Response to reviewer 2, Mikko Äijälä

We thank Dr. Äijälä for carefully reviewing our manuscript and providing very useful comments and suggestions, which improved our manuscript. Below we address each comment point by point. For clarity we have marked the reviewer comment in **blue**, our answers in **black**, and changes to the manuscript in **red**. Page and line numbers (**in black**) in our replies refer to the clean revised manuscript (without tracked changes), and **green** line/page numbers refer to this response.

Summary

Sini Isokääntä and co-authors present a comparison of dimensionality-reductive techniques for mass spectral data analysis, applied to a data set of car exhaust + a-pinene aging experiment in a reaction chamber. The data, collected by a PTR-MS and an AMS, was factorized and clustered using 5 different techniques. The authors present and discuss their analysis procedures and interpretation of results from mathematical and physicochemical viewpoints. For the PTR-MS data, all five techniques produce comparable results, yielding 4 to 5 interpretable factors (or clusters). This is a very important result considering the novelty of applying these methods to PTR-MS and similar mass spectral data of gas phase. For AMS data, only PMF and NMF yielded results that could be compared – PCA, EFA and PAM struggled with the particle data set. Although applied to data from a single experiment, the comparisons presented and the discussion on analysis techniques convey general messages for many similar analyses that abound in atmospheric science.

The manuscript is clearly of interest to AMT readers, and touches the important topic of computational analysis of complex, large mass spectrometric data sets and advanced statistical methods of data analysis. Despite the evident need, a very limited amount of reviews of advanced statistical methods for this type of complex, physicochemical analyses is available, so I see the authors' contribution as extremely welcome to the field. The manuscript is well and clearly written and structured. With some exceptions, detailed in specific comments, the data, analysis techniques and experimental setup is adequately described. I recommend the paper for publication in AMT after addressing the following comments.

Major comments

The authors generally do a good job of introducing and describing their methods, but I would call for discussion on the physicochemical objectives or purpose of the statistical techniques used in this study. Obviously it is about more than reducing the size or dimensions of the data, which is the single common feature of all these SDRTs. However, some of these methods are rather similar (e.g. PMF vs NNMF) while some aren't (e.g. PAM vs factorization methods). Generally, a data analyst should choose a proper tool for the job, depending on the goal. Here it seems the methods are just applied on the data, seemingly without stating beforehand e.g. the difference in categorizing or classifying variables using PAM, studying the main explanative components of variability or doing latent feature extraction by exploratory factorization. Many of these differences are later casually mentioned, but I suggest an effort to further summarize these fundamental differences, and conversely, some of the close similarities could be made in the introduction and conclusions of the paper.

We have now summarized and added discussion of the used SDRTs to a new section in the manuscript (section 4.4.), where we aimed to discuss the objectives and purposes of these techniques in more concise manner than mentioning these properties throughout the manuscript.

4.4 Summary of the SDRTs used in this study

The methods tested in this study have many similarities and many, fundamental or computational, differences between them. Though in literature, they are many times applied to similar problems. In this section we will summarize some of these properties.

EFA has a fundamental difference to the other methods as it is by definition a measurement model of a latent variable, i.e. the factor, (Osborne, 2014) whereas the other methods are basically describing the measured data with linear combinations of measured variables. The latent variables in EFA, i.e. the factors, cannot be directly measured, but instead, they are seen through the relationships they initiate in a set of Y variables, which are measured. In the other methods, in turn, the factors, components or clusters are calculated directly from the measured variables Y (Rencher and Christensen, 2012; Osborne, 2014).

The approach to data reduction in PCA is to create one or more summarizing variables from a larger set of measured variables by retaining as much as possible of the variation present in the original data set (e.g. Jolliffe, 2002). This is done by using a linear combination of a set of variables. The created summarizing variables are called components. The main idea of the PCA is to figure out how to optimize this process: the optimal number of components, the optimal choice of measured variables for each component, and the optimal weights when calculating the component scores.

The objective of cluster analysis is to divide the observations into homogeneous and distinct groups (Rencher and Christensen, 2012). Cluster analysis is a method where the aim is to discover unknown groups in the data, which are not known in advance. The goal of the clustering algorithm is to partition the observations into homogeneous groups by using some measure of similarity (or dissimilarity), such that the within-group similarities are large compared to the between-group similarities. The choice of the similarity measure can have a large effect on the result. One property of cluster analysis is that it will always calculate clusters, even if there is no strong similarity present between the variables in the data (Wu, 2012). This should be noted when interpreting the results, especially if the user has no a priori information about the number of clusters.

NMF and PMF provide an alternative approach to the decomposition assuming that the data and the components are non-negative (Paatero and Tapper, 1994; Lee and Seung, 1999). Thus, all the features learned via NMF and PMF are additive; that is, they are adding together strictly positive features. PCA and EFA tend to group both positively correlated and negatively correlated components together, as they are only looking for the correlations of variables (except SVD-PCA, which can be applied to the data matrix directly). On the other hand, NMF and PMF, by constricting W and H to positive values, find patterns with the same direction of correlation. Thus, NMF and PMF work well for modelling non-negative data with positive correlations. However, if the interest is not only in the positive effects, then PCA and EFA can provide more information of for the investigated system. Cluster analysis is suitable for classifying observations based on certain criteria. The researcher can measure certain aspects of a group and divide them into specific categories using cluster analysis. However, this method is not suitable for data with variables which should show contributions from multiple factors/components (e.g. strongly fragmented signals in AMS data).

Factorization methods, including those used in this paper, operate on the fundamental assumption that the factor profiles (here factor mass spectra) are constant over the investigated period. Often, this has been interpreted in a way that chemical processes occurring in a chamber experiment or the atmosphere violate this assumption. However, this interpretation is based on a too narrow definition of what a factor represents. A factor can be seen as a direct (emission) source of compounds which changes its contribution to the whole signal (e.g. primary emissions from biomass burning as a fire develops and then dies). But a factor can also be interpreted as a group of compounds showing the same temporal behavior. If this group is released together as an emission or if the compounds are formed in the same ratio by some chemical process should not matter. In the latter case, it is important how wide the group is selected, i.e. if we group products of processes together for which the contribution changes with time. This means, choosing the optimal number of factors becomes even more important when chemical processes are occurring. EFA and PCA account for the chemistry happening in chamber measurements with negative loadings, as described above. Same factor can contain educts and products of a chemical process (e.g. oxidation), with the difference that their loadings are negative and positive, respectively.

Taking account of everything above, the most important thing to consider when selecting the SDRT is the interpretation. What are the features the researcher wants to deduct from the data, what are the properties of the data and how the data can answer the research questions. As we have shown in the results section, the methods provide quite similar reconstructions of the time series, but the interpretation of the steps leading to these are quite different. For example, comparing EFA factor FE4 in Fig. 3a and PMF factor FP4 in Fig. 7a reconstructions, they seem to show the same procedure but the first one includes both positive and negative effects and the second one consists only of positive effects.

Secondly: algorithmic methods are often discussed plenty in these types of reviews, but some of the equally significant, but less obvious issues include, among others: data pre-processing (e.g. scaling and weighting), metrics (e.g. for quality evaluation and similarity), and error models. Mostly, the authors cover these topics well enough, but some major questions surrounding e.g. the non-standard PMF error model the authors used and preprocessing of data for PCA/EFA are raised. See specific comments on the details.

