

Supplemental material to
**“Physicochemical properties and origin of
organic groups detected in boreal forest using
an aerosol mass spectrometer”**

by T. Raatikainen et al.

1.2.2010

1 Non-standard AMS adjustments

There were three non-standard instrument adjustments during the campaign. Oven temperature was increased twice (5.4.2005 and 11.4.2005) and the aerodynamic lens was adjusted on 7.4.2005. Usually oven temperature is kept constant as it is known to have an effect on mass spectra. The lens adjustment, which checks particle beam alignment on the oven, indicated that the lens was misaligned for the first 7 days of the measurements. Oven temperature and lens adjustments as well as their effects are shown in Fig. S1. Oven temperature is shown as a black continuous line on the right hand side axis. The first oven temperature increase had a clear effect on potassium signal (m/z 39, e.g. K^+) and the second increase had an effect on signal noise levels (m/z 216) and also some new peaks appeared (m/z 23, e.g. Na^+). The dashed vertical line represents the lens adjustment. It seems that, in comparison with total particle volume from DMPS, the AMS signal was doubled by the improved alignment. It must be kept in mind that even if absolute mass concentrations may be inaccurate, mass fractions are largely unaffected by these adjustments. Increased noise and changes in some peaks are somewhat more problematic in the matrix factorization analysis.

Both organics (e.g. $C_3H_5^+$) and potassium (e.g. $^{41}K^+$) have an effect on m/z 41 signal. Commonly potassium (m/z 39) signal is low and its contribution of m/z 41 (^{41}K is about 7 % of ^{39}K) signal is even lower. When the oven temperature was increased, a huge increase was seen in both potassium signals. As a result, it became clear that organic contribution could not be separated from the m/z 41 signal. Therefore, organic m/z 41 signal was calculated from that of organic m/z 43 signal by using a factor of 0.8.

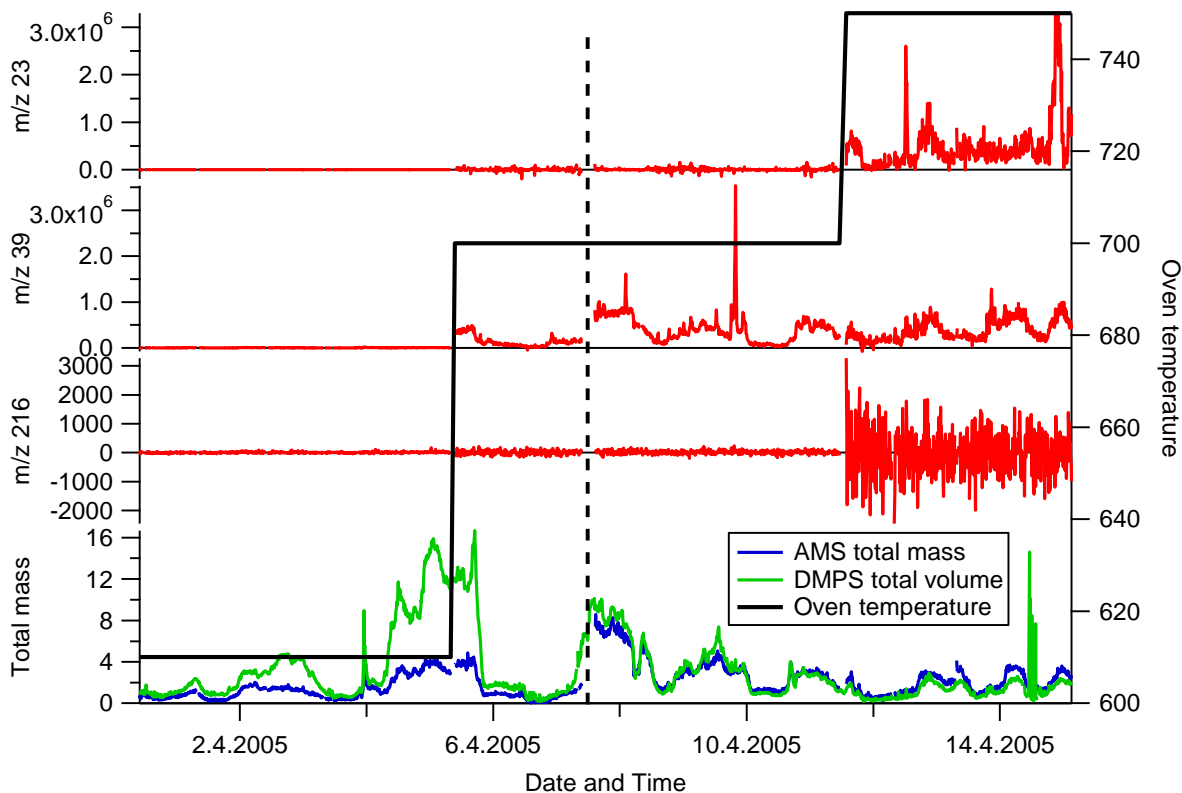


Figure S1: Effect of the non-standard instrument adjustments on measured signals. Oven temperature is shown as a solid black curve on the right hand side axis. The vertical dashed line represents the time of the lens adjustment. The left hand side axis indicates measured signals of m/z 23, 39 and 216 in Hz, and total volume and mass from DMPS and AMS measurements, respectively.

2 Positive matrix factorization

The current positive matrix factorization (PMF) method is described by Paatero and Tapper (1994) and Paatero (1997). Positive matrix factorization algorithm PMF2 version 4.2 (Nov 20, 2006) was used in the calculations. In addition, graphical user interface called PMF Evaluation Toolbox v 2.0 (PET) described by Ulbrich et al. (2009) was used in running the calculations and in interpreting the results. Default PMF2 parameters from the PET were used in the calculations. With the exception of the adjustable parameters and parameters specific to AMS data, PMF2 parameters were essentially the same as the default parameters given in the PMF2 package. Adjustable parameters such as seed option, FPEAK value and matrix sizes are specific for each PMF2 run. For more information about PMF2 and its parameters, see PMF user's guide (Paatero, 2007).

