets and more extensive chamber and/or flow tube measurements.

#### 5 1.4 PMF analysis of a data set combining different SOA types

To study if the different SOA types are really as different as the factors identified in section 3.1 suggest or if this was artificially introduced by pre-grouping the data, we conduct a PMF analysis with the data grouped by sampling conditions. The dry,  $t = 4$  h set was chosen for detailed analysis as we did not want to introduce the added complication of aqueous phase chemistry and the dry low-O : C sample at  $t = 0.25$  h had a large contamination unique to that sample. In the following, we will refer to this as the “combined data set”. The analysis conducted with data pre-grouped by the SOA type will be labelled “pre-grouped”.

The 8- and 13 factor solutions for the combined data set are shown in Figure S 15 and Figure S 16. Based on the change in  $Ratio_{exp}$  values and the residual time series, 8 was the minimum number of factors need to capture the thermograms of all 3 SOA types equally well. But to reach residuals as low as in the pre-grouped data sets, 13 factors were needed. Especially the lower  $T_{desorp}$  regions in the low- and high-O : C case and the high  $T_{desorp}$  part of the medium-O : C sample improved (Figure S 17). Note that with the same criteria applied in the section 2.3.3, we would not select the 8-factor solution as a “best” solution.

There are no factors in the 8-factor solution which are unique to one SOA type. 4 factors (FV1, FV4, FV7, and FB/D1) occur in all SOA types explaining 80% of the signal of the medium-O : C sample and 50% of the low- and high-O : C samples. The other 50% of the signal are explained with two factors each (low O : C - FV2 and FV5, high O : C - FV3 and FV7) which also occur in the medium-O : C sample.

The 8-solution suggests very strong similarities between the three SOA types with a gradual shift in composition with increasing oxidation. But this solution does not capture the detailed thermogram behaviour of single ions very well as the high residuals suggest (Figure S 18 for the example ion also discussed in sections 3.2 and 3.3). When this is improved in the 13-factors solution, the degree of similarity between the SOA types decreases. Here, 20% - 25% of the signal in each SOA type is explained by factors common to all SOA types (FD1, FB1, and FV1). There are some factors with significant contribution to two SOA types (e.g. FV2 and FV10 for low- and medium-O : C cases; or FV3, FV6 and FV11 for medium- and high - O : C cases).

To compare the factor mass spectra derived with the combined and the pre-grouped data sets, we use the spectral contrast angle,  $\theta$ .  $\theta$  is derived from the dot product of two mass spectra (Wan et al., 2002):

$$\cos\theta = \frac{\sum_i a_i b_i}{\sqrt{\sum_i a_i^2 \cdot \sum_i b_i^2}}$$

where  $a_i$  and  $b_i$  are the intensities of ion  $i$  in mass spectrum 1 and mass spectrum 2. Two mass spectra are considered to be similar if  $\theta$  is between  $0^\circ$  and  $15^\circ$ , somewhat similar but with important differences if  $\theta$  is between  $15^\circ$  and  $30^\circ$ , and different with  $\theta$  values  $>30^\circ$  (Bougiatioti et al., 2014).

5 The results from the pairwise comparison of each factor identified in the pre-grouped data sets with all factors from the combined data set are shown in Figure S 19. All type-V factors identified in the pre-grouped data set have a (at least somewhat) similar counterpart in the combined data set (e.g. LV4 and FV7, MV4 and FV8, HV2 and FV6). This shows that the missing similarities between factors identified in the pre-grouped data set are not artificially induced by the pre-grouping but rather stem from the shifts in SOA composition with increasing oxidation. The changes in the groups of correlating ions will cause compounds that occur in all SOA types to be grouped into different factors as explained with the simplified data set in SI section 1.3. The factors FV2, FV3, and FV4 show similarities to two factors within the corresponding SOA types. This suggests that either more factors are needed in the combined data set to resolve the thermogram behaviour of the compounds represented by these factors, or that there was “factor-splitting” in the pre-grouped data set (i.e. too many factors). Also, the combined data set uses only the information from the dry ( $t_{\text{evap}} = 4$  h) sample while the pre-grouped ones contain all 4 sample types. Thus, behaviour unique to a different sample type cannot be correctly captured (e.g. LC1 and 2, or HV3).

This case study shows that the same overall conclusions can be drawn by using pre-grouping of the data according to the information of the sampling conditions (here the SOA type) or the combined data set. With the combined data set, a higher number of factors (here at least 13) has to be chosen to cover all details in the data set equally well. For ambient data, a combined data set approach is favourable as limited information are available for any pre-grouping, and such an extra step is not desirable. For a detailed study of, e.g., particle phase processes with designed SOA types (as presented in this paper), a pre-grouping can be beneficial to highlight the fine details hidden in the data set.

## 1.5 SI References

25 Buchholz, A., Lambe, A. T., Ylisirniö, A., Li, Z., Tikkanen, O. P., Faiola, C., Kari, E., Hao, L., Luoma, O., Huang, W., Mohr, C., Worsnop, D. R., Nizkorodov, S. A., Yli-Juuti, T., Schobesberger, S. and Virtanen, A.: Insights into the O: C-dependent mechanisms controlling the evaporation of  $\alpha$ -pinene secondary organic aerosol particles, *Atmos. Chem. Phys.*, 19(6), 4061–4073, doi:10.5194/acp-19-4061-2019, 2019.

Decarlo, P. F., Kimmel, J. R., Trimborn, A., Northway, M. J., Jayne, J. T., Aiken, A. C., Gonin, M., Fuhrer, K., Horvath,

- T., Docherty, K. S., Worsnop, D. R. and Jimenez, J. L.: Field-deployable, high-resolution, time-of-flight aerosol mass spectrometer, *Anal. Chem.*, 78(24), 8281–8289, doi:10.1021/ac061249n, 2006.
- Donahue, N. M., Robinson, A. L., Stanier, C. O. and Pandis, S. N.: Coupled partitioning, dilution, and chemical aging of semivolatile organics, *Environ. Sci. Technol.*, 40(8), 2635–2643, doi:10.1021/es052297c, 2006.
- 5 Kang, E., Root, M. J., Toohey, D. W. and Brune, W. H.: Introducing the concept of Potential Aerosol Mass (PAM), *Atmos. Chem. Phys.*, 7(22), 5727–5744, doi:10.5194/acp-7-5727-2007, 2007.
- Lambe, A. T., Ahern, A. T., Williams, L. R., Slowik, J. G., Wong, J. P. S., Abbatt, J. P. D., Brune, W. H., Ng, N. L., Wright, J. P., Croasdale, D. R., Worsnop, D. R., Davidovits, P. and Onasch, T. B.: Characterization of aerosol photooxidation flow reactors: heterogeneous oxidation, secondary organic aerosol formation and cloud condensation nuclei activity measurements, *Atmos. Meas. Tech.*, 4, 445–461, doi:10.5194/amt-4-445-2011, 2011.
- 10 Lee, B. H., Lopez-Hilfiker, F. D., Mohr, C., Kurtén, T., Worsnop, D. R. and Thornton, J. A.: An iodide-adduct high-resolution time-of-flight chemical-ionization mass spectrometer: Application to atmospheric inorganic and organic compounds, *Environ. Sci. Technol.*, 48(11), 6309–6317, doi:10.1021/es500362a, 2014.
- Lopez-Hilfiker, F. D., Mohr, C., Ehn, M., Rubach, F., Kleist, E., Wildt, J., Mentel, T. F., Lutz, A., Hallquist, M., Worsnop, D. and Thornton, J. A.: A novel method for online analysis of gas and particle composition: description and evaluation of a Filter Inlet for Gases and AEROsols (FIGAERO), *Atmos. Meas. Tech.*, 7, 983–1001, doi:10.5194/amt-7-983-2014, 2014.
- 15 Yan, C., Nie, W., Äijälä, M., Rissanen, M. P., Canagaratna, M. R., Massoli, P., Junninen, H., Jokinen, T., Sarnela, N., Häme, S. A. K., Schobesberger, S., Canonaco, F., Yao, L., Prévôt, A. S. H., Petäjä, T., Kulmala, M., Sipilä, M., Worsnop, D. R. and Ehn, M.: Source characterization of highly oxidized multifunctional compounds in a boreal forest environment using positive matrix factorization, *Atmos. Chem. Phys.*, 16(19), 12715–12731, doi:10.5194/acp-16-12715-2016, 2016.
- 20 Yli-Juuti, T., Pajunoja, A., Tikkanen, O. P., Buchholz, A., Faiola, C., Väisänen, O., Hao, L., Kari, E., Peräkylä, O., Garmash, O., Shiraiwa, M., Ehn, M., Lehtinen, K. and Virtanen, A.: Factors controlling the evaporation of secondary organic aerosol from a-pinene ozonolysis, *Geophys. Res. Lett.*, 44(5), 2562–2570, doi:10.1002/2016GL072364, 2017.

