

Interactive comment on “Interpretation of organic components from positive matrix factorization of aerosol mass spectrometric data” by I. M. Ulbrich et al.

I. M. Ulbrich et al.

Received and published: 1 April 2009

Response to Comments (cont.) by P. Paatero

Our response has been divided into two parts to fit within the allowed page length for Author Comments prescribed by ACPD (15 pgs.). This part concludes our response and includes responses from items 23 to 33 of P. Paatero's comments. Excerpts from his comments appear below in italics. Our Responses follow each excerpt. Changes to the manuscript appear in bold.

23. - p.6740, ll 26-27: "Note that "solid body" geometric rotations of the factors are only a subset of the possible linear transformations." The remark is OK but the wording is unusual. The word "orthogonal" is typically used, e.g. "Note that orthogonal or "solid

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



body" rotations of the factors are ..."

[Response]: We have adopted this suggestion and changed the text as described.

24. - in Eq(4), what does the term 0.001 mean? Do you mean that in all x_{ij} values (in Hz), a fixed amount of 0.001 was added. Why? Or do you mean that the std-dev of the Gaussian variate was increased by 0.001? If yes, correct the equation accordingly!

[Response]: This equation was erroneous in the ACPD version. The correct equation is

$$x'_{ij} = (\text{Poisson}(\text{open}_{ij}) - \text{Poisson}(\text{closed}_{ij})) * CF + \text{Gaussian}(0, 0.0002)$$

where x_{ij} is an element of the synthetic data matrix, random noise is generated from a Poisson random number generator (Igor Pro v6.03) with a mean and variance of the number of ions observed in the open and closed MS of that point (open_{ij} and closed_{ij} , respectively), CF is the conversion factor from ions per m/z per averaging period to $\mu\text{g}/\text{m}^3$, and electronic noise is estimated from a normal Gaussian distribution with a standard deviation of $0.0002 \mu\text{g}/\text{m}^3$. The amount of $0.0002 \mu\text{g}/\text{m}^3$ is an estimate of the electronic noise present during periods of low signal in several m/z 's > 239 . The Poisson terms are explained further in our response to the next item (25).

25. - p.6744, ll 3-4: "any ions of importance have enough counts to reach a Gaussian distribution to good approximation" This is true. However, the value of Q also encompasses a huge number of ions that are -not- important. If the simulations for those are not properly done, the obtained Q may deviate from Q_{exp} simply because of the poor approximation used in the simulation. I am not sure if this risk is or is not important. Please watch out!

[Response]: We thank the reviewer for this useful pointer. The difference between the Poisson and Gaussian distributions is quantified by calculating the total area between the PDFs of each distribution ($\sum |\text{Poisson} - \text{Gaussian}|$) where e.g. the Gaussian value

S11945

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



at 5 is calculated as the sum of the continuous PDF between 4.5 and 5.5. For 5, 10, and 15 observed events, the Poisson and Gaussian distributions differ by 13.9%, 10.7%, and 8.7%, respectively. For our purposes, we will adopt the conservative assumption that the Gaussian approximation introduces only a small error when these differences are less than 15%. Therefore any measured matrix element x_{ij} with more than 5 measured ions is satisfactorily represented by a Gaussian distribution. Any x_{ij} with fewer than 5 measured ions will have increased error due to the Gaussian approximation. We then convert 5 ions in the 5-minute sample to $\mu\text{g}/\text{m}^3$ using the conversion factor described in Appendix A, and calculate this limit by

$$\left(\frac{5\text{ions}}{300\text{sec}}\right) \left(\frac{1}{0.03\%}\right) \left(\frac{1.1 \times 10^{-5} \mu\text{g}/\text{m}^3}{1\text{Hz}}\right) = 0.6 \text{ng}/\text{m}^3$$

where 0.03% represents the duty cycle for a single m/z in the Q-AMS (DeCarlo et al. 2006).