In our manuscript we did use the standard AMS error model for PMF, and we understand that this was clearly not emphasized enough in the manuscript as our error description was misunderstood. The main point using the standard AMS error was indeed to show how "reference" error works for PMF when compared to others. We have addressed this issue in detail in the specific comments (see our reply to comment "p.7.l.26" in page 10) and modified the manuscript accordingly. We have also now named the "Poisson style error" to "standard AMS error" to emphasize this.

We have also clarified the data preprocessing steps for EFA and PCA as suggested, see our replies to specific comments "p.5.l.30" and "p.6.l.14" in pages 6 and 8, respectively.

Finally, most factorization methods, including those used in this paper, assume constant factor profiles, i.e. that chemical reactions do not take place. This is fundamentally at odds with the approach of using them to model data of reaction chamber chemistry. In practice, the methods can work despite this (as is shown in this work), but this fundamental violation of basic assumptions should be explicitly stated and discussed.

The assumption of constant factor profiles is indeed a very important one for PMF (and NMF). The idea that chemistry in the atmosphere/chamber violates this assumption is however a misconception stemming from a too narrow definition of what a factor represents. A factor can be seen as a direct (emission) source of compounds which changes its contribution to the whole signal (e.g. primary emissions from biomass burning as a fire develops and then dies). But a factor can also be interpreted as a group of compounds showing the same temporal behavior. If this group is released together as an emission or if the compounds are formed in the same ratio by some chemical process should not matter. In the latter case, it is important how wide the group is selected, i.e. if we group products of processes together for which the contribution changes with time.

As an example, consider a data set of gas phase measurements in the atmosphere. There is a mix of precursors and let there be OH and O3 chemistry during the day and NO3 and O3 chemistry at night. All of these reactions can form Highly Oxygenated Material (HOM) among other compounds. The ratio of the reaction pathways (and thus the observed HOM composition) will depend on the ambient conditions (e.g. O3 and NOx concentration, irradiation). If we do not constrain the PMF analysis, we will find more than one HOM-type factor (i.e. a daytime and a night-time factor). Each of these factors then has a constant profile, only the contribution to the data set changes with time, i.e., the time series of the night-time factor will be close to 0 during the day and vice versa for the day-time factor. As there is no strict criterion for the correct number of factors, it is important that the scientist analyzing such a data set compares solutions with different number of factors and uses additional information (such as other measurements).

We also could perform a “rolling window” style PMF (looking at a short period of the data set at a time (e.g an hour) and then shifting this window by e.g. 10 min). If we set the number of factors to only find 1 HOM type factor in the whole data set, this factor would have a different factor profile during the day and at night. But this behavior was created by the choice of the scientist to have a certain number of factors.

In the analysis presented in our manuscript, we did not use any such assumption about the type of factors we were expecting or even their number. Rather we inspected the solutions for multiple factor numbers and chose the number of factors based on the presented criteria and the interpretability, i.e., we first ran the analysis and then interpreted the results. We have now stated the assumptions for constant factor profiles in the new section 4.4 with further discussion. See reply to the first major comment in pages 1 and 2.

Specific comments:

p.2.l.2.: “…sharp change at the beginning of experiments, e.g., switching on UV-lights) may present additional challenges, as PMF was originally developed for field measurement data sets where real changes in factors are expected to be much slower than e.g. the noise in the data.”. In my understanding PMF is not assuming anything about order of measurement points or rates of change in loadings – please provide a reference or expanded rationale for this line of thinking!

Indeed, PMF does not make any (mathematical) assumptions of the order of measurement points or rates of change. However, most of the PMF applications, so far, have been with ambient measurement data where, for most part of the data, the instrument noise is indeed more rapidly changing than the actual observed concentrations, which must be also assumed e.g. in the study by Ulbrich et al. (2009), where the AMS error scheme was introduced. We have now reworded the referred sentence to clarify that:

page 3, lines 5-8

[…] The special conditions in lab experiments (sharp change at the beginning of experiments, e.g., switching on UV-lights) present an additional test scenario, as PMF has been mostly used for field measurement data sets where the main focus is often in the long-term trends and real changes in factors are often expected to be more subtle, than e.g. the variations in the noise in the data. […]

We also adjusted section 4.1.5 accordingly, please see reply to comment (“p.17.l.25”) in page 15.

p.2.l.14. “In our study we chose a set of SDRTs having fundamental differences.” Aside from data reduction or simplification, the objective of SDRTs differ. Often, clustering is applied to classify or categorize non-correlated variables, whereas factor analysis aims to uncover latent features the combination of which would best explain the variation in data**.** Can the authors please comment on their objective of the analysis outside of reduction of data size) of the use of SDRTs for this type of analysis? Especially regarding clustering.

We have added a new section 4.4 to the manuscript to include more discussion of the fundamental differences and objectives the various SDRTs might have. Please see our answer to the major comment in the beginning of this reply.

p.2.l.16: "On the other hand, clustering might be more suitable for a more simplified or preliminary approach, or when the chemical compounds in the data are already known." It is not trivial to envision a case where clustering would be preferable, even as a preliminary approach – please expand on this reasoning or provide an example.

Please see our reply to specific comment "p.10.l.5." in page 13 for added examples. We modified the referred sentence to

section 1, page 3, lines 19-21

[…] On the other hand, clustering might be more suitable for a more simplified or preliminary approach (as it is computationally less demanding), or when the chemical compounds in the data are already known or if strict division between variables is preferred. […]

p.3. : Please describe the PTR-MS and the SP-AMS in some more detail here. Specifically their mass analyzer resolution (C/HR/L –ToF), as this strongly affects mass analysis and type of data.

Both of these instruments are extensively described in Kari et al. (2019) and in the references given there. However, we added the requested details of the mass analyzer resolutions shortly to the manuscript along with the other details requested by reviewer#1;

section 2, page 4 lines 13-22

[…] Detailed setup, calibration procedure and data analysis of the used high resolution PTR-MS have been explicitly presented in Kari et al., (2019b). In the campaign, the high mass resolution of the instrument (>5000) enabled the determination of the elemental compositions of measured VOCs. The instrumental setting was intended to minimize the fragmentation of most compounds, so the quantitation of the VOCs was possible. The chemical composition of the particle phase of the formed SOA was monitored with a soot particle high resolution aerosol mass spectrometer (SP-AMS, Aerodyne Research Inc., USA, hereafter referred to only as AMS, Onasch et al., 2012). In brief, the SP-AMS was operated at 5 min saving cycles, switching between the electron ionization (EI) mode and SP mode. In EI mode, the V-mode mass spectra were processed to determine the aerosol mass concentration and size distribution. The mass resolution in this mode reaches ~2000. The SP-mode mass spectra were used to obtain black carbon concentration. As the used chamber was a collapsible bag, the volume of the chamber decreased over time due to the air taken by the instruments. […]

p.3.l.9.: A note. As the authors state, if one does not account for mass spectra baseline, this may give rise to "background factors" or cause variables to get polluted or mixed (depending on instrument resolution). While theoretically these could be separated by factor analysis, in practice their presence in data hinders the analysis, often notably. For hard clustering, this effect of may even more problematic. Can you give an order of magnitude estimate of this effect? Can you separate baseline factors / clusters in some cases? What is their fraction of total explained variability?