## 1.6 SI Tables

**Table S 1:** Mass concentration in RTC after filling and estimated collected sample mass on FIGAERO filter.

OH exposure	condition	mass conc in RTC after filling / $\mu\text{g m}^{-3}$	sampled mass / ng	
			$t_{\text{evap}}=0.25\text{h}$	$t_{\text{evap}}=4\text{ h}$
low	dry	0.67	178	33
	RH80%	0.52	186	22
medium	dry	1.0	239	72
	RH80%	1.1	258	50
high	dry	0.69	138	46
	RH80%	0.98	172	30

**Table S 2:** Overview on oxidation condition and oxidative state of formed SOA.

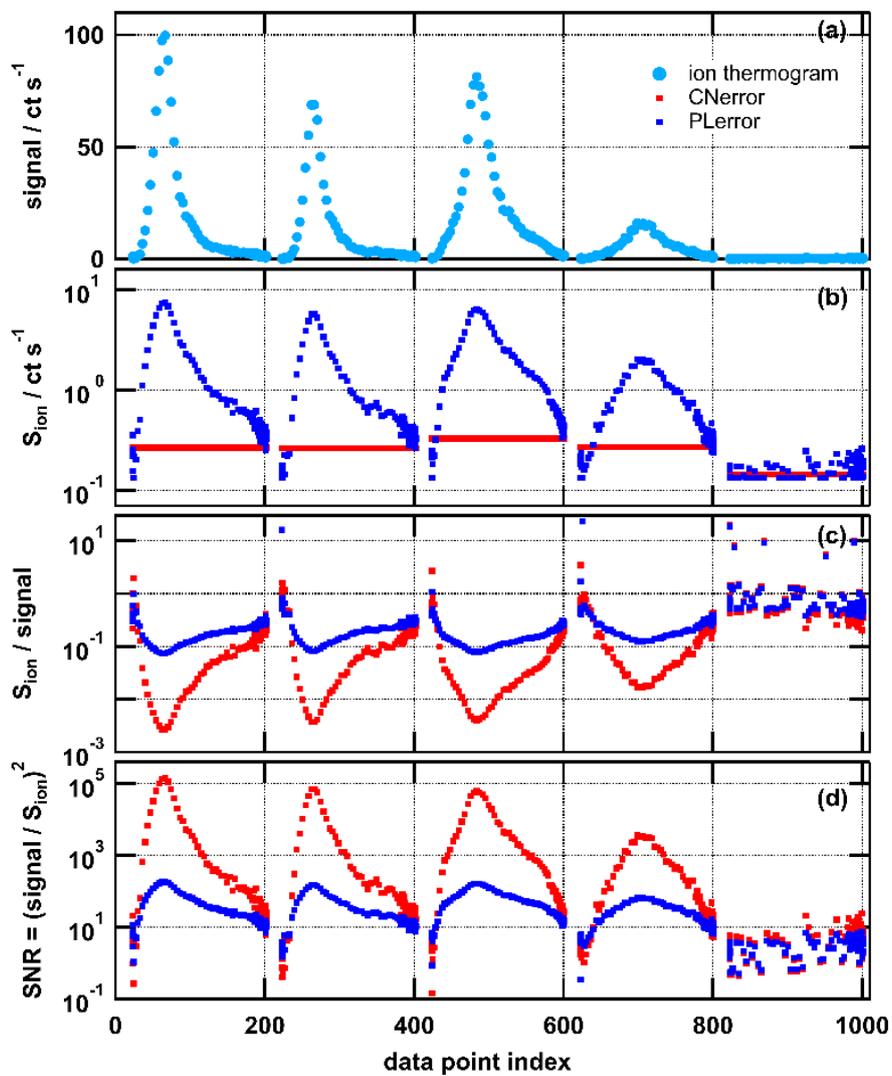
	low O : C	medium O : C	high O : C
$[\text{O}_3]_{\text{inlet}} / \text{ppm}$	6.6	25	25
$[\text{O}_3]_{\text{outlet}} / \text{ppm}$	6.4	22.2	16
OH exposure / $\text{cm}^{-3} \text{ s}$	2.54e11	6.85e11	2.45e12
photochemical age / days	2.0	5.3	18.9
O : C(AMS)	0.53	0.69	0.96
O : C(FIGAERO-CIMS)	0.66	0.75	0.84

5

**Table S 3:** Fitting parameters for fitting Eq. S4 to the  $S_{ij}$  data presented in Figure S 9.

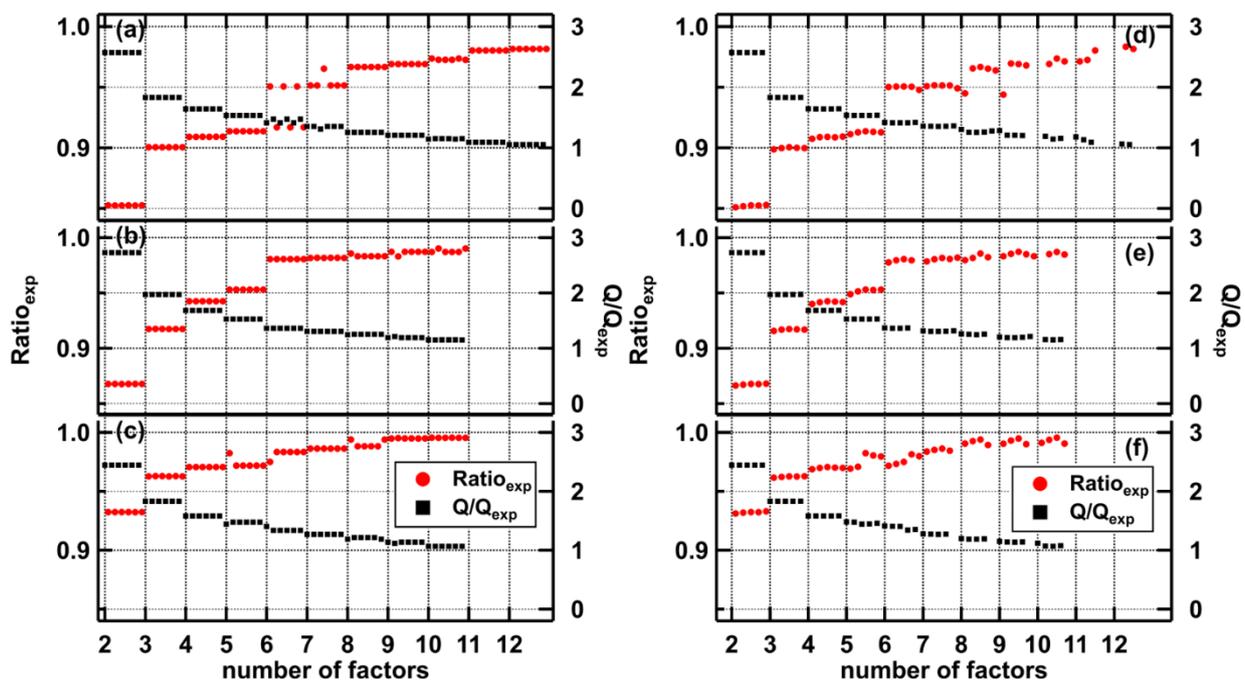
	<b>a'</b>	<b>b</b>	<b>c</b>
free fit	$0.260 \pm 0.003$	$0.056 \pm 0.002$	$0.726 \pm 0.009$
square root fit	$0.303 \pm 0.002$	$0.009 \pm 0.001$	0.5

