Response to Comments by Anonymous Referee #3 on "Three-dimensional factorization of size-resolved organic aerosol mass spectra from Mexico City" by I. M. Ulbrich et al.

Author's note: The response from the reviewer (R3) has been reproduced here in full, shown in black text. Each comment has been numbered. The authors' responses are shown below each comment in blue text.

R3.0 This paper presents the first size-resolved factor analysis of AMS organic mass spectra. Factor analysis techniques are used to deconvolve a 3-dimensional matrix (time x composition x size) into a set of static factor mass spectra, their time-dependent intensities, and, depending on the selected model, their static or dynamic size distributions. The work is a logical extension of the 2-dimensional factor analysis techniques that have been applied to AMS data with great success and are now widely used. The content of this paper is therefore of high interest to the atmospheric community. In the course of the analysis, the authors have made some simplifying assumptions that direct them to a particular set of solutions and results. This limits the method utility, but does not render it invalid. However, the conclusions drawn from the analysis are held to apply generally to the 3-D factorization problem. This is not justified, and leads to a disconnect between the actual analysis and the scope of the conclusions and recommendations for future researchers. This should be rectified before publication.

My principal concern with the analysis is the assumption that any valid 3-D solution should include the 2-D HR-MS factors. There are several reasons why this might not be the case, including evaporation time distributions, representation of aerosol components as linear combinations of multiple factors, signal-to-noise considerations, and general differences in the nature of the correlations required to extract factors from the respective models. These are discussed below in the specific comments. However, the authors make this assumption, then force the 3-D solutions to fulfill it by (1) using it as a criteria for discarding solutions as non-physical and (2) constraining the 3-D factor mass spectra to resemble the 2-D solution. This method does provide useful data, though limits the model's explanatory power. However, by its forced nature it only explores a small subset of the 3-D solution space. Conclusions drawn from such a forced analysis cannot be held to apply to the entire 3-D factorization problem, and the method should certainly not be recommended as a general approach to 3-D AMS factorization; the authors have confronted a different problem (i.e. they have addressed how best to obtain size distributions corresponding to known 2-D HR-MS spectra). I would recommend that the language of the manuscript be adjusted to explicitly reflect the scope of the present analysis as framed above. Alternatively, the authors could extend their analysis to include a more detailed exploration of the 3-D solution space, which critically should include a detailed analysis of the deviations from the 2-D solution and an evaluation of whether they are physically meaningful that does not use similarities to the 2-D solution as the main criteria.

A3.0 We thank the reviewer for his/her careful and detailed review of the paper. The reviewer has three main criticisms of the paper.

A first main criticism centers on our assumption that we should find the same factors in the 3D dataset as in the HR-MS dataset, and that we constrain the 3-D solutions to fulfill the assumption. We rebut this criticism by noting that the 3-vector model was solved without any constraints, and solutions of this model did produce factors that strongly resembled those of the 2-D PMF solution (see e.g., Table 3, second row of the 4-factor solution description; Sect. S5.2 of the Supplement, lines 438-448). Thus, the underlying factors of the 3-D dataset are indeed those found in the 2-D PMF solution. Only the vector-matrix model was solved with constraints based on the 2-D PMF solution. We feel that the analysis presented in this paper (principally, constraining factors in the vector-matrix mode) was the best way to solve the vector-matrix model for the present dataset. However, we did not mean to imply that constraining factors should be practiced for all other datasets; rather, we see this approach as an option when the unconstrained results are unsatisfactory. Our detailed response to the comments below include the changes we have made to the manuscript to clarify this position. (Items A3.8, A3.14, and A3.16)

Second, the reviewer argues that some solutions that we show but have rejected could possibly be “physically meaningful”. This is possible. However, we feel that we cannot satisfactorily support these factors, and therefore must reject these solutions. We appreciate the reviewer’s critical comments in this area, which have helped us recognize that “physically meaningful” factors must meet two criteria that were not explicitly explained in the submitted manuscript: “physically meaningful” factors must (1) have a link to a physical source or process (not just be a mathematical artifact), *and* (2) must “supportable” for publication. (See more detailed responses under items A3.8, 3.9, A3.11, and A3.15)

Finally, the authors and the reviewer may have different understandings of how deviations between the data and the model would appear in the calculated factors and residuals. We present some arguments below about why we think the reviewer may be mistaken in some instances. Understanding these behaviors would require factorization of carefully constructed synthetic datasets (or laboratory data) that model the different assumptions/understanding of the vaporization process in the AMS. Such a study is outside the scope of the present paper. (See more detailed responses under items A3.4, A3.5, A3.9, and A3.11.)

Specific comments

R3.1. Page 4564, lines 15-20: The ToF-AMS is also capable of single particle measurements, and yields quantitative rather than qualitative results. See Cross et al. (ACP 9, 7769-7793, 2009).

A3.1 The ToF-AMS can be used for quantitative single-particle measurements in “Brute Force Single Particle” (BFSP) mode, or with an internal light scattering unit (LS-ToF-AMS). However, BFSP mode is very limited in duty cycle as described by DeCarlo et al. (Anal. Chem. 2006) and is rarely used outside of instrument calibration. The LS-ToF-AMS is a specialized modification of the ToF-AMS that requires extra hardware and manpower, and has so far been used in a small minority of the field campaigns conducted with the ToF-AMS. (We would guess that LS-ToF-AMS mode has been used in less than 5% of such campaigns worldwide). The 3-D factorization techniques could be applied to data from these modes, but neither type of data was acquired by our instrument during this study.

R3.2. Page 4571, line 9: Ions of different sizes will have different IToFs. This can (and probably should) be corrected by instrument calibration. Note also that for determining whether IToF has a significant influence on the measured particle size, the relevant quantity is the delta_IToF between two ions of interest compared to the PToF resolution (i.e. measurement time step dictated by the ToF pulser frequency), not IToF vs. PToF.

A3.2 We are unclear about the reviewer's comment. The reviewer may be referring to IToF in the mass spectrometer (orthogonal extraction to detection); however, this IToF does not influence particle size because all ions from a single extraction are assigned to the same size bin, corresponding to the timing of the extraction pulse. Instead, the reviewer could be referring to the transfer time from the ionization region through the focusing optics to the pulsed extraction region. This latter case was discussed in response to a comment by Reviewer 1 (R1.7).