Current PMF analysis was performed in two separate experiments. First of all, the number of factors was chosen based on calculations with rotational parameter FPEAK fixed to zero and by varying the number of factors. As a result, it was concluded that there are only two separate organic groups. The second part was focused on examining rotational variability by

varying FPEAK values for the two factorial solutions. In addition, different seeds (random numbers) were used, but the same solutions were always found.

2.1 Finding the number of factors

Before the PMF runs, the original organic mass spectrum matrix was smoothed, spikes were removed, m/z 19 and m/z 20 peaks were removed, minimum error criteria was applied, weak columns were downweighted and m/z 44 related peaks (m/z 16, 17 and 18) were downweighted as suggested in Ulbrich et al. (2009). In addition, m/z 27, 29, 31 and 41 peaks were ignored, because m/z 41 was calculated from m/z 43 and the other peaks were noisy due to interferences from air nitrogen (m/z 28) and oxygen (m/z 32) peaks.

PMF runs were done for one to five factors. The relative residual (Q/Q_{expected}) of the solutions is shown in Fig. S2. There is a significant decrease in the residual, when the number of factors is increased from one to two, but after that the decrease is lower. It seems that measured mass spectrum can be explained well just by having two factors. Therefore, we will look at the three factorial solutions to see if the factors are meaningful.

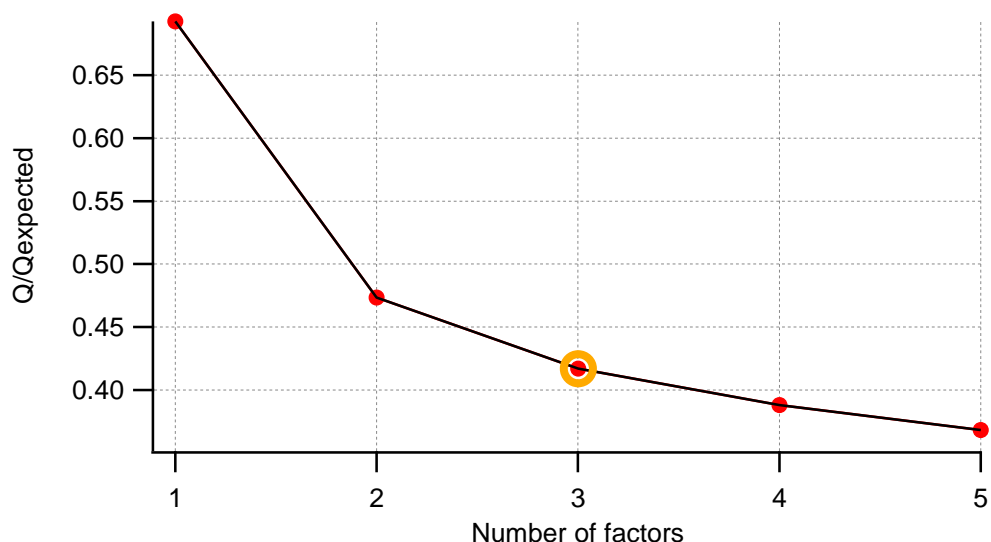


Figure S2: Relative prediction errors for PMF runs with one to five factors.

Figure S3 shows mass spectra for the three component case. With the exception of peaks at m/z 57 and 85, the first and last mass spectra are very similar (correlation $r=0.94$). This indicates that these are not separate factors. Factors like hydrocarbon-like (HOA) and wood/biomass burning organic aerosol (e.g. Lanz et al., 2007) are often found, but the current factors have clearly different spectra. Furthermore, characteristic peaks of the HOA (m/z 57) and wood burning aerosol (m/z 60) are practically missing from the current mass spectrum; here m/z 57 and m/z 60 peaks are only 1.1 % and 0.2 % of the total organic signal, respectively. This does not mean that these species are completely missing, just that their concentrations are low (compared to the total loading). In general, there are methods for finding factors with low or strongly correlated concentrations (e.g. Lanz et al., 2008), but these are not important for our case. For example, growth factor correlations described

in the main text are difficult for species with low mass fractions, because they do not have noticeable effects on mixture growth factors. Finally, interferences due to low total organic mass and the non-standard instrument adjustments are additional reasons for ignoring the three factorial solutions.

