

We thank the referees for their comments, which helped improving the quality of our manuscript. A point by point response (in blue) to the reviewers' comments (in black, italics) will follow. Changes in the text are indicated in in black.

### **P. Paatero (Referee)**

*This work develops first a mathematical machinery whereby they attempt to formulate a small number of factor spectra (2 assumed "known" mass spectra plus 4 fitted mass spectra) whereby a large number (819) of mass spectra, measured in 9 locations in different times of the year, could be successfully fitted. In the second part, they perform a very large number of repetitions of the modeling, so that different details of the model are varied. From variation of the results, they deduce reliability estimates for the obtained spectral profiles and contribution time series.*

*The mathematical operations are not adequately explained. It is impossible to know what they have done (and why) in different stages of the work. One of their key concepts is a constraining of F factor elements. Unfortunately, the defining equation (3) is so unclear (and possibly contains a typo) that I cannot even guess what they might mean by this equation. This review is limited to the first part of the work, modeling the measured mass spectra by a bilinear model. Although the second part is important, and has required a large amount of work, it cannot be analyzed in the time that would be available for such analysis. Thus I do not comment on the second part. I recommend that this manuscript should be published after the listed problems in mathematical presentation and elsewhere have been corrected.*

#### *Problems in the mathematical presentation*

*The fundamental principle of science is repeatability. Used methods should be defined clearly and in sufficient detail so that colleagues will be able (at least in principle) to repeat what was done by the authors. In the following, I discuss mathematics that have not been described in an understandable manner. In general, equations are the language of mathematics. Mathematical work should be described using equations. Verbal explanations may only help in understanding the equations, they cannot replace equations. In order to use equations, it is necessary to define symbols for various quantities. If necessary, use two-letter symbols.*

*Consider typographic questions: The prime should be avoided. Complicated notation in subscripts or in superscripts is often very difficult to read.*

We have taken into account the reviewer comments, clarified the mathematical operations used and their purpose and added/modified mathematical equations when needed. In the following we respond to the different points raised by the reviewer and indicate the changes in the text we have made.

*P5 L9-11 (page 5, lines 9-11)*

*.. by the estimated organic matter (OM) concentration, calculated as the product of the OC concentrations .. and the OM/OC ratios from the AMS measurements.*

*Reading this verbal explanation, it is very hard to understand what was done. In fact, I would need to try to write the equations in order to understand. Please do this equation writing for the benefit of your readers!*

We have added an additional equation (now equation 3), to clarify the quantification operation. The text reads as follows:

“... ”

Input data and error matrices consisted of 202 organic ions. The organic fragments,  $x'_{i,j}$ , obtained from offline AMS analyses do not directly represent ambient concentrations. Therefore, the signal of each fragment was converted to such an ambient concentration ( $x_{i,j}$  in  $\mu\text{g m}^{-3}$ ), by multiplying the fraction of this signal with the estimated organic matter (OM) concentration. The latter was calculated as the product of the OC concentrations measured by the Sunset OC/EC analyzer and the OM/OC ratios from the offline AMS measurements  $(\text{OM/OC})_{\text{oAMS}}$  (Eq. 3). Note that such scaling does not change the outcome of Eq. 2 since both data and error matrices are scaled in the same manner and the fingerprints ( $f_{k,j}$ ) are not changed.

$$x_{i,j} = \frac{x'_{i,j}}{\sum_i x'_{i,j}} * OC * (\text{OM/OC})_{\text{oAMS}} \quad (3)$$

“... ”

*P5 EQ(3): I do not understand eq(3) at all.*

*- what are  $f_{kn}$  and  $f'_{kn}$  (or is it  $f_{kn}$ '), what difference is there between them?*

*- is there a typo in the equation? As shown, the equation appears impossible.*

*This is a key detail in the manuscript as it describes constraints that perhaps have never been published in similar work. It must be described so that it is understandable.*

We do agree with the reviewer that the equation is misleading. The constraints were performed in the same way as in Canonaco et al. (2013) and the equation in the text was adapted to the one in the mentioned publication. The modified text reads as follows:

“... The PMF algorithm was solved using the multilinear engine-2 (ME-2, Paatero, 1999). Normalization of the PMF solution during the iterative minimization process is disabled as implemented in SoFi (Canonaco et al., 2013). ME-2 enables an efficient exploration of the solution space by a priori constraining the  $f_{k,j}$  elements within a certain range defined by the scalar  $a$  ( $0 \leq a \leq 1$ ) from a starting value  $f'_{k,j}$ , such that the modelled  $f_{k,j}$  in the solution satisfy Eq. 4:

$$f_{k,j} = f'_{k,j} + a * f'_{k,j} \quad (4)$$

$f'_{k,j}$  is the starting value used as a priori knowledge from previous studies and  $f_{k,j}$  is the resulting value in the solution. ...”

*What was included in matrix X, to be modeled by PMF. It remains unclear what information was contained in X. Section 2 lists a number of variables whose concentrations were measured. Were all these included in X, or were some of these included, or none of them?*

P5 L8-9 say: *Input data and error matrices included 202 organic ions. Do you mean that input data consisted of 202 organic ions? The formulation you used may also be interpreted so that among other data, input data also included 202 organic ions.*

This sentence was indeed misleading. **X** consisted of 202 ions (variables) measured over n time points. The text has been adapted accordingly:

“... Input data and error matrices consisted of 202 organic ions. ....”

*Also regarding matrix X:*

*Input data ... were rescaled by the estimated organic matter (OM) concentration ... In PMF, one is allowed to rescale data rows in any way, provided that error rows are also scaled in the same manner. Thus your scaling is OK. However, it would help the reader a lot if you state briefly why to scale, what advantage was achieved by scaling. Your scaling (which does not change profile spectra in any way) is useful for plotting figure 9. On the other hand, your scaling influences (improves or worsens) correlations between factor values and marker concentrations. Please explain and/or correct. Note that this scaling does not change the computations performed according to Eq(4) in any way.*

The scaling procedure does not influence in any way the PMF calculations. However, as the offline AMS signals do not directly represent ambient concentrations the factors retrieved should be scaled to the real ambient concentrations, before any correlation with external markers and data interpretation is made. Thus the scaling could indirectly affect the result by influencing the solution selection. The scaling could be achieved before or after running the PMF (as it was mentioned correctly by the reviewer the scaling does not change the computations performed according to Eq. 4). We have scaled the data before running PMF for computation reasons as in this way we do not need to scale the results obtained from every PMF solution repeatedly. The text related to this section has been adapted as follows:

“... ”

Input data and error matrices consisted of 202 organic ions. The organic fragments,  $x'_{i,j}$ , obtained from offline AMS analyses do not directly represent ambient concentrations. Therefore, the signal of each fragment was converted to such an ambient concentration ( $x_{i,j}$  in  $\mu\text{g m}^{-3}$ ), by multiplying the fraction of this signal with the estimated organic matter (OM) concentration. The latter was calculated as the product of the OC concentrations measured by the Sunset OC/EC analyzer and the OM/OC ratios from the offline AMS measurements  $(OM/OC)_{oAMS}$  (Eq. 3). Note that such scaling does not change the outcome of Eq. 2 since both data and error matrices are scaled in the same manner and the fingerprints  $(f_{k,j})$  are not changed.

$$x_{i,j} = \frac{x'_{i,j}}{\sum_i x'_{i,j}} * OC * (OM/OC)_{oAMS} \quad (3)$$

“... ”

P5 L5-7 (input errors):

> *The input errors ... include ... the uncertainty related to ion counting statistics and ion-to-ion signal variability at the detector.*

*I understand that you counted the ions. Then, ion-to-ion signal variability is \*not\* a source of uncertainty. If ion current is measured, instead of counting ions, then ion signal variability \*is\* a source of error. Please correct or clarify.*

The AMS does not count ions, but rather an integrated signal (bits) that can be related to ion counts by means of a single ion calibration. Thus single ion signal variability is a source of uncertainty for PMF analysis, in addition to counting statistics. Both sources of errors are included in Allan et al. (2003, alpha and I, respectively).