We were not able to separate, or at least qualitatively justify that the background factor, e.g. FE4 from EFA applied to the gas-phase data, would indeed be pure instrument and/or chamber background, as this factor also included ions related to the car exhaust, thus being slightly mixed. The lack of identified compounds for our gas phase measurements hinders the detailed analysis and identification of possible background factors. In principle, by adding more factors, we might be able to find one or more background factors, but as we would not have been able to identify those explicitly, we did not apply this approach here but rather acknowledged the possible uncertainty in our factors. Therefore, also quantifying this effect here is impossible.

p.3.l.18.: This type of truncation correction is likely detrimental to analysis and should be avoided in statistical analysis (especially factor analysis, as it creates an artificial, non-random, variability component). If only few points had this problem, would it not be preferable to omit them?

The number of observations which needed to be elevated to small positive values was not large, and so was the number of time points Removal of one negative value from one variable would require the removal of the complete row, and as the negative values occur in a few variables, but not at the same time point, this would remove a significant amount of the data. The small negative values in the data are in cases when the concentration of the compound was in practice zero, but the application of the baseline correction caused them to be very small negative numbers instead. We acknowledge the possibility of increased uncertainty due to this kind of procedure, but due to the very small size of the AMS data set, omitting the negative data points was not possible. PMF can be actually applied to some extent even with the existence of some negative values, but the NMF algorithm we used, does not allow the input of any negative values. In addition, also for PMF the data should be positive by definition and without extensive inspection of the source code of the PMF Evaluation tool we do not know how the software handles the negative numbers. Thus, to give all SDRTs the same data as an input, we replaced the negative values. To acknowledge this issue for the AMS data more clearly, we modified the end of the section 4.2.5 (please also note that this section has now been modified to include the 2-factor solution from PMF instead of the 4-factor one)

page 24 lines 17-23

[…] This, on the other hand, assigns different emphasis between the matrices, possibly causing NMF to use more effort to reconstruct the data matrix with factor time series. However, the reader should keep in mind that for detailed chemical analysis of such a small data set, especially with PMF, the downweighing is advisable. In addition, the replacement of very small negative numbers with very small positive numbers is not mandatory for PMF, as it can run containing a few negative values. However, we did the replacement here, as the NMF algorithm used here, does require strictly positive input data. Acquiring a balance between statistically good results and realistic factors might be challenging, and to achieve more robust results, testing different error schemes may be beneficial, especially for a data set with such a small size. […]

p.5.l.30: "[…]where X includes the analysed variables (centred and scaled by their standard deviations)" In their examination of errors in PCA (and PMF) in environmental applications Paatero and Hopke (2003) strongly advice against "autoscaling":

> "*5.1. Do not autoscale noisy variables in PCA*
> In PCA, it is customary to scale columns of X so that in the scaled matrix, all of the columns have the

> same variance. This procedure means that the sum of the two components of the variance (signal and noise) is constant over all variables. It follows that for the weakest variables, having the smallest amount of signal, the noise variance is much larger than for the strong variables. This behavior is in severe conflict with the recommendations found in this work: the exaggerated noise in a few noisy variables will cause the small principal components to be undetectable in the analysis and will increase the noise in other principal components. The recommendation is clear: "do not auto-scale noisy variables in PCA modeling" […]

Is this type of scaling also done in here? Please reflect and add some discussion on the issue of error model in PCA. Perhaps recommend the readers to note this for future analyses?

The sentence to which the reviewer is referring, explains how the scores (factor time series) are usually calculated. However, in the end of the paragraph, we say "the scores are calculated without standardizing the data matrix to achieve more interpretable component time series".

However, we assume the paper from Paatero and Hopke means if the data matrix X is scaled before applying PCA (or inside the algorithm). We used 2 variations of PCA, EVD-PCA, which uses eigenvalue decomposition into the correlation matrix (calculated from unscaled X, though centering for example, would not have an effect on the correlations), and SVD-PCA, which uses singular value decomposition into the data matrix X or the correlation matrix. Thus, no scaling, whatsoever, was applied when using EVD-PCA (which produced interpretable results shown in the manuscript). We also tested SVD-PCA, and applied it to the scaled data matrix, as this was the recommendation given for the algorithm we used. The SVD-PCA results were only shown in the supplement as they were not interpretable. However, as the reviewer explains, this might affect negatively into the results and is contradicting to the recommendations given by Paatero and Hopke (2003). Therefore, we now also looked into the SVD-PCA applied to the unscaled data matrix X, but those results did not show any improvement. We added one example of this to the supplementary material (see section S4.1 Figure S13) and modified sections 4.1.2 and S4.1 accordingly. We modified sections 3.1.1 and 3.1.2 to emphasize that no scaling was applied to the data for EVD-PCA at any point and explain scaling/not scaling situation for SVD-PCA.

section 3.1.1
page 5 lines 23-25
[...] It should be noted, however, that Eq. (1) describes the theory behind the PCA model, not the actual calculation process, which is described below. Thus, for example, centering of variables is not required. […]
page 5 lines 29-31
[…] In our study we applied EVD-PCA to the correlation matrix (calculated from the unscaled data matrix) and SVD-PCA was applied to the data matrix without and with the scaling (centered and scaled by their standard deviations). In addition, the acquired […]

section 3.1.2, page 7 lines 6-8
[…] As for the PCA explained in the previous section, Eq. (4) here describes the form of an EFA model based on literature, not the direct calculation in the algorithm. Thus, no scaling (what e.g. the subtraction of the mean $\bar{y}_1$ from $\boldsymbol{y_1}$ in Eq. (4) essentially is) is applied here. […]

section 4.1.2, page 16 lines 10-14
[…] SVD-PCA (when applied to the scaled data matrix) was not able to separate α-pinene as its own component, but instead created two factors which were dominated by the unreacted α-pinene and its fragments (see Fig. S12 in SI section S4.1). In addition, the unrotated solution included a

large number of negative loadings, which complicated the interpretation of the components. No improvement was achieved when SVD-PCA was applied to the data matrix without any scaling (see. Fig. S13). […]

section S4.1, page 6 lines 14-17
[…] components from SVD-PCA (applied to the scaled data matrix) and the unrotated component time series and original loadings (scaled eigenvalues) are shown in Fig. S12 with 4 components. The unrotated results from SVD-PCA when applied to the unscaled data matrix are shown in Fig. S13, and Fig. S14 shows the original loading values for the Oblimin rotated EVD-PCA with 4-components. […]

p.5.l.24. : While potentially simplifying the chemical interpretation of factors, rotating the variables in a direction where ions are more separated between the factors, does this not equally degrade the time series interpretability (loadings' time series get more similar)?