R3.3. Page 4571, lines 10-13: "The measured PToF and the length of the particle flight region give the particle velocity, which is inversely proportional to the square root of particle vacuum aerodynamic diameter." This statement is inaccurate. Velocity is inversely proportional to the quantity (1+(dva/d0)ˆb), where both d0 and b are empirically determined from instrument calibration.

A3.3 This topic was also discussed by Reviewer 1 (see response to comment R1.18).

R3.4. Page 4572, lines 12-14: "As long as an OA component has a constant distribution of evaporation times, its retrieval in 3-D factorizations will not be affected." I am not sure I understand precisely what the authors mean by "retrieval." Is this retrieval of particular factor profile without bias due to evaporation time distributions, or simply the ability to retrieve a (potentially biased) factor from the model? The first of these is incorrect and conflicts with the authors' statement later in the paragraph ("But the mass spectrum of each organic aerosol component sampled in PToF mode can be slightly biased toward the fast-vaporizing species"). I suspect the second interpretation is also not accurate, as a distribution of evaporation times requires multiple factors to accurately represent, meaning less mass per factor (either by distributing mass across multiple factors or transferring it to the residuals), which the authors previously showed to be detrimental to factor retrieval in 2-D factor analysis (Ulbrich et al., 2009).

A3.4 We use "retrieval" in the sense of the second definition given in the reviewer's comment: "the ability to retrieve a (potentially biased) factor from the model".

We have changed the phrase "distribution of evaporation times" to "changes in evaporation timescale" and have revised this text for clarity (see response to comment R1.7).

As discussed in response to a related point raised by Reviewer 1, we have added a subsection to Sect. 3.4.1, Models for factoring the 3-dimensional array, at pg. 4581 line 4, entitled "**When the model does not match the data**". The new text can be found in in response to comment R1.10.

In addition, we have added a subsection to the Discussion as Sect. 5.3 entitled, "**Fitting of variations in mass spectra and size distributions**":

"Fitting of variations in mass spectra and size distributions

We explained in Sect. 3.4.1 that atmospheric processing of aerosol could change the mass spectrum, and if the processing resulted in slower particle vaporization, the apparent size distribution of OA components. We now discuss whether such changes were fit as additional factors in the solutions of either factorization model. The most likely place to find such additional factors is in solutions of the 3-vector model, whose requirement of constant mass spectrum and time series for each factor restrict this model's ability to fit variations. Only a few solutions of the 3-vector model included a pair of factors with similar mass spectra and time series but different size distributions, indicating a mathematical rather than physically meaningful separation.

In Sect. 3.4.1 we also noted a 2-D factorization case in which the time series of the residuals and Q/Q_{exp} contributions were highly correlated with the time series of the semi-volatile OOA-II factor, suggesting that there were changes in the composition of OOA-II that could not be fit with additional factors (Ulbrich et al., 2009). However, no such correlations were observed in the factorization of this dataset, nor did we observe other indicators that spectral variation was significant enough to interfere with the factorization using these models."

R3.5. The issue of evaporation time is quite important for factor evaluation and interpretation, so additional discussion by the authors would be welcome. This is an issue not just for species that vaporize on a slower timescale than the PToF cycle, where I agree with the authors' present treatment, but for species with "fast-but-different" evaporation times. The difficulty is as follows. If a factor consists of components with a distribution of evaporation times, these can only be represented by multiple size distributions (which broaden towards large sizes for slower-evaporating components). However, both 3-D factor models studied here assume that at a given point in time, a single size factor size distribution corresponds to a single factor mass spectrum. Therefore, even if the multiple size distributions needed to describe evaporation remain constant with time, retrieval of the OA components will depend on vaporization. As an example, one could consider a case in which the "true" factor consists of 50% NH4NO3 and 50% (NH4)2SO4. In 2-D (composition and time) factor analysis, this can be represented by a single factor. However, internally mixed NH4NO3 and (NH4)2SO4 will yield different size distributions because of the slower vaporization of (NH4)2SO4. The 3-D factor models require a minimum of 2 factors to accurately represent this, e.g. a small-size factor with enhanced NH4NO3 (representing faster evaporation) and a large-size factor with enhanced (NH4)2SO4 (slower evaporation), or alternatively, separate factors with unique size distributions for NH4NO3 and (NH4)2SO4. In practice, when such a factor is resolved from a complex ambient dataset, it could also appear as a single factor with the differences in evaporation timescales transferred to the residuals. Note, however, that in this case the retrieved factor is likely be biased towards roughly the geometric mean of the mass distribution (where the signal-to-noise is highest), and may therefore contain biases in both composition and size distribution (whether static or time-varying). For ambient OA, a broad range of evaporation timescales is expected and distinguishing between true factors and evaporation-based split factors is potentially quite difficult. The authors touch on this, noting "the mass spectrum of each organic aerosol component sampled in PToF mode can be slightly biased toward the fast-vaporizing species" (page 4572, lines 14-16). This is not quite true; as discussed above the bias is roughly towards the geometric mean of the factor mass distribution, which depending on the precise nature of the factor is not necessarily the fastest-vaporizing species. Further, the size distribution may also be biased. However, the authors' broader point (biases induced in factor spectra and size distributions due to evaporation time distributions) is well taken. As this manuscript appears intended in part as a blueprint for future 3-D AMS factor analysis, some additional discussion on the magnitude of these biases and methods for identifying them is needed.

A3.5 We agree with the reviewer that, in principle, multiple size distributions might be needed to fit factors with "fast but different" vaporization times. Then the question is whether these differences cause significant enough variation in the size/chemistry distributions to cause the algorithm to fit another factor.

The reviewer notes that "both 3-D factor models studied here assume that at a given point in time, a single … size distribution corresponds to a single factor mass spectrum." While this is technically true, there is an important distinction between the two models. The 3-vector model forces a factor to have the same size distribution for every time step; thus, changes to a factor mass spectrum, which might be caused by atmospheric variation or by changes in vaporization times, are either fit with a separate factor or could remain in the residuals. In contrast, the vector-matrix model allows a factor's size distribution to vary with time, so the single size distribution at each sampling time can capture changes in the real size distribution or evaporation times.