Allan, J. D., Jimenez, J. L., Williams, P. I., Alfarra, M. R., Bower, K. N., Jayne, J. T., Coe, H., and Worsnop, D. R.: Quantitative sampling using an Aerodyne aerosol mass spectrometer 1. Techniques of data interpretation and error analysis, *J. Geophys. Res.*, 108, 4090, doi:10.1029/2002JD002358, 2003.

P5 L20-24

*describe a complicated method for filling knowledge gaps in the known/assumed fixed factor spectra ("reference profiles") of HOA and COA. I am not fully convinced about the performance of this method. The "natural" alternative method is to leave the unknown elements in reference profiles (profiles of HOA and COA) as ordinary factor elements, to be fitted by ME-2 together with all "normal" non-fixed F factor elements, as explained in detail below. Please include this remark in the corrected ms, so that future colleagues are encouraged to follow the safer and simpler method instead of your complicated method.*

We added the following statement to the manuscript, based on the reviewer's suggestions:

"... Alternatively, such missing ions can be also treated as ordinary factor elements, to be fitted by ME-2 with all other ordinary factor elements. ..."

*Using "constrained factors" based on known profiles of HOA and COA. This topic was very difficult to understand at first. It was not clear what is "constrained" by what. Now I assume that you mean the following: In all PMF runs, two constant F factors were used, i.e. two rows of factor F were defined as a-priori fixed, so that the values of these constant factors were set equal to previously known mass spectra of HOA and COA. Is this what you mean? – Using constant or constrained factors is not familiar to PMF users, not at all. Such unusual methodology must be carefully explained so that all readers have the possibility to understand what you have done. In particular, you should explain that using fixed factors is not the same as using inequalities in order to constrain factors to lie between upper and lower limits, set very close to each other. Also, you should go into technical details here, because it is possible to implement constant factors in two different ways in ME-2. You should guide your readers to the optimal usage. It is possible to use "constant factors" that reside in a different matrix, which is clumsy. The alternative is to keep all F factors in the same matrix but define that the elements in two first (say) rows of F are "locked", not allowed to change during the fitting process. These elements are set equal to the known profile values before initiating the fit. If there are gaps in the knowledge of HOA and COA, then those unknown elements in "locked" F rows should simply not be locked at all, so that they may obtain their best possible values during the fit.*

For a better readability we adapted the text to:

“...  $f_{k,j}'$  is the starting value used as a priori knowledge from previous studies and  $f_{k,j}$  is the resulting value in the solution. In all PMF runs (unless mentioned otherwise), we used the high resolution mass spectra for HOA and COA (cooking OA) from Crippa et al. (2013b) as constraints, i.e. two rows of  $f_{k,j}$  were set equal to the mass spectra of HOA and COA. Ions that were present in our datasets but not in the reference profiles for HOA and COA were inferred from published unit mass resolution (UMR) profiles (Ng et al., 2011 and Crippa et al., 2013c). For this purpose, the fraction of signal at a specific  $m/z$  in the UMR reference spectrum ( $f_{UMR,m/z}$ ) was compared to the fraction of signal of all ions at this  $m/z$  in the HR reference spectrum ( $f_{HR,m/z}$ ). The difference  $f_{UMR,m/z} - f_{HR,m/z}$  was used as entries in  $f_{k,j}'$  for such missing ions. For these ions, an  $\alpha$ -value of unity was set. For the other factors, the factor elements were fitted by ME-2. Alternatively, such missing ions can be also treated as ordinary factor elements, to be fitted by ME-2 with all other ordinary factor elements.

...”

*Use of constant factors or constrained factors often causes so-called "normalization conflicts". How did you protect your bilinear model against normalization conflicts? This is another important detail that should be communicated to colleagues who might follow your example.*

We believe that the reviewer comment on normalization conflicts is referring to the normalization equations for the factor contribution present in the default ME-2 instruction file. If so, when calling ME-2 from SoFi, these normalization equations are disabled.

We added this technical aspect to the manuscript (P5 L18-23):

“... The Source Finder toolkit (SoFi v.4.9, Canonaco et al., 2013) for Igor Pro software package (Wavemetrics, Inc., Portland, OR, USA) was used to configure the PMF model and for post- analysis. The PMF algorithm was solved using the multilinear engine-2 (ME-2, Paatero, 1999). Normalization of the PMF solution during the iterative minimization process is disabled as implemented in SoFi (Canonaco et al., 2013). ME-2 enables an efficient exploration of the solution space by a priori constraining the  $f_{k,j}$  elements within a certain range defined by the scalar  $\alpha$  ( $0 \leq \alpha \leq 1$ ) from a starting value  $f_{k,j}'$ , such that the modelled  $f_{k,j}$  in the solution satisfy Eq. 4: ...”

P7 L16-17

> For each of the four PMF datasets, 2420 PMF runs were performed for  
> evaluating the sensitivity of the model to the chosen  $\alpha$ -value and the seed.

*This statement mentions sensitivity of the model to random seed. The random seed determines the pseudorandom initial values of PMF fit. In plain language, this statement says that there were local solutions so that depending on seed, PMF iteration converged to different local solutions. Presumably, these solutions had comparable Q values because otherwise, Q values would be used for selecting between solutions. Now these different solutions are somehow pooled together and their presence is otherwise ignored.*

*The presence of multiple solutions should be properly reported (e.g. how many of PMF runs had multiple solutions, how many different individual solutions per PMF run were obtained at most and on the average, are the solutions rotationally equivalent having identical residuals of fit, etc.) There are no fixed rules on what to do with multiple solutions. On one extreme, it has been suggested that scientists may at will pick the one solution they like most and ignore the others. At the other extreme,*

*PMF modeling of such data may be considered failing if there are several local solutions with comparable Q values.*