In PCA and EFA the rotation is applied to the loading matrix only, as the loading's time series are calculated after adjusting the actual loading values. This means in principle, that also the time series get more "different/separated" meaning rotation does not degrade, but in contrast even increase, the time series interpretability. In general, as pointed out, rotations potentially simplify the chemical interpretation of the factors, but they might decrease the overall fit, as the rotation is applied after the solution has reached the optimal value for the condition to be minimized.

p.6.l.14.: Same question on EFA error model as for PCA above, as PCA and EFA are very similar methods. Please discuss error weighting and accounting signal-to-noise (Paatero and Hopke, 2003).

Please see also our replies above ("p.5.l.30" and "p.5.l.24"). Regarding the error weighing. In EFA (or in PCA) no kind of weighing is applied based on measurement error as it is done in PMF. We now noticed from our formulation for EFA, that we use the term "the error term ε_1", which might cause misunderstanding. The term in question is not the error in the same sense as for PMF (which needs the error matrix), but a residual error arising from the model fit. We renamed it to "the residual term ε_1" to avoid further confusion.

Regarding the signal-to-noise, please see our reply to comment "p.7.l.21" in page 9.

p.6.l.6.: Does the limit of 0.3 apply to AMS data as well? This means any m/z signal explaining under 0.3 ug/m3 was set to zero? How many variables does this affect – I would imagine it is a very large fraction? How does this truncation affect mass (signal) conservation in data?

After applying PCA to the correlation matrix, the acquired eigenvectors were scaled by the square root of the eigenvalues (see page 5, lines xx-xx) to acquire similar loading values compared to EFA. The limit of 0.3 given here, was applied to those loading values, which are (and should be in most cases) between -1 and 1. It does not mean that signals explaining less than 0.3 µm/m^3 are set to zero, but that the contribution of a variable into that specific factor is very small, and when setting the loading value to 0 for that factor, we enhance the separation between the factors. The limit of 0.3 for the absolute value of the loadings is widely applied in literature, and was thus selected (e.g. Field, 2013, Izquierdo et al., 2014). We adjusted the paragraph and references to justify our selection of the limit 0.3:

section 3.1.1, page 6 lines 26-29

[…] Very small loading values (absolute value less than 0.3) are suppressed to zero to enhance the separation of ions between the components. The limit of 0.3 was selected, as this is often given as a reference value for insignificant loadings (see. e.g. Field, 2013; Izquierdo et al., 2014). […]

Regarding the mass signal conservation. As for EFA and PCA the factor scores can be calculated in various ways, and as the loading values can be negative, EFA and PCA do not conserve the original mass signal in that sense, but instead aim to conserve as much information (i.e. we could calculate how much of the correlations are preserved with different number of factors) from the original data as possible. Even though there are no practical obstacles to back-calculate the original signal from the obtained loadings and scores (which are calculated with the loadings), it does not make sense in a similar manner as for NMF and PMF, where the "scores" and "loadings" are optimized at the same time inside the algorithm. Due to this, the concentration of the factors between these methods is not fully comparable, but this is not an issue in this type of analysis.

p.7.l.15 On the error model of st.dev / sqrt(n): as written in the text, st.dev reflects the changes in the concentrations of compounds in the chamber, during the experiment, and not measurement uncertainty (or counting statistics error of the instrument) in a repetition measurement? How does dividing it by (square root of) data series length make it a more relevant error metric? Please explain. Also: Why not use the standard error model computed by the AMS analysis software (Squirrel/Pika; e.g. Ulbrich et al., 2009)?

The PMF model was first introduced by using the standard deviations as an error metric, see Paatero and Tapper, (1994). The error metric applied here (st.dev / sqrt(n)) is commonly called "standard error". It takes into account that deductions from measurement with less observation contain more uncertainty. As the standard deviation may be high in this type of data (containing the step change at the onset of photochemistry), it may be a too conservative estimate for uncertainty in PMF analysis. Accounting for the number of observations gives a more reasonable estimate for total uncertainty. We clarified this in section 3.1.3:

page 8 lines 8-11

[…] But as a reference, the standard error of the mean (the standard deviations of the ion traces divided by the square root of the number of observations, i.e., length of the ion time series) was used as an error for both PTR-MS and AMS data. It considers that measurements with less observations contain more uncertainty. These error values are constant for each ion throughout the time series […]

We did use the "standard" AMS error in our study (named Poisson-like error). For more details see our reply to comment ("p.7.l.26") in page 10.

p.7.l.21. (and in the supplementary referred): In calculation of the signal-to-noise-ratio (SNR), you identified weak and bad variables (defined by their low SNR). However in contrast to usual practice (in PMF), you do not down-weight these signals. Especially since you are using a custom (not thoroughly validated) error model, I would listen by the advice, Paatero and Hopke (2003):

> *"Regarding weak/bad variables, the main result of this work is that even a small amount of overweighting is quite harmful and should be avoided. In contrast, moderate downweighting, by a factor of 2 or 3, never hurts much and sometimes is useful. Thus, it is recommended to routinely downweight all weak variables by a factor of 2 or 3. This practice will act as insurance and protect against occasions when the error level of some variables has been underestimated resulting in a risk of overweighting such variables. Regarding bad variables (where hardly any signal is*

> *visible from the noise), the recommendation is that such variables be entirely omitted from the model."*

The reasoning for not down-weighting only states that low SNR data has high noise-to-signal rate, which is only stating the evident, and not very useful. Also, the small number of variables is not really a good excuse to deviate from the practice (without at least some sensitivity analysis). While I agree the overall error from not doing this properly is likely smallish (for PMF), I recommend you acknowledge the issue and cite this "best practice", the Paatero and Hopke (2003) recommendation, for the readers – not to proliferate a deficient data pre-processing practice.

We have added more justification for omitting the downweighing. Please also see our reply to Reviewer#1 comment ("page22 line 20")

section 3.1.3, page 8 line 27 – page 9 line 2

[…] . In contrast to the suggested best practice (Paatero and Hopke, 2003), we did not downweighed any ions in our data sets. This approach was used in order to give each SDRT an equal starting point for the analysis, as e.g. for NMF or PCA similar downweighing is not possible, as we do not have any error estimates to calculate the signal-to-noise ratios in similar manner. However, to avoid misguiding the reader to omit recommended data pre-processing practice for PMF, we also tested PMF with downweighing. This, as expected, did not change our results significantly, but we acknowledge it should be indeed applied if aiming for a more detailed chemical interpretation of the PMF factors. […]

section S2.3, page 5, lines 1-5

[…] However, no removal or downweighing of the variables was applied in the results presented in the main manuscript, as for the weak ions the error values were already in the range of the ion signal itself, and for the bad variables the error was usually way above the signal throughout the ion time series. In addition, the number of weak or bad ions for both AMS and PTR-MS data were rather small. To justify our decision, we did run the PMF analysis with the downweighing. This, as expected, did not change our results or interpretation, and thus those results are omitted. […]

Regarding the AMS error, see our reply to the next comment (to comment "p.7.l.26").

p.7.l.26., also Figure S4 (should be: "S7"?). The standard AMS error model is composed of a minimum (Gaussian) error (Ulbrich et al., 2009), related to electrical background noise of the instrument (background at zero signal) plus a counting statistics uncertainty that follows the Poisson distribution (error is proportional to the square root of signal intensity; Allan et al., 2003). This model seems different from the ones used in this paper. Please comment on this, and why you chose not to directly use the approach of Allan et al and Ulbrich et al. (2009), readily available from the AMS analysis software.
From the plot S4 (should be "S7"?) "PMF error schemes" it seems the constant error scheme ("Static error") overestimates error, especially for low concentrations, and the counting statistics (Poisson) error seems negligible (underestimated), even for the most abundant aerosol ion, at m/z 44, at highest concentrations. Please double check and report the error calculations here against what you get from the "Squirrel" AMS analysis software.
Why was the "signal following error" not used/tested for the AMS, since it seems to work for PTR-MS?