The reviewer gives an example of a "true" factor with 50% NH_4NO_3, 50% $(NO_4)_2SO_4$ particles and states that slower vaporization of ammonium sulfate would require a second factor. However, a second factor would be required only if the apparent size distribution of ammonium nitrate is significantly different from that of ammonium sulfate. We can quantitatively estimate the effect of ammonium sulfate's slower vaporization on its recorded size bin in the following manner. First, the difference in vaporization time between ammonium nitrate (80 μs) and ammonium sulfate (150 μs) is about 70 μs (pg. 4572, line 6). Second, the ions from sulfate would travel from the vaporization region to the extraction region (see R1.7) slower than those of nitrate, taking ~9 μs and ~14 μs for NO^+ and SO_3^+ ions, respectively, for a difference in transit time of ~5 μs. Thus the total delay for sulfate ions from vaporization + drift time is about 75 μs. In this study, ions were extracted (pulsed) into the mass spectrometer every 50 μs, so the total delay time is equivalent to about 1.5 extractions. Furthermore, during this study two adjacent extractions were co-added together on the data acquisition card before being saved, so spectra were recorded every 100 μs. Thus, the difference between ammonium nitrate and ammonium sulfate ions from the same particle should be slightly smaller than one size bin.

The situation is similar for vaporization of OA components, the most and least volatile of which have volatilities similar to those of ammonium nitrate and ammonium sulfate, respectively (Huffman et al., 2009). Therefore, we expect such evaporation time effects to be mostly limited to about one size bin. This difference is not significant enough to change the shape of the typically broad ambient size distributions, since even the narrowest factor size distribution (that of BBOA) spans about 30 of the 36 size bins included in the factorization dataset.

The quantitative estimates of ion transfer time and vaporization timescale, and their effects on apparent particle sizing have been included in revisions of the paragraphs beginning at pg. 4571 line 25, which are shown in full in response to comment R1.7.

The reviewer's comment about the potential bias in factor mass spectra and size distributions was addressed in our response to the previous comment (A3.4). We also note that when we consider "fast" vs. "slow" vaporization, we mean in the sampling timescales of PToF and MS modes, respectively, while the reviewer might have only considered the scale of PToF sampling.

Finally, the reviewer requests additional discussion on the magnitude of these potential biases and how to identify them in other datasets. At the present time, we can only note that we have not observed factors that appear to have split in this manner and are physically meaningful. (See the responses to comments R3.0 and R3.8 for additional discussion of "physically meaningful" factors.) Discussion of this topic was added as Sect. 5.3 as described in our response to comment R3.4 above.

Better understanding of the relationship between aerosol changes and the number of factors needed to represent the data would require a study with carefully constructed synthetic data, which is outside the scope of this work, but would be useful as a future study.

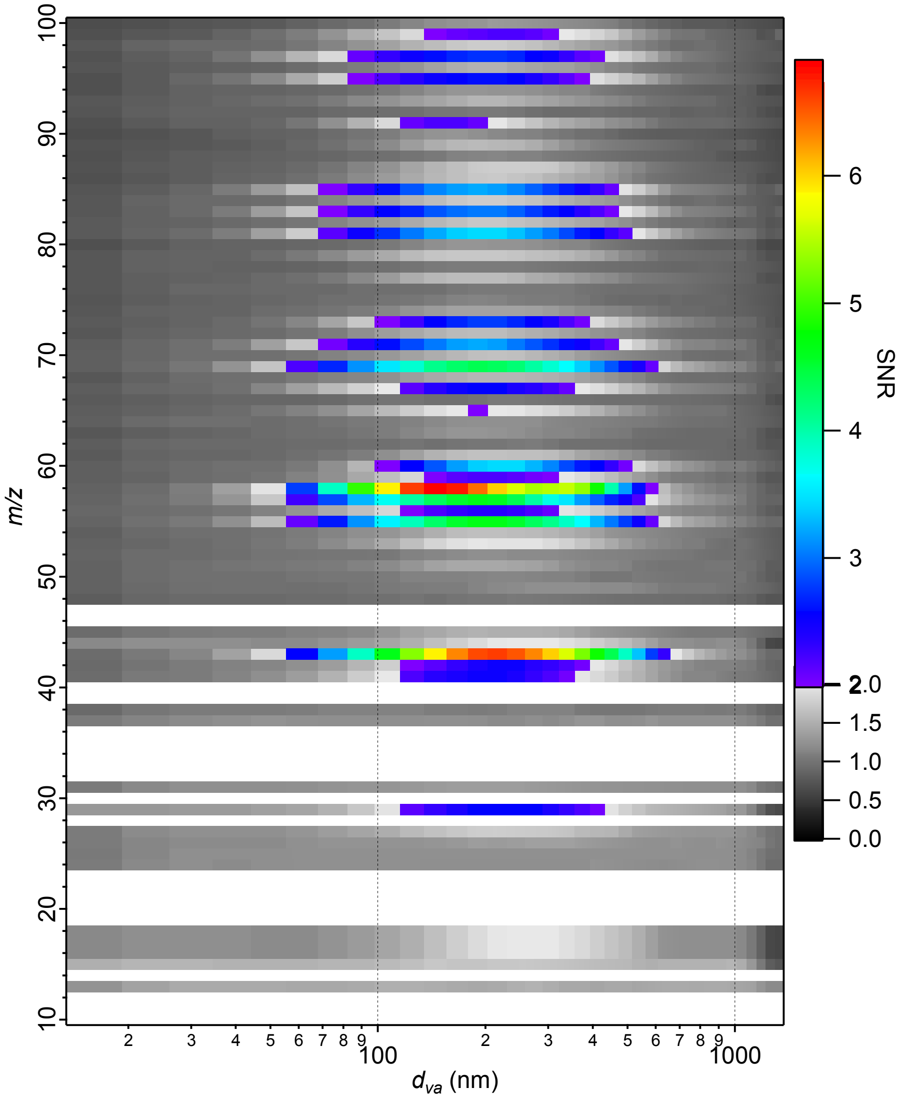
R3.6. Section 3.3.1: Why are uncertainties in the measured size not included in the error estimation? I realize that it is not possible to account for different evaporation times. However, because the chopper slit is not infinitely small, it introduces uncertainty into the particle sizing measurement. This uncertainty is not constant as a function of size, and so may affect the model results. If IToF is not explicitly corrected for as discussed in Comment 2, then this becomes an additional uncertainty that should be considered.

A3.6 The reviewer is correct that there is uncertainty in the particle sizing measurement. However, these uncertainties are not random errors (precision of the measured signal), which are appropriate for inclusion in the error estimation for PMF [Eq. (3), p. 4574)]. Uncertainties in the particle sizing measurement form a transfer function that converts the real distribution into the measured distribution identically for each datapoint in time. Therefore these uncertainties can only cause systematic errors. We attempt to correct for such systematic effects with the particle size calibration of the AMS. More sophisticated quantification, evaluation, and corrections of these effects are possible and have been discussed e.g. at AMS Users Meetings, but are outside the scope of this work.