For each of the four PMF datasets, 2420 PMF runs were performed for evaluating the sensitivity of the model to the chosen  $\alpha$ -value and the seed. The quality of each of the 2420 PMF runs was individually assessed using the criteria lined out in Sec. 3.6. We assessed the environmental interpretability based on the correlation of factor time series and markers of the respective source. All solutions fulfilling these criteria are accepted and considered plausible. Using these criteria 331 PMF runs were selected for PMF<sub>block</sub> (for PMF<sub>zue,isol</sub> 230, for PMF<sub>zue,reprs</sub> 99, and for PMF<sub>1filter/month</sub> 269). This information was added to the text. We found that 80% of the accepted solutions have an  $\alpha$ -value $\leq$ 0.3 for HOA and an  $\alpha$ -value $\leq$ 0.5 for COA. The output HOA and COA factor profiles are therefore not significantly variable and very similar to the input profiles, indicating that similar solutions were selected. Furthermore, the yearly average factor concentrations of all selected PMF<sub>block</sub> solutions after  $R_k$  correction are now shown for the case of Zurich as an illustration in Fig. S5. The distributions of each of the different factors do not show more than 1 distinct mode, indicating that we do not have several populations of solutions. This information was added to the supplementary. In a last step, all accepted solutions are pooled together. In Sec. 3.3, we added a remark linking to Sec. 3.6:

The modified Sec. 3.6 reads:

“... Monte Carlo simulations were performed and simulations for which res-OC<sub>i</sub> distributions were significantly different from 0 ( $Q_{25}<0<Q_{75}$ , details in SI) were discarded until 500 acceptable simulations were found. Thereby, 331 PMF runs were selected for PMF<sub>block</sub> (230 for PMF<sub>zue,isol</sub>, 99 for PMF<sub>zue,reprs</sub>, and 269 for PMF<sub>1filter/month</sub>). Median factor time series and recovery parameters from all retained simulations were then determined and the interquartile range (IQR) represents our best estimate of the uncertainties for the single PMF datasets. The Monte Carlo process was repeated for the four different PMF datasets described above and the resulting median time series of their estimated uncertainties were compared. The resulting uncertainty estimates and the method are described in Sec. 4.2.1. and in the SI. ...”

The added information in the SI reads:

“... ”

Cumulative density functions for the  $\alpha$ -values of HOA and COA are presented for the accepted solutions in Fig. S4. We found that 80% of the accepted solutions have an  $\alpha$ -value $\leq$ 0.3 for HOA and an  $\alpha$ -value $\leq$ 0.5 for COA. The output HOA and COA factor profiles are therefore not significantly variable and very similar to the input profiles, indicating that similar solutions were selected. Furthermore, the yearly average factor concentrations of all selected PMF<sub>block</sub> solutions after  $R_k$  correction are shown for the case of Zurich as an illustration in Fig. S5. The distributions of each of the different factors do not show more than 1 distinct mode, indicating that we do not have several populations of solutions.

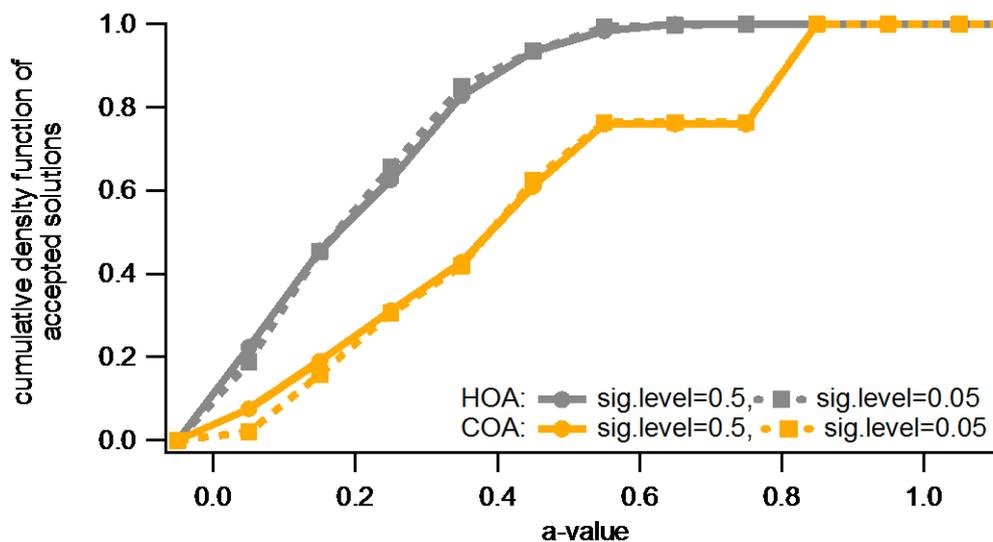


Figure S4: Cumulative density functions of a-values for HOA and COA for the accepted solutions.

The yearly average factor concentrations of all selected  $PMF_{block}$  solutions after  $R_k$  correction are shown for the case of Zurich as an illustration (Fig. S5). The distributions of each of the different factors do not show more than 1 distinct mode.

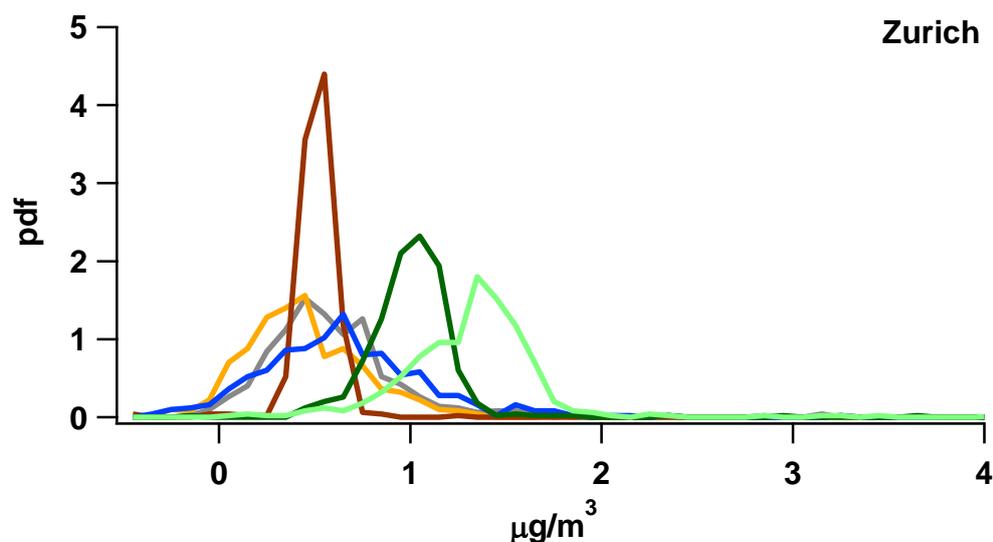


Figure S5: Histograms of yearly average factor concentrations of all selected  $PMF_{block}$  solutions (after  $R_k$  correction)....”

*If DISP is used for uncertainty estimation of a case with several local solutions, one often obtains the outcome that the model is "Not Well Defined" or "NWD". I would not suggest what the authors should do with their many-solution cases in addition to discussing them. Whatever they opt to do, they should describe it: what was done and why.*

The procedure that we have adopted aims at assessing the sensitivity of the PMF to the constrained profiles (a-value), to the pseudo-random starting point (seed), to the measurement uncertainties (by

including repeated measurements from one of the sites) and to the input data/rotational ambiguity (bootstrap). We have discarded a number of solutions based on a set of criteria. The PMF runs satisfying these criteria were all considered as plausible and discussed in the manuscript. Differences amongst these solutions are used as an estimate of the combined uncertainties mentioned above. The process is described in Sec. 4.2.1. and in the Supplementary Information. We added the following sentence at the end of Sec. 3.6 linking to Sec. 4.2.1.