## 1.7 SI Figures

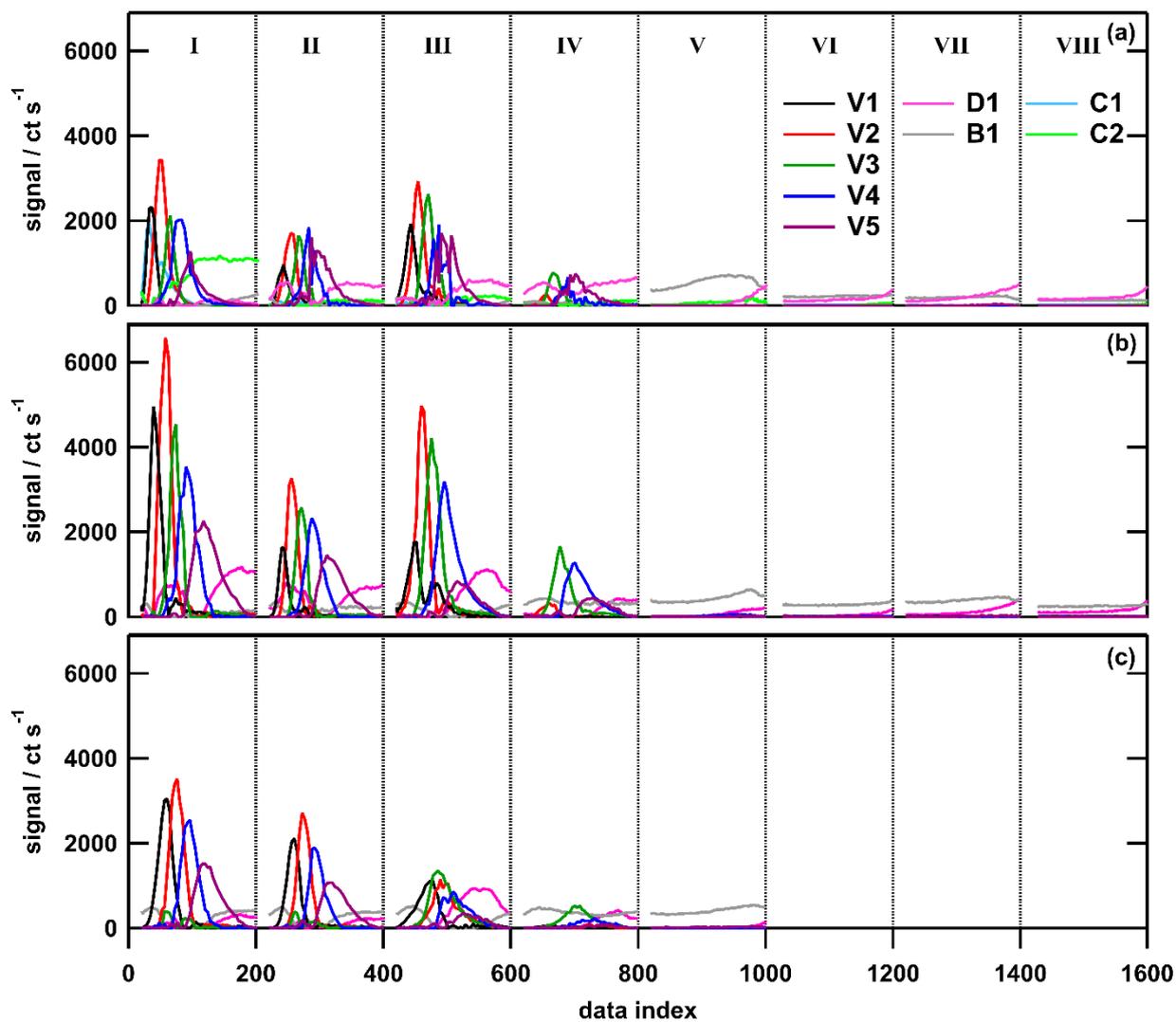


**Figure S 1:** Single ion thermogram (a) and time series of errors ( $S_{\text{ion}}$ , b), relative error ( $S_{\text{ion}} / \text{signal}$ , c), and signal-to-noise-ratio (SNR, d) for  $\text{C}_5\text{H}_5\text{O}_6^-$ . Values calculated for CError are marked in red and those for PError in dark blue. Note that panels b, c, and d have logarithmic y axis scaling.

5

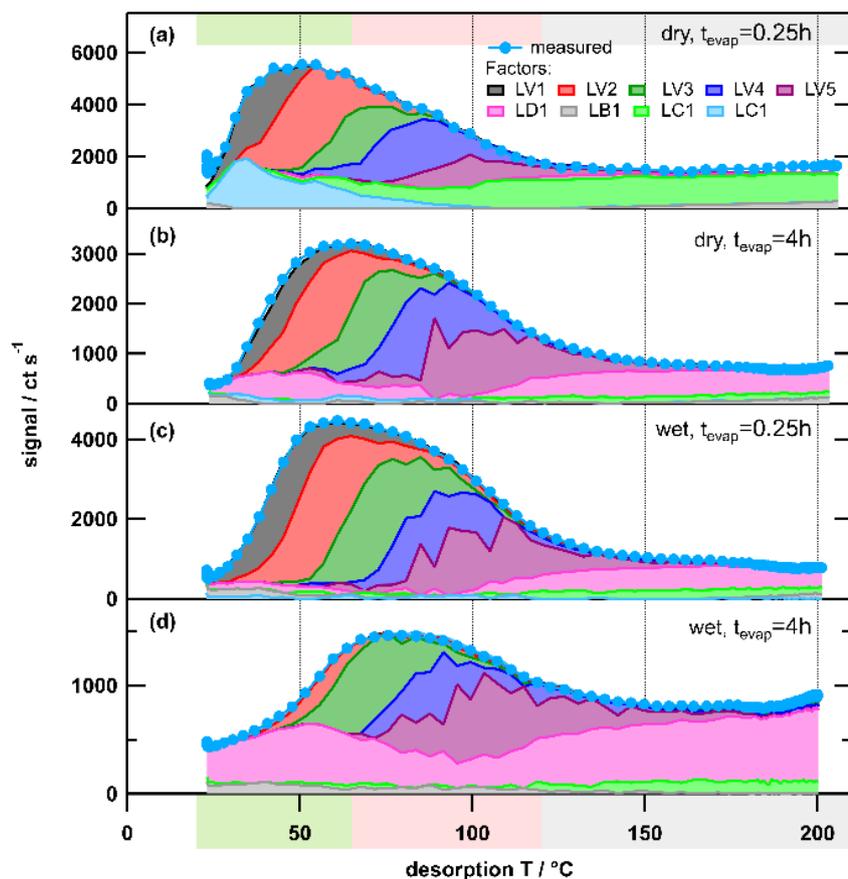


**Figure S 2:** Fraction of explained variance ( $Ratio_{exp}$ , red, left axis) and  $Q/Q_{exp}$  values (black, right axis) for all PMF solutions with seed values varying from 1 to 6 (a, b, c) and  $f_{peak}$  values varying from -1.0 to +1.0 (d, e, f) for all SOA types (low (a&d), medium (b&e), and high O : C (c&f)).



