Supplementary information for manuscript

Spatial variation of chemical composition and sources of submicron aerosol in Zurich: Factor analysis of mobile aerosol mass spectrometer data

C. Mohr[1], R. Richter[1], Peter F. DeCarlo[1,*], A. S. H. Prévôt[1], U. Baltensperger[1]

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

[*] now an AAAS Science and Technology Policy Fellow hosted at the US EPA, Washington DC, USA

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

1 Air mass back trajectories

Four-day backward trajectories were calculated based on 3-dimensional wind fields of the regional weather prediction model COSMO using the trajectory model TRAJ (Fay et al., 1995). The fields were taken from hourly "analyses" operationally generated by the Swiss weather service MeteoSwiss at a resolution of 7 km x 7 km x 60 vertical levels for a domain covering large parts of Europe.

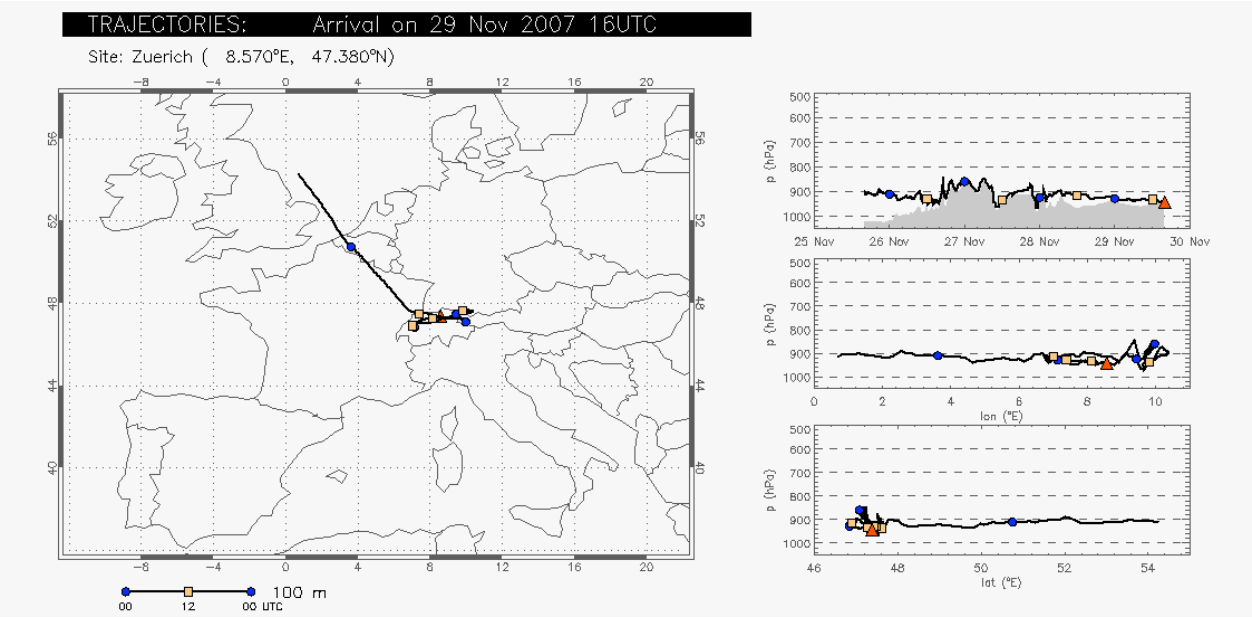

Figure SI- 1: Air mass back trajectories for 29 November 2007. Air masses moved from Belgium/Germany to Switzerland and stagnated over the Swiss plateau, residing there for about 3 days prior to reaching the receptor site Zurich Kaserne.

2 Representativeness plot of mobile measurements

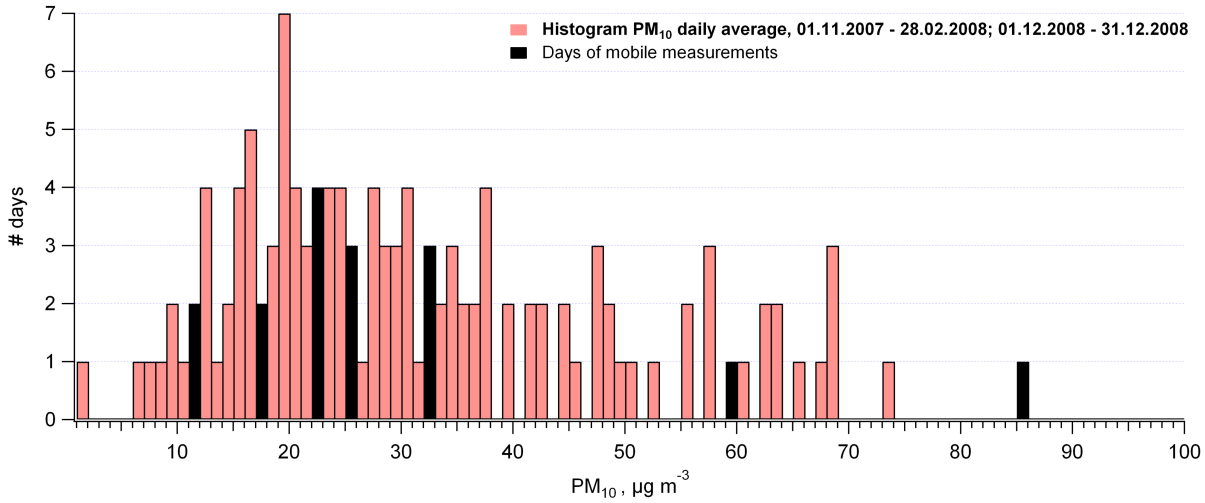

Figure SI- 2: Histogram of PM_{10} daily mean values for the periods of 01 November 2007 – 31 February 2008 and 01 December 2008 – 31 December 2008. Values of days when mobile measurements were performed are colored in black.

3 PMF diagnostics

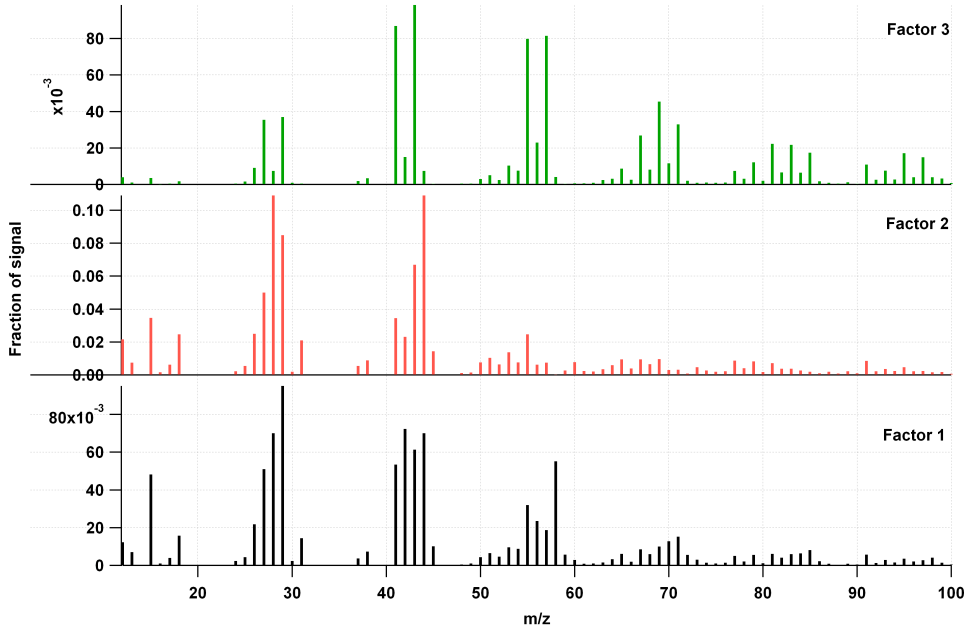

Figure SI- 3: 3-factorial solution of running the PMF2 algorithm on the organic data matrix where the *m/z*'s directly proportional to *m/z* 44 were downweighted.