"... The resulting uncertainty estimates and the method are described in Sec. 4.2.1. and in the SI. ..."

### Section 3.3, sensitivity analysis

*I cannot comment more on this analysis because I do not understand what the  $\alpha$ -values are and how they were used.*

We do agree with the reviewer that the equation is misleading. The constraints were performed in the same way as in Canonaco et al. (2013) and the equation in the text was adapted to the one in the mentioned publication. The modified text reads as follows:

"...

The PMF algorithm was solved using the multilinear engine-2 (ME-2, Paatero, 1999). Normalization of the PMF solution during the iterative minimization process is disabled as implemented in SoFi (Canonaco et al., 2013). ME-2 enables an efficient exploration of the solution space by a priori constraining the  $f_{k,j}$  elements within a certain range defined by the scalar  $\alpha$  ( $0 \leq \alpha \leq 1$ ) from a starting value  $f'_{k,j}$ , such that the modelled  $f_{k,j}$  in the solution satisfy Eq. 4:

$$f_{k,j} = f'_{k,j} + \alpha * f'_{k,j} \quad (4)$$

$f'_{k,j}$  is the starting value used as a priori knowledge from previous studies and  $f_{k,j}$  is the resulting value in the solution.

"..."

P6 L26-28 say:

> Paatero et al. (2014) compare the effectiveness in estimating modelling errors using two different approaches: the displacement (DISP) and bootstrap analysis (BS), respectively.

*Here seems to be a terminological problem: in the quoted paper, Paatero et al. estimated the uncertainties of estimated F factor elements in the situation when no modelling errors are present. These F uncertainties depend mainly on random error in X and on rotational freedom of factor matrices G and F. It was specifically emphasized that the obtained uncertainties do not cover effects of modelling errors in the results. (Examples of modelling errors: non-constant factor profiles, wrong uncertainties assigned to data values.) Thus modelling errors were not estimated in the quoted paper.*

P6 L29:

*DISP involves running PMF several times using randomly perturbed factor profile elements of a reference solution*

*In fact, DISP estimation does not involve any randomness at all. F factor elements are perturbed in a systematic fashion by DISP. Perhaps here is confusion with Monte Carlo methods where random perturbations may be applied. DISP is not Monte Carlo.*

We apologize for the terminological mistake. The respective part of the manuscript has been corrected.

“... While this approach has been proven very effective in selecting a range of environmentally relevant solutions (Elser et al., 2016a, 2016b and Daellenbach et al., 2016), the resulting uncertainties may be underestimated. Paatero et al. (2014) compared the effectiveness in estimating uncertainties of factor elements using two different approaches: the displacement (DISP) and bootstrap analysis (BS). BS involves applying the model to input matrices consisting of a subset of the entire dataset. DISP involves running PMF several times using systematically perturbed factor profile elements of a reference solution, but allowing a defined difference in Q from the reference solution. Both approaches are computationally intensive, especially DISP. Because of such computational limitations the combination of BS and DISP was not feasible for the dataset presented here, especially in combination with  $\alpha$ -value sensitivity tests. Therefore, we chose to perform 4 sensitivity tests performing PMF runs using 4 different input datasets, presented in the following. These sensitivity tests allow conclusions on the stability of PMF analysis when reducing the temporal or spatial resolution as well as the influence of the measurement repeatability.

..”

*In table (3), uncertainty estimates of percentage concentrations are not correctly computed.*

The overall uncertainties on the mass concentrations were estimated as the sum in quadrature of two error terms. The first relates to the variability in the yearly average mass concentrations for the different factors at the different sites calculated by changing the COA and HOA  $\alpha$ -values using the  $PMF_{block}$  setting (where all data points are considered,  $\sigma_a$ ). The second error term relates to the additional variability in the yearly average mass concentrations for the different factors at the different sites due to the change in the input matrix ( $PMF_{block}$ ,  $PMF_{zue,isol}$ ,  $PMF_{zue,repr}$ ). As the latter was determined only for the case of Zurich, then the same additional relative error ( $\sigma_b$ ) was assumed for the other sites. This is indeed only a best estimate of the errors. In the corrected version of the manuscript we clarify the way these errors were determined (in Caption to Tab. 3) and we do not show errors on the percentage concentrations.

**“...Table 1: Yearly average contribution and uncertainty of resolved factors for  $PMF_{block}$  run for the different sites and the average for all sites. The uncertainty is calculated based on the variability of the yearly averages from  $PMF_{block}$  and the variability between the sensitivity tests. ...”**

*Notation:*

*In supplement, subscript "i" is used as a subscript of Q. It is not defined what "i" means here. Does it mean number of factors? If yes, then the symbol used for number of factors should be used. If i does not have a definable meaning, then it might be clearer to omit the subscript in this case. In general, systematic use of subscripts would be a help for the reader. E.g. use i only as the index of sample (time), j as index of column of X, k index of factor. For other quantities, select other symbols and define what they mean.*

The index  $i$  in this figure and the text was removed.

*In different places, factor elements are denoted as  $G_{ik}$  and as  $g_{ik}$ . This may confuse readers. Eventually they will recognize that this difference does not mean anything, but first they will waste time trying to understand. Either, select one notation (preferred), or, in the section "Notation" (to be written), specify that  $G_{ik}$  and  $g_{ik}$  mean the same, and also  $F_{kj}$  and  $f_{kj}$  mean the same.*

$G_{ik}$  was replaced by  $g_{ik}$ .

*Eq (2) is incorrect. If you wish to use matrix element notation, then summation over  $k$  must be indicated. If you wish to use vector-matrix notation, then its use should be defined, especially because many of your readers are not familiar with such notation. In vector-matrix notation, index  $k$  would not be visible.*

Equation 2 and the adjacent text have been corrected:

“...

Source apportionment of the organic aerosol is performed using positive matrix factorization (PMF, Paatero, 1994). PMF is a statistical un-mixing model explaining the variability of the organic mass spectral data ( $x_{i,j}$ ), as linear combinations of static factor profiles ( $f_{j,k}$ ) and their time-dependent contributions ( $g_{i,k}$ ), see Eq. 2 (where  $p$  is the number of factors). The index  $i$  represents a specific point in time,  $j$  an ion, and  $k$  a factor. The elements of the model residual matrix are termed  $e_{i,j}$ .

$$x_{i,j} = \sum_{k=1}^p g_{i,k} f_{k,j} + e_{i,j} \quad (2)$$

...”

P5 L10: sunset → Sunset OC/EC analyzer

Has been corrected.

P5 L21

> Fitted ions in our datasets missing in the reference ...

What do you mean by "fitted ions"? I do not understand this sentence.

The sentence was rephrased to:

“...Ions that were present in our datasets but not in the reference profiles for HOA and COA were inferred from published unit mass resolution (UMR) profiles (Ng et al., 2011 and Crippa et al., 2013c). ...”