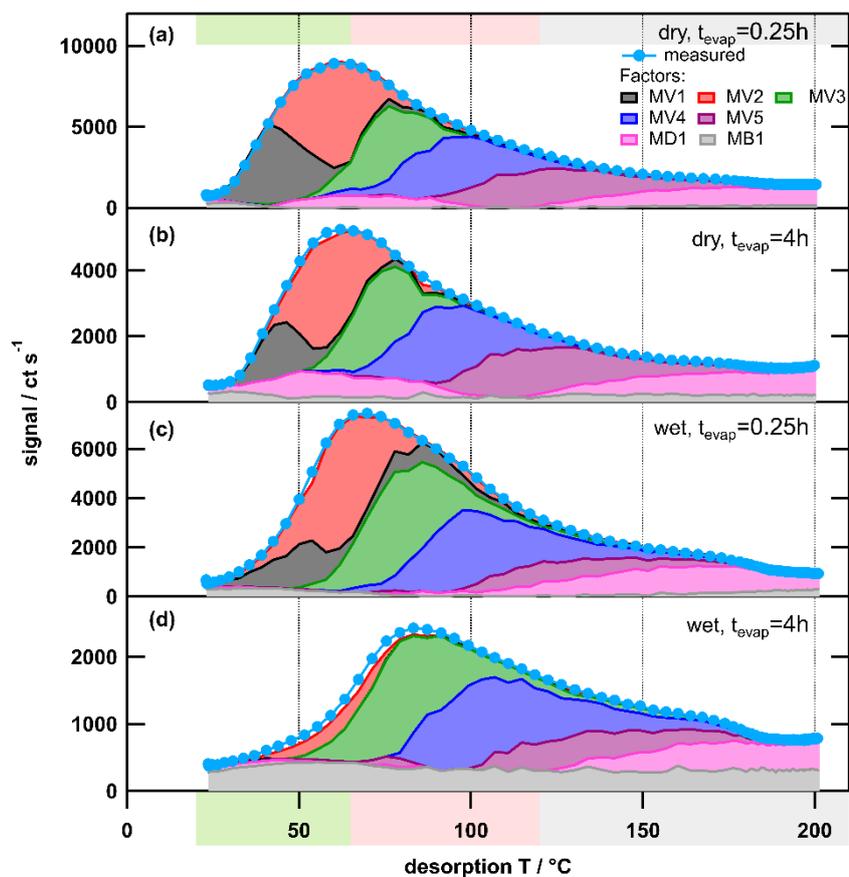
**Figure S 3:** Full data set of factor thermograms for samples (I - IV) and 1 or 4 filter blank measurements (V - VIII) for low- (a), medium- (b) and high-O : C (c) cases plotted against the data index. Note that factors are different between SOA types (i.e. factor V1 in panel a and b have different factor mass spectra, see Figure 5 and Figure 6 in main text). Colour code for the factors is the same as in main text figures.

5



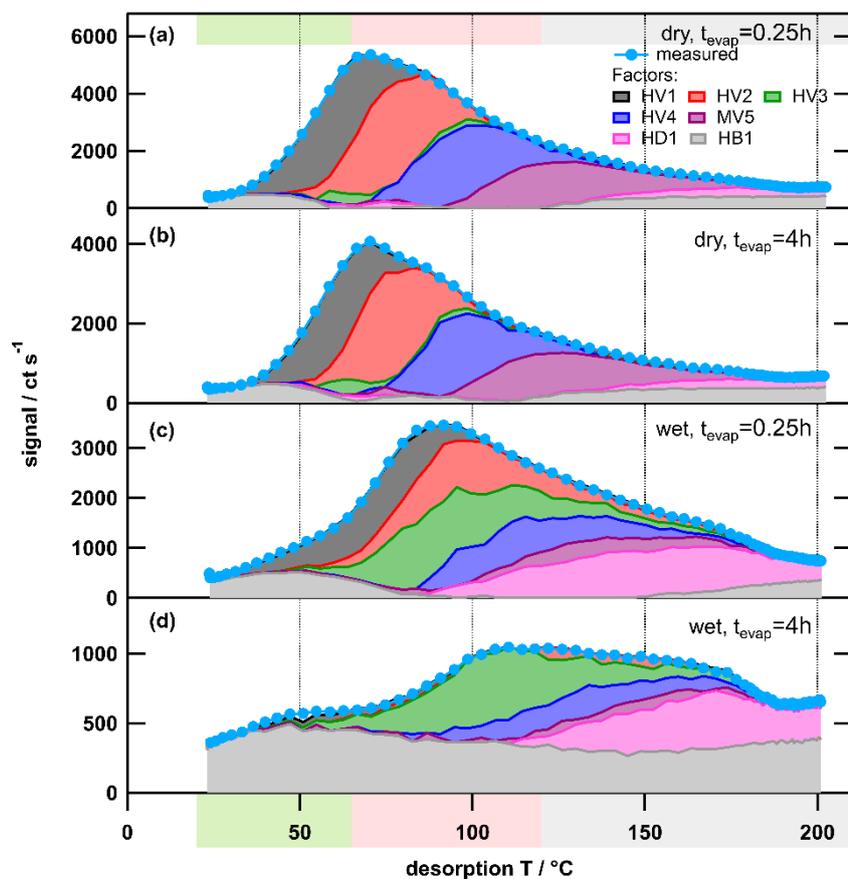
**Figure S 4:** Factor thermograms for the 9-factor solution for the low-O : C case same as in main manuscript Figure 5 but stacked to highlight reconstruction of measured total thermogram signal. The colour code is the same for all panels. Background colours indicate volatility classifications according to Donahue et al. (2006) derived from  $T_{\max}-C_{sat}^*$  calibrations (green: SVOC, red: LVOC, grey: ELVOC). Note the different scaling for y-axes in panels a-d.

