

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

P. Paatero (Referee)

Pentti.Paatero@helsinki.fi

Received and published: 16 May 2008

My first review submission has caused misunderstandings. (1) I wrote that a second review is needed. This means the following: when a number of changes are requested in a manuscript, then the editors often wish to know if the reviewer(s) wish to see the final manuscript before it is accepted for publication. In this case, I request a number of changes. My remark about the second review means that it would be good to see the enhanced manuscript before publication. (2) I wrote that it may not be possible to report all problems in the first review round. This does not mean that I would intentionally withhold something. The question is about the capabilities of the human

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



mind. When there are a large number of problems in the manuscript, then the attention of the reviewer may get "saturated" so that even an essential problem slips through unnoticed. This has happened to me recently – then I only noticed and reported one essential problem when reviewing the improved version of the manuscript.

How to evaluate a data analytic method

In my first review, I reported that the determination of the best F_{peak} value was based on circular reasoning. In their preliminary response item (d), the authors report that with their data, the assumed true factors are recovered well with $F_{\text{peak}}=0$ even if the true factors were obtained from a run with a significantly non-zero F_{peak} value. Such a finding does not, however, make my remark less relevant. It is essential to understand what it means that a data analytic method is to be evaluated.

When the results of a physical or chemical experiment are evaluated and published, then it is only claimed that the data were evaluated with a method that was adequate for this data set. When a data analytic method is evaluated, however, the situation is different. Then the applicability of the method to all possible data sets within a certain branch of science is examined. If it is known that the method does not work with some data, then the method cannot be recommended unless such criteria can be given that enable the prospective user to know when the method can and cannot be applied.

In the present case, there are two conflicting aspects. (1) The authors report that $F_{\text{peak}}=0$ recovers the true factors that were used for generating the simulated matrix X , irrespective of the F_{peak} that was used for generating the true factors. (2) Theory says that exactly the same X is obtained from different sets of true factors if the different sets are obtained from each other by an exact rotation. As the different incarnations of X are now identical, it is impossible that the analysis of these X with $F_{\text{peak}}=0$ would produce sets of factors having different rotational shapes so that these computed factors could match the different rotational shapes of different sets of assumed true factors.

The conflict is this: on one hand, the theory says that with some true data sets, the

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

method (i.e. $F_{\text{peak}}=0$) will fail to reproduce the true data. On the other hand, there is empirical evidence from one data set where the method appears to work. The resolution is that such a method cannot be recommended. In any new application, the scientists would not know if their data belong to the "good" group (presumably of little rotational ambiguity) where the method would work, or to the "bad" group, presumably of significant rotational ambiguity. Thus the prospective users would not be able to use the method. It follows that there is no basis for recommending $F_{\text{peak}}=0$ although it happened to work in the present case.

It should be noted that the question of selecting the best F_{peak} is moot if rotational ambiguity is negligible. The question becomes more and more important with increasing rotational ambiguity. For this reason, results obtained with (almost) unambiguous data sets, such as the present one might be, are irrelevant when studying how to determine the best F_{peak} .

Centered or uncentered R?

The authors claim that the centered "Pearson R" is not significantly different from the uncentered R so that the benefit of using a much-used correlation measure outweighs other considerations.

This question is similar to the previous " $F_{\text{peak}}=0$ " question: the evaluation of a method should be directed towards different data sets. In the present data set, all time series have strong variation. Hence, the two R's are equivalent and it does not matter which one is used. However, some atmospheric concentrations may have little variation, of course depending on the time scale of the measurement. E.g. such regional pollutants as SO₄ often have slow variation.

If there is even one significant component that has only minor variation during the study, then there is essential difference between the two R measures. Then the centered R may lead to quite wrong conclusions as the non-varying component becomes invisible in the correlations.

Scientists should not be afraid of using the best available tool (uncentered R) only because others have not used it before! Comparisons with prior work will be valid in all such cases where the use of the centered R was well justified in prior work. It would be good to use the uncentered R in the present analysis. In this way, future authors will be better equipped to use it when it is really needed.

It should be noted that the centered R was originally devised because many measurements used scales without a natural zero, such as temperature, IQ, pH, etc. For comparisons of such quantities, the centered R is still the right choice. For mass spectra, both R's are equally valid unless there is a possibility of a uniform distribution of all possible m/z values.

Using rotmat for determining the correct number of factors.

One of the good features of this work is that criticism is presented against the use of rotmat values as a guide for determining the "right" number of factors. I encourage that this criticism might be made even sharper:

There is no theoretical reason whatsoever that would support the use of rotmat. On the contrary, simulations should indicate that whenever the true data are selected such that there is rotational ambiguity, then the rotmat criterion will favor a wrong number of factors. The popularity of the rotmat tool may depend on the false belief that true data in principle should not possess rotational ambiguity.

A small number of published analyses may seem to support the rotmat criterion. Here may be a "publication bias": whenever the rotmat criterion clearly fails, then the manuscripts may not even mention that rotmat was tried.

