Response to Comments by Referee J. Allan 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 (R1) 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.

R1.1 This paper presents the application of two 3D factorisation techniques to an AMS dataset. This is a very timely and relevant paper, given the fact that no-one has published an analysis technique such as this before. While many important papers have been published using 2D factorisation, these have intrinsically been limited to the integrated mass concentrations with no size-resolved information being delivered (although I am aware of people attempting this on a few occasions). While there is no great scientific insights provided in this instance, the work has the potential to be of high significance if applied to other datasets. Overall, the paper is very thorough and certainly appears to be very diligently prepared. I have no hesitation in recommending this be published in AMT.

A1.1 The authors thank the reviewer for his positive comments on our work.

I would ask, however, that the authors take the following (mainly technical) comments under consideration first:

General comments:

R1.2 The error model as presented will not be completely sufficient for the 3-vector data model because it does not include particle counting statistics. This is not included in the 2D TOF-AMS error model for PMF because it is a multiplicative error in m/z space, however it will have an effect in PToF space if the aerosol is polydisperse (even if it is monodisperse, there will still be a smaller effect due to chopper broadening). Successive samples of a given aerosol will produce differently shaped distributions if the number of particles within a sample is small. However, this is unlikely to be an issue in a high signal environment such as Mexico City and is likely to be insignificant compared to the problems associated with 'real' changes in particle size with time. Given how much work would be involved in putting this in the error model, I am not asking the authors to include it. However, it should be mentioned as a caveat for the sake of someone attempting to repeat this work in a low signal environment. Note that this will not affect the vector-matrix model, as this treats the PToF points independently so random variations in the size distribution caused by particle counting statistics will not be discernible from 'real' variations.

A1.2 We agree with the reviewer that particle-counting statistics could be important to the uncertainty estimates in some cases, including low particle number concentration environments. We have added the following text on pg. 4574 at line 15:

"A fourth term could be added to the estimation of σ to reflect particle-counting statistics when particle number concentrations are low."

R1.3 I feel a little uneasy about how the beta parameter is being used to constrain the factors present. The difference in vaporisation times (which, to be fair, the authors point out on several occasions) presents a physical reason why the factor profiles from MS and PToF modes might not be expected to agree. Furthermore, the level of disagreement may vary factor to factor, so hence different levels of relaxation may be appropriate for different factors (although I accept that implementing and testing this would be a little much to ask for this paper). The discussion of the choice of beta to be 0.06 on pages 4592-3 justifies it empirically, but I would be more comfortable if the authors could provide some more discussion in this section on the competing physical or statistical effects that would dictate the optimal value of this parameter, if only on a conceptual level. This will be invaluable to anyone attempting to repeat this work on other datasets because my gut tells me the optimal value is not likely to be the same.

A1.3 We agree that a separate β for each factor might be the most appropriate way to constrain the factorization, but that exploring the solution space of factorizations to choose the correct combination of βs would be an enormous amount of work for any dataset, not just for this paper.

Our approach to choosing the “best” β for the factorization solution is similar to the approach for α used by Lanz et al. (2008) and is purely empirical. We do not know of any physical or statistical effects that might dictate an optimal value of beta. However, we can posit that using constrained spectra that better match the “true” factors in the dataset might allow smaller values of β, while constrained spectra that don’t match as well would require more “wiggle room”, i.e., larger β, to get to a good solution. We have modified the beginning of the paragraph on page 4591 line 22 to read:

“We know of no physical or statistical constraints that could be used to predict the optimal degree of relaxation (β) of the constrained factors. However, we posit that starting guesses that better match the dataset might allow smaller values of β, while constrained spectra that are less similar to the true factors would require a larger deviation from the original factors, i.e., larger β, to obtain a good solution. Thus, we take an empirical approach to determining the most appropriate value of β for this dataset. Two options for choosing β are considered here."

We did not mean to imply that the value of β we selected for this dataset should be correct for other datasets. To the contrary, we would encourage other researchers who adopt our approach to carefully examine the residuals of the fit and other available metrics to choose an appropriate value of β for their datasets. We have revised the paragraph on pg. 4604 at line 12 to include a third recommendation to researchers who factor size-resolved AMS datasets, adding the following text at the end of that paragraph (line 19):

“However, we recommend that researchers who constrain factors using a parameter such as α or β carefully examine residuals of the fit and other available metrics to choose an appropriate degree of relaxation for their datasets.”

R1.4 The figure captions are, in general, a little on the long side. I would suggest the authors revise them for brevity and rely on the body text to convey some of the more technical concepts.