The second uncertainty mentioned by the reviewer is that of ion flight times (IToF). Uncertainties in IToF are the same at all times during a campaign, as long as the ion source voltage is not changed — the typical practice during a field campaign. Thus these uncertainties are systematic and should not be included in the estimates of the precision. Rather, uncertainties in IToF are handled with the particle size and *m/z* calibrations of the instrument, and uncertainties in ion transfer time could influence a systematic correction for the PToF mode (see response to comments R1.7 and R3.5.)

R3.7. Page 4578, lines 1-3: How sensitive are the model results to the value selected for the "weak" variable SNR threshold?

A3.7 When using the SNR threshold for "weak" variables of 2, most of the "good" signal is at "picket fence" *m/z*'s common to HOA, BBOA, and LOA in the HR-MS solution, and long-chain hydrocarbons (figure below; compare to Fig. S3). Notable exceptions include the LOA marker, *m/z* 58, which has the highest SNR in the dataset, and the BBOA marker, *m/z* 60, which has "good" SNR over a large size range. In addition, the oxidation marker, *m/z* 44 has no "good" signal. The differences in the factors between the datasets with the weak thresholds set to 1.5 and 2 are minor.



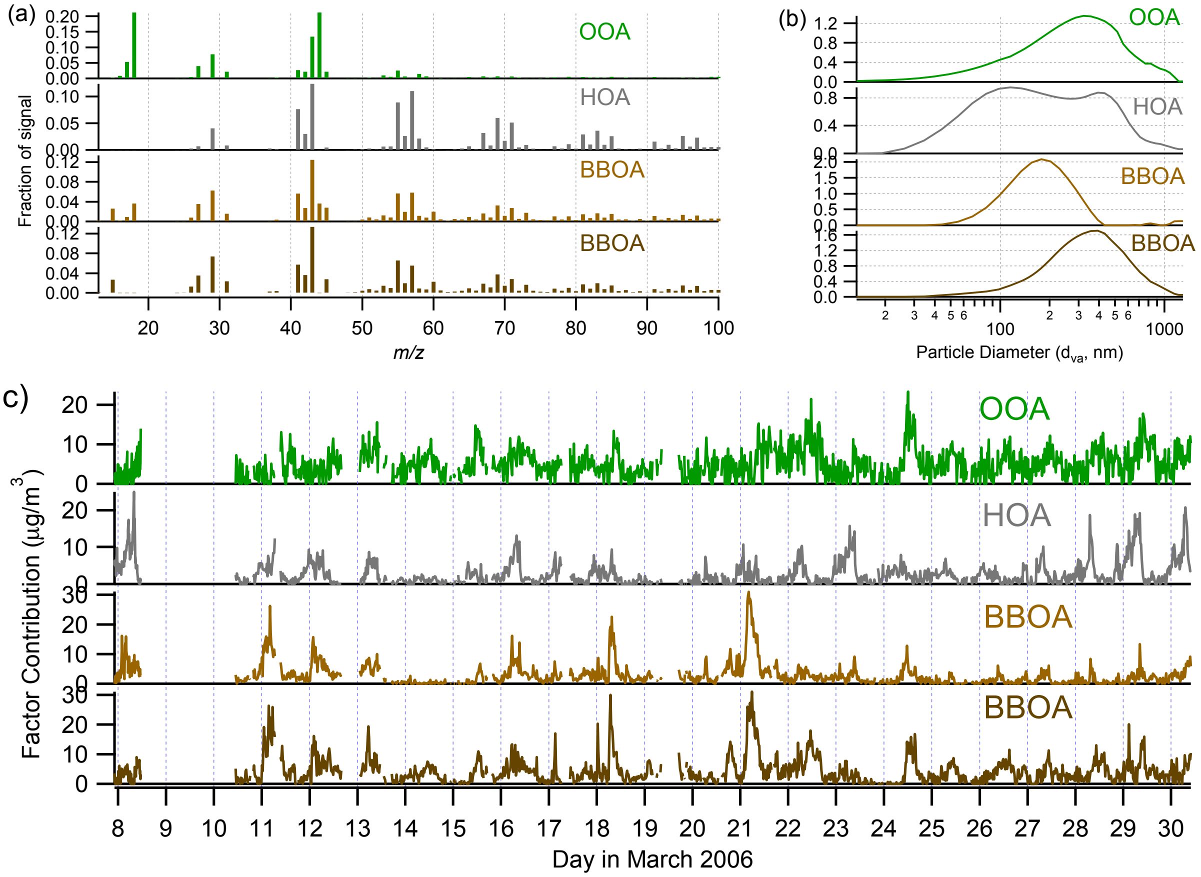
Early in our analysis we had applied downweighting to entire *m/z*'s (layers or "pages" of the 3-D dataset), similar to the approach of Pere-Trepat et al. (2007). In this case, *m/z*'s that were "weak" over all times and sizes were downweighted, so even more of the dataset was downweighted. Solutions using this downweighting scheme included unusual representations of the BBOA factor, which usually occurred with extremely high fractions of *m/z* 60 (f_{60} ~ 0.6). The high f_{60} resulted from smaller contributions from the "picket fence" *m/z*'s that are common to HOA, BBOA, and LOA. The change to the downweighting scheme used in the papers (i.e., downweighting "weak" *m/z*-size combinations) retained more of the "good" information and improved the interpretation of the BBOA factor mass spectra.

We note that the sensitivity of model results to different thresholds for "weak" SNR is very likely to depend on the particular dataset, and which and how many *m/z*'s have SNR near the threshold. If changing the threshold affects many *m/z*'s, the resultant Q/Q_{exp} value may change significantly. However, we would expect that factorization of datasets with overall higher SNR would be less sensitive to small changes in the "weak" SNR threshold.

R3.8. Page 4585, lines 2-8: "We presume that the bulk aerosol has the same composition whether measured in the MS or PToF modes… We therefore hypothesize that we should find the factors identified in the HR-MS analysis in the 3-D analysis, and may identify additional factors." This appears to govern the application of the authors' metrics for evaluation of the model solutions, so "hypothesize" is perhaps too weak a word. For example, taking together Table 3 and the evaluation of the unconstrained vector-matrix model solutions in section S6, it appears that the criteria for a "physically meaningful" solution include the requirement that all 4 factors from the HR-MS analysis be resolved. I do not find this a compelling argument, as there are reasons one could imagine retrieving a meaningful solution without fulfilling this requirement: (1) As discussed in Comment 5, a distribution of vaporization times within an HR-MS factor could lead to a substantial change or non-retrieval for the size-resolved analysis. Note that because of this, "split" factors having similar time series are not necessarily indicative of non-physical solutions. (2) Factors with distinct mass spectra and time series but similar size distributions could be inherently more difficult to distinguish in the vectormatrix model than in the 2-D HR-MS analysis. (3) The decreased signal/noise in PToF data means that the data pretreatment (e.g. smoothing, downweighting) may affect the resolution of the HR-MS and vector-matrix models differently, making specific factors harder or easier to resolve.