P7 L21:

> The identity of HOA and COA were identified first as their mass spectra were initially constrained.

Why do you need to identify HOA and COA? I would assume that they are on preselected rows of  $F$ , such as rows 1 and 2. No identification is needed for factors that are in known positions. What is wrong here? Am I understanding all this completely wrong?

Yes, indeed HOA and COA are on preselected rows of F. The text has been adapted to:

“... Since HOA and COA were initially constrained on preselected rows of F, they did not need to be identified. ...”

*Supplement*

*Figure SI.1*

*shows ratios of obtained Q vs. expected Q<sub>exp</sub>. How were Q<sub>exp</sub> computed? Did you take into account that downweighted columns of X contribute very little to Q<sub>exp</sub>? How many downweighted columns were present?*

The results displayed in FigS1 refer to the distributions of  $Q_{i,j}$  (median and quartiles and not to  $Q/Q_{exp}$ .  $Q_{i,j}$  is computed as:

$$Q_{i,j} = \left( \frac{e_{i,j}}{s_{i,j}} \right)^2$$

We corrected the text and axis labels accordingly and refer only to  $Q_{i,j}$ .

In PMF<sub>block</sub>, only 7 of 202 ions were downweighted by a factor 3. Therefore, the bias in Q between using the non-downweighted or the downweighted  $s_{i,j}$  is smaller than the variability in Q for the same set of solutions. Thus we do not think that this bias has an influence on the results, all the more since the absolute Q-values are not used in the solution selection.

*Obtained ratios Q/Q<sub>exp</sub> are of the order of 10 for 6 factors. This indicates that there are one or several significant modeling errors. It has been assumed among atmospheric scientists that a ratio of 4, say, would not be significant, that it could be caused by random variations from the expected Q. This assumption is totally wrong. Such ratios (>1.5, say) always have a cause that preferably should be understood in a project where careful mathematical analysis is attempted.*

*Possible causes of modeling errors are: underestimation of random errors in mass spectra, variation of factor profiles with time and between sites, systematic errors in preprocessing m/z spectra, and spurious sporadic local sources that cannot be modeled by PMF. An attempt should be made at understanding and discussing those errors, even if the effect to obtained results cannot be eliminated any more at this final stage. One useful diagnostic is to examine contributions to Q from different m/z values, from different times of day and days of week, and so on.*

Variation of factor profiles between sites:

We have explicitly represented the average  $Q_{i,j}$  per site and per season to examine the performance of the model and reasons behind the observed  $Q_{i,j}$  values. As we have already pointed out in the manuscript the model does get the data at sites in the north better than at sites in the South during winter. We did attribute this difference to the variability in the wood burning source profiles, here represented by only one factor. Indeed, we do agree that additional reasons may result in the high  $Q_{i,j}$  observed. These are discussed below.

First, as pointed out above we do represent sources at all sites with a single profile and the examination of the site-to-site median  $Q_{i,j}$  values do indicate a discrepancy in the model performance between sites in the north and south of the Alps. In comparison to PMF<sub>block</sub>, the average  $Q_{i,j}$  for Zurich is slightly reduced for the same number of factors when only including 1 site in PMF (PMF<sub>zue,isol</sub>, PMF<sub>zue,reprs</sub>, average  $Q_{i,j}$  6). The difference between the site that is best explained and

the site that is least explained is also approximately 6 (Fig. SI.1). Therefore, we do indeed attribute a significant part of the unexplained variability to variation in source profiles at different sites.

Variation of factor profiles with season:

We note that these differences can also occur because of representing SOA at different seasons with only two factors: For example the difference in Q values between the best explained season (April-Mai-June) and the least well explained season (January-February-March) is approximately 4 (Fig. SI.1). A consistent difference between Q for week-days and weekends cannot be found (Fig. S2).

Underestimation of random errors in mass spectra

As mentioned in the manuscript, the samples from Zurich were measured at two instances, 5 months apart. When performing PMF on either batch separately (PMF<sub>zue,isol</sub> and PMF<sub>zue,rep</sub>, respectively) their average  $Q_{i,j}$  is comparable (Q=6). Therefore, there does not seem to be a difference in the representation of random errors between these measurements.

Systematic errors in preprocessing m/z spectra

As we have mentioned in the manuscript the error matrix considers the detector counting statistics and the variation in the background. However, additional uncertainties can result from peak fitting of the high resolution AMS data (attribution of signal at a nominal mass to several ions). This procedure is affected by errors in the m/z calibration accuracy and precision and by the peak width. These errors are not currently taken into account, but may result in a significant underestimation of the measurement uncertainties especially for overlapping peaks, which may explain (at least partially) the high average  $Q_{i,j}$ .

In Fig. S3, we present the average  $Q_{i,j}$  of individual ions. There is no clear dependence on the ion molecular weight and therefore, m/z. However, the average  $Q_{i,j}$  for ions with a small mass defect (nominal mass – exact ion mass) is higher (~10) than for the other ions (~3). There are a lot of ions with small mass defects close to 0.03 a.m.u. in our dataset. This makes these peaks prone to overlap with other ions and thus their error subject to underestimation because this effect is not considered in the  $s_{ij}$  calculation.

Sporadic local sources that cannot be modeled by PMF:

The ion families CHN and especially CS show higher average Q than CH, CHO1, and CHOgt1. Since CH<sub>3</sub>SO<sub>2</sub><sup>+</sup> shows an event-driven time series, at least the high Q related to CS ions could be related to the inability of PMF to resolve these events accurately.

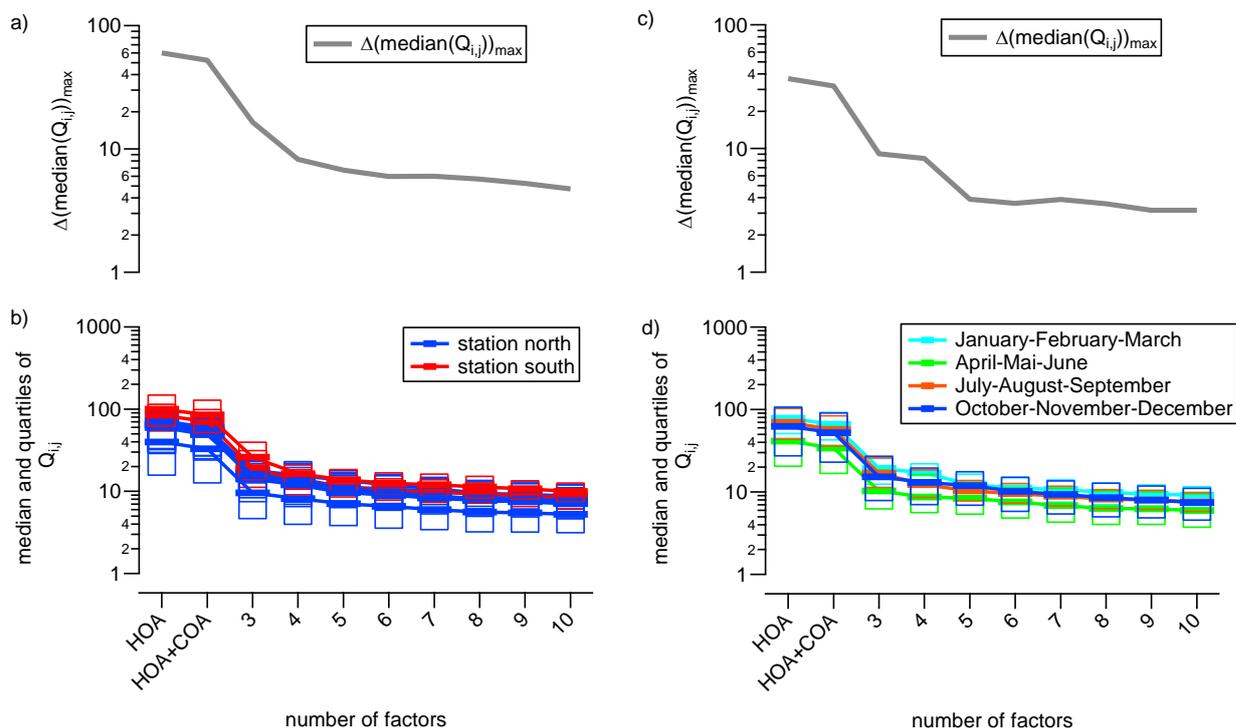
The corresponding section in the SI has been thoroughly reworked and reads now:

“... ”

Based on the input data for PMF<sub>block</sub>, we evaluate the influence of the number of factors, p, on  $Q_i$ . For this experiment, both the traffic and cooking signatures were constrained using adapted reference profiles from Crippa et al. (2013b) as described in section III.1. Based on this evaluation, we chose to perform PMF using 6 factors.

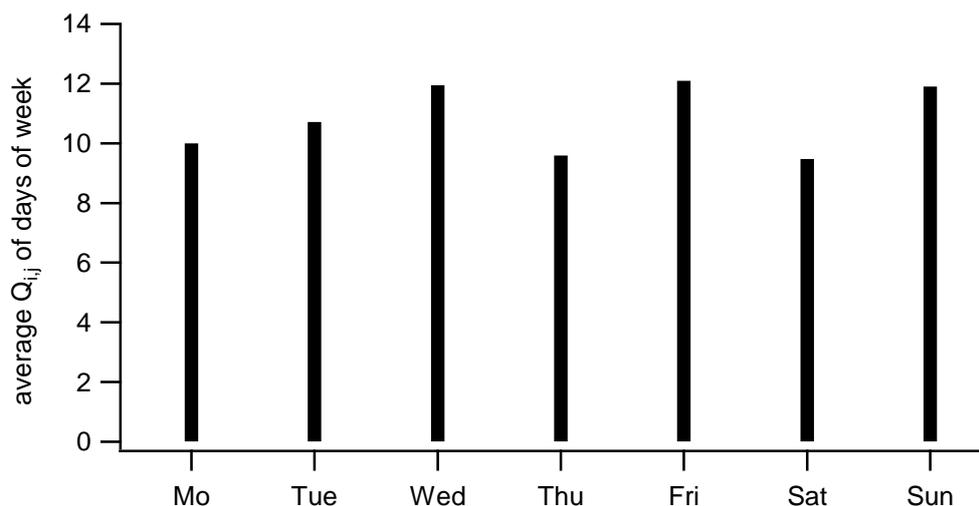
$Q_{i,j}$  is computed using the PMF residuals ( $e_{ij}$ ) and the PMF input errors ( $s_{i,j}$ ):

$$Q_{i,j} = \left( \frac{e_{i,j}}{s_{i,j}} \right)^2 \quad (\text{S1})$$



**Figure S1:  $Q_{i,j}$  as a function of the number of factors for a reference experiment with all data used in PMF (9 sites, full year 2013, HOA and COA constrained with  $a=0.0$  (b and d).  $\Delta(\text{median}(Q_{i,j}))_{\text{max}}$  is evaluated for the different periods during the year 2013 (January-February-March, April-Mai-June, July-August-September, October-November-December) and for all sites (a and c). The grey line depicts the difference between the category (geographical or season) with the highest and the lowest median  $Q_{i,j}$ .**

Fig. S1 shows  $Q_{i,j}$ s as a function of the number of factors for different sites (b) and seasons (d) and the difference between the highest (a) and lowest (c) median to evaluate the maximal difference in the mathematical quality of the solutions. As expected, forcing PMF to explain the variability in the dataset only with the 2 constrained factors ( $p=2$ ), results in very high median  $Q_{i,j}$ .  $\Delta(\text{median}(Q_{i,j}))_{\text{max}}$  shows the difference in the median  $Q_{i,j}$  between groups of points like sites or season. The smaller the  $\Delta(\text{median}(Q_{i,j}))_{\text{max}}$ , the smaller are the differences in the mathematical quality of the PMF solution for the different seasons/sites. To explain the temporal and geographical variability at least 5 factors are required. However, the difference between the site that is best explained and the site that is least explained is approximately 6 when using 5 or 6 factors. When increasing to 6 factors, also a factor explaining the variability of sulfur-containing organic ions (especially,  $\text{CH}_3\text{SO}_2^+$ ) is resolved. Therefore, we opted to perform PMF using 6 factors. Using 6 factors, there is also no difference between the average  $Q_{i,j}$  on week-days and weekend (Fig. S2).