5



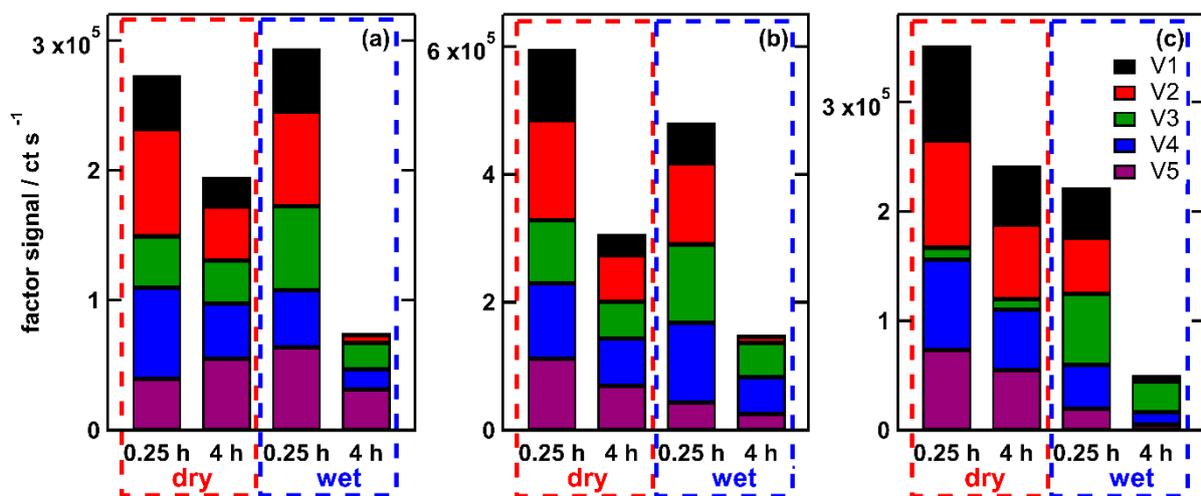
**Figure S 5:** Factor thermograms for the 7-factor solution for the medium O : C case same as in main manuscript Figure 6 but stacked to highlight reconstruction of measured total thermogram signal. The colour code is the same for all panels. Background colours indicate volatility classifications according to Donahue et al. (2006) derived from  $T_{\max}\text{-}C_{sat}^*$  calibrations (green: SVOC, red: LVOC, grey: ELVOC). Note the different scaling for y-axes in panels a-d.

5

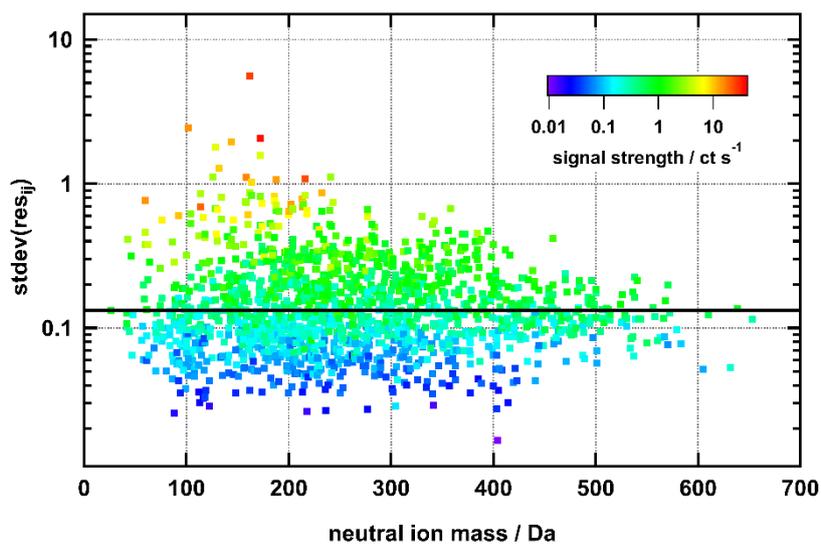


**Figure S 6** Factor thermograms for the 7-factor solution for the high-O : C case same as in main manuscript Figure 7 but stacked to highlight reconstruction of measured total thermogram signal. The colour code is the same for both panels. Background colours indicate volatility classifications according to Donahue et al. (2006) derived from  $T_{\max}-C_{sat}^*$  calibrations (green: SVOC, red: LVOC, grey: ELVOC). Note the different scaling for y-axes in panels a-d.

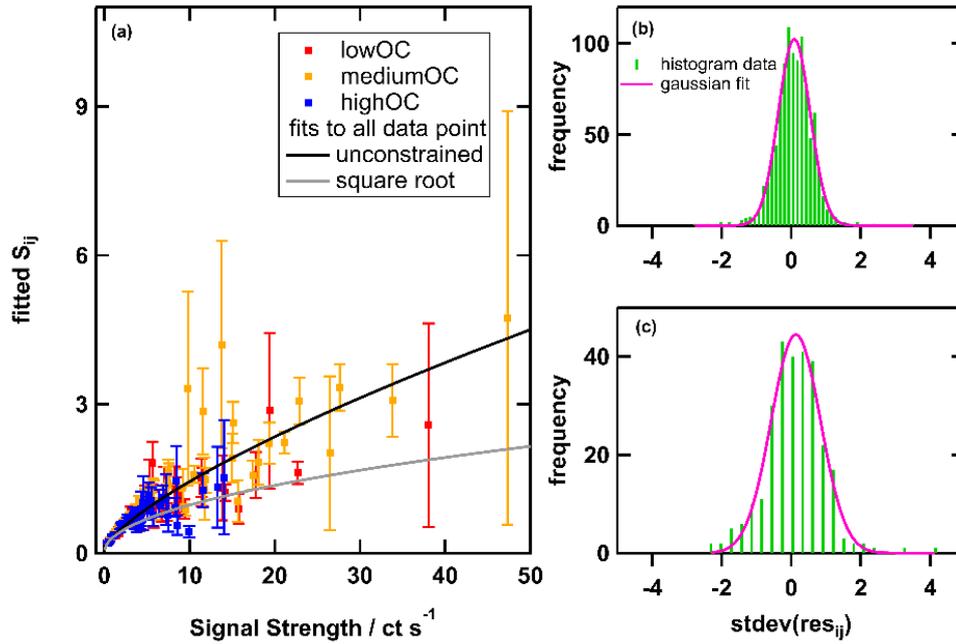
5



**Figure S 7:** Absolute contribution of type-V factors to the measured signal for low- (a), medium- (b), and high-O : C cases (c). Factors are sorted by their  $T_{\max}$  values from V1 to V5. Colour code of the factors is the same as in main text.

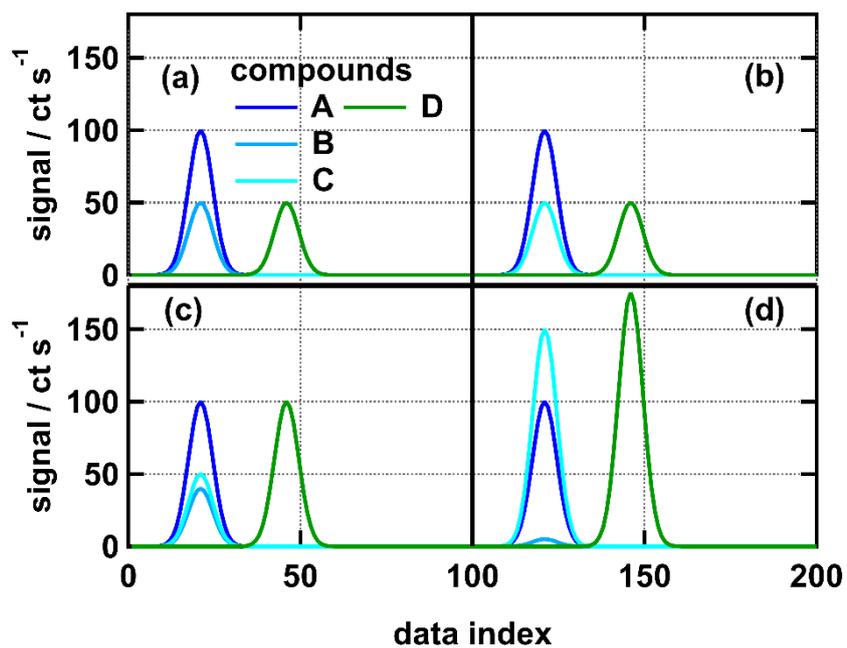


5 **Figure S 8:** Standard deviation of residual for one thermogram (one point for each ion). Colour code is the average signal strength in the last 400 sec. Horizontal line indicates the median of the values which is defined as  $\sigma_{\text{noise}}$ .