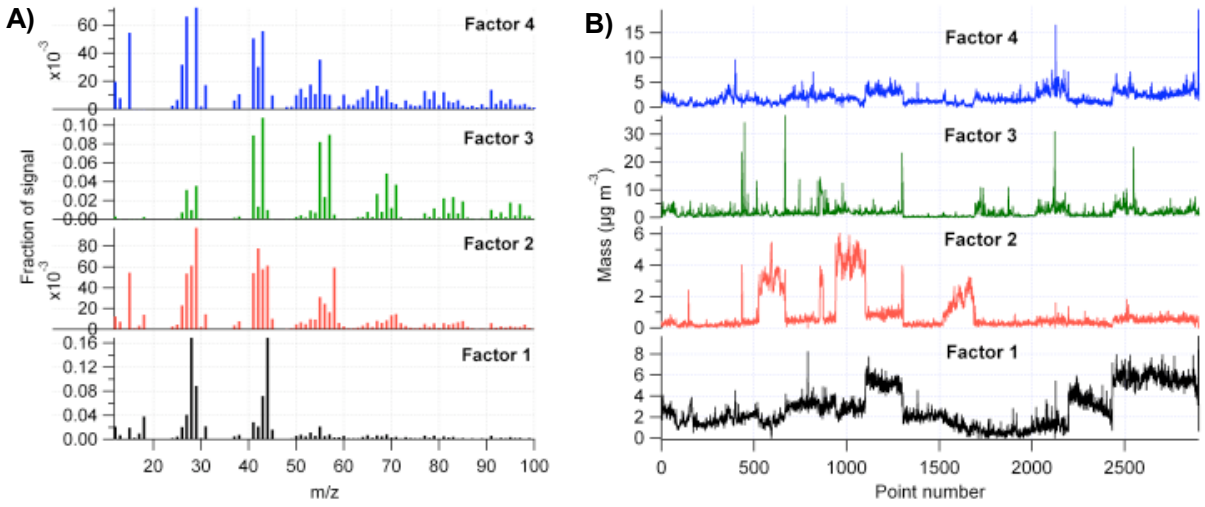

Figure SI- 4: 4-factorial solution for part 1, source spectra (F, panel A), and time series (G, panel B).

Choosing $p > 3$ did not yield meaningful results. For part 1, $p = 4$ resulted in an additional factor correlated with BBOA and OOA and a prominent peak at *m/z* 58 related to amines, most likely an

artifact of inlet contamination during overnight parking of the mobile laboratory at a garage of public transport buses. Adding an additional factor for part 2 led to a split of the HOA factor (Fig. SI-5).

For part 1, $p = 4$, the resulting additional factor 2 (Fig. SI-4) shows high similarity with factor 1 (Pearson's $R = 0.74$) and factor 4 (Pearson's $R = 0.77$) and can be interpreted as a recombination of OOA and BBOA. Interestingly, it features a few distinct peaks relating to the ion series ($C_nH_{2n+2}N$) characteristic for amines, e. g. *m/z* 58 (Silva et al., 2008). As shown by the time series of factor 2 in panel B), there were 3 measurement drives with substantial factor 2 mass loadings – drives following a night when the mobile laboratory had been parked in a garage of public transport buses in Zurich. The punctual occurrence of this factor and the missing analogies in volatile organic compounds (VOC) time series measured at Zurich Kaserne (not shown) lead to the hypothesis that the amine signal could be explained by emissions related to SCR (selective catalytic reduction, a NO_X abatement technology using an aqueous urea solution (Koebel et al., 2000)) systems the buses are equipped with to meet the EURO V legal emission standards (implemented in Switzerland on 01.09.2009).

Running PMF excluding the amine-influenced periods yielded the same 3 factors as for the complete part 1 dataset ($R^2 > 0.99$ for all 3 factors). The 3-factorial solution of the full part 1 dataset exhibits elevated total residual masses for those 3 measurement drives (Fig. SI-13), mostly due to *m/z* 58 (compare non- normally distributed scaled residuals for *m/z* 58 in the inset of Fig. SI-14).

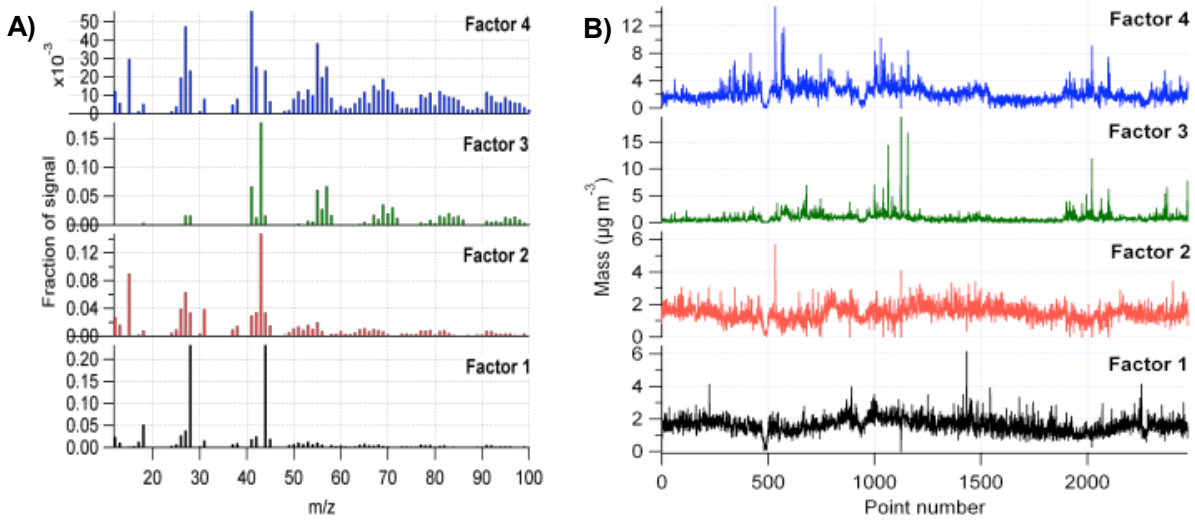

Figure SI- 5: 4-factorial solution for part 2, source spectra (F, panel A), and time series (G, panel B).

The *Q*-value is one mathematical criterion for the quality of the fit (compare Eq. 2 in main paper). If the model is appropriate for the problem at hand and the data uncertainties estimations are accurate, then $(e_{ij} / \sigma_{ij})^2$ is ~1 and the expected Q (Q_{exp}) = mn-$p(m+n) \approx mn$, the degrees of freedom of the fitted data. $Q/Qexp >> 1$ indicates an underestimation, $Q/Qexp << 1$ an overestimation of errors in the input data (Paatero et al., 2002). Each added factor introduces more degrees of freedom allowing more data to be fit and hence decreases *Q*. From a mathematical point of view, the correct value of *p* in PMF is where the line changes the slope in the plot of a series of *p* values versus their respective minimized *Q*. However, PMF solutions of ambient datasets also have to be feasible in an ambient context and hence it is the subjective task of the modeler to choose a set of factors able to explain real world phenomena which may or may not correspond to the mathematically correct value of *p*.

Another parameter to explore the quality of the PMF fit is *max(rotmat)*, the largest element in **RotMat** where PMF2 reports the standard deviation of possible values of the transformation matrix **T**. PMF solutions are not unique since linear transformation still conserving the non-negativity constraint may be possible (**GF** = **GTT**$^{-1}$**F**). This rotational indeterminacy is a significant problem in the use of factor analysis (Paatero et al., 2002). Generally, the best fit demands a minimal *max(rotmat)*, since larger values in **T** imply greater rotational freedom of a solution. However, it has been stated clearly (e. g. Lanz et al., 2007) that "**RotMat** values […] are not suited as a unique criterion for the determination of the number of factors" (compare

Figure SI-6, right panel: For part 2, a solution purely based on *max(rotmat)* would not include $p = 3$).

Once p has been defined, the rotational freedom of the chosen solution may 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. F*peak* and was chosen to be -0.1 for part 1, and 0 for part 2, based on a trade-off between "high" signal at *m/z* 60 ($C_2H_4O_2^+$, among others a fragment of levoglucosan which in turn is a pyrolysis product of cellulose an hence a marker of biomass burning emissions (Alfarra et al., 2007)), and non-zero signal at *m/z* 44 (predominantly non-gaseous CO_2^+) in the BBOA spectrum.