We did use the standard AMS error model, as mentioned in our reply to major comment in page 3 line 27. We have now modified the manuscript accordingly to avoid this misunderstanding. In addition, we renamed the "Poisson style error" to "standard AMS error" to emphasize this.

Section 3.1.3, page 8 lines 17-25

[…] For AMS, we also applied a standard error that is frequently used by the AMS community. The standard AMS error consists of the minimum error related to the duty cycle of the instrument and the counting statistics following the Poisson distribution (Allan et al., 2003; Ulbrich et al., 2009). Shortly, the standard AMS error for signal I can be formulated as

$$\boldsymbol{I_{err}} = \alpha\sqrt{\frac{I_O+I_C}{t_s}}, \tag{9}$$

where α is an empirically determined constant (here $\alpha = 1.2$, generated by the AMS analysis software PIKA, http://cires1.colorado.edu/jimenez-group/ToFAMSResources/ToFSoftware/index.html, last visit on 9.9.2019), I_O and I_C are the raw signal of the ion of particle beam (ions s^{-1}) for the chopper at open and closed position, respectively, and t_s is the sampling time at a particular m/z channel (s). […]

The signal following error was indeed tested for AMS data, but the results were omitted as they did not show any clear improvement and were rather similar to those acquired with the standard AMS error and thus would have lengthened the manuscript unnecessarily. We now mention this in the manuscript section 4.2.4:

page 23 lines 29-31

[…] The signal following error, used for PTR-MS, was also tested for particle phase data. However, as this type of error showed very similar behaviour as a time series as the Standard AMS error, and produced very similar outcome, those results are omitted from this manuscript. […]

p.8. l.6.: How much does the Q (or Q/Q exp ratio) increase with these fpeak values?

Figure 1 below show the Q/Qexp for PTR-MS data with the 5-factor solution with static error (left) and signal following error (right). The change with fpeak is rather small, as seen from the figures. The actual factorization also does not change (for PTR-MS) with different fpeak, indicating the 5-factor solution is quite robust.

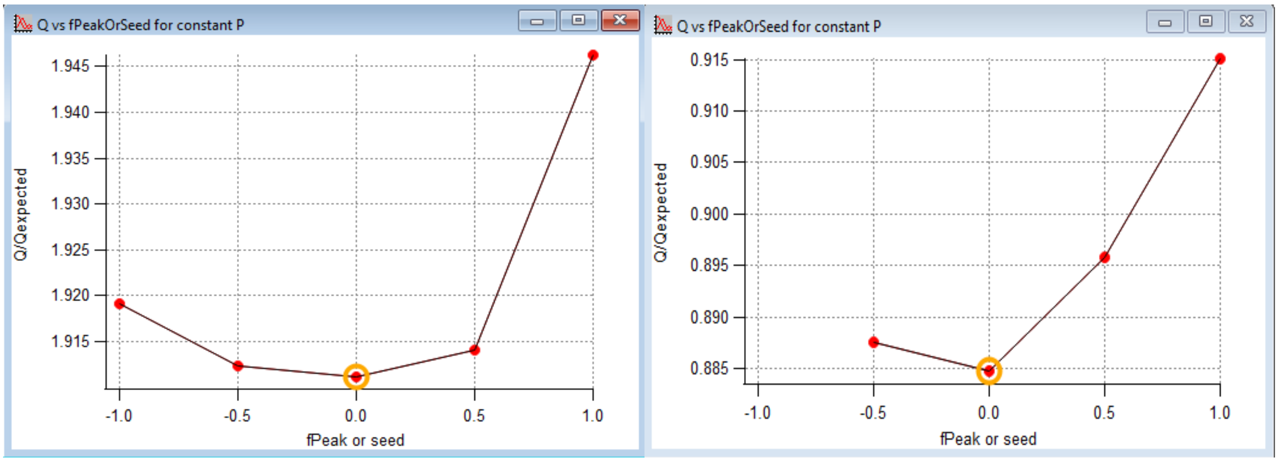

Figure 1 Q/Qexp for PTR-MS data with the 5-factor solution with static error (left) and signal following error (right).

p.9.l.10.: How do you interpret the negative loadings in EFA and PCA? Why were these solutions deemed physically sound, if they feature negative mass loadings?

These negative loadings refer to the feature of EFA and PCA that the factor loadings can be negative. Negative factor loading may have different interpretations. It may indicate that the compound has a decreasing effect on the factor, i.e. it acts as a sink for the compounds with positive loading in the same factor. In chamber experiments, negative loading may also refer to decreasing concentration of the compound within chemical reactions. For example, a compound that can be rather easily oxidized, or forms in the later state of the experiment, might get negative loading in a factor, that mostly contains later generation products. An example of this is benzene (detected as $C6H7^+$), assigned to FE1. When inspecting the original loading values, it has a negative loading in FE1 (identified as later/slowly forming products), and a positive loading in FE4 (identified as precursors from car exhaust/background). As benzene originates from the car exhaust, it contributes positively to FE4. However, as it oxidizes over the course of the experiment (thus it has a decreasing concentration), it contributes negatively to FE1, which mostly includes those later generation products.

We understand our approach does indeed conceal some additional information EFA/PCA can provide, but as we did not have enough detailed information of the measured compounds (e.g. the sum formulas were not identified for all compounds), the interpretation of the negative loadings separately, for the most part, was not possible. We added more clarification about this in the end of section 3.1.5, and also discuss this property of EFA/PCA in the new section 4.4 (see pages 1-3 in this reply).

section 3.1.5, page 10 lines 9-19
[…] However, this type of approach conceals the information of the negative factor loadings in EFA and PCA (which are included in the calculation of factor/component time series as weights), but instead visualizes the general contribution of an ion to a factor. Negative factor loadings may have different interpretations. It may indicate that the compound has decreasing effect on the factor, i.e. it acts as a sink for the compounds with positive loading in the same factor. In chamber experiments, negative loading may also refer to decreasing concentration of the compound participating in chemical reactions if it acts as a precursor for other compounds in the same factor. One example of this is benzene (C6H6), observed with PTR-MS. When inspecting the original loading values from EFA, for example, it has negative loading in FE1 (identified as later/slowly forming products), and positive loading in FE4 (identified as precursors from car exhaust/background). As benzene originates from the car exhaust, it contributes positively to FE4. However, as it oxidizes over the course of the experiment (thus has decreasing concentration), it has a strong correlation with oxidation products but appears negative in FE1, which mostly includes those later generation products. […]

p.10.l.5.: *"Depending on the aim of the study 5 and the type of the data, this property of cluster analysis may be considered either as an advantage or disadvantage."*. This is an important point in this paper overall, as is the difference between hard vs soft divisions of variables. Please elaborate, and e.g. give the reader some examples of data analysis of objectives where cluster analysis would be at an advantage or disadvantage.