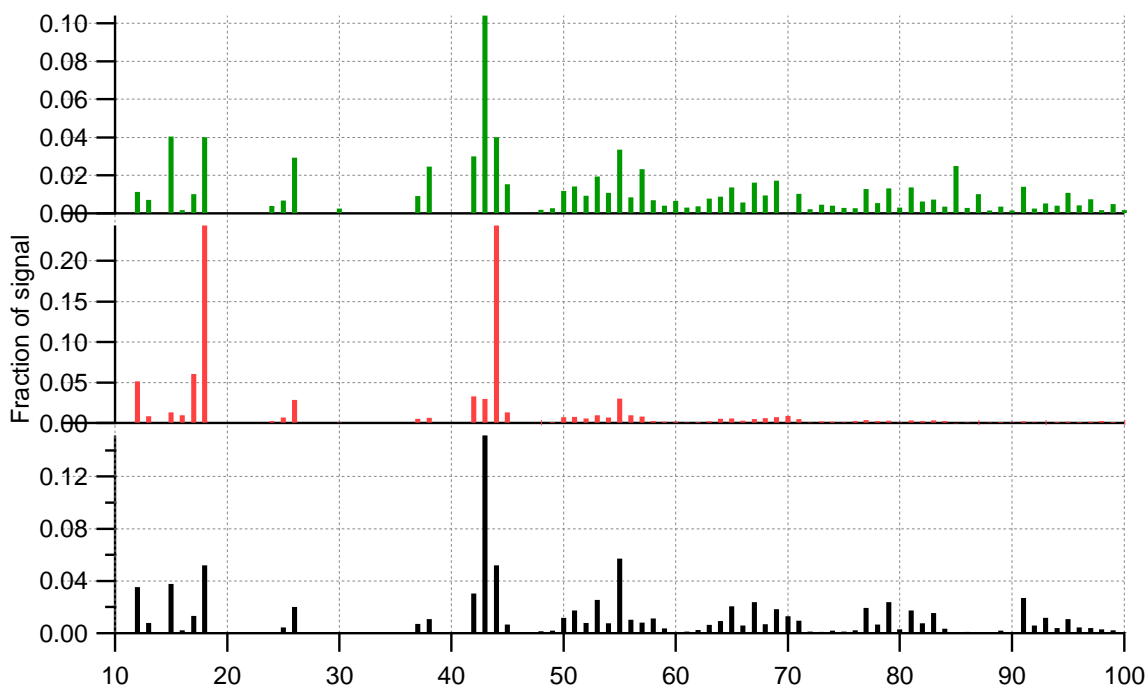


Figure S3: Mass spectra for the three component PMF run.

2.2 Varying FPEAKs of the two factorial solution

The two factorial solution is the best for the Hyytiälä case. In the previous calculations, problematic peaks (m/z 27, 29, 31 and 41) were just ignored. This is not possible for the final results as it is highly desirable that the sum of concentration time series of the factors is equal to the measured total organic mass. For the same reason, data pretreatment such as smoothing and removing spikes was not performed at this time. The other standard adjustments (Ulbrich et al., 2009) were done: the negligible m/z 44 related peaks (m/z 19 and 20) were ignored, minimum error criteria was applied, and weak columns as well as m/z 44 related peaks (m/z 16, 17 and 18) were downweighted according to the suggestions. In addition to the standard procedures, m/z 43 related peaks (m/z 41, which had to be calculated from m/z 43) were properly downweighted. Finally, PMF was run with the default settings for different FPEAK values.

The downweighting can be justified by the fact that some peaks have fixed values, which are not based on measurements. For example, normally organic m/z 41 signal can be measured, so it is not affecting m/z 43. In this case, it had to be calculated from m/z 43 signal by using an average factor of 0.8. In the first place, these kinds of linearly dependent data could be left out from the PMF calculations; correct values could be calculated

afterwards from the parent peaks. On the other hand, if this m/z 41 is included without downweighting, it will upweight m/z 43 signal. As the residual to uncertainty (e_{ij}/σ_{ij}) ratios are the same for the both m/z 's (error is calculated by the same factor), their combined error norm is a factor of square root two higher than the original m/z 43 error norm. Thus m/z 43 would be upweighted by a factor of square root two if m/z 41 were included, and this is not desirable. The same goes with m/z 44 related peaks. For example, m/z 18 signal can not be measured due to a huge water background signal, so based on laboratory experiments organic m/z 18 is set equal to the organic m/z 44 signal.

As mentioned above, the number of factors was fixed to two and FPEAK parameter was varied in order to estimate the rotational variability of the factors. For all possible parameters, the factors were clearly identified as OOA1 and OOA2 (Lanz et al., 2007). When FPEAK was increased from zero up to 1.0, only minor changes were seen in the concentration time series. On the other hand, even a small decrease in FPEAK had clear changes to concentration time series. This can be seen in Fig. S4, where the green (FPEAK=0) and red (FPEAK=-0.2) curves have the largest differences. When FPEAK was decreased from -0.4 to -1.0 or increased from 0.4 to 1.0, only minor changes were seen. With the exception of time period from 4.4.2005 to 5.4.2005, concentration time series are quite similar.

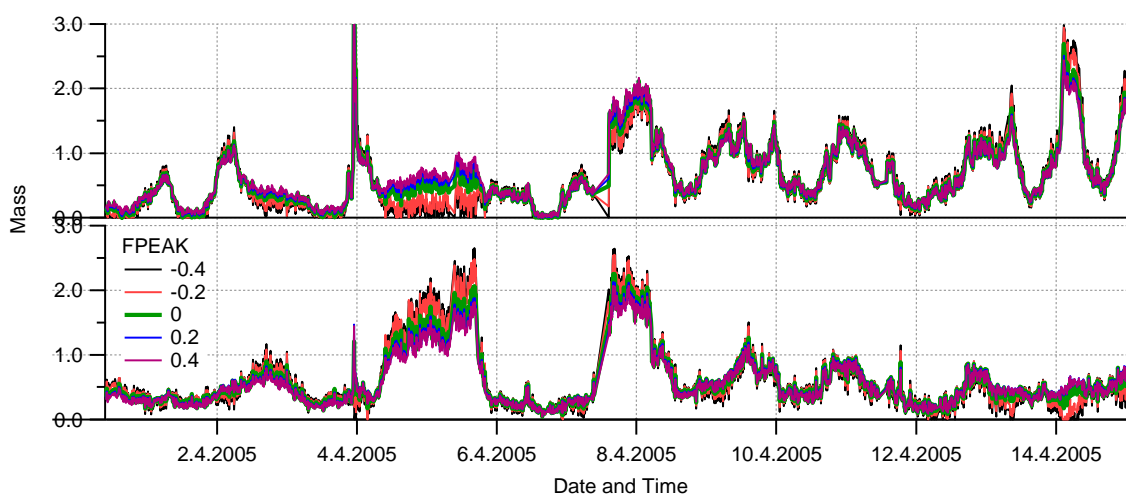


Figure S4: Concentration time series for different FPEAKs

As we are more interested in concentrations, negative FPEAKs leading to larger discrepancies are examined more carefully. Figure S5 shows mass concentration time series as well as mass spectra for three different FPEAK values. These concentration time series as well as mass spectra are quite reasonable and also the residuals (Fig. S6) are the smallest ones. Therefore, we conclude that FPEAKs from 0 to -0.3 are all possible. The mean value (FPEAK=-0.15) is selected for the calculations. Some diagnostics plots for the FPEAK=-0.15 solution are shown in Fig. S7.