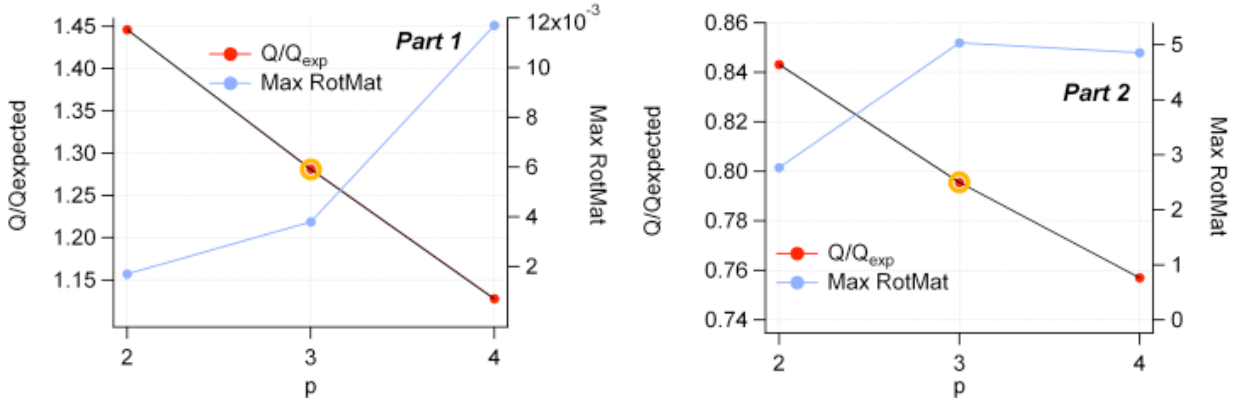

Figure SI- 6: *Q/Q*$_{exp}$ and the maximum value of the rotational matrix versus the number of factors for part 1 and part 2. The chosen solution is denoted by the orange circle.

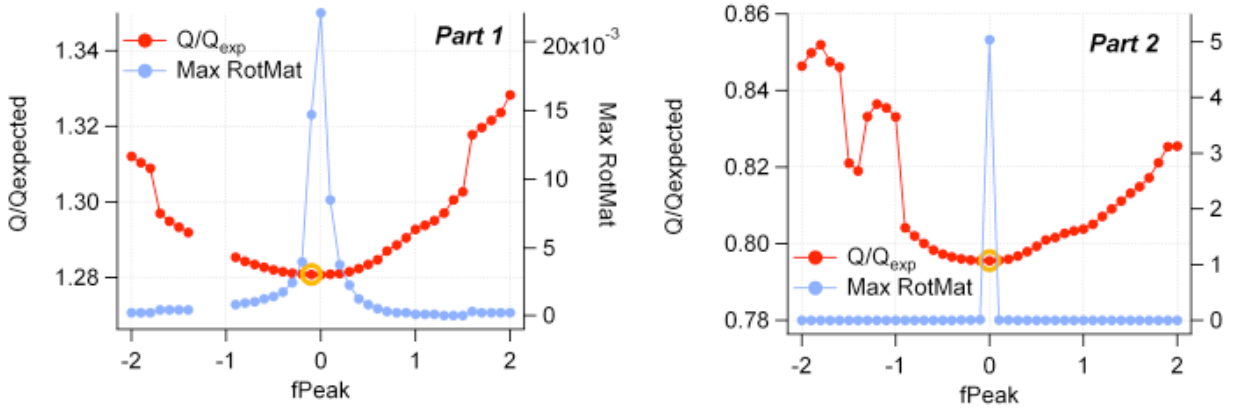

Figure SI- 7: Q/Q_{exp} and the maximum value of the rotational matrix versus *fpeak* for part 1 and part 2. The chosen solution is denoted by the orange circle.

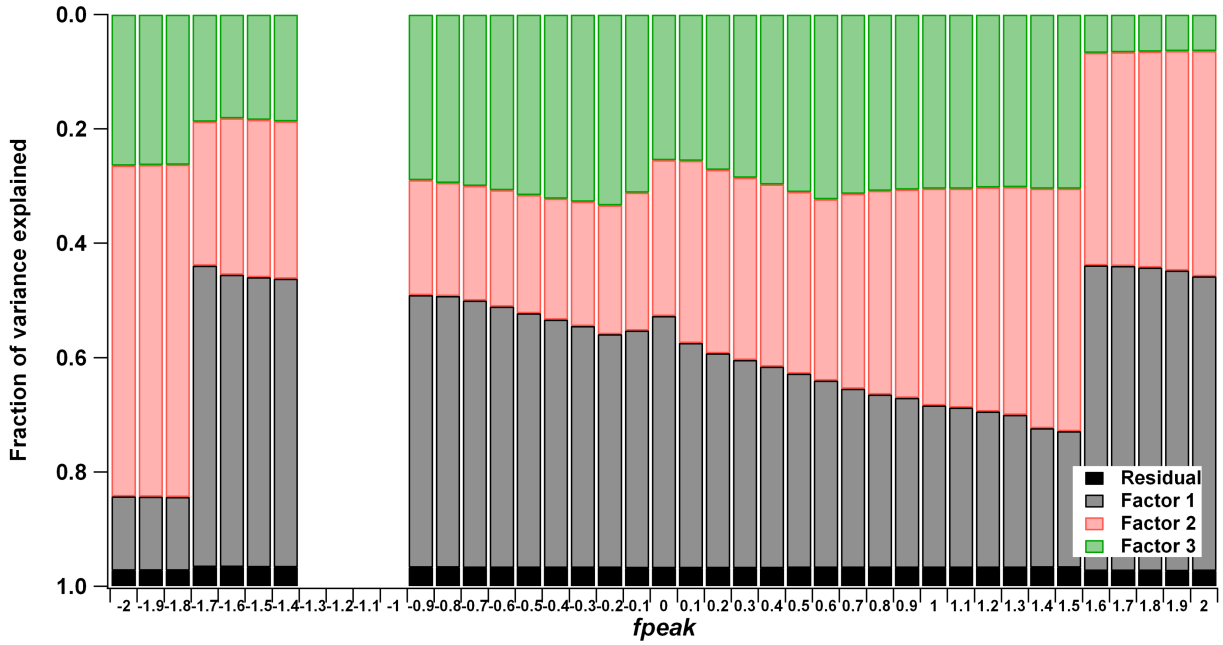

Figure SI- 8: Part 1 (data from 27 November 2007 – 19 February 2008) - variance explained by p = 3 as a function of rotational parameter *fpeak*. *fpeak* was chosen to be -0.1 for this part of the campaign.