Obvious disadvantage is in cases where one compound affects multiple factors by physical or chemical definition. Then clustering loses information by forcing the compounds on only one cluster. For example, when identifying pollution sources, black carbon can be emitted from both wood burning emissions and traffic. There, clustering might not be the best approach. On the other hand, computational time (as shown also in this study) is one clear advantage for PAM or hard division techniques in general, especially when analyzing very long ambient data sets. Hard division techniques have also been shown to work efficiently when distinguishing between different coffee types (i.e. strict separation between clusters is needed), as shown by hierarchical clustering by Sánchez-López et al. (2014).

We added few examples to section 3.2.1, page 11 lines 15-21

[…] One obvious advantage of cluster analysis (or hard division techniques in general) is computational time, especially if analyzing long ambient data sets. For laboratory measurements, this most likely is not an issue. Hard division techniques have also been shown to work efficiently for VOC measurements when distinguishing between different coffee types (espresso capsules), where strict separation between clusters is needed, as shown in Sánchez-López et al., (2014). For source apportionments studies, where one variable might originate from multiple sources, cluster analysis using hard division technique is probably not as suitable as softer division techniques, which can assign one variable to multiple sources/factors. […]

p.10., Section 3.3.: As you state, interpretability is key. Chemical interpretability is discussed in a concise way. However, in addition to chemical interpretation, looking at loading time series is an important indicator. Do the time series reflect the kind reaction kinetics (in experiment chamber) and take place in reasonable timescales? This relates to e.g. Fig.13.
Comparing e.g. the PMF results in Figs. 13, S26, S27 – only the rank 3 Poisson-model (Fig S26-c) solutions' loadings (and factor 3 in Fig 13d) behave in a realistic way. The others anti-correlate highly, usually signaling they are over-resolved (unrealistically split) and usually then less components should be used. See for example f1 andf2 in Fig 13-a: correlation is undoubtedly close to -1 and the dynamics do not make sense – this simply seems like a bad factorization solution.

Following the suggestion regarding the PMF solution for AMS in comments below (see reply to comment "p.21., Section 4.2.4" in page 15) we now present a 2-factor solution in the PMF results. The figure below shows the evolution of the two PMF factors and α-pinene. In the chamber, we assume a steady-state OH concentration of 10^7 molecules/cm^3, and the reaction rate constant OH-α-pinene of 52.3×10^{-12} cm^3 molecule^{-1} s^{-1}, it is very likely to get α-pinene consumed (25ppb) in a time scale of half an hour. If we assume a mass yield of a-pinene SOA of 30% in this study (Kari et al., 2019), the consumed α-pinene (about 140 µg m^{-3}) can be transformed to a SOA mass concentration of 52 µg m^{-3}, This is consistent with what is observed in the figure. Thus, we can assume the time series from the SDRTs used in this study reflect the reaction kinetics in the chamber and the processes take place on reasonable timescales.

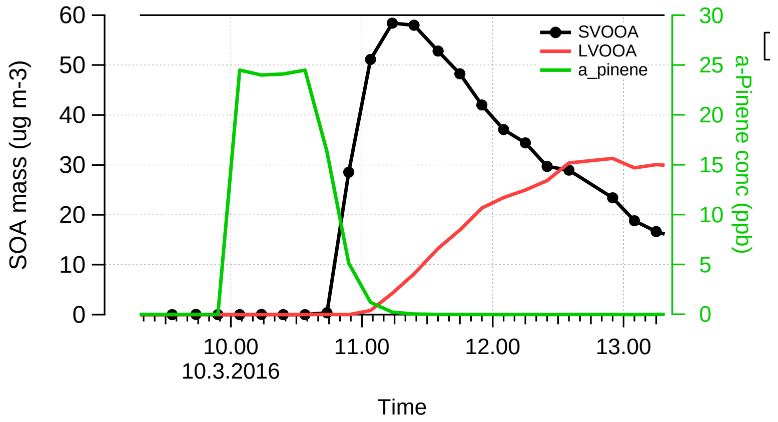

Figure 2 Two-factor solution for particle-phase data from PMF (fpeak = 0) with the injected α-pinene.

p.13, Section 4.: the factor profile figures (Figs 3 through 7) are really difficult to read this way! Please separate each factor to its own sub-plot, similar to Figure 12-b. Maybe put the fractional plots to supplementary material?
Figure 1, Figure 2. Please highlight in these figure factor numbers (e.g. a larger dot?) that were selected according to the evaluation metrics.
In all figures, for quick reference, please state if it is gas or particle (or AMS vs PTR-MS) data.

Figures 1, 2 and 11 are now modified as suggested, RSS in Figure 2 and 11 is changed to more NMF related metric as suggested by Reviewer #1. Captions in each figure were adjusted to include the information of the data (AMS or PTR-MS), and we added factor description (those in table 1 & 2) to the figures in the manuscript (when applicable), as requested by the reviewer#1.

With the factor contribution figures, we looked for a way of displaying the factor/compound "compositions" in a way allowing direct comparison of the different methods. Normalizing the factor contribution for each ion was chosen to avoid the very different weighting methods in NMF and PMF between the factorization matrices and also to be able to show EFA/PCA/PAM results in a similar manner. We now added figures with the factor mass spectra in separate panels in the SI material for easier comparison (section S5, renamed to "Contrast angle and factor spectra comparison", Figures S20, S21), and pointed them out in the manuscript:

section 4.1.6
page xx, lines xx-xx
[…] Table 1 summarizes the acquired results from different SDRTs for the gas phase composition data measured with PTR-MS, and Figs. S20 and S21 in SI section S5 show separate factor contributions for each of the SDRTs. […]

p.14. Section 4.1.2.: The difficulty to interpret PCA (negative) highlights that PCA separates principal components of variability, whether positive or negative, which can not [necessarily] directly be interpreted as physical [concentration] components of the system – the variability could be equally connected to losses of signal due to it chemically reacting away. Again, this ties to the objective of the analysis and should be discussed more clearly.

Please see our reply to specific comment ("p.9.l.10") in page 12. In addition, see our reply to the major comment in the beginning of this reply (we have now included a new section 4.4, to the manuscript with more discussion of the objectives or the analyses).

p.17.l.25. "This is caused by the used error scheme, where errors are larger for the fast changes in the data (Fig. S4b)." I am very confused. Is this "feature" intentional? This also relates to the question of p.7.l.15. Is the variability of ion concentration in chamber is indeed used as a metric for measurement uncertainty (even when scaled by sqrt(n)?. This seems a peculiar, to say the least, a very un-orthodox way to do error modeling in PMF. It could explain why PMF does surprisingly poorly compared to the other models for AMS data.
Importantly, please include a comparison of your error models versus "the standard" error model in-build in AMS data analysis software(s) (Squirrel / Pika / PET) and list the differences between the standard practice (see e.g. Allan et al., 2003; Ulbrich et al., 2009; Zhang et al., 2011) and your data pre-processing procedure.