A3.8 The reviewer is correct that we relied on the HR-MS to guide our interpretation of the 3-D factorization results. Our hypothesis that we should find the same factors in both analyses was not intended to be a criterion for choosing a solution. However, we found that many of these factors were present in unconstrained solutions of the 3-vector model (Table 3); thus the underlying factors of the 3-D dataset are indeed those found in the 2-D PMF solution. But as we examined solutions of the 3-D factorizations, we were unable to convince ourselves (let alone a reader) that many of the other factors we saw were "real" because we could not support them with external information or assign them meaningful names that are chemically consistent with the mass spectra — what we would call "not physically meaningful." Thus the reviewer has helped us realize that our stated goal of identifying "physically meaningful" factors is really a convolution of finding "physically meaningful" factors that can be "supported" and linked to atmospherically relevant sources or processes. We discuss specific examples of rejected solutions in response to this reviewer's comments R3.9 and R3.11 below, and provide a third, related example here.

The reviewer points out that split factors with similar mass spectra and time series but different size distributions could be physically meaningful. But what would such factors mean, and how could we support them? One such case occurs in the four-factor solution of the 3-vector matrix model shown in Fig. S8 and discussed in Sect. S5.2 of the Supplement (lines 447-470) and reproduced below. The time series of the two factors are similar enough that we would not be able to support the separation based on correlations with other measurements. We wrote (Lines 451-455), "The two BBOA factors have quite different size distributions. One factor has a somewhat narrow size distribution with particle d_{va} that ranges ~40 nm to ~400 nm and has a mode at d_{va} ~180 nm. The other BBOA factor has a broader distribution with d_{va} that extends to 1150 nm and has a mode at a larger d_{va} near 370 nm." The difference in modes might suggest that the factor with the larger mode represents more aged BBOA particles. However, that factor lacks signal at *m/z*'s 18 and 44, which usually represent oxidation, while the mass spectrum for the factor with the smaller mode has contributions from *m/z*'s 18 and 44 of 3.5%. Thus the interpretations based on the factor size distribution and mass spectra are contradictory, and we find that we cannot support this solution. We could recombine the two factors to get a single size distribution, but that approach would not increase our understanding of the aerosol over a solution with a single BBOA factor."



We have updated the text in the Supplement in which we discuss the 3-vector solution with two BBOA factors to highlight the differences in the contributions of *m/z*'s 18 and 44 in the factor mass spectra. In addition, we have clarified our definition of "physically meaningful" by adding the following text on pg. 4585 line 2:

"Criteria (4) and (5) form the basis for claiming that factors are "physically meaningful" — that is, that physically meaningful factors can be linked to atmospherically relevant sources or processes and assigned meaningful names that are chemically consistent with the mass spectra."

This challenge of interpreting factorization solutions is one reason why we presented a number of solutions in the Supplement of the AMTD version. This practice, though unusual, makes this sort of discussion possible, and allows other researchers to reevaluate the interpretation of the solutions when new insights or solutions for related datasets become available in the future. We encourage other researchers to include additional solutions in supplements to their factorization manuscripts to continue these discussions.

The reviewer also states that "(2) Factors with distinct mass spectra and time series but similar size distributions could be inherently more difficult to distinguish in the vector-matrix model than in the 2-D HR-MS analysis." We disagree. Two factors with distinct mass spectra and time series could not be fit by one factor because both the mass spectrum and the time-dependent size distribution of the resultant factor would have to be linear combinations of the two factors. If the two factors have sufficiently distinct mass spectra and time series, a single factor could not fit the data well.

Finally, the reviewer suggests that pretreatments to the dataset may cause differences in the ability to retrieve factors in the 2-D and 3-D cases. Reviewer 2 (comment R2.8) noted "that linear smoothing before PMF may enable the recovery of weak factor(s) that would not be detectable without such smoothing." Similarly, downweighting also decreases the influence of noisy data and increases the influence of good data (Paatero & Hopke, 2003). The reviewer is correct that the resulting weighting might be different in the HR-MS and PToF datasets, and thus, the same factors may not be retrievable in both. However, the factors in danger of not being retrieved are those that are the "weakest." We would put LOA in this category since it contributes only 10% of the HR-MS organic mass and its HR-MS mass spectrum has significant contributions from many *m/z*'s also common to HOA and BBOA. Nevertheless, a factor representing LOA was found in unconstrained solutions of the 3-vector and vector-matrix model. In general, we would expect that most factors that could be identified from the noisier PToF dataset would have been found in the less noisy HR-MS dataset, unless adding the third dimension — here, size — gives a weak factor a distinctly different structure. Our hope in performing the present analysis was that adding size data would help identify factors that were not separable in the HR-MS dataset, though this did not happen for this particular dataset.

R3.9. Supplement S6, lines 527-533: I do not understand why the authors prefer a single OOA factor to OOA described by a combination of m/z 44- and 43-dominated factors. Such a split is consistent both with (1) evaporation timescales in the AMS and (2) current descriptions of OOA as consisting of components with a range of volatilities, where high-43 OOA is expected to undergo repeated evaporation/condensation with temperature cycles. This could give a high-43 OOA distribution more influenced by the low-volatility particle surface area distribution. For both cases, one would expect a size distribution skewed towards larger sizes for the high-44 OOA, which appears to be the case in Fig. S7 (and perhaps also in Fig. S12).

A3.9 First, we note that the separation of OOA into one or two factors depends on the dataset. For example. a summary of many campaigns (Fig. 1 of Jimenez et al., 2009) shows that in some campaigns the authors were able to separate SV- and LV-OOAs, while in others it was only possible to retrieve total OOA. In the 2-D factorization of this dataset, two types of OOA could be separated when analyzing subsets of the entire time period [as reported in e.g. Aiken et al. (ES&T 2008)]. However, tuning changes in the mass spectrometer detector that were made a few times during the campaign prevent the separation of two OOA factors when factoring the entire 2-D dataset. So it is possible that the same cause might prevent good separation of two OOA factors in the 3-D dataset, too.