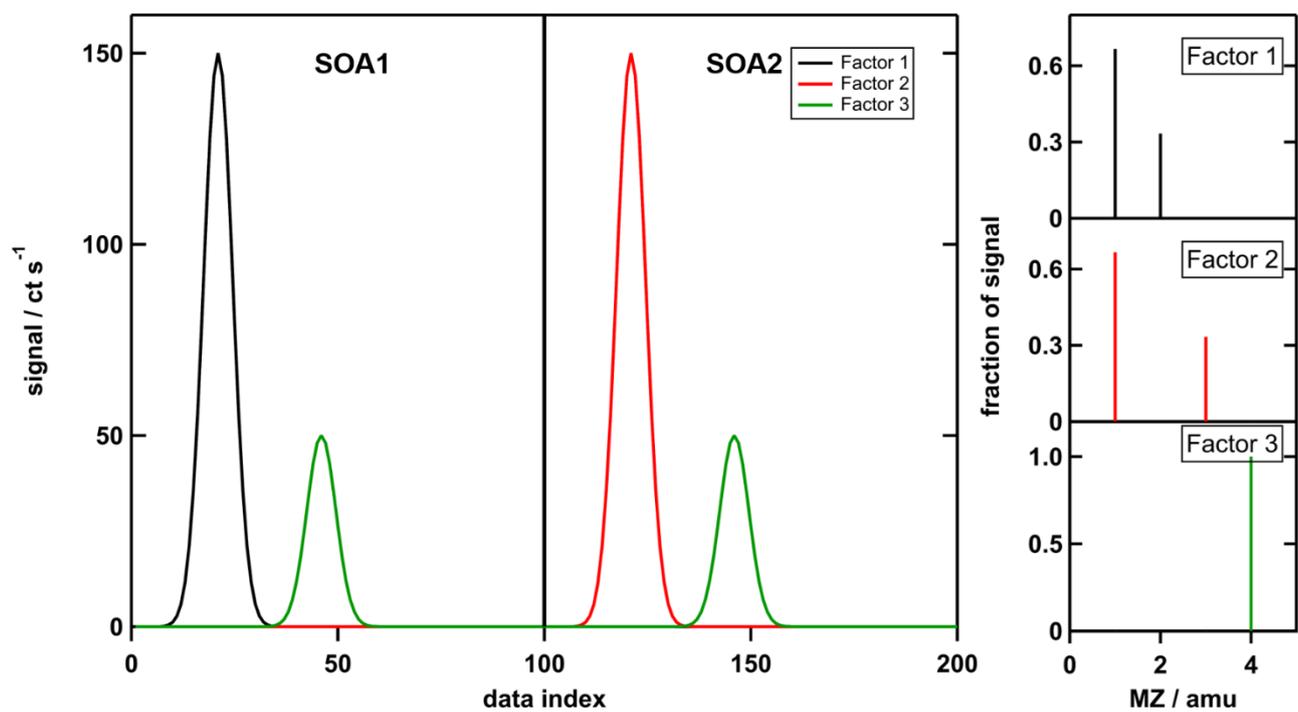


**Figure S 9:** (a)  $S_{ij}$  from gaussian fits to histograms plotted against the average signal strength for all thermograms for all O : C cases. Symbol colour stands for low-, medium-, and high-O : C cases. (b) and (c) histograms and gaussian fit for 2 of the 12 classes defined for calculation of  $S_{ij}$ . The fitted  $S_{ij}$  values in panel (a) are the halfwidth of these gaussian fits and the error bars are the 95<sup>th</sup> percentile ranges.

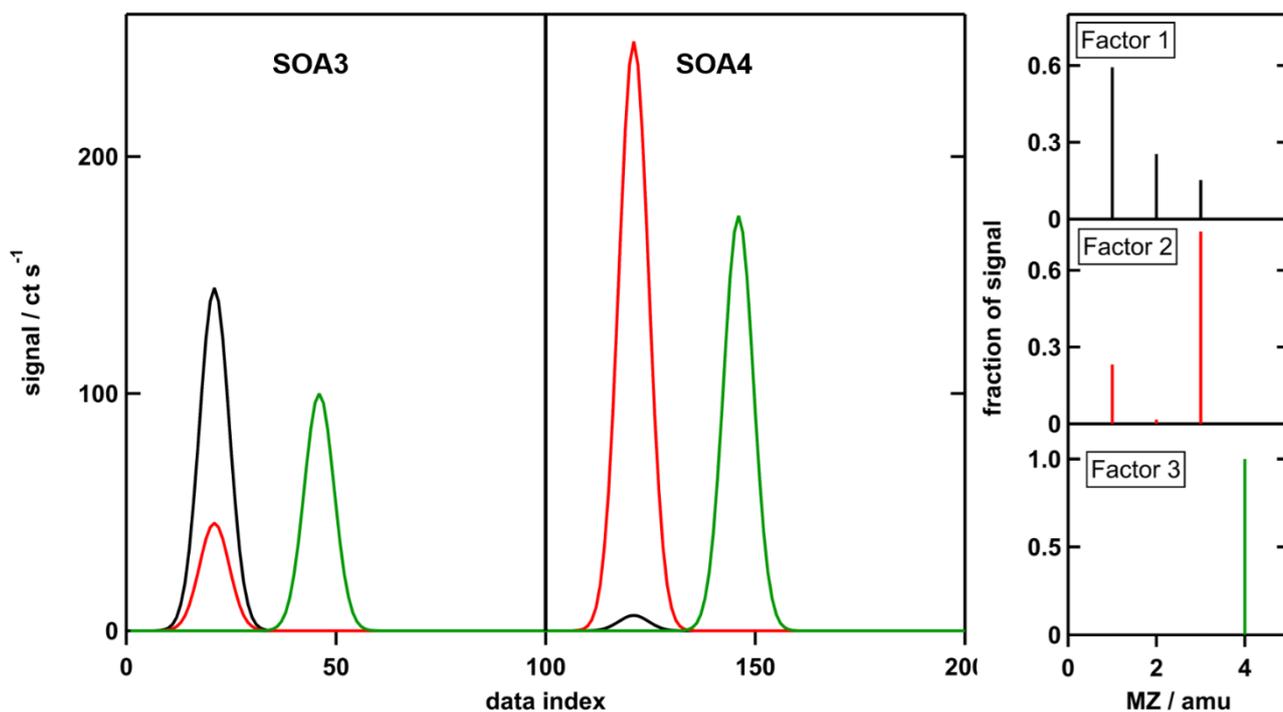
5



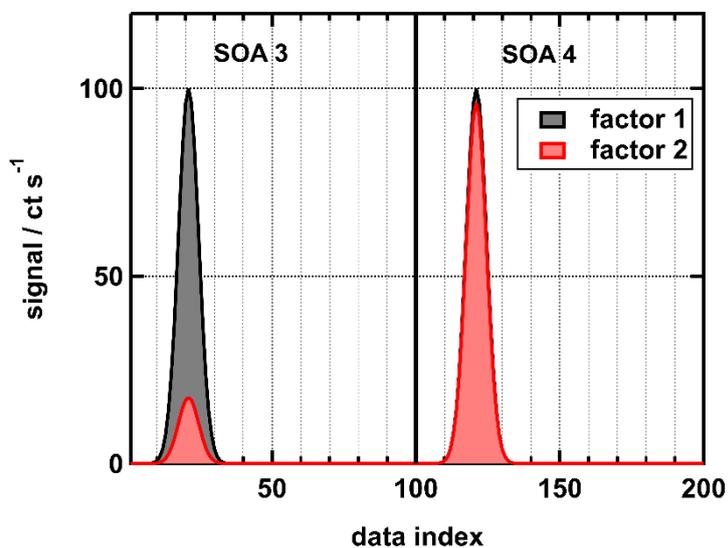
**Figure S 10:** Artificial thermogram data for four SOA types. (a) SOA1, (b) SOA2, (c) SOA3, (d) SOA4. SOA1 and SOA2 are combined in one data set for PMF analysis and so are SOA3 and SOA4. Note that the thermograms are plotted against data index. Compounds A, B, and C have the same  $T_{\max}$  values in all SOA types.