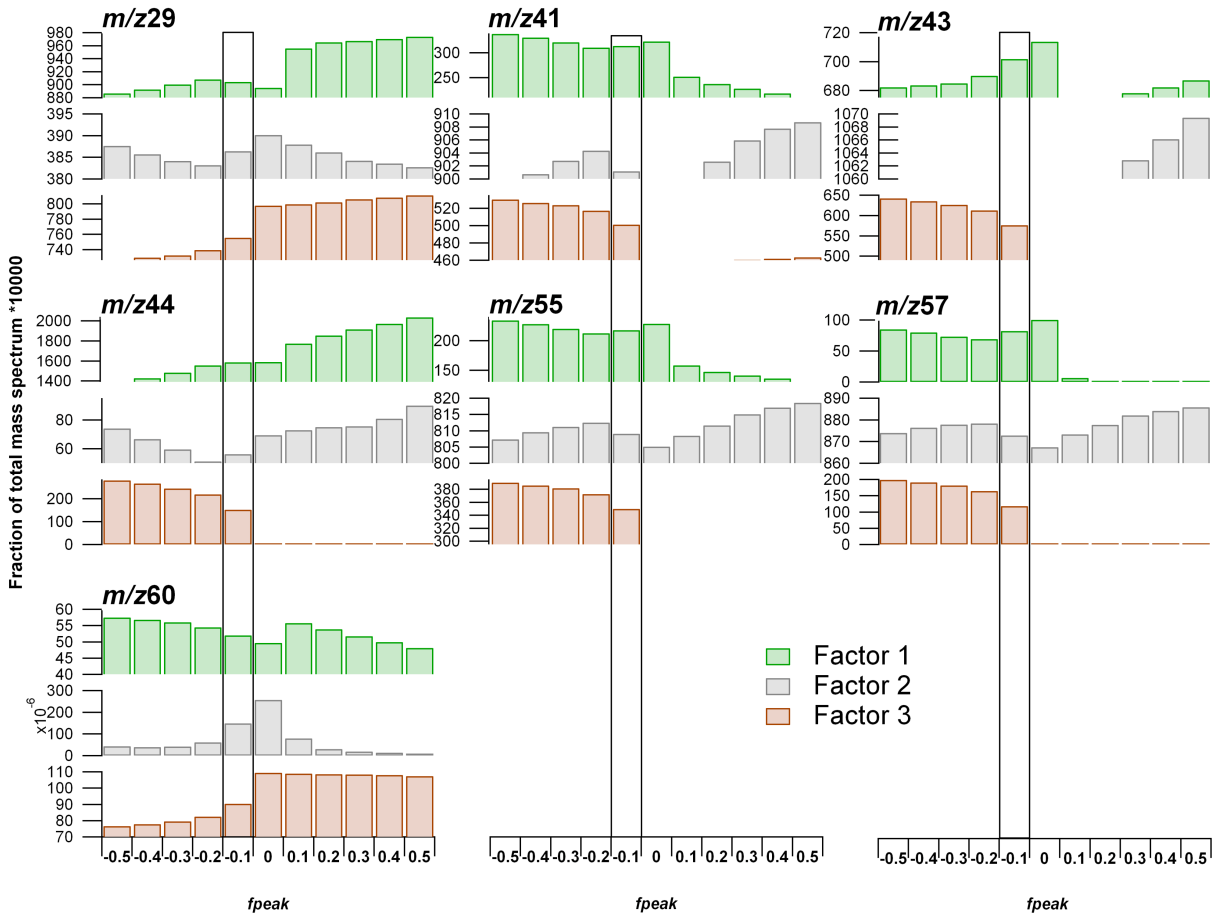

Figure SI- 9: Part 1 (data from 27 November 2007 – 19 February 2008) - fraction of organic m/z's 29 (CHO^+, $C_2H_5^+$), 41 (pre-dominantly $C_3H_5^+$), 43 ($C_2H_3O^+$, $C_3H_7^+$), 44 (pre-dominantly CO_2^+, also $C_2H_4O^+$, $C_2H_8^+$), 55 (pre-dominantly $C_4H_7^+$), 57 ($C_3H_5O^+$, $C_4H_9^+$), and 60 ($C_2H_4O_2$+) as a function of *fpeak* [-0.5,0.5] for the 3-factorial solution. Note the different scaling of the y-axes.

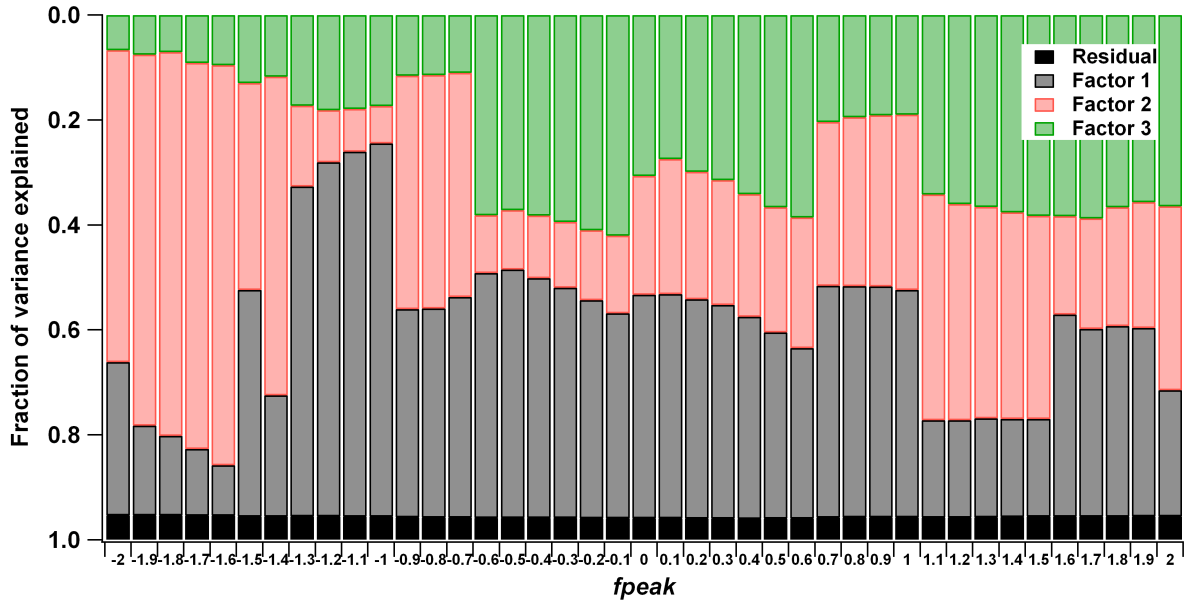

Figure SI- 10 Part 2 (data from 14 December 2008 – 16 December 2008) - variance explained by $p = 3$ as a function of rotational parameter *fpeak*. . *fpeak* was chosen to be 0 for this part of the campaign.

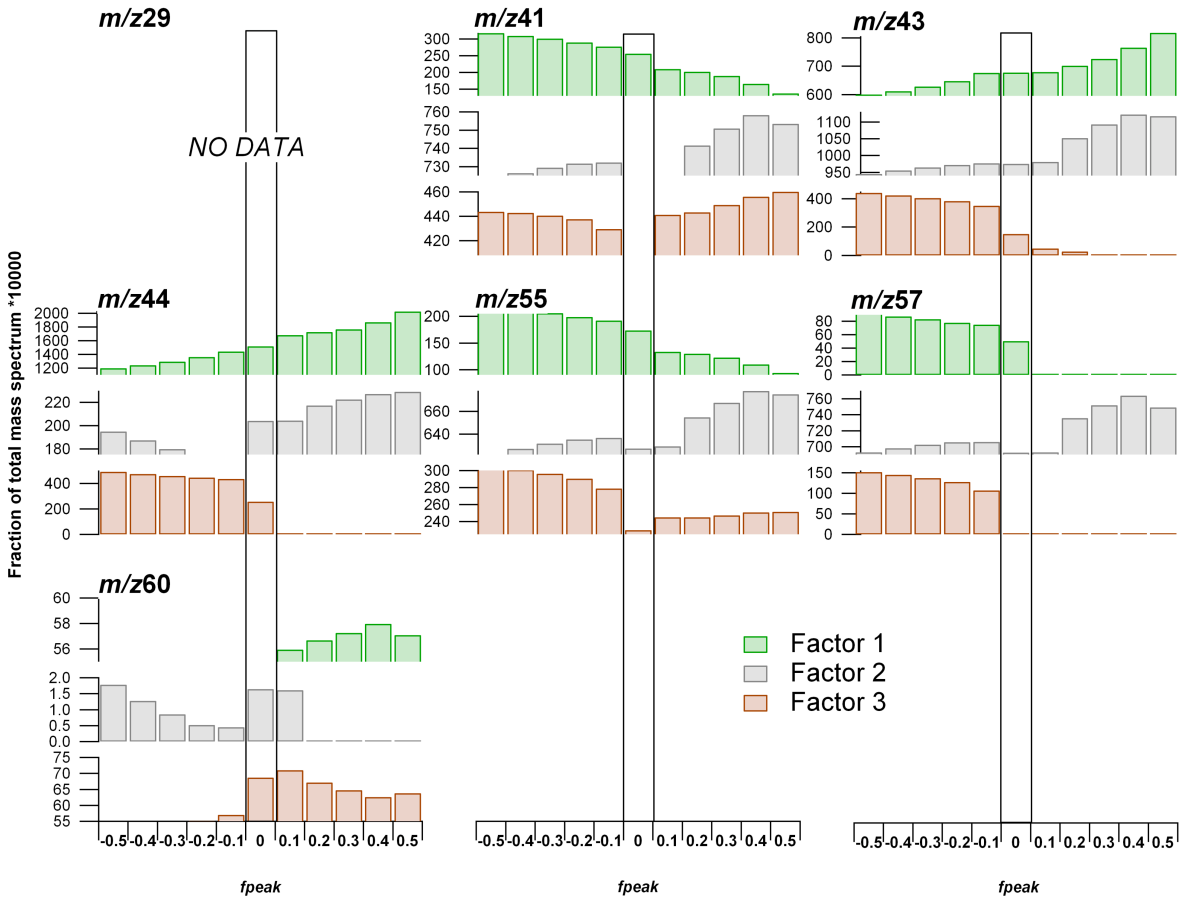

Figure SI- 11: Part 2 (data from 14 December 2008 – 16 December 2008) - fraction of organic m/z's 41 (pre-dominantly $C_3H_5^+$), 43 ($C_2H_3O^+$, $C_3H_7^+$), 44 (pre-dominantly CO_2^+, also $C_2H_4O^+$, $C_2H_8^+$), 55 (pre-dominantly $C_4H_7^+$), 57 ($C_3H_5O^+$, $C_4H_9^+$), and 60 ($C_2H_4O_2$+) as a function of *fpeak* [-0.5,0.5] for the 3-factorial solution. Note the different scaling of the y-axes.