A1.4 The figure captions are written so that the reader may skim the figures and understand the outline of the paper without needing to refer to the text. However, we have tried to shorten the captions where possible, while still maintaining the narrative.

R1.5 Sections 1 and 2 of the supplementary material are informative, but most of it does not seem to be relevant to the work presented here (the 2D work in particular seems a little out of place, given 2D factorisation of AMS data is already well established). Unless the authors can provide more context, I don't see why these sections are necessary.

A1.5 These sections summarize previous studies that factor datasets that include particle size information. Such studies are less common than those performed on composition-only data, and their inclusion in this work gives an overview of the current state of the field, so we prefer to keep these sections in the paper. Reviewer 2 also suggested that these sections remain in the text (see Response to Comments by Pentti Paatero, comment R2.1).

Specific comments:
R1.6 P4563, L4: The comment about being applicable to other techniques is not really adequately supported (see later). I would suggest removing or toning down this comment.

A1.6 Our complete response to this comment is given below in A1.11.

R1.7 P4571: The authors are slightly incorrect in their discussion of PToF vs IToF times. IToF is the time taken to transit between the orthogonal extraction region and the detector, however all the data collected on an individual mass spectrum is organised according to the time of the extractor pulse, which is fixed in PToF space, so therefore IToF in itself will not cause any artefacts at all on the PToF distribution. However, what could potentially cause a problem is the time delay between ionisation and transport to the orthogonal extraction region. This is not measured and could, in principle, vary according to m/z ratio. This should be discussed.

A1.7 We have revised the beginning of the first paragraph on pg. 4571 (lines 3–10) to read

"Particle size is determined from particle velocity in the particle-flight region. Particle velocity is calculated from the length of the particle-flight region from the chopper to the vaporizer (0.293 m) and particle time of flight (PToF), measured as the time between the midpoint of the opening of the chopper slit intersecting the particle beam and the pulsed extraction of ions into the mass spectrometer. The measured PToF and the length..."

We have also added details about the extraction frequency for this study, updating the text at pg. 4571 line 17 to read

"Mass spectra are recorded, e.g., 100 times during each chopper cycle, then reduced to 50 spectra by co-adding two adjacent spectra, to obtain one spectrum every 50 µs in this study."

Finally, the description of the contributors to measured PToF and their relative lengths has been incorporated into the paragraphs that begin at pg. 4571 line 25, which have been rewritten to read

“This description of PToF sampling describes “ideal” particle behavior in which particles stick to the vaporizer and vaporize quickly. However, the measured particle size can be affected if non-ideal behavior changes any part of the measured PToF. The measured PToF includes three components: the true PToF from the chopper to the vaporizer, the vaporization and ionization time, and the time for ions to transfer from the ionization region to the TOF mass spectrometer extraction region.

We discuss possible deviations from each component and their effects here. First, the true PToF can change if a particle bounces off the vaporizer, then impacts another surface and evaporates; thus the vaporization is “delayed.” If this particle’s vaporized molecules are ionized and the ions are mass analyzed, they are recorded at a nominal size larger than the true particle size. Second, different aerosol species have different vaporization rates. “Fast” vaporization happens on the timescale of <200 µs for several aerosol constituents [NH_4NO_3 and $(NH_4)_2SO_4$ vaporize in 80 and 150 µs, respectively, at 600°C (Cross et al., 2009)]. In contrast, some inorganic and organic compounds in ambient aerosol can have longer evaporation timescales (1/e decay), e.g., $<10^5$ µs (0.1 s for PbCl, Salcedo et al., 2010), $>5\times10^6$ µs (5 s for 5% of the total organic mass, Huffman et al., 2009) , or 1.4×10^8 µs (2.4 min for other Pb compounds, Salcedo et al., 2010). Ions from these longer vaporization timescales are measured as part of the background signal in PToF mode, while ions from moderate vaporization timescales are recorded at larger nominal sizes than the true particle size. The final component of the measured PToF is the transfer time from the ionization region to the pulsed extraction region of the TOF mass spectrometer. For typical voltages of the transfer ion optics (of the order of a few tens of volts; we will use 20 V for this example calculation), the time to travel 96 mm between the ionization and extraction regions (Drewnick et al., 2005) is 6 and 27 µs for *m/z*’s 17 and 300, respectively.

Thus, transfer time << vaporization time << measured PToF (~3000 µs), and only variations in the vaporization time could potentially be large enough to shift the recorded particle size. Note that these deviations broaden the size distribution only towards nominally larger particle sizes. These possible deviations should be considered part of the “transfer function” of the instrument, and could be further evaluated and corrected, although such techniques are outside of the scope of this paper.”