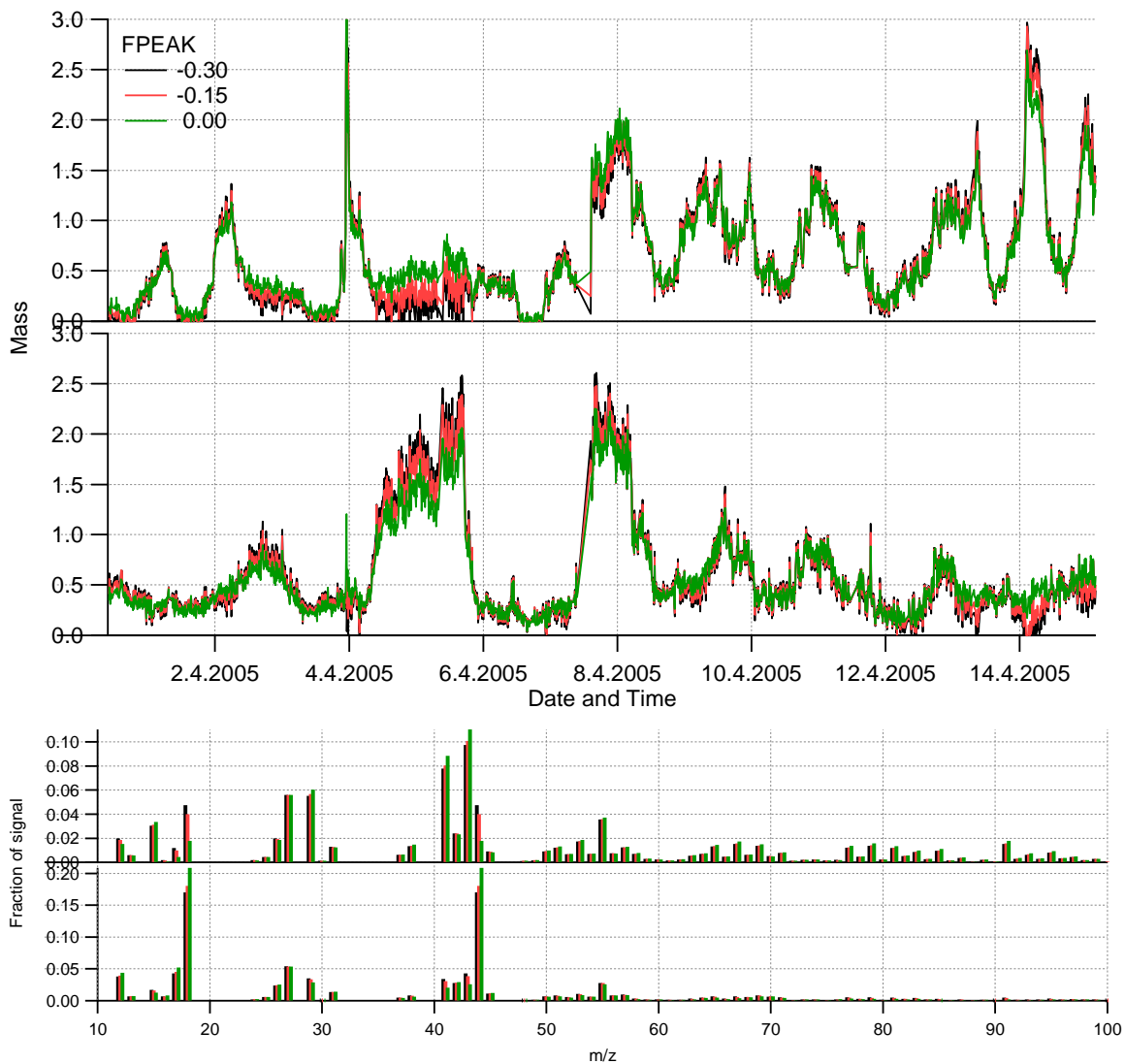


Figure S5: Mass concentration time series (upper panel) and mass spectra (lower panel) for three FPEAK values.

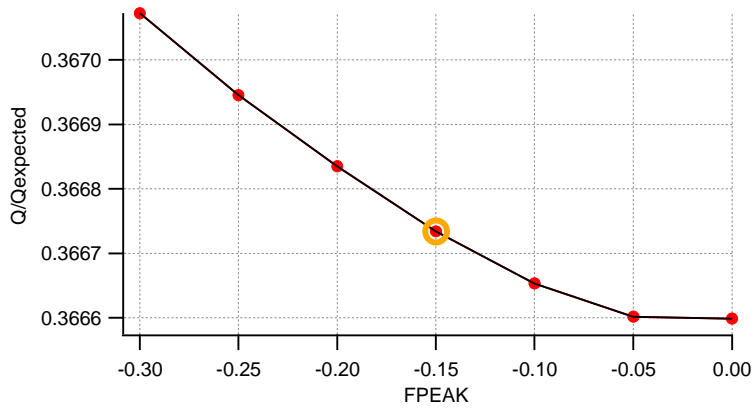


Figure S6: Relative residual as a function of FPEAK.