To our surprise, we found that the Q/Q_{exp} -contribution for the points below this threshold is 45% of the total Q/Q_{exp} in the 3-factor solution of the 3-factor base case. As indicated by the reviewer, this may introduce greater error in the proper estimation of noise for these small signals, so for this reason we have updated our calculation of the noise for these data using a Poisson distribution instead of a Gaussian distribution. The paragraph describing the addition of synthetic noise (including Eq. (4)) has been replaced with the following text:

"Synthetic Poisson noise was added to the difference spectrum synthetic data such that the noisy synthetic data, x'_{ij} , were calculated by

$$x'_{ij} = (\text{Poisson}(\text{open}_{ij}) - \text{Poisson}(\text{closed}_{ij})) * CF + \text{Gaussian}(0, 0.0002)$$

where x_{ij} is an element of the synthetic data matrix, random noise is generated from a Poisson random number generator (Igor Pro v6.03) with a mean and vari-

ance of the number of ions observed in the open and closed MS of that point ($open_{ij}$ and $closed_{ij}$, respectively), CF is the conversion factor from ions per m/z per averaging period to $\mu\text{g}/\text{m}^3$, and electronic noise is estimated from a normal Gaussian distribution with a standard deviation of $0.0002 \mu\text{g}/\text{m}^3$. The amount of $0.0002 \mu\text{g}/\text{m}^3$ is an estimate of the electronic noise present during periods of low signal in several m/z 's > 239 . Poisson noise is used for ion counting noise instead of Gaussian noise because many of the signals do not have sufficient counts to reach a Gaussian distribution to a good approximation. The sum of ion counting and electronic noises represents most of the noise in a Q-AMS dataset, but does not reflect "particle counting statistics noise" from events when a large particle present in low number density is vaporized and "extra" (much greater than average) signal is detected at only one m/z during the scanning of the quadrupole across the m/z range (Zhang et al., 2005a)."

26. - p.6752, l 6: *Was the SVD computed of the unscaled residual matrix? If not, what form of scaling was used? Note that the SVD of the unscaled residuals is quite inefficient because the std-dev of different values are so different. Also cf. p.6748, l 10.*

[Response]: SVD of residuals matrices described in the discussion paper was performed on unscaled residuals (i.e., on the matrix \mathbf{E} from Eq. (1)). We are unsure of the meaning of "inefficient" in this context. We have performed SVD on the scaled residuals matrix (i.e., \mathbf{E} / σ) which results even stronger conclusions from this analysis. In this case, 186 factors would be necessary to explain 95% of the variance of the scaled residuals matrix obtained after fitting 3 factors in PMF. We mention this briefly in the revised manuscript.

27. - p.6754, lines 17-18: *"As the simulated noise is white and thus has a large number of degrees of freedom..." This contains two errors: first, the noise is not white. Second, the form of distribution of the noise does not influence the number of degrees of freedom, hence the word "thus" is wrong. The correct formulation might be "The simulated noise has a large number of degrees of freedom..."*

[Response]: According to Shanmugan and Breipohl (1988), "For random sequences, white noise $N(k)$ is a stationary sequence that has a mean of zero and $E\{N(n)N(m)\} = 0$ for $n \neq m$, and σ_n^2 for $n=m$." Gaussian noise may be white, but is not necessarily white. The simulated noise may not be strictly white by this definition. These lines have therefore been changed as suggested by the referee.

28. - p. 6755, ll 17-18: *"The third factor lies 6 degrees out of the plane of the HOA and OOA factors." How was the "6 degrees" defined and computed? Was some form of weighting used in order to compensate for the different magnitudes of low and high m/z profile values?*

[Response]: The distance between the OOA-II MS and the plane created by the HOA and OOA-I MS was calculated by projection of the OOA-II MS vector onto this plane. We have added the following text as an explanation of this geometric consideration of linear combinations of vectors to a Supplementary Info section of the revised manuscript.

"The factor MS can be considered as vectors in a 270-dimensional space. Any two vectors in this space define a plane. Here we consider the plane defined by the HOA and OOA-1 MS vectors. Any 3rd vector in that 270-dimensional space will have an angle with respect to that plane. Any vector which is the linear combination of HOA and OOA-1 would lie exactly in that plane and have an angle of zero with respect to the plane. Any vector not in the HOA-OOA-1 plane is not a linear combination of HOA and OOA-1, and its "distance" from the HOA-OOA-1 plane can be quantified by the angle α between the plane and the vector.