However, the report of only one OOA factor in the 2-D HR-MS dataset was not the reason that we preferred solutions with a single OOA factor in the 3-D factorization solutions. Rather, we found that factors dominated by *m/z* 44 were not physically meaningful by the definition clarified in A3.8 above. Our explanation for why these factors are unacceptable was given in the Supplement in lines 403–409 and is repeated here:

"One factor has a mass spectrum dominated by signals at *m/z* 44 and related peaks, as defined in the fragmentation table (*m/z*'s 28, 18, 17, and 16). The fraction of signal contributed by *m/z* 44 (f_{44}) is proportional to the aerosol oxygen-to-carbon ratio (O:C), which is also proportional to aerosol age, so high f_{44} represents highly oxidized, aged aerosol (Aiken et al., 2008). This factor might therefore represent OOA, and in fact has a high correlation with the OOA factor from the three-factor solution (R = 0.86). But this high-*m/z*-44 factor contributes only half of the mass of the OOA factor in the three-factor solution. Curiously, this mass spectrum has $f_{44} = 0.39$, and the group of related peaks contributes 90% of the signal in this mass spectrum. This spectrum represents aerosol with an elemental oxygen-to-carbon ratio (O:C) of 1.6 (Aiken et al., 2008) — much higher than O:C measured in ambient or laboratory aerosol by the AMS (Ng et al., 2010). O:C can be combined with the estimated hydrogen-to-carbon ratio (H:C, Ng et al., 2011a) to estimate the average carbon oxidation state (OS_C) of the aerosol from this factor (Kroll et al., 2011). This factor has $OS_C = +2.0$. However, the OS_C of atmospheric aerosols rarely exceeds +1 because compounds with higher oxidation states tend to decompose (Kroll et al., 2011). Thus the factor in this solution with $f_{44} = 0.39$ is not physically meaningful, even though it improves the fit of the factorization, as evidenced by its appearance in the family of solutions with the lowest Q/Q_{exp}."

Thus we would describe this factor as explainable, but not supportable. We think that most reviewers and readers would not accept the mass spectrum of this factor as the full representation for a more-oxidized OOA.

R3.10. Supplement S6, lines 534-536: An m/z 15-dominated factor is observed to appear repeatedly in the unconstrained 3-vector and vector-matrix solutions. This is taken as evidence of a non-physical solution. It may simply indicate a problematic m/z for which the sources of error have not been completely assessed. The authors should explore removing or downweighting this element if they believe it impedes their ability to retrieve meaningful results.

A3.10 Upon reviewing the signal and noise at *m/z* 15, we saw that this was the only *m/z* with “good” SNR at $d_{va} < 45$ nm (Fig. S3), which is near the lower range of particles transmitted by the aerodynamic lens (~35 nm, pg. 4570 line 18). Thus this signal is unlikely to represent particles and may instead represent gas-phase contributions of ^{15}N. Therefore, we could downweight the signal in this range, as happened with *m/z*’s 29 and 44 (see A1.8). If we increase the “weak” SNR threshold for *m/z* 15 to 1.6 (recall that the rest of the dataset has a “weak” SNR threshold of 1.5), then the signals of *m/z* 15 for $d_{va} < 125$ nm and $d_{va} > 412$ nm are “weak.”

We factored the dataset with this revised error matrix using the 3-vector and unconstrained vector-matrix models. The results were similar to those in the AMTD version of the manuscript. The factor dominated by *m/z* 15 did appear in some solutions of the 3-vector model with 8 or more factors, or in solutions of the unconstrained vector-matrix model with 7 or more factors. Solutions with fewer factors showed further splitting of factors such as HOA and BBOA, and many solutions included multiple HOA or BBOA factors with similar mass spectra and time series but different size distributions (similar to those described in A3.8). However, none of these split factors can be supported as physically meaningful.

We have updated the manuscript with the results from the dataset with downweighted *m/z* 15. This change also required us to run the constrained vector-matrix model with the LOA factor from these new studies. However, the differences between the factors with and without downweighting *m/z* 15 are quite small. Factor mass spectra from the two cases are virtually indistinguishable from each other. The time-evolving size distribution matrices for each factor show slight differences for some factors. The greatest differences in these matrices are observed for OOA, the noisiest factor, and then mainly in the points with the smallest (noisiest) signals. Even so, the correlation between the time-evolving size distribution matrices for the OOA factors is R=0.998 with a slope of 0.9996 ± 0.0003. Thus, the constraint on the solution has more influence than the weighting of *m/z* 15 for this dataset.

R3.11. Supplement S6, lines 536-539: The authors observe a factor having a prominent set of m/z’s (67, 81, and 95) corresponding to specific hydrocarbon classes, but discard the factor as nonphysical because “the spectra of real compounds contain more than just these three fragments.” First, Figure S12 clearly shows that many more fragments are present. In fact, many of the characteristic HOA m/z are present, albeit at reduced intensities and with the absent peaks grouped at low m/z, so this statement is rather misleading. But more importantly, it is not necessary that the factor mass spectra resemble the spectra of an actual molecule or an “aerosol type,” to be meaningful. For example, it is possible for factor analysis to represent a complex or variable factor as the linear combination of two other factors; in this case, only the combination of factors need have a spectrum resembling a molecule or aerosol type. The OOA subtypes are well-known examples of such linear combinations in two-dimensional factor analysis. The OOA spectra are not understood to represent discrete aerosol types, but rather serve as a basis set for describing the dynamic OOA composition. It is possible that the 67/81/95 factor and HOA shown in Fig. S12 form an analogous basis set. Other interpretations of this and other solutions may be possible, and the authors may be able to demonstrate they are correct in considering the factor physically unmeaningful. However, this requires careful consideration and discussion of the factor, which is currently not included. At present, the analysis structure seems overly governed by the authors' expectation that the resolved factors should follow a particular pattern analogous to the HR-MS solutions (see Comment 8), leading to the exclusion of potentially meaningful factors.