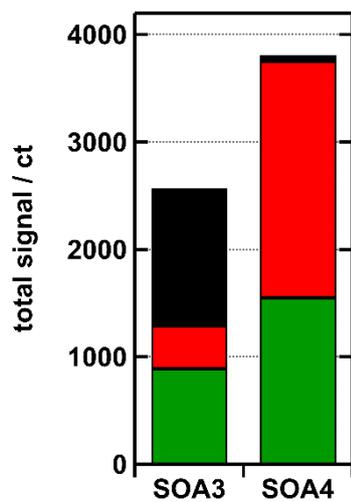
**Figure S 11:** Factor thermogram (left) and factor mass spectra (right) for the 3-factor solution for scenario X. For plotting, the nominal MZ values 1, 2, 3, and 4 are assigned to the compounds A, B, C, and D, respectively.



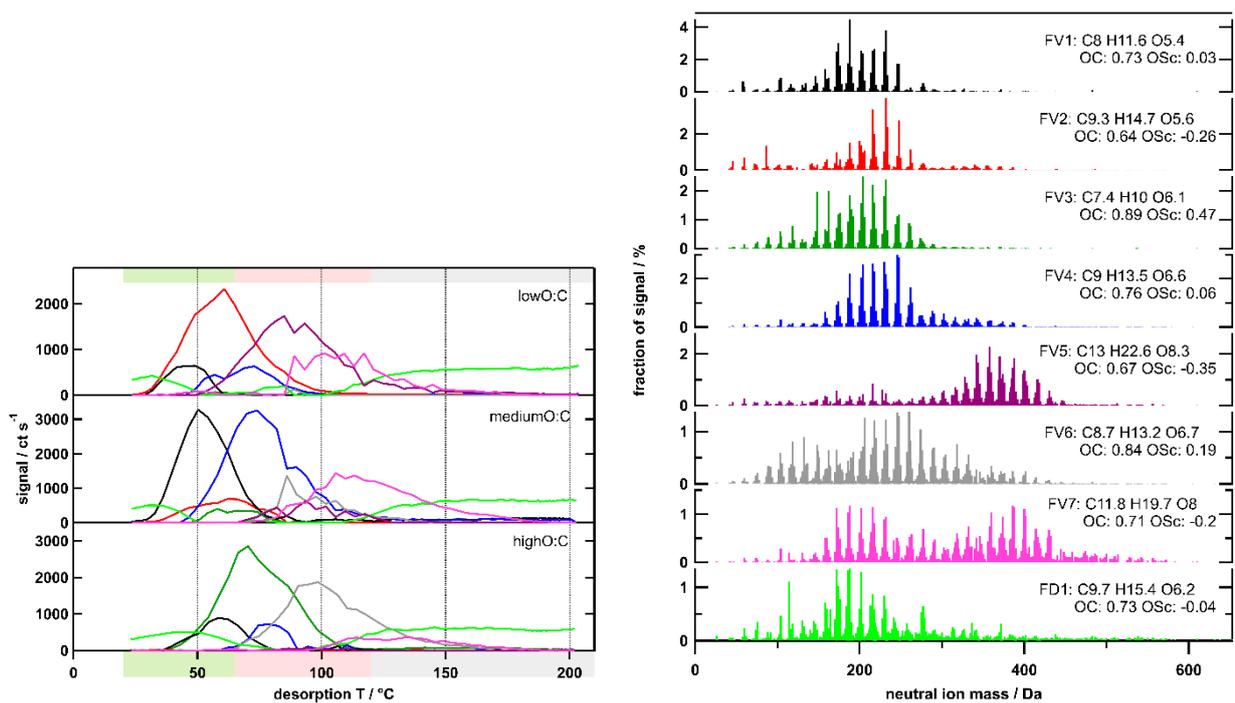
**Figure S 12:** Factor thermogram (left) and factor mass spectra (right) for the 3-factor solution for scenario Z. For plotting, the nominal MZ values 1, 2, 3, and 4 are assigned to the compounds A, B, C, and D, respectively.



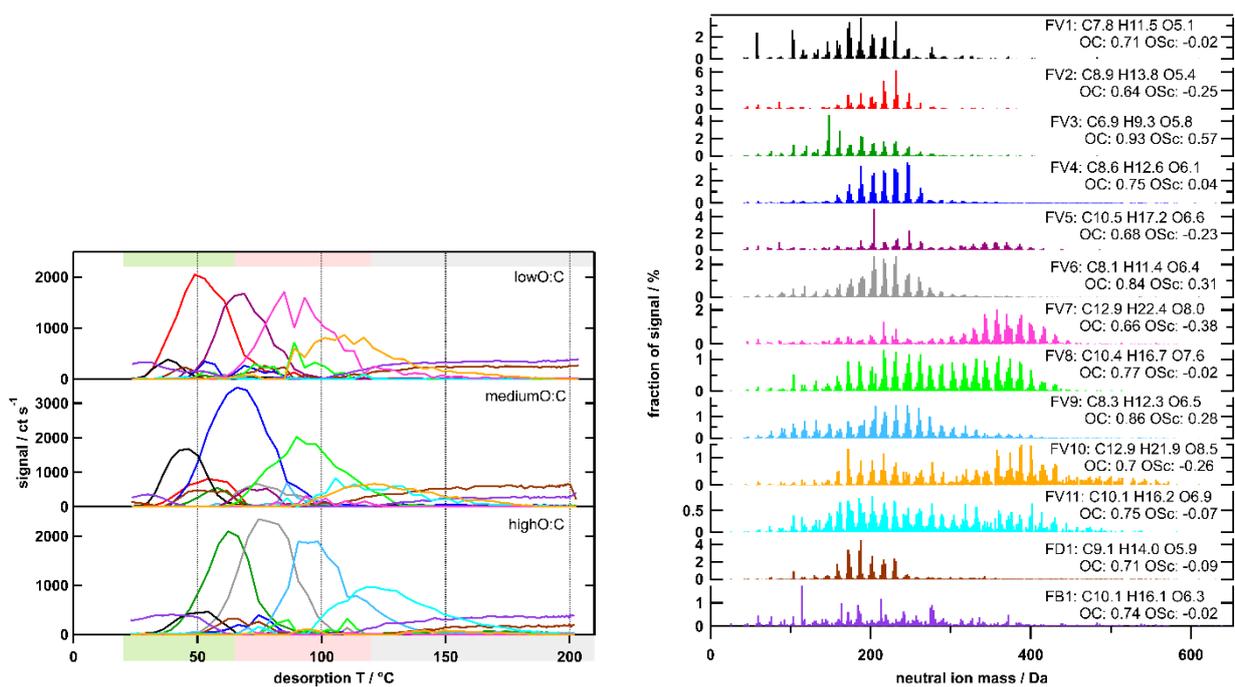
5 **Figure S 13:** Stacked factor thermograms for compound A in scenario Z.