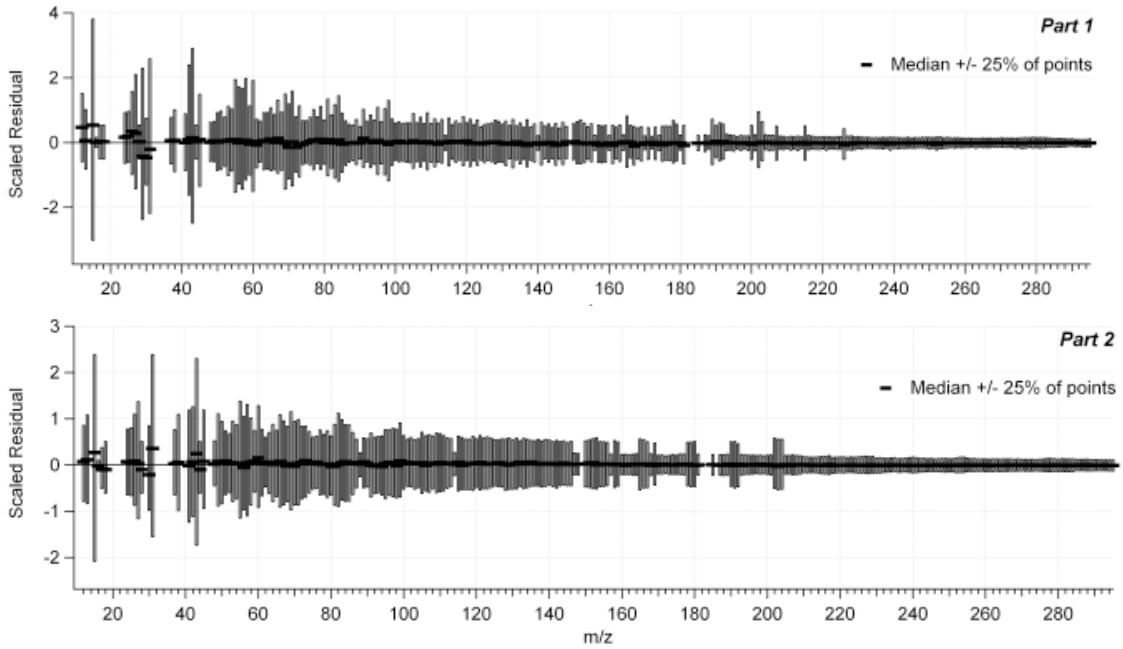

Figure SI- 12: Boxplots of scaled residuals (only median and 25%-percentiles shown) as a function of *m/z*.

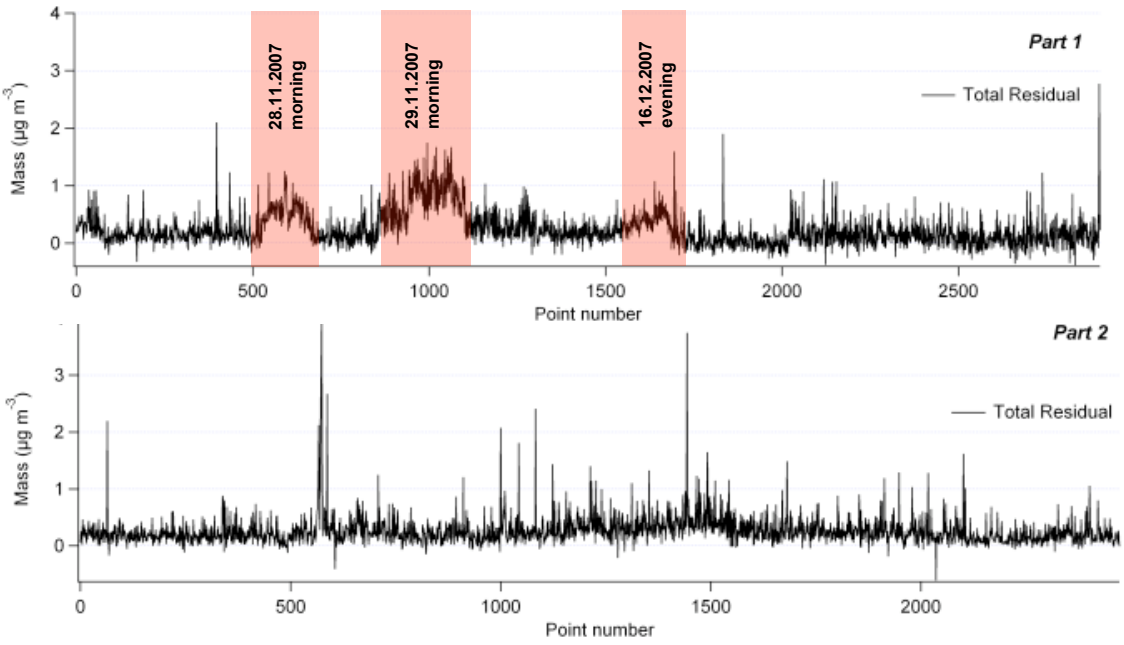

Figure SI- 13: Time series of summed total residuals. Red bars in part 1 panel denote periods influenced by amine-like factor.

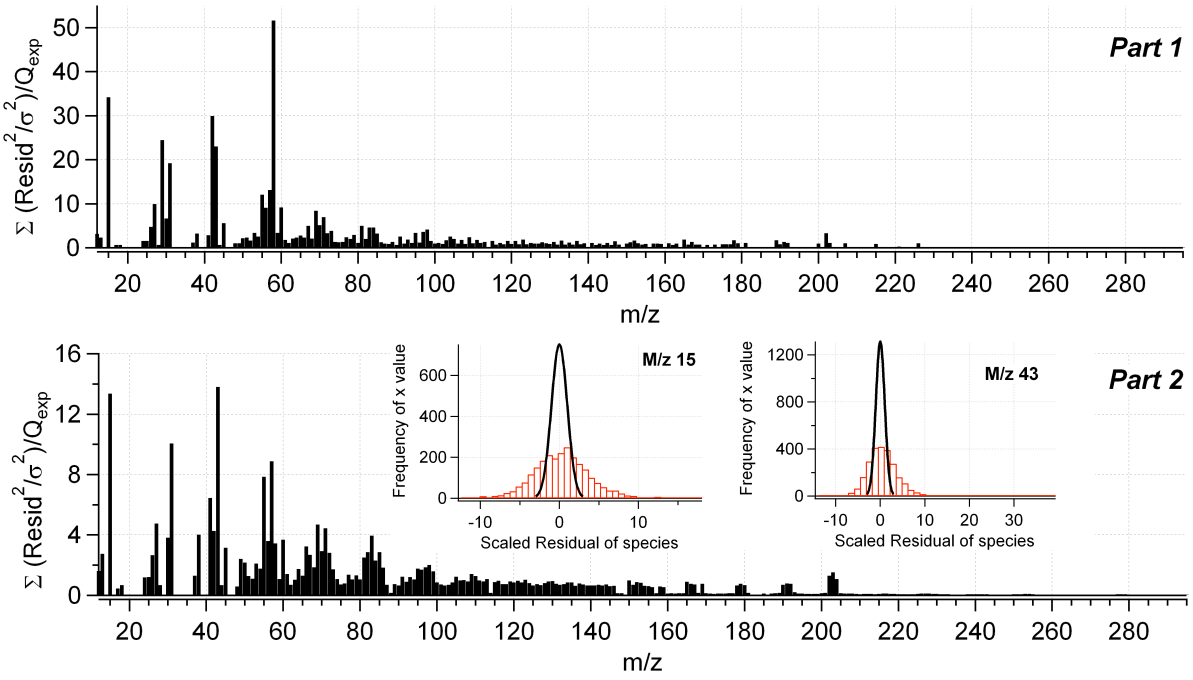

Figure SI- 14: Q/Q_{exp} as a function of *m/z*. Insets show normal distribution of scaled residuals for individual peaks. Note positive bias of distribution of residuals of *m/z* 58.