A3.11 We disagree with the reviewer's statement that a factor mass spectrum does not have to resemble that "of an actual molecule or 'aerosol type' to be meaningful." In fact, it was the similarity of factor OOA spectra to those obtained from laboratory standards [e.g., fulvic and humic acids for OOA and LV-OOA (Zhang et al., ES&T 2005)], laboratory chamber oxidation studies [e.g., ozonolysis of b-caryophyllene and photooxidation of diesel exhaust for SV-OOA (Ulbrich, 2009)], and field observations of aged organic aerosol [e.g., in urban and rural Canadian sites (Zhang et al., ES&T, 2005)] that convinced us and others that the early factorization results were believable. If a factor mass spectrum does not resemble an actual group of molecules or aerosol type, by what process could we name and interpret it?

Thus we disagree that the spectra of previously reported OOA subtypes (low-volatility and semi-volatile OOA, LV- and SV-OOA respectively) "are not understood to represent discrete aerosol types." To the contrary, LV- and SV-OOA are in fact distinct aerosol types — representing less and more oxidized OA — and also, as the reviewer states, "a basis set for describing the dynamic OOA composition" as it undergoes processing in the atmosphere. Each OOA subtype has a real, physical interpretation, and the subtypes describe endpoints of an atmospheric continuum.

It is true that a basis set could be formed from the solution shown in Fig. S12 that contains HOA and a factor mass spectrum dominated by *m/z*'s 67, 81, and 95. But what aerosol space would this basis set span? We are not sure how to describe the space without understanding both endpoints. The major fragments in the 67/81/95 mass spectrum are prominent in the spectra of alkynes in which the triple bond is near the center of a long chain (e.g., 3-octyne) or dienes in which the two double bonds share a carbon and are near the center of a long chain (e.g., 3,4-heptadine) [NIST Webbook]. Perhaps this factor represents such compounds. However, we feel that such a specific source assignment would be an overinterpretation of the data without additional evidence to link the factor to an atmospheric source or process. In addition, these solutions usually lack a BBOA or LOA factor, and have correspondingly large Q/Q_{exp} contributions from *m/z* 58 or *m/z* 60, respectively. Given the high SNR in each of these *m/z*'s (see A3.7), their high Q/Q_{exp} values are a strong indicator that important information in the data matrix has not been fit.

Nevertheless, the reviewer is correct that this mass spectrum does contain "more than just these three fragments," and the relevant text has been updated.

R3.12. Page 4590, lines 18-20: "Thus, the vector-matrix model captures the real variability in the size distribution of each factor better than the 3-vector model when using the same spectra." The case for this statement is built on the lower Q/Qexp values observed for the vector-matrix model. However, Q/Qexp was previously dismissed by the authors as a metric for judging a solution's explanatory power for the unconstrained vector-matrix model, which yields solutions with a lower Q/Qexp but is judged physically unmeaningful. The authors indicated that the decreased Q/Qexp for the unconstrained vector-matrix model was due to fitting the noise (Supplement, lines 518-526). Why is Q/Qexp now used in the vector-matrix model to indicate an improved description of real variability? It is more likely that the apparent improvement is again due to fitting the noise via increased degrees of freedom. If the authors wish to make this case, they should be consistent in the usage of their metrics or demonstrate that Q/Qexp is an applicable metric in only the selected case.

A3.12 The reviewer has misinterpreted the text from the Supplement, which states,

"The low Q/Q_{exp} values for the vector-matrix solutions imply that the vector-matrix model can fit more of the data than the 3-vector model. In other words, the two vectors in the 3-vector model fit the size-time information less well than the matrix in the vector-matrix model. The better fit of the vector-matrix model is consistent with this model's greater degrees of freedom [Eqns. (6) and (7)]. However, the large number of degrees of freedom in the vector-matrix model also allows the model to fit a large amount of noise. The ability to fit noise in this low SNR dataset distorts the mass spectra in the model solutions."

"No solutions of the vector-matrix model have mass spectra that are all physically meaningful (Table 3)."

The text then gives examples of factors that we could not support as physically meaningful using the definition that was clarified in response to comment R3.8.

The text in the Supplement describes two separate criteria for choosing good solutions: Q/Q_{exp} values, and whether factors are physically meaningful. The unconstrained vector-matrix solutions have low Q/Q_{exp} values; this criterion supports the choice of an unconstrained vector-matrix solution. However, upon examining the factors in these solutions for physical meaningfulness, we find them lacking, and reject the solutions based on this second criterion. Thus the reviewer is incorrect that the Q/Q_{exp} value was "dismissed by the authors as a metric for judging a solution's explanatory power for the unconstrained vector-matrix model." As we discussed in Sect. 3.4.3, Guidelines for choosing a solution, Q/Q_{exp} is only one of several metrics used to choose a solution: "(1) $Q/Q_{exp} \sim 1$, (2) decrease in the rate of change of Q/Q_{exp} with increasing number of factors, (3) little structure in the solution residuals, (4) strong correlation between component time series and diurnal cycles with those of tracers not included in the factorization array, and (5) plausibility of the factor mass spectra and their similarity to observed spectra of real-world sources."

R3.13. Page 4595, line 26 to page 4596, line 3 and Fig. 5: Overall, Q/Qexp values show that the best solution of the constrained vector-matrix model fits the data better than the best 3-vector solution, although the differences are small (0.95 vs. 0.97, respectively; Fig. 5)." I do not understand how the 3-vector Q/Qexp values are represented in this figure. The legend in the top panel denotes the "range of 3-vector solutions" with a horizontal black bar; in the figure, this extends from beta = 0 to beta 0.05, at a constant value of 0.965. This doesn't seem right; am I misreading the figure or is the 3-vector solution data not included?

A3.13 We have found the representation of the 3-vector Q/Q_{exp} values in this figure challenging. The thin black line is the correct representation of the range of 3-vector Q/Q_{exp} values for the solutions described, similar to the grey box for the unconstrained vector-matrix solutions. The range of Q/Q_{exp} values for the 3-vector solutions is quite small, as shown in Fig. 3a (blue /; the value is the midpoint of the marker). The black line is only shown across beta = 0 to 0.05 so that it doesn't interfere with the figure legend in the upper figure. We have changed the line so that it represents the range of 3-vector solutions with four factors, which makes the line thicker.

R3.14. Page 4599-4600: The authors compare results obtained with the 3D models to estimated size distributions constructed from tracers for the 2D HR-MS factors. Given the similarity of the tracer size distributions to the vector-matrix model, what do the authors see as the advantage of the 3D analysis over the tracers? As presently formulated, the 3D analysis requires the existence of constrained 2D factor mass spectra, from which tracer size distributions could presumably be constructed. In the case of a disagreement between the vector-matrix model and the 2D size distributions, how would the authors demonstrate that the vector-matrix model yielded useful data? Certainly for the mass concentration data, the 2D HR-MS solution has here been treated as superior. Given this, why should vector-matrix size distributions be superior to a high-resolution tracer (e.g. vector-matrix HOA vs. C4H9+-based HOA)?