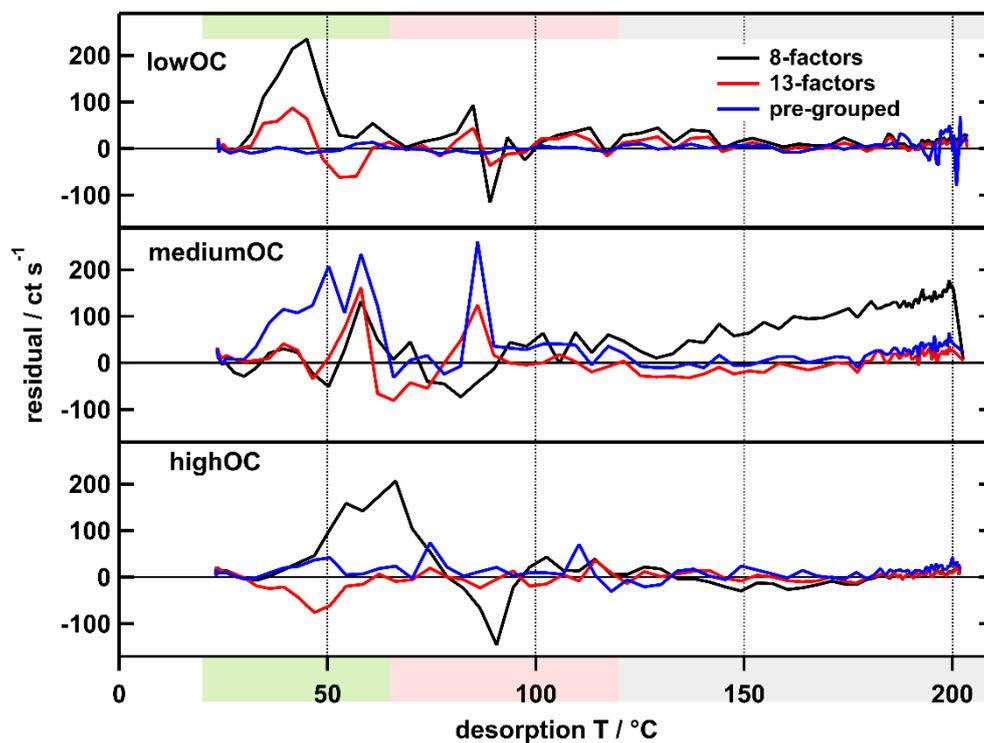
**Figure S 14:** Absolute contribution of factors to the signal of SOA3 and SOA4 in the scenario Z. Black: factor 1, red: factor 2, green: factor 3.



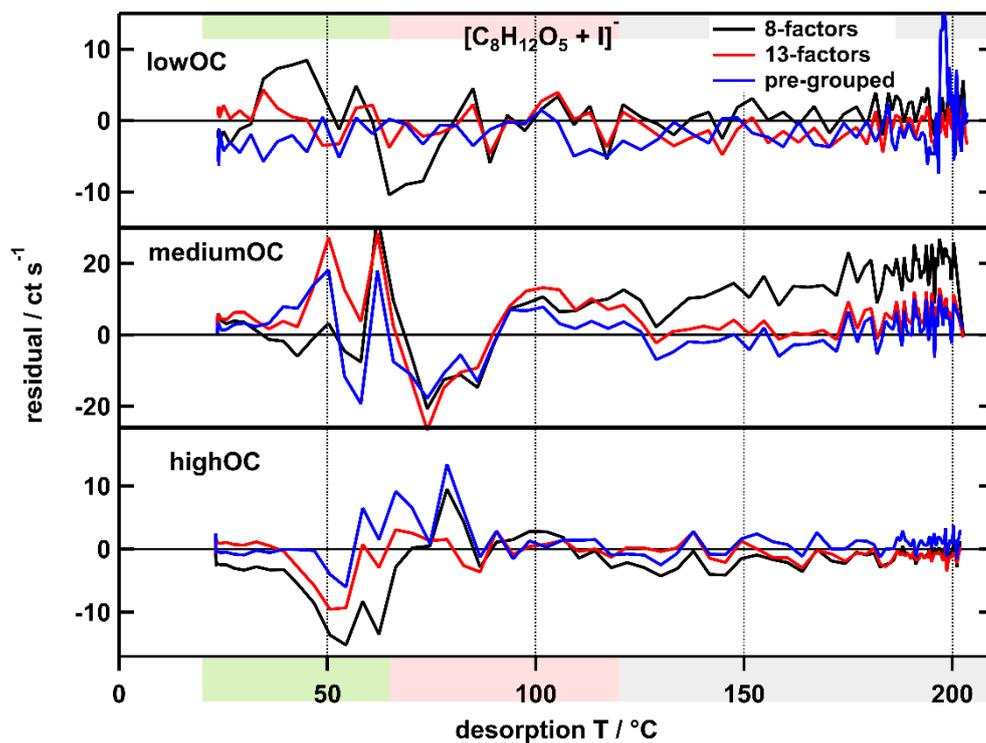
**Figure S 15:** Factor thermograms (left) and factor mass spectra (right) for the 8-factor solution for the combined data set for dry,  $t = 4$  h samples. Each factor mass spectrum is normalised. The colour code is the same for both panels. Background colour in the left panel indicates volatility classification derived from  $T_{\max}-C_{sat}^+$  calibrations (green: SVOC, red: LVOC, grey: ELVOC).



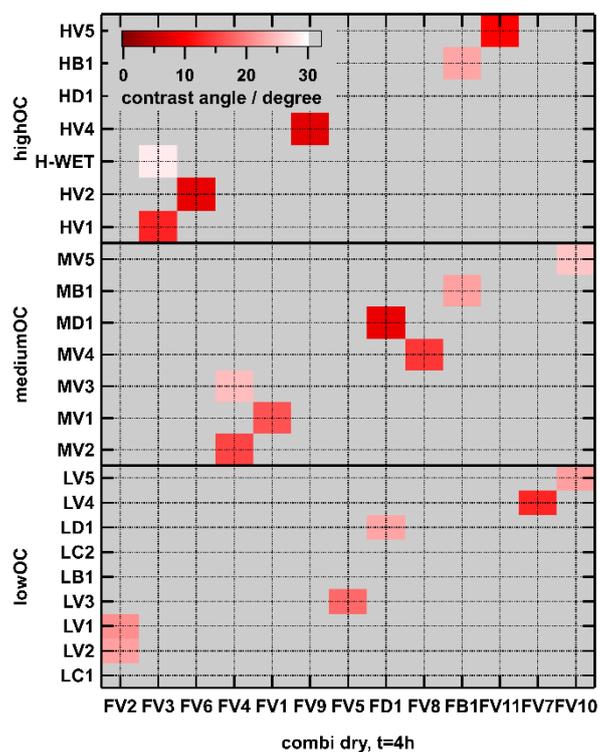
**Figure S 16:** Factor thermograms (left) and factor mass spectra (right) for the 13-factor solution for the combined data set for dry ( $t = 4$  h) samples. Each factor mass spectrum is normalised. The colour code is the same for both panels. Background colour in the left panel indicates volatility classification derived from  $T_{\max}$ - $C_{sat}^*$  calibrations (green: SVOC, red: LVOC, grey: ELVOC).



**Figure S 17:** Time series of residuals for the 8-(black) and 13-factor (red) solutions for the combined data set and the corresponding pre-grouped data sets (blue).



**Figure S 18:** Time series of residuals of the ion  $[C_8H_{12}O_5 + I]^-$  for the 8-(black) and 13-factor (red) solutions for the combined data set and the corresponding pre-grouped data sets (blue).



**Figure S 19:** Contrast angle plot comparing the factor mass spectra from the separate PMF analysis of each SOA type (y-axis) with those from the combined analysis with 13 factors (x-axis). Grey areas indicate no similarity (contrast angle > 30°) while shapes of red indicate decreasing degree of similarity from dark to light.