R1.8 P4579, L5: It is worth pointing out that the PToF distributions of the gases in the AMS have been found to roughly follow Boltzman distributions according to the kinetic velocities of the gas molecules (ch. 2 in: "Chemical Kinetics and Microphysics of Atmospheric Aerosols", James Morris, Ph.D. Thesis Boston College, 2002). This will cause the CO2 molecules to arrive slightly later on average than the nitrogen molecules.

A1.8 We thank the reviewer for highlighting this resource, which describes the Boltzmann distribution of molecular velocities in one component direction, with a drift velocity term.

The differences between the Boltzmann distributions of $N^{15}N$ and CO_2 and the reference N_2 distribution increase as the difference between the masses of the molecules increases. Therefore, using the N_2 distribution to estimate that of $N^{15}N$ is a reasonable assumption, but this assumption is not as good for estimating the CO_2 distribution. The heavier CO_2 molecules tend toward slower velocities and thus later arrival times in the ionization region; therefore, estimating the CO_2 distribution from the N_2 distribution leads to an oversubtraction at the peak of the CO_2 distribution and an undersubtraction in the tail of the distribution of later-arriving CO_2 molecules. Thus, even after truncating most of the gas-phase signal, some of the tail of the gas-phase CO_2 signal remains in the prepared 3-D dataset. A better approximation of the $N^{15}N$ and CO_2 distributions should be explored as a direction of future research.

However, while reviewing some of our data preparation, we realized that we had not actually applied this downweighting in the final preparation of the factorization matrices. Therefore we have removed the text describing Step 6 (pg. 4578 line 9 to pg. 4579 line 14), the accompanying line in Table 2, and Table S1. The region of potential gas-phase influence at *m/z*'s 29 and 44 had low SNR and was downweighted by a factor of 2 in the previous data preparation step; this downweighting appears to be strong enough to address most of the interference. Furthermore, as described in response to a comment from Reviewer 3 (R3.10), we have increased the weight on the gas-phase-interference region of *m/z* 15. However, some solutions of the 3-vector model with seven or more factors include factors dominated by *m/z* 15 or 29, so better estimation and downweighting of gas-phase interference may be helpful for future studies.

R1.9 P4585, L23: The authors need to expand on what they mean by "too small" and what they are comparing them to.

A1.9 This sentence was actually incorrect due to a misunderstanding among some of the authors, and referred to a different version of the error estimates. This sentence has been removed from the text.

R1.10 P4598: One line of discussion notably absent here is whether the authors think that there is any 'real' variation in the mass spectral profile of factors taking place (BBOA, for instance, is known to vary according to fuel, burn conditions and atmospheric processing). This may be part of the cause for the variability and ambiguity in the solutions.

A1.10 The reviewer is correct that there can be variability in the source spectra, and the variability can take several forms. For example, there is variability among the spectra of particles from individual cars or fires; however, the sum of the exhaust from tens of thousands of cars, or smoke from wildfires, in the real world tends to show much less variability in our experience when measuring the ensemble average of the variability of the individual sources. We did not discuss this type of variability explicitly in the manuscript because we feel that the issues are the same in the 2-D and 3-D factorizations and thus did not warrant further discussion in this manuscript.

A second type of variability in factor spectra can occur because of chemical processing in the atmosphere. This type of variability is present at some level in all ambient AMS datasets. However, we still chose the vector-matrix model in which the vector contains the factor mass spectrum because, from our experience in analyzing ambient data, "we expect the component size distributions to change more rapidly than their chemical composition" (p. 4569 line 12-13).

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**" with the following text:

"When the model does not match the data

Both models used in this study — and the bilinear model in 2-D factorization studies — assume that there is no variability in the mass spectrum of each factor. However, as we discussed in the Introduction, aerosol undergoes chemical processing in the atmosphere and thus their spectra can change, so the assumption of constant factor mass spectra cannot be strictly correct. These variations might appear in two forms in the factorization results. The variation could be fit by an additional factor, or the variation may not be fit well and appear in the residuals. For example, Ulbrich et al. (2009) reported a dataset 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. Unfortunately, we are not aware of criteria that might predict when changes in the underlying data are significant enough to cause separate factors, and so must instead consider this possibility carefully when choosing a particular factorization solution.

Despite the limitations of the assumptions of the models used in this work, we believe that they are the most appropriate for the data being analyzed here. Models that allow spectral variation should be considered in future research."

In addition, we have added a section to the Discussion in which we summarize how these differences between the data and the model appear to been fit. This new text appears in the response to a comment from Reviewer 3 (R3.4).