In linear algebra terms, let A be a 270 x 2 matrix formed by combining the non-orthogonal, non-collinear column vectors HOA and OOA-1 such that

$$A = (HOA, OOA - 1)$$

and rank(A) = 2. A projection matrix P may be constructed by

S11948

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

$$P = A(A^T A)^{-1} A^T$$

which gives the projection (w) of a third vector (v , which is OOA-2 in the quoted example) onto the plane by

$$Pv = w$$

A right triangle between v and the HOA-OOA-1 plane can be identified in which w is the base (and lies in the HOA-OOA-1 plane), v is the hypotenuse, and a vector z could be drawn perpendicular to the plane to define the height (Olver and Shakiban, 2006). The cosine of the angle α between v and the plane can be defined by

$$\cos \alpha = \frac{\|w\|}{\|v\|}$$

and α can be calculated with the inverse cosine function.

These calculations are made without weighting the MS. If we weight each of the MS vectors by the average estimated error of each m/z over the campaign the angle for the OOA-II in the real dataset is reduced from 23 degrees to 13 degrees or 8 degrees (unweighted, weighted by error with no additional error treatments or errors treated with minimum error and 3-point box smoothing, and weighted by error with all error treatments (minimum error, 3-point box smoothing, and downweighting weak and m/z -44-related peaks, respectively). The "mixed" factor in the 3-factor solution of the 2-factor case lies 3 degrees, 1 degree, or 0.8 degrees out of the plane (unweighted, weighted by error with no additional error treatments or with minimum error, and weighted by error with all error treatments (minimum error and downweighting weak and m/z -44-related peaks, respectively).

The concept we are highlighting here is that any vector which is a perfect linear combination of two basis vectors lies within the plane of those two basis vectors (i.e., has an angle of 0 to the plane), so that this angle may then be used as one test of "splitting" or "mixing" of factors. Note that the range of \mathbf{P} is identical to the range of \mathbf{A} , and that for any vector \mathbf{v} that is a linear combination of the basis of \mathbf{A} , $\mathbf{P} \mathbf{v} = \mathbf{v}$. Two different types of linear combinations are possible. First, consider $\mathbf{v} = a \cdot \mathbf{HOA} + 0 \cdot \mathbf{OOA-1}$, a "split" HOA factor. Second, consider $\mathbf{v} = a \cdot \mathbf{HOA} + b \cdot \mathbf{OOA-1}$, a "mixed" factor comprised of elements of both the HOA and OOA-1 factors. In either case, the vector \mathbf{v} lies within the plane spanned by HOA and OOA-1 and the angle between the plane and the vector is 0.

29. - after Eq(A1): *The shape of distribution of single-ion signals is not relevant. Omit "Gaussian". Write e.g. "account for the random variation of the heights of single-ion signals,"*

[Response]: We have made the suggested change to the revised manuscript.

30. - Eq(A2): *Under the square root, the expression is a sum of products, such as in (matrix by vector). Please write the correct mathematical expression, with summation sign.*

[Response]: We have changed equation A2 and the accompanying description to read:

"The error for a particular species (e.g., organics), $\sigma_{species}$, is calculated by propagating the application of an $n \times n$ "fragmentation matrix" (Allan et al., 2004) to the total difference error σ_{diff} , such that

$$\sigma_{species} = \sqrt{\sum_{k=1}^n (\sigma_{diff,i,k}^2 \cdot \text{FragmentationMatrix}_{kj}^2)}$$

Other details

31. I do not understand figure A2. It says "errors". Does this mean error estimates σ_{ij} , computed for real and for synthetic values x_{ij} ? Or does it mean residuals? Use precise math notation, once again! What is the difference between "Time Variants" on x-axis and "Time Varying" on y-axis? For some points, the errors are much larger in horizontal direction, i.e. for the synthetic data set. Why is this so? I thought that the same expression was used in both cases for computing the sigma values.

[Response]: We have revised the axis labels in this figure to read, for the left axis, "Error estimates (σ_{ij}) from the real Pittsburgh data set ($\mu\text{g}/\text{m}^3$)" and for the bottom axis, "Error estimates (σ_{ij}) for the 2-factor synthetic data set calculated using campaign averaged conversion factors and closed spectrum ($\mu\text{g}/\text{m}^3$)".

The referee is correct that some of the data points have much larger values in the synthetic case than in the real case. The outlying red points in the figure are from the first sample (row of the matrix). This sample was made over less than the standard 5-minute averaging time. The error values calculated in the real case reflect the actual measurement time while the error values calculated in the synthetic case use the longer campaign-averaged sample time, decreasing the error for these points. The large band of green outlying points is mainly from m/z 41, which had much higher than average background levels in the instrument during two events. Again, this background is included in the calculation of errors the real case and its effect is diminished by using campaign-averaged background levels. We have explained this in the caption of the figure.