We did use the "standard" AMS error in our study (named Poisson-like error). For more details see our reply to comment ("p.7.l.26") in page 10.

In principle the feature of having larger errors for fast changes is intentional, as if we would have ambient data, such large variations most likely would be considered as instrument malfunctions (“spikes” in the data), thus applying a larger error for those data points is reasonable. Our aim using the “signal following error” was indeed to use an error scheme, that could also be used for ambient measurements. In addition, for ambient data we are not that interested in the brief period with rapid and fast changes, as long as the new values after these changes are captured well by PMF. However, in chamber measurements this type of changes can be relatively large (as in this study) and should be assessed in a different way. We wanted to test two very different error schemes here to investigate how this affects to the results. The static error stays constant over time, thus giving equal weight to time points whereas the signal following error assumes those fast changes are not so important. For example, if we would be interested in long term changes in ambient measurements, signal following error would be the choice. We have now reworded the corresponding part in section 4.1.5 to be consistent:

page 19, lines 17-24
[…] This is caused by the used error scheme, where errors are larger for the fast changes in the data (Fig. S4b). In ambient data not measured at instant proximity of strong emission sources, for which PMF is often used, this type of error is beneficial as there the fast changes are more likely to be noise or instrument malfunctions (excluding, for example, sudden primary emission plumes), and we are more interested of the long-term changes instead. For laboratory data, where large changes are often caused by rapid changes in actual experimental conditions, e.g. due to injecting α-pinene or turning the UV lights on, the static type of error is most likely preferable. Usage of the static error scheme helps to avoid overcorrecting intentional (large) changes in experimental conditions and confusing them with real variation taking place during the experiment and typically being much less pronounced. […]

p.21., Section 4.2.4.: I would have inspected the 2 factor solution, is it had percentage-wise the largest decrease in Q/Qexp. Usually it is also good practice to start from lower number that you can certainly interpret and continue to higher numbers. This could explain also why EFA and PAM suggest 2 factors (clusters). Please show these 2 factor (cluster) solutions (in the supplementary material).
The analysis on the error scheme issues seems correct.
I have to disagree on the interpretability of the PMF solutions presented here. This solution seems a mathematical one rather than physically meaningful. See also comment to p.10., Section 3.3., Figure 13 etc. The time series indicate over-splitting of factors (extreme anti-correlation of f1,f2,f3) and the most of spectral profiles have extreme positive correlation (extreme high similarity by visual inspection). I don’t really see many interpretable features in Figure 13. Mainly the HOA spectrum in Fig 13b, LV(f1,f2,f3)/SV(f4) split in 13 c&d.

Regarding EFA and PCA. We would not trust the number of factors – tests here, as the data from particle phase measurement has more variables than time points, which creates significant calculational issues (see e.g. page 21 lines 19-24) for the eigenvalue decomposition used with PCA and for the EFA algorithms due to the issues with the correlation matrix. We can still run EFA and PCA, but in both cases some smoothing is applied inside the algorithms due to the non-positive-definite correlation matrix. Solutions with 2 factors/clusters were already shown in Figs. S24 and S26/S27. The PCA solution with 2 components was omitted, as it was almost identical to the EFA solution thus bringing no new information to the manuscript.

We do agree with the reviewer about the interpretability of the presented PMF solution. We have conducted PMF on the entire experiment dataset (compare to only one day experiment in this study), which produced four meaningful factors (see Figure 3 below). The 4-factor solution is nicely consistent with those in NMF (Sec. 4.2.3). HOA (hydrocarbon OA) was dominated by the prominent CnHC2n-1 and CnHC2n+1 family ions and appeared as primary emission in the experiments that included the vehicle, the αP_SOA_SVOOA and αP_SOA_LVOOA factors are present only in the pure a-pinene SOA experiments. The mixed_LVOOA factor appears in both mixed car exhaust and α-pinene SOA experiments. The four factors are physically and chemically meaningful. However, in this study, measurement data of only one day (i.e. one chamber experiment) was included for the analysis. This clearly caused issues for PMF, and as suggested by the reviewer, a two-factor solution may be preferable for this subset of data. Hence, we have now changed the AMS PMF results in the main manuscript to the 2-factor solution and added the results with 4 factors to the SI. The factors from 2-factor solution can be interpreted as SVOOA and LVOOA.

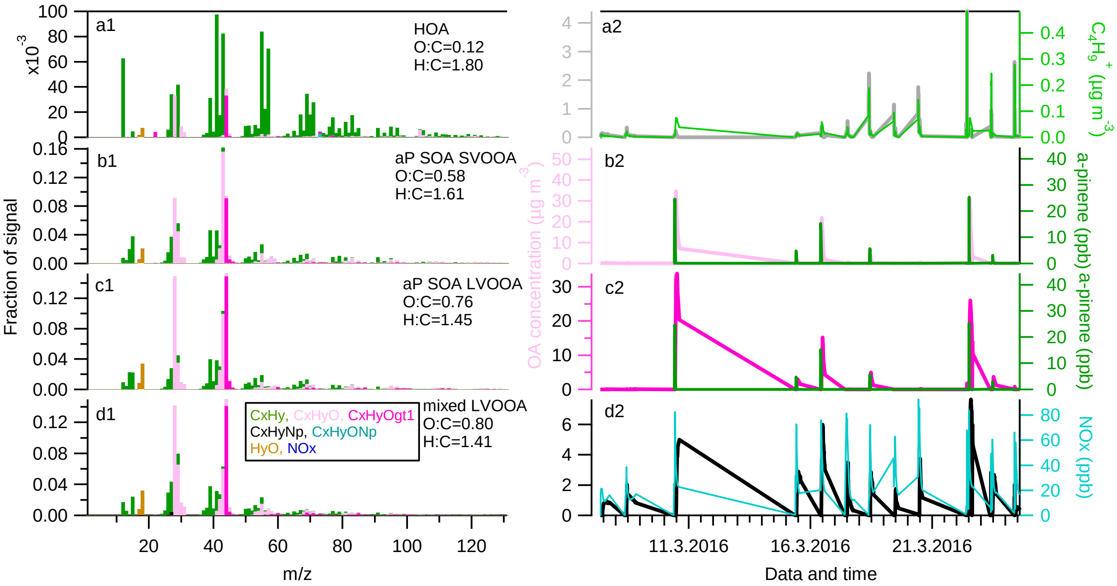

Figure 3 The mass profiles and time series of a 4-factor solution of PMF analysis on the entire experiment (Kari et al., 2019)

The results from PMF with particle phase data were modified, and the sections 4.2.5, 5 and the corresponding SI material, abstract and figures were modified accordingly.

corrected section 4.2.4

The Q/Q_{exp} for the two error schemes are shown in Fig. 11b. Neither of the error schemes show a clear inflection point. Examples of behaviour of the errors as a time series are shown in Fig. S7 (SI Sect. S2.2). With the standard AMS error, the Q/Q_{exp} values do not reach 1 (with 10 factors Q/Q_{exp} = 1.76) whereas with the static error the values decrease below 1 for 7 factors. The solutions with 2-5 factors were inspected, and the 2-factor solution (Fig. 13) is presented here as the most interpretable one (summarized in Table 2). The primary OA factor, separated by NMF (Fig. 12, FN2) was only found if using 4 factors and the static error scheme in PMF (see SI Sect. S6.2, Fig. S31a-b). However, interpretation of the time series for that solution was found to be very difficult due to the extreme anticorrelation between the time series, and thus the 2-factor solution was selected. The two factors were interpreted as SVOOA and LVOOA. In addition, the largest relative decrease in the Q/Q_{exp} was observed with the 2-factor solution.