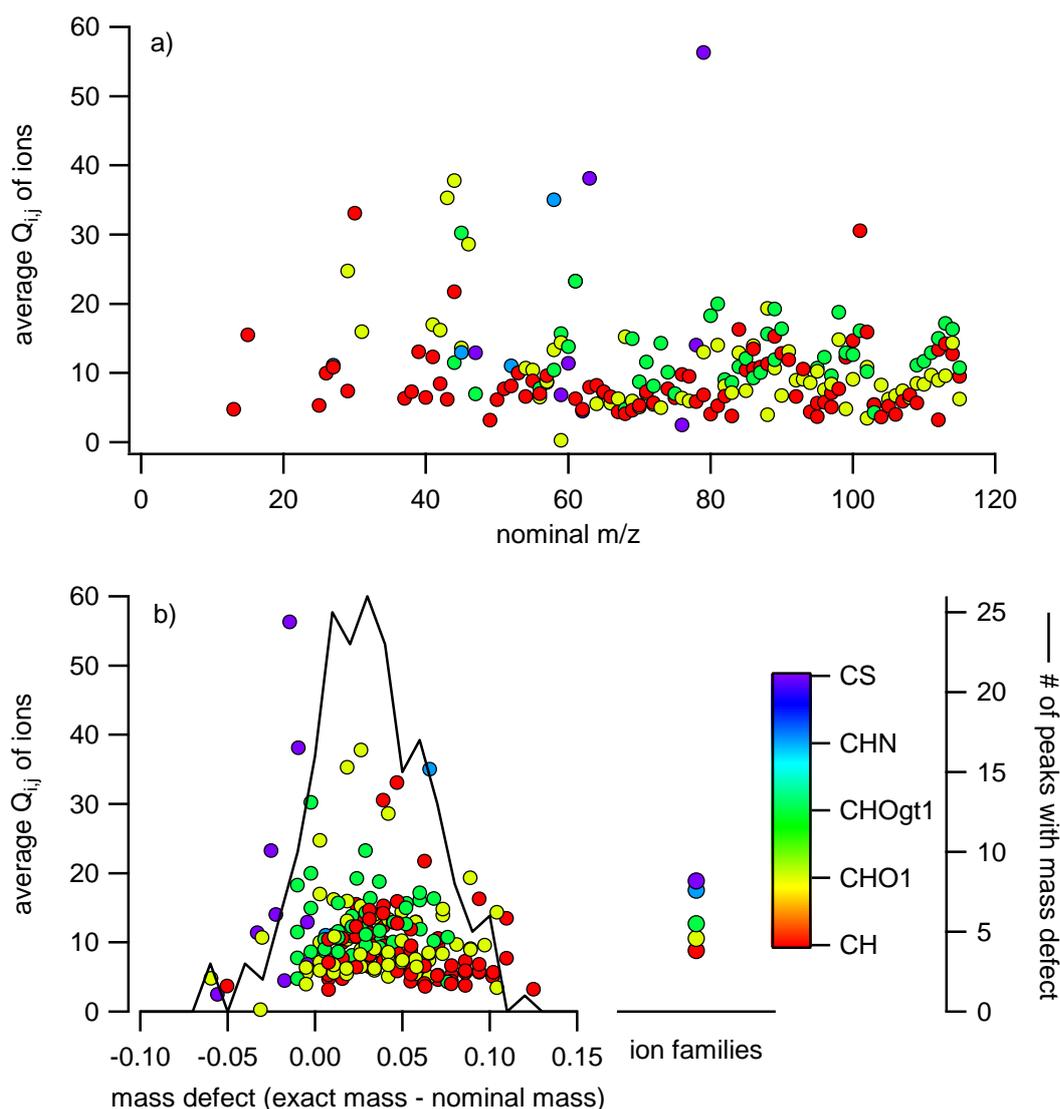


**Figure S2:  $Q_{ij}$  as a function of the day of the week.**

However, for  $PMF_{block}$  also with 6 factors, the average  $Q_{ij}$  is clearly larger (7 only the Zurich data points) than the ideal value of 1, i.e. the PMF residuals are larger than the measurement uncertainties. In comparison to  $PMF_{block}$ , the average  $Q_{ij}$  for Zurich is slightly reduced for the same number of factors when only including 1 site in PMF ( $PMF_{zue,isol}$ ,  $PMF_{zue,reprs}$ , average  $Q_{ij}$  6). In this study, we analyse yearly cycles and, thereby, assume constant factor profiles throughout the year which can contribute to  $Q > 1$ .

Another possible reason for  $Q > 1$  is an underestimation of the measurement uncertainty. A main contributor in high-resolution AMS data treatment (attribution of the signal at a nominal mass to several ions) stems from errors in the  $m/z$  calibration which could not be incorporated in the current data analysis. Recent studies demonstrate that for overlapping peaks (ions) the measurement uncertainties are strongly underestimated (Cubison et al., 2015; Corbin et al., 2015). For  $PMF_{block}$  using 6 factors, average  $Q_{ij}$  do not depend on  $m/z$  but rather on the ion family (Fig. S3): ions consisting of C, H, S, (and O) summarized under the name (CS) and ions consisting of C, H, N, (and O) summarized under the name CHN have a higher  $Q_{ij}$  than hydrocarbon ions (CH, only C and H) and oxygenated ions ( $CHO_{z=1}$  with 1 oxygen and  $CHO_{z>1}$  with more than 1 oxygen). Since the time series of  $CH_3SO_2^+$  is event-driven, the high  $Q_{ij}$  of this ion hints to the fact that PMF is unable to accurately resolve all of these events.

The average  $Q_{ij}$  for ions with a mass defect (nominal mass – exact ion mass) around 0.03 a.m.u. is higher than for the other ions (Fig S3). Mass defects in this range are most common in our dataset. This makes these peaks prone to overlap with other ions and thus their error prone to an underestimation because this effect is not considered in the  $s_{ij}$  calculation (described above).



**Figure S3: a) Average  $Q_{i,j}$  of ions in  $PMF_{block}$  as a function of their mass-to-charge ratio ( $m/z$ ). The ions are color-coded with their composition (CH: ions consisting only of C and H; CHO1: ions consisting of C, H, and 1 O; CHOgt1: ions consisting of C, H, and more than 1 O; CHN: ions consisting of C, H, N, (and O); CS: ions consisting of C, H, S, (and O)). b) Average  $Q_{i,j}$  of the ions in  $PMF_{block}$  as a function of their mass defect (exact mass – nominal mass) as well as a histogram of the number of ions with a certain mass defect. The mean  $Q_{i,j}$  of the ion families is displayed separately.**

Cumulative density functions for the  $a$ -values of HOA and COA are presented for the accepted solutions in Fig. S4. We found that 80% of the accepted solutions have an  $a$ -value  $\leq 0.3$  for HOA and an  $a$ -value  $\leq 0.5$  for COA. The output HOA and COA factor profiles are therefore not significantly variable and very similar to the input profiles, indicating that similar solutions were selected. Furthermore, the yearly average factor concentrations of all selected  $PMF_{block}$  solutions after  $R_k$  correction are shown for the case of Zurich as an illustration in Fig. S5. The distributions of each of

the different factors do not show more than 1 distinct mode, indicating that we do not have several populations of solutions.

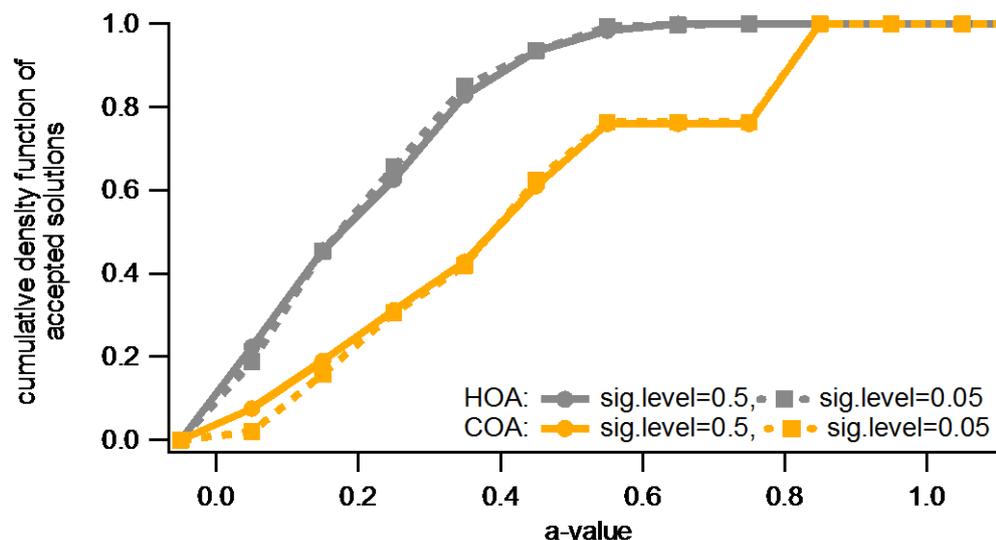


Figure S4: Cumulative density functions of a-values for HOA and COA for the accepted solutions.