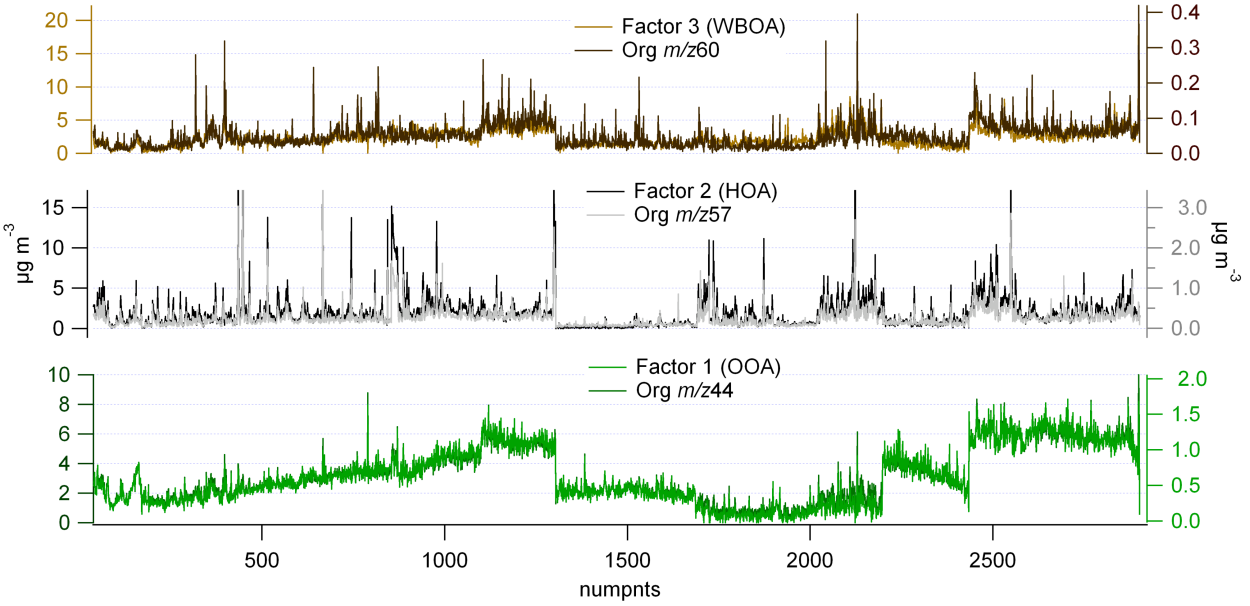

Figure SI- 15: Part 1 – time series of factors and organic marker masses 60, 57, 44.

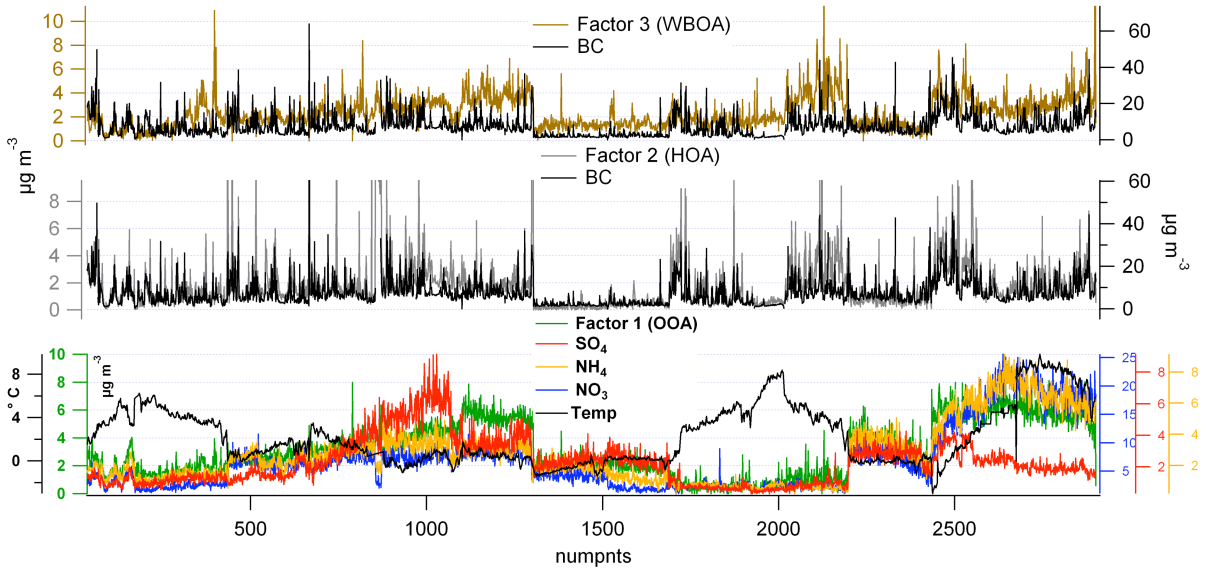

Figure SI- 16: Part 1 – time series of factors and ancillary data.

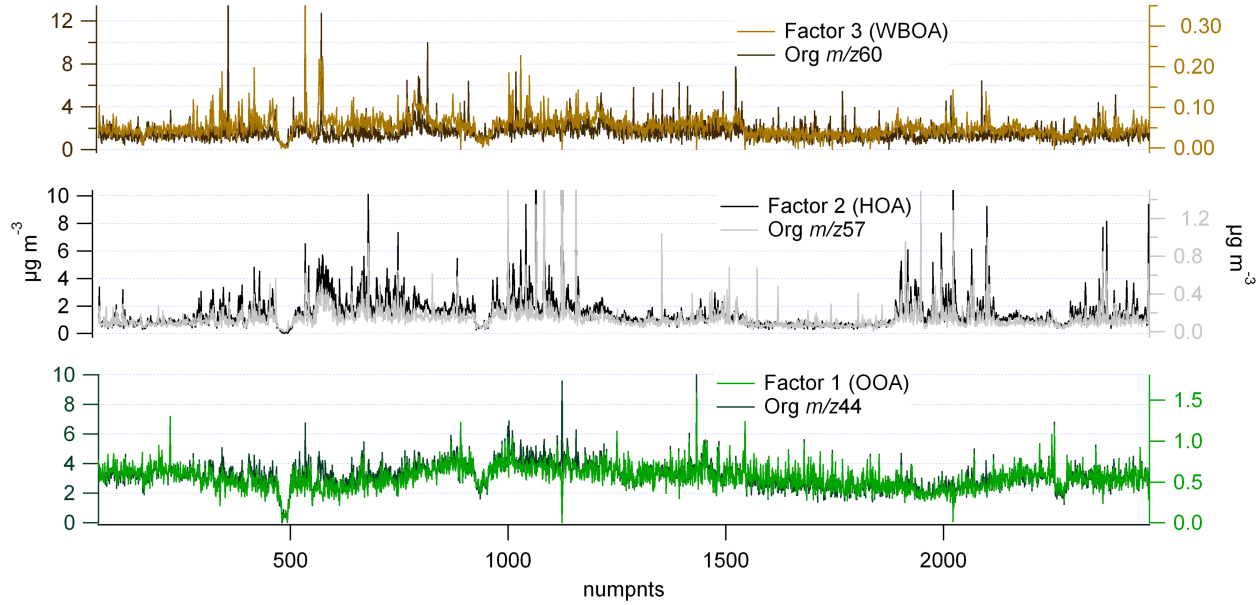

Figure SI- 17: Part 2 - time series of factors and organic marker masses 60, 57, 44.