32. Figures 9 are very informative when they are plotted in large scale. Instead of discussion of correlations, please pay attention to the details in these figures. It would be good to go through the details of how these figures change because of rotations. Also, connect the changes with the additions and subtractions that occur because of different FPEAK. Such discussions would be useful for your readers.

[Response]: We already described the changes in this figure in the ACPD version

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

(P6752/L21) by stating, "Changes in TS occur more in some periods than others. Mass concentrations of all factors remain fairly constant at all FPEAKS during periods in which at least one factor has a mass concentration near zero, but periods in which all factors have non-zero mass concentration show more variation as FPEAK is changed." We will expand the discussion of Fig. 9 in Sect. 3.1.2. Specifically, the following text have been inserted at P6752/L21 of the ACPD version of the manuscript:

"Overall, the effect of positive FPEAK is to create more near-zero values in the MS and decrease the number of near-zero values in the TS. The effect of negative FPEAK is to create more near-zero values in the TS and decrease the number of near-zero values in the MS. Note for example that the TS of the FPEAK = -1.6 solution have periods of zeros that do not correlate with any interpretable events, likely indicating that this solution represents rotation beyond the range that gives useful insight for this dataset."

33. *Better graphics are needed*

This section is addressed both to the authors and the publishers! Many of the diagrams of this work contain a wealth of information. Unfortunately, this information is hidden by the graphical technique. It is true that by expanding the graphics on the computer screen, one may see the details. But many of us wish to have paper copies, e.g. in order to make annotations. In their original size, the diagrams are much too small for seeing the details (this is true both for the html version and for the printer-friendly version). On the other hand, I did not find any method at all for producing enlarged prints out of Adobe Acrobat Reader. The only method of producing enlarged prints seems to be via screen capture! In this way, it was possible to produce A4-sized (same as letter-sized) pictures of e.g. each half of figure 9. Only with this magnification, the important details in these figures became clearly visible. The authors should do their best in order to help the readers in this respect. Figures should be expanded horizontally so that they fill the full width of the page. In some cases, the figures might need to be split in two horizontal halves. Use enough markers on coordinate axes, etc.

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

The publishers should work hard in order to create a presentation that allows the output of expanded diagrams directly out of Adobe Acrobat or other presentation format. After all, computers should enhance the flexibility of our work. With paper copiers, it was easy to enlarge the figures we copied. Why should this be impossible when using computers?

[Response]: The small size in the ACPD version is an unfortunate effect of fitting the figures onto the landscaped, half-page format used by ACPD. We will be sure to ask ACP to print the figures in as large and clear format as possible in the final version. In future papers we will try to break up and size the figures to minimize this problem in the ACPD version.

REFERENCES

DeCarlo, P. F., Kimmel, J. R., Trimborn, A., Northway, M. J., Jayne, J. T., Aiken, A. C., Gonin, M., Fuhrer, K., Horvath, T., Docherty, K. S., Worsnop, D. R., and Jimenez, J. L.: Field-Deployable, High-Resolution, Time-of-Flight Aerosol Mass Spectrometer, *Analytical Chemistry*, 78, 8281-8289, 2006.

Dzepina, K., Arey, J., Marr, L. C., Worsnop, D. R., Salcedo, D., Zhang, Q., Onasch, T. B., Molina, L. T., Molina, M. J., and Jimenez, J. L.: Detection of Particle-Phase Polycyclic Aromatic Hydrocarbons in Mexico City Using an Aerosol Mass Spectrometer, *International Journal of Mass Spectrometry*, 263, 152-170, 2007.

Shanmugan, K. S. and Breipohl, A. M. *Random Signals: Detection, Estimation and Data Analysis* John Wiley & Sons, Inc., pp. 314-316, 1988.

Zhang, Q., Canagarata, M. R., Jayne, J. T., Worsnop, D. R., Jimenez, J. L.: Time- and size-resolved chemical composition of submicron particles in Pittsburgh: Implications for aerosol sources and processes, *Journal of Geophysical Research*, VOL. 110, D07S09, doi:10.1029/2004JD004649, 2005.

Interactive comment on *Atmos. Chem. Phys. Discuss.*, 8, 6729, 2008.

S11953

ACPD

8, S11944–S11953, 2009

Interactive
Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