The yearly average factor concentrations of all selected  $PMF_{block}$  solutions after  $R_k$  correction are shown for the case of Zurich as an illustration (Fig. S5). The distributions of each of the different factors do not show more than 1 distinct mode.

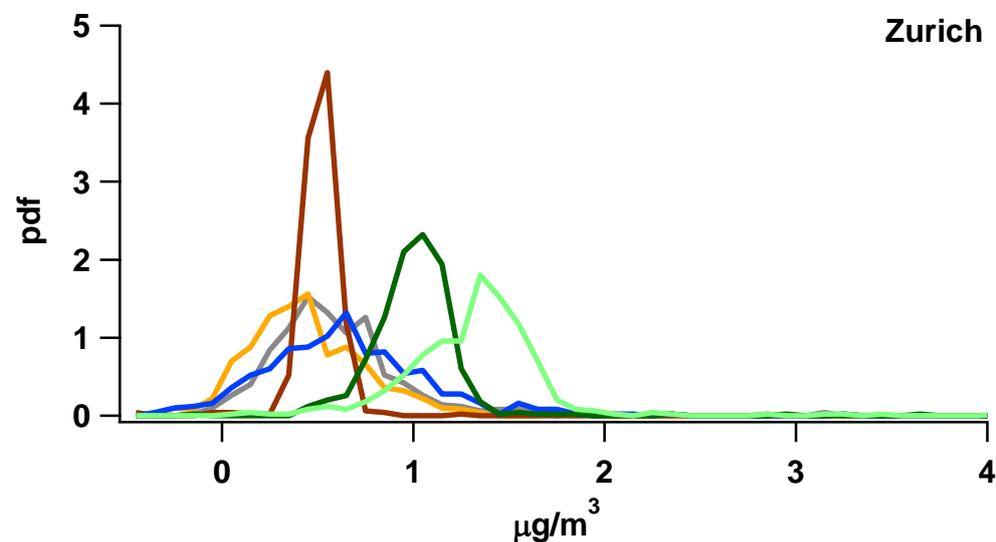


Figure S5: Histograms of yearly average factor concentrations of all selected  $PMF_{block}$  solutions (after  $R_k$  correction).

... “

Table SI.1

Why is there a table for mass closure criteria (used for rejecting bad solutions) when all entries in this table are identical? The criteria seem to concern distribution of residuals of OC fitting. Is it really so

*that the fit is rejected if 1st quartile point is negative and 3<sup>rd</sup> quartile point is positive? In other words, the fit is rejected if residuals are symmetrical around zero. Usually, such residuals would be considered desirable.*

Symmetrical residuals around 0 were considered desirable and solutions were accepted if the first quartile was smaller than 0 and the third quartile larger than 0. The title of the Table was changed to:

**“Table S1: set of acceptance criteria used.  $r$  is the correlation coefficient between a factor time series and the respective marker. Q25 is the 1<sup>st</sup> quartile and Q75 the 3<sup>rd</sup> quartile.”**

## References:

Allan, J. D., Jimenez, J. L., Williams, P. I., Alfarra, M. R., Bower, K. N., Jayne, J. T., Coe, H., and Worsnop, D. R.: Quantitative sampling using an Aerodyne aerosol mass spectrometer 1. Techniques of data interpretation and error analysis, *J. Geophys. Res.*, 108, 4090, doi:10.1029/2002JD002358, 2003.

Bozzetti, C., Daellenbach, K. R., Hueglin, C., Fermo, P., Sciare, J., Kasper-Giebl, A., Mazar, Y., Abbaszade, G., El Kazzi, M., Gonzalez, R., Shuster Meiseles, T., Flasch, M., Wolf, R., Křepelová, A., Canonaco, F., Schnelle-Kreis, J., Slowik, J. G., Zimmermann, R., Rudich, Y., Baltensperger, U., El Haddad, I., and Prévôt, A. S. H.: Size-resolved identification, characterization, and quantification of primary biological organic aerosol at a European rural site, *Environ. Sci. Technol.*, 50, 3425-3434, doi:10.1021/acs.est.5b05960, 2016.

Bozzetti, C., Sosedova, Y., Xiao, M., Daellenbach, K. R., Ulevicius, V., Dudoitis, V., Mordas, G., Byčenkienė, S., Plauškaitė, K., Vlachou, A., Golly, B., Chazeau, B., Besombes, J.-L., Baltensperger, U., Jaffrezo, J.-L., Slowik, J. G., El Haddad, I., and Prévôt, A. S. H.: Argon offline-AMS source apportionment of organic aerosol over yearly cycles for an urban, rural and marine site in Northern Europe, *Atmos. Chem. Phys. Discuss.*, doi:10.5194/acp-2016-413, 2017a.

Bozzetti, C., El Haddad, I., Salameh, D., Daellenbach, K. R., Fermo, P., Gonzalez, R., Minguillón, M. C., Iinuma, Y., Poulain, L., Müller, E., Slowik, J. G., Jaffrezo, J.-L., Baltensperger, U., Marchand, N., and Prévôt, A. S. H.: Organic aerosol source apportionment by offline-AMS over a full year in Marseille, *Atmos. Chem. Phys. Discuss.*, doi:10.5194/acp-2017-54, in review, 2017b.

Chen, Q., Miyazaki, Y., Kawamura, K., Matsumoto, K., Coburn, S. C., Volkamer, R., Iwamoto, Y., Kagami, S., Deng, Y., Ogawa, S., Ramasamy, S., Kato, S., Ida, A., Kajii, Y., and Mochida, M.: Characterization of chromophoric water-soluble organic matter in urban, forest, and marine aerosols by HR-ToF-AMS analysis and excitation emission matrix spectroscopy, *Environ. Sci. Technol.*, 50, 10,351–10,360, 2016.

Corbin, J. C., Othman, A., Allan, J. D., Worsnop, D. R., Haskins, J. D., Sierau, B., Lohmann, U., and Mensah, A. A.: Peak-fitting and integration imprecision in the Aerodyne aerosol mass spectrometer: effects of mass accuracy on location-constrained fits, *Atmos. Meas. Tech.*, 8, 4615-4636, doi:10.5194/amt-8-4615-2015, 2015.

Cubison, M. J. and Jimenez, J. L.: Statistical precision of the intensities retrieved from constrained fitting of overlapping peaks in high-resolution mass spectra, *Atmos. Meas. Tech.*, 8, 2333–2345, doi:10.5194/amt-8-2333-2015, 2015.

Sun, Y., Zhang, Q., Zheng, M., Ding, X., Edgerton, E. S., and Wang, X.: Characterization and source apportionment of water-soluble organic matter in atmospheric fine particles (PM<sub>2.5</sub>) with high-resolution aerosol mass spectrometry and GC-MS, *Environ. Sci. Technol.*, 45, 4854–4861, 2011.