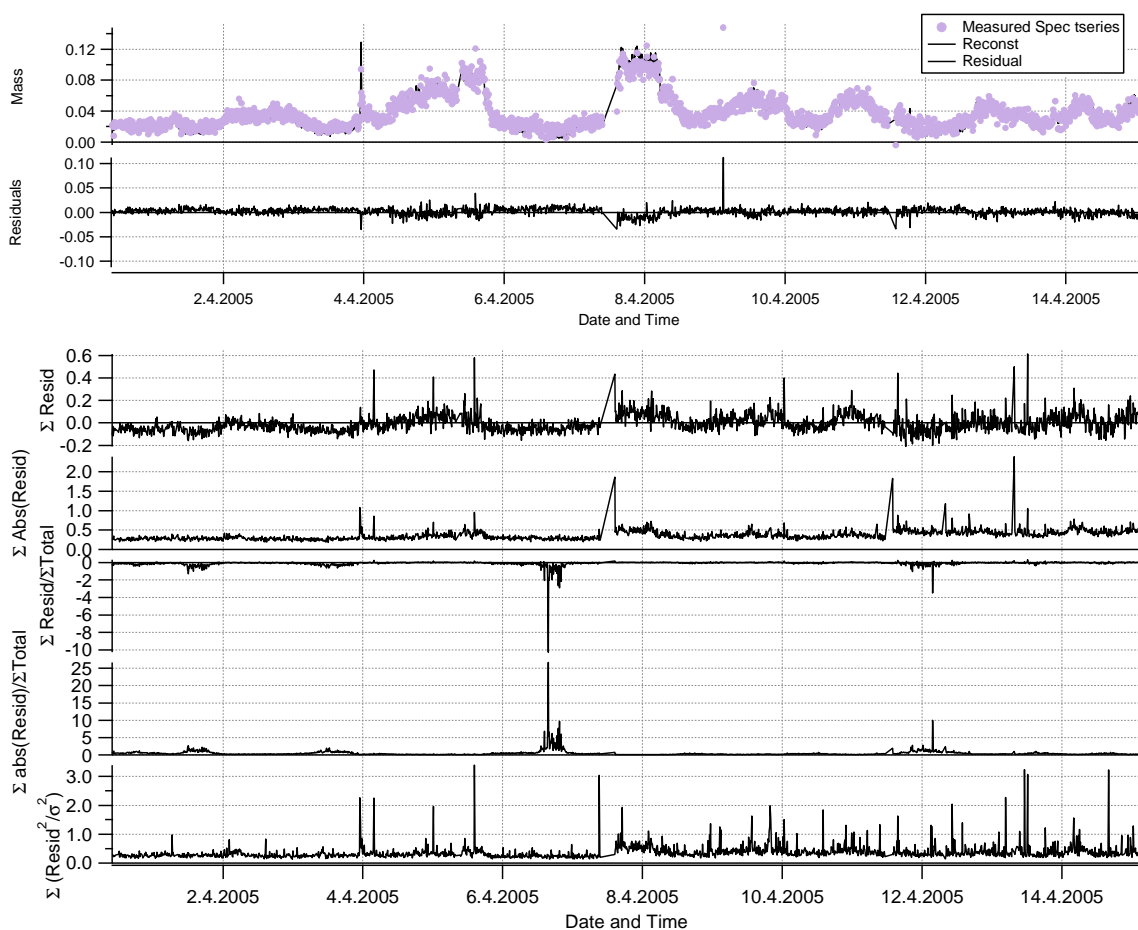


Figure S7: Diagnostic plots for the selected FPEAK=-0.15 solution from the PET panels (Ulbrich et al., 2009). The upper graph shows measured and reconstructed total organic mass as well as residual. The lower graph shows different total residuals.

Uncertainties of the mass concentrations of the current factors (FPEAK=-0.15) can be estimated by calculating how much they differ from those of the extreme FPEAKs (FPEAK=0 and FPEAK=-0.3). Table S1 shows average mass concentrations, fraction of m/z 44 peak and average absolute and standard deviations between mass concentrations from FPEAK=-0.15 and those of the other solutions. In this case, factor 1 is OOA1 and factor 2 is OOA2. As a result, we can say that OOA1 average mass concentration is $(0.61 \pm 0.02) \mu\text{g}/\text{m}^3$ and OOA2 average mass concentration is $(0.64 \pm 0.02) \mu\text{g}/\text{m}^3$. It must be kept in mind that these uncertainties are well below their real average mass concentration (e.g. the effect of non-standard AMS calibrations). OOA1 average mass fraction from total organic mass are 47, 48 and 49 % for the FPEAKs 0, -0.15 and -0.30, respectively.

Table S1: Average mass concentrations, fraction of peak m/z 44 and average absolute (AAD) and standard (STD) deviations between mass concentrations from the selected (FPEAK=-0.15) and from the other two solutions (FPEAK=0 and -0.3).

FPEAK	Factor	Average ($\mu\text{g}/\text{m}^3$)	m/z 44	AAD ($\mu\text{g}/\text{m}^3$)	STD ($\mu\text{g}/\text{m}^3$)
0.00	1	0.591	0.209	0.061	0.091
	2	0.660	0.018	0.060	0.090
-0.15	1	0.606	0.180	0	0
	2	0.643	0.040	0	0
-0.30	1	0.614	0.170	0.035	0.051
	2	0.635	0.048	0.034	0.050

3 Different growth factor fits

There are several different ways in converting OOA1, OOA2, SO_4^{2-} , NO_3^- and NH_4^+ mass concentrations to the volume fractions (ξ) needed in the ZSR equation ($\text{GF}^3 = \sum \xi_i \text{GF}_i^3$). Groups can be considered individually or two or more groups can be combined, but obviously, all of them must be included. For example, organics can be considered as one group (OOA1+OOA2), inorganic species can be considered as one group ($\text{SO}_4^{2-} + \text{NO}_3^- + \text{NH}_4^+$), and it is possible to estimate concentrations of common salts and acids such as $(\text{NH}_4)_2\text{SO}_4$, NH_4NO_3 , H_2SO_4 and HNO_3 . In the first place we are interested in the accuracy of the prediction, so densities can be selected quite freely as they have a greater effect on pure component growth factors. For simplicity, we take all inorganic species to have density of $1.7 \text{ g}/\text{cm}^3$ as this is close to that of most sulfate and nitrate salts and acids. Organic densities have great variability, but an average value of $1.3 \text{ g}/\text{cm}^3$ is chosen. Because OOA1 is expected to be denser than OOA2, their densities are set to $1.4 \text{ g}/\text{cm}^3$ and $1.200 \text{ g}/\text{cm}^3$, respectively.