Improved standard deviations for x_{ij}

In my earlier comments, I already remarked that the standard deviations $\text{std}(x_{ij})$ were not properly computed. The wrong std cause that interpretation of the obtained Q values is not correct. Besides this main effect, too large std have another effect in

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

the analysis: the scaled residuals are computed too small, so that the classification of poorly fitted values as outliers becomes distorted: fewer values x_{ij} are classified as outliers. This effect is not dramatic, it corresponds to a different choice of outlier threshold limit, viz. choosing 6.9 instead of 4 as the outlier threshold.

There may be another problem with $\text{std}(x_{ij})$. The true values l_o and l_c should in principle be used in Eq(A1) for computing $\text{std}(x_{ij})$. If measured values are used instead, a bias is caused: on the average, values with negative residuals have too much weight in the fit. Although the bias may be small in the present case, it would nevertheless be good to avoid the bias. With real data, there are two alternative possibilities:

(1) instead of the measured values, use the initial approximate fitted values in Eq(A1) in order to compute better values for $\text{std}(x_{ij})$. – This is not practical with Q-AMS because l_o and l_c are not separately fitted.

(2) Calculate preliminary " $\text{stdp}(x_{ij})$ " from Eq(A1), as is done in the manuscript. Then calculate the improved $\text{std}(x_{ij})$ values as the averages of the preceding and following stdp values:

$$\text{std}(x_{ij}) = \max(1/t_s, (\text{stdp}(x_{kj}) + \text{stdp}(x_{hj}))/2)$$

where $k=i-1$ and $h=i+1$ and x_{ij} is expressed in Hz. – In this way, the bias is eliminated. The lower limit $1/t_s$ makes sure that an uncertainty of one ion is computed even in those rare cases where all ion counts are exactly zero. – If the boxcar smoothing of width three is used, then $k=i-3$ and $h=i+3$ must be used for complete removal of the bias.

There may be a third problem in $\text{std}(x_{ij})$. For the highest peaks, the Poisson uncertainty from Eq(A1) may be well below one percent of the data value, i.e. $\text{std}(x_{ij}) < 0.01 * x_{ij}$. In many other applications of PMF, it has been observed that there are other errors (often called "modeling errors") that clearly exceed the one percent level. Variation of chemical profiles with time often causes modeling errors. Instability of the

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

instrument may well exceed 1%, especially considering how complicated and delicate the AMS is, and so on. Thus I would expect that some fraction of x_{ij} should be added to $\text{std}(x_{ij})$. (The amount to be added may be determined by observing the scaled residuals with different added fractions. Please create scatter plots of scaled residuals vs. data values. If such plots appear informative, they might be included in the supporting information.) Information about the short- and long-term stability of the AMS instrument is probably available. That info should also be considered when specifying $\text{std}(x_{ij})$.

More information about Q contributions

Q contributions from different m/z values offer one important diagnostic tool, connected both with the mathematical method and also with the physical and chemical reality. These contributions should be included in the paper itself, not in supporting information.

The distributions of scaled residuals with different m/z should be displayed in a more informative way. The present Figure S3 only illustrates the largest positive and largest negative residuals. Instead, it might be more useful to scale the figure e.g. from -10 to 10 and display the more interesting percentiles, e.g. 2,5,10,25,50,75,90,95,98. The simplest way to generate such a plot is to sort each column of the matrix of scaled residuals and then plot selected rows of the sorted matrix. This approach is probably better than the well-known box-and-whisker plots available in statistical program packages.

How to handle bad data values

Figure S3 suggests that there are some bad data values (unless the uncertainties were wrongly computed). There is no plausible "good" explanation for the scaled residuals that are below -10, say. You should explore all these suspect values and try to find explanations. If bad values x_{ij} can be identified, they should be eliminated from the fit (when using ME-2) or strongly downweighted (when using PMF2). – It is true that robust mode offers some protection against bad data. However, this protection is not

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

100% reliable, especially if the number of bad values is large. It is always better to eliminate or downweight bad data if it can be identified.

Correlations

How were missing m/z values handled when computing correlations? Were they replaced by zeros, say? Or were they truly omitted so that shortened or truncated vectors were used in correlation computations?

Math details

p.6740, l.25. The word "opposite" is ambiguous and unclear. What is opposite to what. As the whole sentence is now, it is impossible to understand what the words after "while" mean (unless one already knows the meaning, of course). Either, clarify the sentence or omit the words after "while".

p.6744, l.24. $F'=FS$ is wrong. It should be $F'=SF$.

Figures

Please order factors (e.g. OOA-I, OOA-II, HOA) in the same ordering in different figures (e.g. Figs 7 and 9). Also, it would be good to align different subplots in figures S1 and S5 so that the same peaks are located above each other in different subplots. It is much easier to study the spectra if the plots are aligned. – Unfortunately, some graphics programs may be hostile towards such alignment.

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

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