R1.11 P4604-5: While it is tempting to think that this technique will be directly applicable to other aerosol mass spectrometric techniques, many of the examples cited will carry with them their own inherent issues that should not be trivialised. For instance, the GC/MS-FID will have the problem of column bleed, which could potentially play havoc with the data model presented here. Also, the AMS thermal denuder data will need to be binned and normalised to be put on a 3D data scheme, which will need to be taken account of in the error model. While 3D factorisation no doubt has great potential here, I'm a little concerned that the authors are perhaps a little too bullish about its chances of success.

A1.11 We are beginning to explore some of these other datasets and, perhaps luckily, these two examples do not appear to be problematic. Nevertheless, the reviewer is correct that applying the factorization models to new datasets requires careful construction of error estimates and appropriate choice of models that match the underlying processes in that data. We have revised the text regarding other 3-D datasets in three locations.

First, we have revised the last sentence of the abstract (pg. 4563, lines 4-7) to read,

"Furthermore, other fast and chemically complex 3D datasets, including those from thermal desorption or chromatographic separation, could be analyzed with these 3-D

factorization models. Application of these models to new datasets requires careful construction of error estimates and appropriate choice of models that match the underlying structure of those data. Factorization studies with these 3-D datasets have the potential to provide further insights into organic aerosol sources and processing."

Second, we have added the following sentence at the end of the discussion (pg. 4605, line 10):

"Application of 3-D models to new datasets requires careful construction of error estimates and appropriate choice of models that match the underlying structure of that data."

Finally, we have revised the last sentence of the conclusions (pg. 4606, lines 26-29) to read

"Finally, these techniques, with appropriate error estimates and choice of 3-D model, can be applied to other 3-D datasets, especially those obtained by measuring thermal desorption aerosol mass spectra or chromatographically resolved aerosol composition."

R1.12 Fig. 1: I find this figure a little bewildering. Given that b) and c) were not ultimately used and are not analogous to existing 2D factorisations of AMS data, I would strongly urge the authors remove these parts.

A1.12 The models described in Figs. 1a and 1b were used in the paper, while those described in Figs. 1c and 1d were provided to illustrate possible models that might be more appropriate for some AMS and other 3-D datasets. Although this paper focuses on an application of two of these 3-D factorization models to an AMS data, Sect. 2 (Three-dimensional array factorization) is meant to provide a general overview of 3-D array factorization. We agree with the comment from Reviewer 2 (R2.1) that it is important to keep the whole figure in the paper.

R1.13 Fig. 2: The concept described in this figure has been in existence since the advent of the TOF-AMS and as such, I don't consider this figure necessary.

A1.13 We agree that the concept of PToF data acquisition in a ToF-AMS has been described in previous publications. However, we feel that the visualization of the detection process and data acquisition scheme provided in Fig. 2 helps orient the less-experienced reader to the data construction. We agree with the comment from Reviewer 2 (R2.1) that it is important to keep this figure in the paper.

R1.14 Figs. 7&8: Was the data smoothed before or after factorisation?

A1.14 The 3-D dataset was smoothed before factorization in both the time and size dimensions, as described in Sect. 3.3.2 Step 3 (pg. 4575, lines 23-29 and Table 2). The factorization results shown in Figs. 7 and 8 have been further smoothed as described in the figure caption for improved visual clarity. We have updated the figure captions to clarify this detail, changing the appropriate sentence to read,

"The factor size distributions have been binomially smoothed by one point each in time and size for improved visual clarity."

R1.15 S12, L298: The average fraction should be stated

A1.15 The average fraction of the mass omitted by removing *m/z*'s that have organic signal and interferences from air or inorganic ion signals is 7%, and was reported in line 301 of Supplement. We have revised the latter half of the paragraph to move this fraction to line 299, so that the last three sentences now read.

"This omission removes only a small part of the organic aerosol mass, about 7% of the total mass (Aiken et al., 2009). Of the omitted m/z's, the greatest contribution to the organic mass comes from *m/z*'s 39 and 40. Each of these *m/z*'s contributes ~2% of the total organic mass of the MS mode data."

Technical comments:
R1.16 P4565, L28: "only a few researchers" seems an odd choice of words given the three lines of references that follow the sentence. Consider revising.

A1.16 We have replaced "only a few researchers" with "comparatively few researchers", since the 10 references provided here represent, to the best of our knowledge, the complete record from the literature.