The deviations (average absolute and standard deviations) for the different approaches are shown in Table S2. The fitting was done as described in the main text. The first and second row (and the last two rows) of the table shows that inorganic species can be considered as one group without increasing model deviations. On the other hand, the deviations are increased when the organic groups are combined. Interestingly, the deviation is somewhat smaller when the inorganic species are divided into salts and combined (last row) than in the case that ions are combined directly (third row). It seems some noise is removed, when the ions are converted into the salts. Some growth factor predictions as well as experimental data are shown in Fig. S8. Clearly, the predictions are worse for the two component fit ($\text{NH}_4^+ + \text{NO}_3^- + \text{SO}_4^{2-}$ and OOA1+OOA2), but the other fits are nearly identical and equally good. Therefore, we chose the simplest approach of having three groups: OOA1, OOA2 and $\text{NH}_4^+ + \text{NO}_3^- + \text{SO}_4^{2-}$.

Table S2: Deviations of growth factor and volatility correlations for different groups

Groups		HGF	EGF	<150°C	150-280°C	>280°C
5: NH_4^+ , NO_3^- , SO_4^{2-}	aad	0.068	0.023	0.053	0.039	0.028
OOA1, OOA2	std	0.092	0.028	0.067	0.050	0.037
4: NH_4^+ , NO_3^- , SO_4^{2-}	aad	0.073	0.022	0.063	0.045	0.029
OOA1+OOA2	std	0.094	0.027	0.078	0.058	0.038
3: NH_4^+ + NO_3^- + SO_4^{2-}	aad	0.068	0.023	0.053	0.039	0.029
OOA1, OOA2	std	0.092	0.028	0.067	0.051	0.038
2: NH_4^+ + NO_3^- + SO_4^{2-}	aad	0.074	0.022	0.063	0.047	0.029
OOA1+OOA2	std	0.094	0.027	0.078	0.061	0.039
4: NH_4NO_3 , $(\text{NH}_4)_2\text{SO}_4$	aad	0.067	0.022	0.049	0.038	0.028
OOA1, OOA2	std	0.091	0.027	0.062	0.049	0.037
3: NH_4NO_3 + $(\text{NH}_4)_2\text{SO}_4$	aad	0.067	0.022	0.049	0.038	0.028
OOA1, OOA2	std	0.091	0.027	0.062	0.049	0.037

aad=average absolute deviation
std=standard deviation

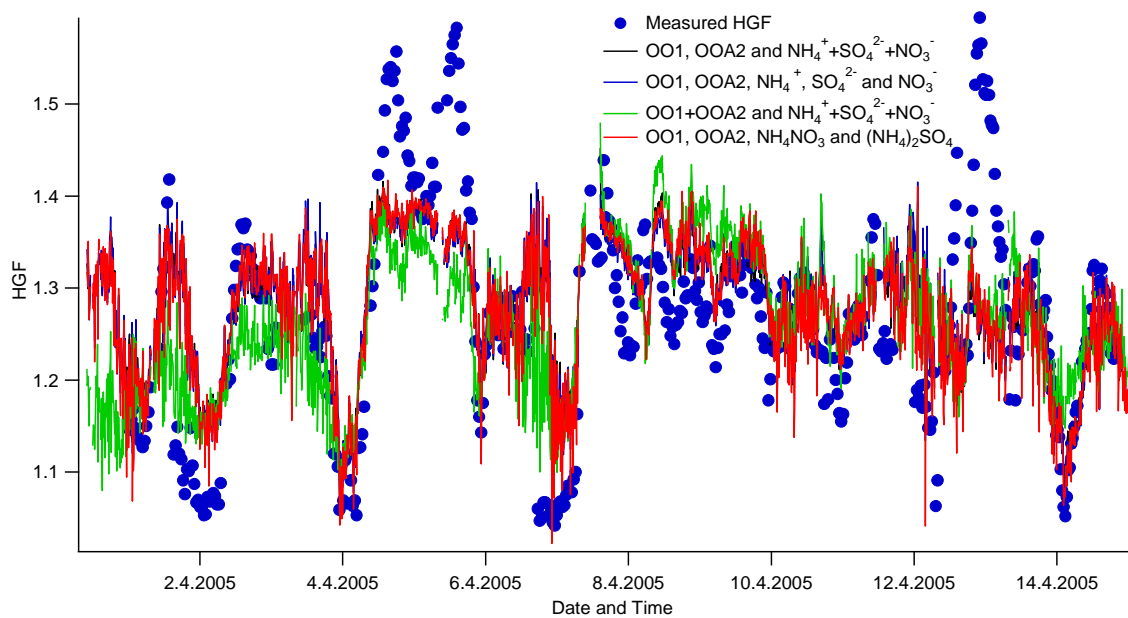


Figure S8: Measured and predicted hygroscopic growth factors. Predictions are made using different groups in the fit to the ZSR equation.

The results were about the same, when different weights such as total mass concentration were used in the fitting and when some noisy data points were ignored based on different criteria (mainly minimum total mass concentration). In addition to the ZSR equation, also linear fitting was tested, but the results were somewhat worse. It seems that at least in this case, particle properties are described well by considering two separate organic groups and one inorganic group.