The residuals for the standard AMS error were smaller, as shown in Fig. 14. This agrees with the analysis of the gas phase data set, where the residual for the signal following error (which has a similar profile in time as the standard AMS error) were generally smaller compared to the static error.

The signal following error, used for PTR-MS, was also tested for particle phase data. However, as this type of error showed very similar behaviour as a time series as the standard AMS error, and produced very similar outcome, those results are omitted from this manuscript.

corrected section 4.2.5

Table 2 summarizes the acquired results from NMF and PMF for the particle phase composition data measured with AMS. PAM was not able to separate distinct clusters due to the inability of clustering techniques to classify an ion into multiple clusters. Comparing the relative factor spectra and fraction of signal from NMF and PMF with the static error (Figs. 12b and 13b), the distribution of ions is similar between the LVOOA in the PMF solution and the mixed LVOOA factor in the NMF solution, and also similar between the SVOOA in the PMF solution and the αP-SOA factor (integrated- αP-SOA-SVOOA and αP-SOA-LVOOA factors). When inspecting the individual ion time series in the original AMS data, most of them have rather “smooth” behaviour, similar to the factors acquired from NMF. It seems, that PMF gives more weight to the background ions (with very small concentration) which do not have that clear structure in their time series, thus including more of their behaviour in the final factors, if the number of factors in PMF is increased from 2 (see SI Sect. S6.2 Figs. S30-S32). Residuals from NMF reconstruction (with 4 factors) were over ten orders of magnitudes smaller (for NMF between $-1.3\cdot10^{-13}$ and $7.1\cdot10^{-14}$) than those from PMF (between -0.06 and 0.13 for the 2-factor solution with standard AMS error, see Fig. 14), indicating better reconstruction of the data with NMF. Most likely PMF struggles with the small data set, thus not being able to recover all the factors found by NMF and construct reasonable time series for those factors (see 4-factor solution in Fig. S31), whereas NMF does not seem to be affected by the data size. In addition, the weighing between the factorization matrices between NMF and PMF is different not only due to the error matrix that is given as a weight in PMF, but also because of the different solving algorithms for each method. This, on the other hand, assigns different emphasis between the matrices, possibly causing NMF to use more effort to reconstruct the data matrix with factor time series. However, the reader should keep in mind that for detailed chemical analysis of such data set, especially with PMF, downweighing is advisable. In addition, the replacement of very small negative numbers with very small positive numbers is not mandatory for PMF, as it can run with a few negative values into some extent. However, we did the replacement here, as the NMF algorithm used here, does require strictly positive input data. Acquiring a balance between statistically good results and realistic factors might be challenging, and to achieve more robust results, testing different error schemes may be beneficial, especially for a data set with such small size.

abstract page 1 lines 23-24

[…] For the particle phase measurements, four factors were discovered with NMF: one primary factor, a mixed LVOOA factor, and two α-pinene SOA derived factors. PMF was able to separate two factors: SVOOA and LVOOA. […]

section 5

page 26 line 32 – page 27 line 1

[…] From the particle phase data, NMF was able to separate four factors, whereas PMF separated two. […]
page 27 lines 3-5
[…] EFA and PCA had computational constraints due to the small data size acquired from the AMS and could not to be applied. In addition, PMF also faced assumedly computational issues with the small particle phase data set, thus not being able to reasonably separate the HOA factor. […]

Technical corrections:
p.3.l.4.: PTR generally refers to the ionization reaction and not the instrument, Please use PTR-MS or similar, common acronym.

We corrected PTR to PTR-MS in the manuscript and supplement and in all figures.

p.7.l.30. ts here is sampling time per m/z channel, when scanning (Allan et al., 2003), not to be confused with the total sampling time of the instrument, so please add this clarification.

We have now clarified this in the manuscript. Please see our reply to specific comment ("p.7.l.26") in page 10.

p.21.l.13.: Please add a reference to m/z 57 and 59 link to HOA.
Supplementary material: please check figure numbering is in numerical order. Similarly to main part figures, please add if data is PTR—MS or AMS.

We added the reference that was missing (Zhang et al., 2005), corrected one other reference (Hoyle et al, 2011 to National Research Council, 1991) and corrected the numbering of the figures and indicate in each figure caption the type of data presented.

section 4.2.3, page 23 line 5
[…] contains marker ions associated with HOA (e.g. m/z 57, Zhang et al., 2005) condensed again in a cooler chamber […]

REFERENCES

Field, A.: Discovering Statistics using SPSS, 4 ed., SAGE, London, 2013.
Izquierdo, I., Olea, J., and Abad, F. J.: Exploratory factor analysis in validation studies: uses and recommendations, Psicothema, 26, 395-400, 10.7334/psicothema2013.349, 2014.
Sánchez-López, J. A., Zimmermann, R., and Yeretzian, C.: Insight into the time-resolved extraction of aroma compounds during espresso coffee preparation: online monitoring by PTR-ToF-MS., Anal Chem, 86, 11696-11704, 2014.
Zhang, Q., Alfarra, M. R., Worsnop, D. R., Allan, J. D., Coe, H., Canagaratna, M. R., and Jimenez, J. L.: Deconvolution and quantification of hydrocarbon-like and oxygenated organic aerosols based on aerosol mass spectrometry, Environ Sci Technol, 39, 4938-4952, 10.1021/es048568l, 2005.

references

Allan, J.D., J.L. Jimenez, H. Coe, K.N. Bower, P.I. Williams, and D.R. Worsnop, Quantitative Sampling Using an Aerodyne Aerosol Mass Spectrometer. Part 1: Techniques of Data Interpretation and Error Analysis, Journal of Geophysical Research- Atmospheres, Vol. 108, No. D3, 4090, doi:10.1029/2002JD002358, 2003. (NB! Includes corrigendum)

I.M. Ulbrich, M.R. Canagaratna, Q. Zhang, D.R. Worsnop, and J.L. Jimenez. Interpretation of Organic Components from Positive Matrix Factorization of Aerosol Mass Spectrometric Data. Atmospheric Chemistry and Physics , 9(9), 2891-2918, 2009.

P. Paatero and P. K. Hopke, “Discarding or Downweighting High-Noise Variables in Factor Analytic Models,” Analytical Chimica Acta. Vol. 490, No. 1-2, 2003, pp. 277-289. http://dx.doi.org/10.1016/S0003-2670(02)01643-4

Q. Zhang, J.L. Jimenez, M.R. Canagaratna, I.M. Ulbrich, S.N. Ng, D.R. Worsnop, and Y. Sun. Understanding Atmospheric Organic Aerosols via Factor Analysis of Aerosol Mass Spectrometry: a Review. Analytical and Bioanalytical Chemistry, 401, 3045-3067, DOI:10.1007/s00216-011-5355-y, 2011.