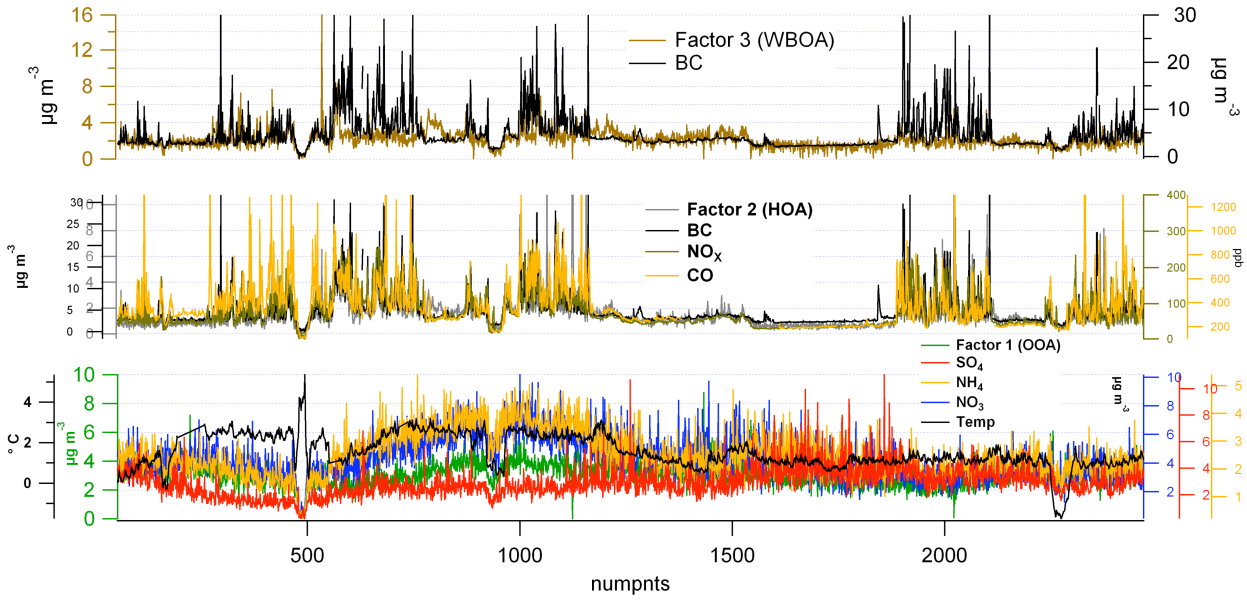

Figure SI- 18: Part 2 - series of factors and ancillary data.

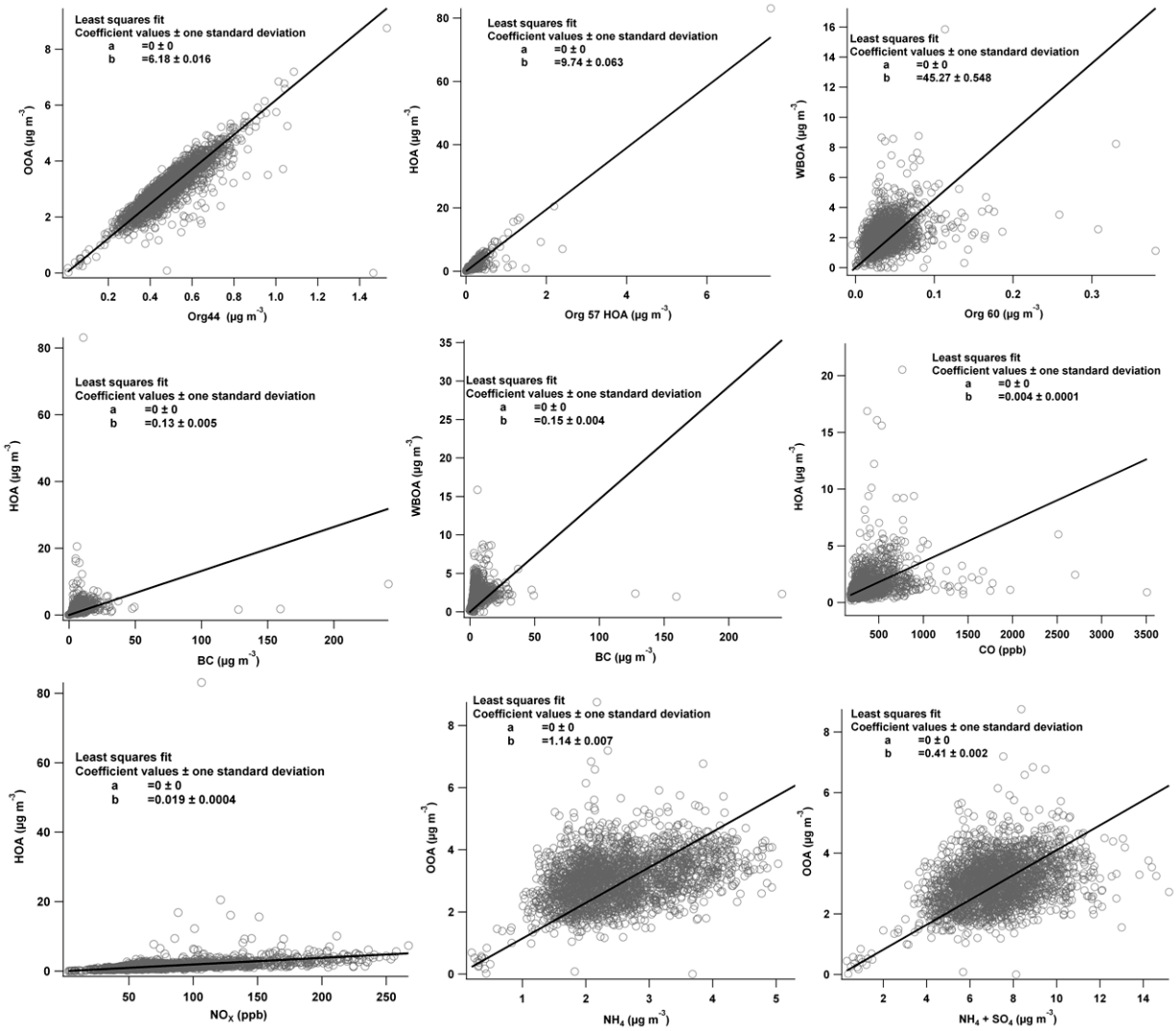

Figure SI- 19: Regression analysis of PMF factor timeseries and ancillary data, no corrections applied.

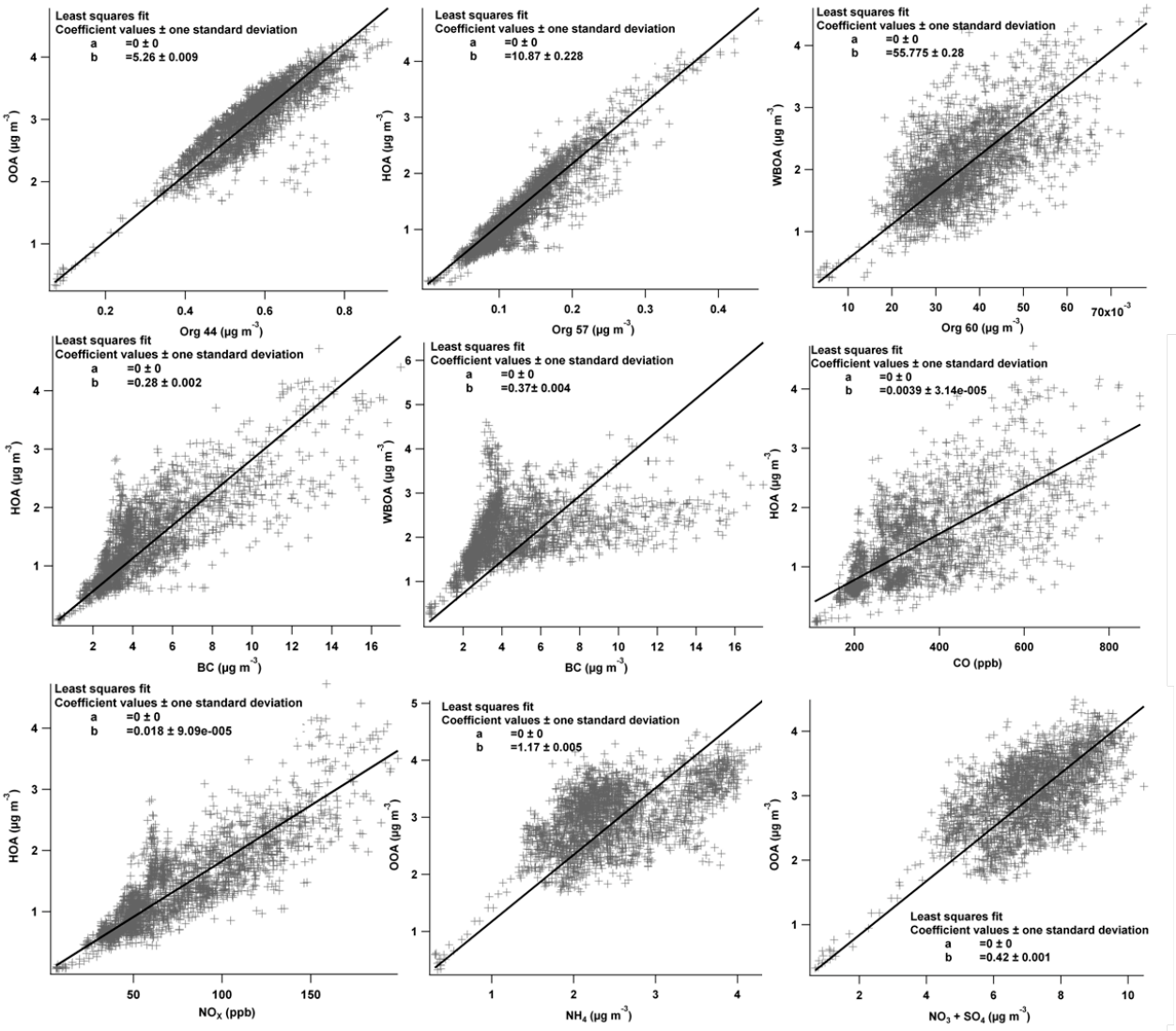

Figure SI- 20: Regression analysis of PMF factor timeseries and ancillary data, after removing the upper 1st percentile of data points and applying a moving average over 5 data points.