It was seen that based on prediction error, the three component fitting with the two organic factors ($\text{NH}_4^+ + \text{NO}_3^- + \text{SO}_4^{2-}$, OOA1 and OOA2) is clearly the best choice. When testing different groups etc, fitted growth factors were always about the same. For example, when fitting simple linear equation based on mass fractions ($\text{GF} = \sum w_i \text{GF}_i$), the fitted hygroscopic growth factors were 1.50, 1.29 and 1.00 for the $\text{NH}_4^+ + \text{NO}_3^- + \text{SO}_4^{2-}$, OOA1 and OOA2, respectively. These are very close to the values given in the main text (1.53, 1.29 and 1.00). Growth factor fitting was not at all sensitive on the FPEAK values used in the PMF analysis. Table S3 contains fitted growth factors of OOA groups from the reasonable FPEAK range (from 0.0 to -0.3). As can be seen, growth factors are well within reported uncertainties. In general, OOA1 growth factors have the greatest variability ranging from 1.24 to 1.36, which is not too much when the nature of these groups is taken into account; it seems that the average properties of the organic groups are nearly constant.

Table S3: Results of the growth factor correlations when OOA1 and OOA2 are based on different FPEAK values.

FPEAK	OOA1		OOA2		$\text{NH}_4^+ + \text{NO}_3^- + \text{SO}_4^{2-}$	
	HGF	EGF	HGF	EGF	HGF	EGF
-0.30	1.27	1.13	1.00	1.16	1.54	1.00
-0.15	1.29	1.12	1.00	1.16	1.53	1.00
0.00	1.29	1.12	1.00	1.16	1.53	1.00

4 Diurnal variations

Diurnal concentration cycles of the ions and organic factors are shown in Figs S9 and S10, respectively. In addition to the simple averages (black markers and lines), also 25 %, 50 % and 75 % percentiles (red markers and bars) are shown. OOA1 has a weak diurnal cycle with maximum in the afternoon and minimum in the morning. OOA2 is the only one with a clear diurnal cycle.

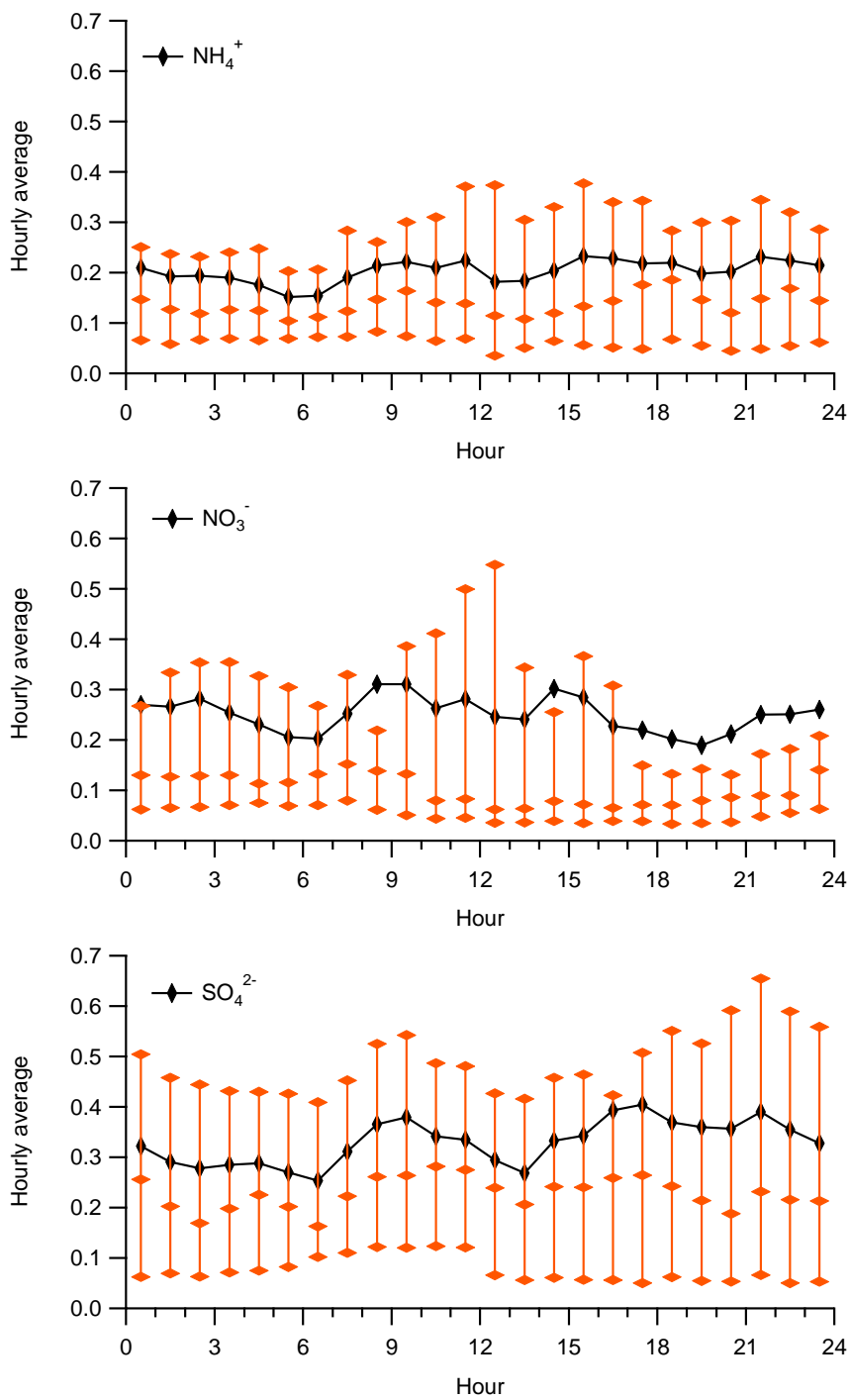


Figure S9: Diurnal cycles for the detected inorganic species. Black lines and markers show the average values, and red markers are the 25 %, 50 % and 75 % percentiles.