R1.17 P4568, L5: Technically, it is "mass spectral profile" rather than "chemical composition"

A1.17 The reviewer is correct that this is true for AMS data, but it is not necessarily true for other datasets. This section of the manuscript provides a general overview of 3-D array factorization, and so the terminology used here is generic. When we discuss application of the 3-D models to the AMS dataset in Sect. 3.4.1 (Models for factoring the 3-dimensional array, p. 4580) of the AMTD version, we state that "When this model is applied to the present [AMS] dataset, the vectors contain the factor's chemical composition (here a mass spectrum), size distribution, and mass concentration time series."

R1.18 P4571, L12: The standard calibration function used with the AMS (eqn. 9 in Allan et al., 2003) has a much more complex relationship than a simple inverse square root dependence.

A1.8 Reviewer 3 also mentioned this relationship (R3.3). We have changed the description of the relationship between PToF and d_{va} at pg. 4571 lines 10-13 to read,

"The measured PToF and the length of the particle flight region (L_p) give the particle velocity (v_p), which is inversely related to the particle vacuum aerodynamic diameter (d_{va}), by

$$v_p = \frac{L_c}{PToF} = \frac{v_s - v_l}{\sqrt{1+(d_{va}/D^*)^b}} + v_l \qquad (3)$$

where v_s and v_l are the asymptotic velocities for very small and very large particles, respectively, and D^* and b are calibration constants (Jayne et al., 2000; Allan et al., 2003).

Particle size (d_{va}) is calibrated to PToF using fast-vaporizing, monodisperse particles such as ammonium nitrate or polystyrene latex spheres (PSLs, at 800°C).”

R1.19 P4573, L23: Allan et al. (2003) did present a method to calculate precision of PToF data in terms of a detection limit, although this would not necessarily be best suited to this application.

A1.19 Allan et al. (JGR 2003) did present an estimate of the detection limit for PToF mode; however, the precision of the PToF measurement includes not only the noise present in the detection limit, but also additional sources of noise. Since Allan et al. (2003) is already cited in several places in our paper we have not modified the paper in response to this point.

R1.20 P4575, L19: The lower particle duty cycle will only affect precision when using the 3-vector model (see general points). The lower duty cycle is not necessarily the more important source of noise in PToF-resolved data; a potentially more significant cause is because the data is distributed through a larger number of channels. Compared to MS mode, less averaging takes place, so the electronic and background ion noise will be greater. The smoothing, in effect, partly reverses this.

A1.20 We agree that decreased precision because of low duty cycle, which results in measuring fewer particles in PToF mode than in MS mode, is compounded by measuring the fewer particle ions over several size bins in PToF mode. We have added the following text at pg. 4575, line 23:

“The SNR of the PToF-mode data is further decreased because the ions produced from particles that reach the vaporizer are sampled over many nominal particle sizes instead of being averaged together as in MS mode.”

R1.21 P4599, L22: Surely this is also as a result of the non-negativity constraint on the data? Also, was the data used in the tracer method still smoothed?

A1.21 We are unsure about what the reviewer is referring to. “This” could refer to similarity between the dynamic OOA size distributions calculated in the 3-vector and vector-matrix models (line 22), or to the noise in the tracer method being tempered in the 3-D factorizations (end of the same paragraph, line 28).

In the former instance, similarity between the size distributions should not be related to the non-negativity constraints of the 3-D factorizations. In the latter case, the non-negativity constraint does have some influence on the residual of the fit, but most of the impact occurs in the tails of the distributions, where there are more negative values. However, the factorization fits the entire distribution, and the regions with particle signal likely have a stronger influence than the non-negativity constraint on the tails. A more detailed exploration of this topic would require additional custom model runs and is outside the scope of this work.

The tracers presented in the AMTD version had not been calculated from smoothed data. We have now smoothed the data before calculating the tracer size distributions; the resulting diurnal averages show only minor differences from those in the AMTD version.

R1.22 P4602, L11: Recommend using something more descriptive than ‘spiky’.

A1.22 We believe that this term is clear enough, and it is commonly used in scientific discussions of time series of environmental variables.

R1.23 Figs. 7-9: If the size distributions have been normalised, the quantity should not be described as dM/dlogdva, as this risks confusion with the non-normalised data.

A1.23 We have changed the axis labels for normalized size distributions to “dM/dlog d_{va} (Normalized to unit area)”. For non-normalized size distributions, the axis labels have been changed to “dM/dlog d_{va} (µg/m^3)”. These changes have been made to figures in the main text and in the Supplement.

R1.24 Fig. 9: dM/dlogdva should have the units µg/mˆ3 in the non-normalised image plot. The normalised distributions should be dimensionless.

A1.24 See response to comment R1.23.

R1.25 Supplement figures: The axis labels should be checked in light of the two points above.

A1.25 See response to comment R1.23.