4 Estimation of local contribution

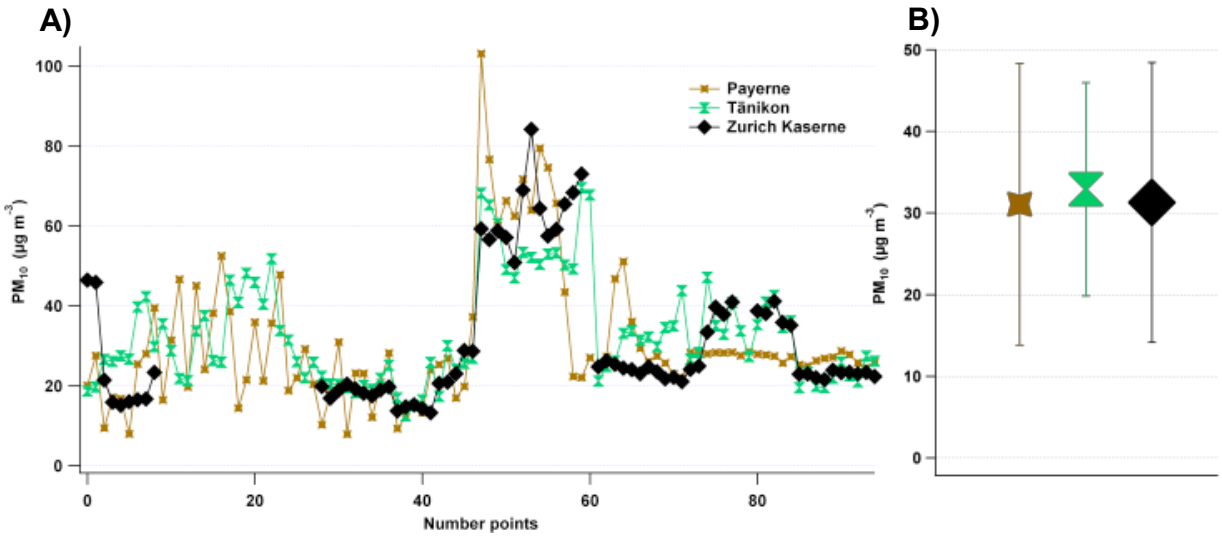

Figure SI- 21: Time series of PM_{10} at Payerne (rural station), Tänikon (rural station), and Zurich Kaserne (urban background station) (panel A) during the same time intervals as the mobile measurements. Panel B shows the mean value and standard deviation of the time series in panel A.

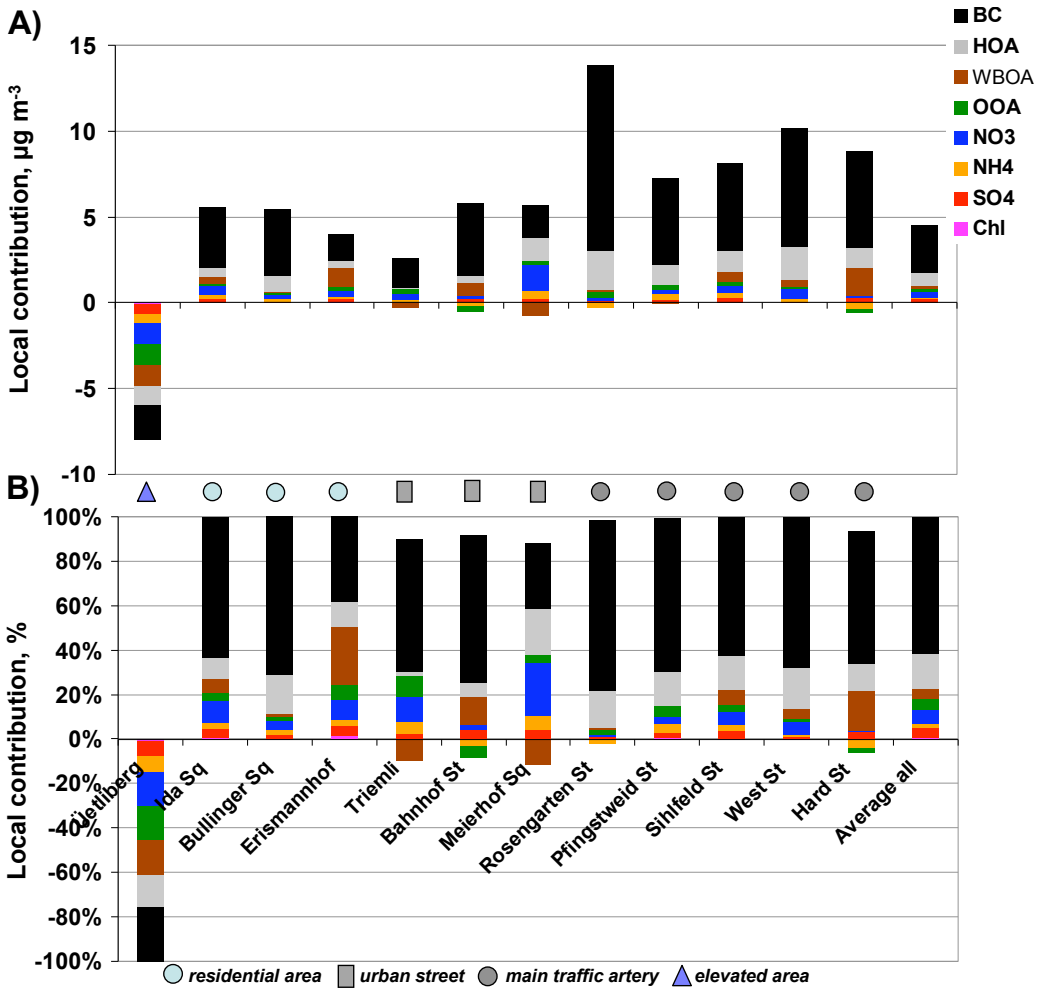

Figure SI- 22: Local concentrations calculated by subtracting the concentration of component *S* measured at Kaserne from the concentration of component *S* measured on-road at the same time. For the time series of Kaserne data, the interpolated median value of 2 subsequent Kaserne visits was used. The "average all" bar is the mean value of the local contribution of all data.

References

Alfarra, M. R., Prevot, A. S. H., Szidat, S., Sandradewi, J., Weimer, S., Lanz, V. A., Schreiber, D., Mohr, M., and Baltensperger, U.: Identification of the mass spectral signature of organic aerosols from wood burning emissions, Environ. Sci. Technol., 41, 5770 - 5777, 2007.

Fay, B., Glaab, H., Jacobsen, I., and Schrodin, R.: Evaluation of Eulerian and Lagrangian atmospheric transport models at the Deutscher-Wetterdienst using Anatex surface tracer data, Atmos. Environ., 29, 2485-2497, 1995.

Koebel, M., Elsener, M., and Kleemann, M.: Urea-SCR: a promising technique to reduce NOx emissions from automotive diesel engines, Catal. Today, 59, 335-345, 2000.

Lanz, V. A., Alfarra, M. R., Baltensperger, U., Buchmann, B., Hueglin, C., and Prevot, 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.

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

Silva, P. J., Erupe, M. E., Price, D., Elias, J., Malloy, Q. G. J., Li, Q., Warren, B., and Cocker, D. R.: Trimethylamine as precursor to secondary organic aerosol formation via nitrate radical reaction in the atmosphere, Environ. Sci. Technol., 42, 4689-4696, 10.1021/es703016v, 2008.