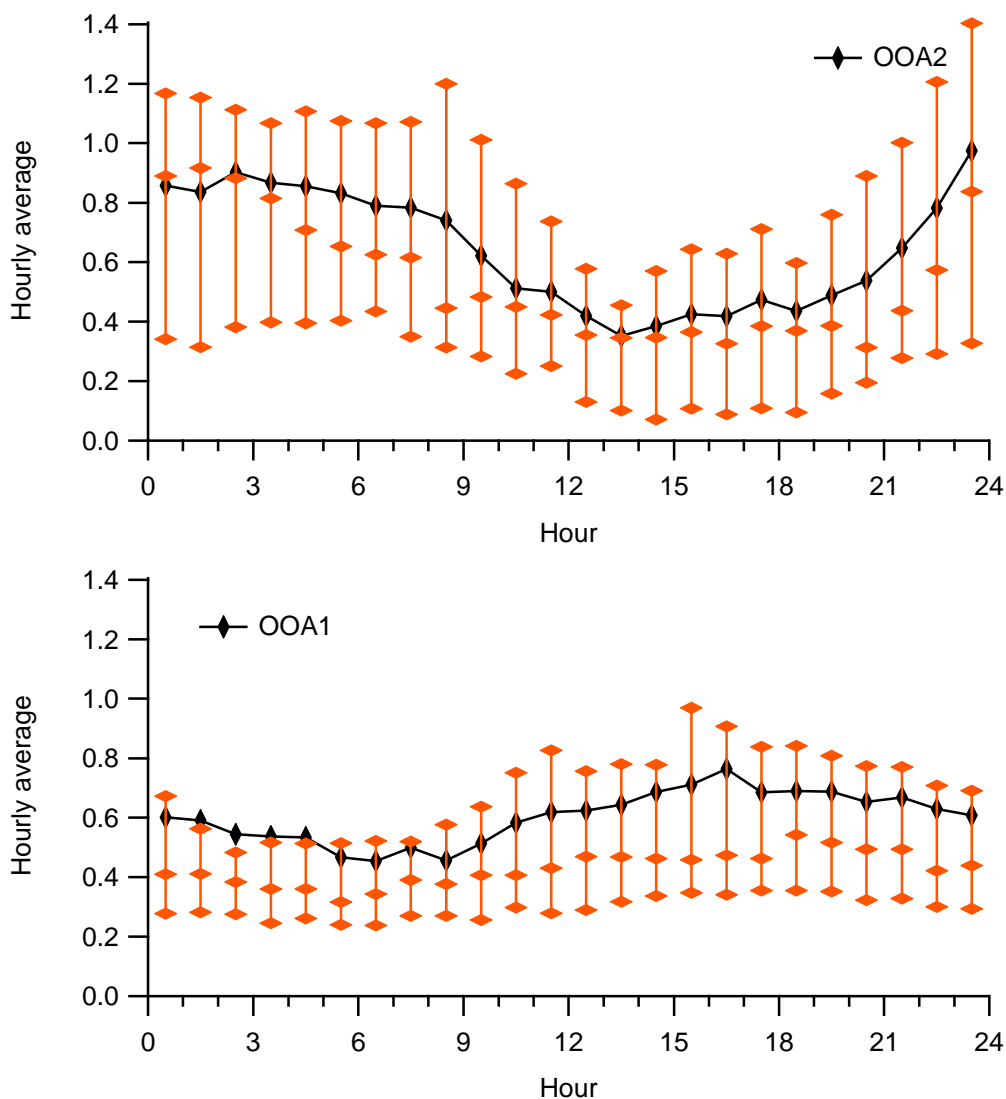


Figure S10: Diurnal cycles for the organic factors. Black lines and markers show the average values, and red markers are the 25 %, 50 % and 75 % percentiles.

References

- Canagaratna, M., Jayne, J., Jimenez, J., Allan, J., Alfarra, M., Zhang, Q., Onasch, T., Drewnick, F., Coe, H., Middlebrook, A., Delia, A., Williams, L., Trimborn, A., Northway, M., DeCarlo, P., Kolb, C., Davidovits, P., and Worsnop, D.: Chemical and microphysical characterization of ambient aerosols with the Aerodyne aerosol mass spectrometer, *Mass Spectrom. Rev.*, 26, 185–222, 2007.
- Lanz, V. A., Alfarra, M. R., Baltensperger, U., Buchmann, B., Hueglin, C., and Prévôt, A. S.H.: Source apportionment of submicron organic aerosols at an urban site by factor analytical modelling of aerosol mass spectra, *Atmos. Chem. Phys.*, 7, 1503–1522, 2007.

- Lanz, V.A., Alfarra, M.R., Baltensperger, U., Buchmann, B., Hueglin, C., Szidat, S., Wehrli, M.N., Wacker, L., Weimer, S., Caseiro, A., Puxbaum, H., Prevot, A.S.H.: Source attribution of submicron organic aerosols during wintertime inversions by advanced factor analysis of aerosol mass spectra, *Environ. Sci. Technol.*, 42, 214–220, 2008.
- Paatero, P. and Tapper, U.: Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values, *Environmetrics*, 5, 111–126, 1994.
- Paatero, P.: Least squares formulation of robust non-negative factor analysis, *Chemom. Intell. Lab. Syst.*, 37, 23–35, 1997.
- Paatero, P.: User's Guide for Positive Matrix Factorization programs PMF2 and PMF3, Part 2: reference, Last changed on July 23, 2007.
- Ulbrich, I. M., Canagaratna, M. R., Zhang, Q., Worsnop, D. R., and Jimenez, J. L.: Interpretation of organic components from Positive Matrix Factorization of aerosol mass spectrometric data, *Atmos. Chem. Phys.*, 9, 2891–2918, 2009.