A3.14 The 3-D factorization approach has two advantages over the tracer method alone. First, the 3-D method has the potential to find factors not found in the 2-D factorization, as described in A3.8. A tracer method could not provide size distributions for such otherwise-unknown factors. Second, the 3-D approach does not require coefficients for the tracer method. In this study, we used the OOA tracer equation from Ng et al. (2011) instead of that derived by Aiken et al. (2009) because the Aiken relationship subtracted a high *m/z* 44 background signal not observed in the PToF data and resulted in negative OOA size distributions. In fact, if the tracer and vector-matrix size distributions disagreed, we would suspect that the coefficients in the tracer method were inappropriate for that dataset.

As we discussed in response to comments R3.0 and R3.8, we disagree with the reviewer that "the 3D analysis requires the existence of constrained 2-D factor mass spectra." However, we do agree with the reviewer that we treat the 2-D HR-MS solution as superior in this case. The HR-MS dataset has much higher SNR and contains high-resolution spectral information from the separation of ions at the same nominal *m/z*. Because the 3-D PToF dataset was so noisy in this case, constraining the spectra in the vector-matrix model to those from the HR-MS factorization helped make the 3-D results more meaningful. However, there is no reason that this treatment should be necessary for every dataset.

High-resolution tracer size distributions (e.g., based on the $C_4H_9^+$ size distribution) are difficult to implement because extracting and fitting high-resolution mass spectra has not yet been implemented in the standard HR-ToF-AMS analysis software. Calculating such high-resolution size distributions can be done manually, e.g. in Figure 6 of DeCarlo et al. (Anal. Chem., 2006), but this is a large amount of work and is outside of the scope of this study. Once high-resolution PToF distributions become available, it will be very interesting to apply 3-D factorization methods to these datasets. Once such studies are performed and reliable coefficients for a tracer method are established, we expect that HR tracer size distributions might be a simpler approach than 3D factorization. However, the tracer method would never identify factors that might be separated only in the 3-D factorization if they have a distinctive size distribution.

R3.15. Page 4601, lines 8 to 10: “OOA has... an asymmetric distribution with strong extension to smaller particles... consistent with the condensation of semivolatile gasphase molecules onto smaller particles.” Note that this is also consistent with the description of OOA as a combination of a more volatile (m/z 43-dominated) OOA and a less volatile (m/z 44-dominated) OOA, as discussed in Comment 9.

A3.15 We agree with the reviewer that the OOA factors dominated by *m/z* 43 and *m/z* 44 are consistent with more-volatile and less-volatile OOA, but we still find the mass spectra for these factors not physically meaningful, as discussed in response to comment R3.9.

R3.16. Page 4604, lines 12 to 17, also page 4606, lines 21 to 24: The authors provide some suggestions for future researchers who wish to investigate 3D factorization of size-resolved AMS data. They suggest beginning with 2D factorization of HR-MS data to better understand the dataset, which makes sense. However, the utility of this step is described in terms of diagnosing unsatisfactory solutions and identifying split factors. Unfortunately, this evaluation appears to depend in large part on the unproven assumption that the 3D model should yield essentially the same set of factors as the 2D model. As discussed above (esp. Comments 8-11 and 15), there are many reasons why this assumption could break down. Discarding 3D model solutions containing significant deviations from the 2D solution as non-physical or erroneous both severely limits the analysis and renders it circular. To summarize, the authors assume that the 3D solution should resemble the 2D HR-MS, discard solutions that do not satisfy this condition as non-physical, and finally conclude that the 2D solution was vital in interpreting the 3D (after constraining the 3D model such that the output spectra closely resemble those of the 2D solution). While the data obtained by this method may indeed prove quite useful in certain cases, it is far from a comprehensive analysis of the 3D solution, and the recommendation for future researchers to follow the same method is not supported by the analysis. To justify such a recommendation would require a much deeper analysis, focused specifically on (1) characterization of the 3D model behavior with respect to factor identification, mixing, splitting, description of dynamic factor composition (e.g. OOA), etc., and (2) development and evaluation of criteria for evaluating non-physical factors on their own merits (i.e. without invoking the 2D solution), and allowing for the possibility of complex and/or dynamic composition to be characterized by linear combinations of factors.

A3.16 We repeat here, from A3.8, that the underlying factors of the 3-D dataset are indeed those found in the 2-D PMF solution, since the same main factors were identified in solutions of the unconstrained 3-vector model. Therefore, it is not an “unproved assumption” that the factors from the 2-D factorization should be present in the 3-D solutions. We disagree with the reviewer that the analysis method is circular. Indeed, the spectra in the vector-matrix model were constrained to those from the 2-D factorization and we relied heavily upon the 2-D factorization results to understand the unconstrained vector-matrix model factors. However, most of those factors were determined not to be physically meaningful based on other arguments, e.g., too-high fraction of signal from *m/z* 44 (A3.9 and Supplement lines 403–409), or unusual fragmentation patterns (A3.11 and Supplement S6, lines 536-539).

Our recommendations to other researchers suggest using the 2-D factorization results to understand the main trends and factors from 3-D model solutions. However, we do not recommend that the 3-D solution must match the 2-D solution, nor that researchers constrain the 3-D solutions. We have clarified this position by adding the following text at pg. 4604 line 19:

“**Finally, we do not recommend constraining the factors in any model unless the unconstrained solutions are not useful.**”

Finally, we address the reviewer’s suggestions for deeper analyses of the 3-D solutions in reverse order. First, we do support linear combinations of factors as valid representation of aerosol types, but we did not find them to be beneficial in this case (A3.11). Second, our criteria for evaluating the physical meaningfulness of factors in the 3-D analysis include “(4) strong correlation between component time series and diurnal cycles with those of tracers not included in the factorization array, and (5) plausibility of the factor mass spectra and their similarity to observed spectra of real-world sources” (A3.12), and are not strictly dependent on the 2-D factorization results. Finally, an extensive characterization of 3-D factorization model behaviors would require a separate, detailed study with synthetic data, and are outside the scope of the current study.

Minor comments
R3.17. Page 4585, line 5 “MS and PToF modes” should be “MS or PToF modes”

A3.17 We have corrected this typo.

R3.18. Page 4586, line 9: “solutsions”

A3.18 We have corrected this typo.