Atmos. Chem. Phys. Discuss., 9, C114–C123, 2009
www.atmos-chem-phys-discuss.net/9/C114/2009/
© Author(s) 2009. This work is distributed under
the Creative Commons Attribute 3.0 License.

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Simultaneous factor analysis of organic particle and gas mass spectra: AMS and PTR-MS measurements at an urban site" *by* J. G. Slowik et al.

P. Paatero (Referee)

Pentti.Paatero@helsinki.fi

Received and published: 30 March 2009

— General discussion of the manuscript

This is a very important paper that suffers from one fundamental error in data analysis and from severe shortcomings in the presentation. The paper definitely deserves to be published but not in its present form. The whole analysis must be recomputed so that the inherently non-robust nature of PMF modeling is taken into account. The presentation must be re-written so that (1) all mathematical details are shown by means of equations and (2) all needed concepts and notations are defined. In the present version, the presentation is so unclear that after studying the paper a lot, I still do not

know what the authors actually did. More of these questions below, among the detailed comments. Sometimes, the referee cannot voice strong conclusions. In this case, the opposite is true: the manuscript -must- be published, it -must-not- be published without a major reworking.

This paper attempts to analyze results from two different instruments so that their results are merged into one large matrix. This is a very sensible and potentially useful approach. It has been tried repeatedly with PMF but several attempts have failed and therefore remained unpublished. No formal analyses of the reasons of failures have been published. The likely reason is the following: A measured matrix of AMS, PTR-MS, or other similar data may contain multiplicative errors or effects that affect uniformly entire rows of the matrix. All the values on any one row will become multiplied by an "erroneous coefficient" that is near unity but not quite equal to one. Such variation does not harm a customary PMF analysis. The composition factors remain unchanged and the erroneous variation is smoothly propagated to all time series factors. When matrices from two different instruments are combined, then the erroneous multiplicative variations in the two member matrices are generally different. This effect may distort the obtained composition factors. No resolution of this dilemma has been published so far. In practical work, the scientists should watch for signs of this problem in the residuals of the fit. If some rows of the PTR-MS part of the matrix mostly display positive residuals while other rows mostly contain negative ones, then this problem is indicated. In the present ms, the authors cannot possibly "solve" this problem but they should mention it and perhaps report about their findings regarding the residuals in their analysis.

Weighting the member matrices differently is one important contribution in this ms. As far as I can remember, such weighting has not been published in connection with PMF analyses. It is difficult to know if this part of the analysis has been carried through properly because the presentation of the weighting procedure is not adequate. There are many sentences of verbal explanation and only one equation (3). All new mathemati-

cal procedures must be presented, first of all, by giving all the equations. All different quantities must have their unique symbols. One source of confusion is that the symbol s_ij denotes both the original data value uncertainties and also the "weighted uncertainties". Any verbal explanations should be understood as aiding in understanding the equations, not vice versa. If no one among the authors is able to formulate the equations and define the needed symbols, then a mathematically/statistically oriented scientist should be invited to join the team.

— What was done

It remains unclear what the authors did. They state that "the data from the two instruments was combined into a single dataset, and PMF was applied to this unified dataset"

What does this mean? No mathematical equations or graphical illustrations are shown. Elsewhere in the ms, they write that the X matrix consists of time series of AMS and/or PTR mass spectra. At first, I thought that some columns of the combined X come from AMS while others come from PTR-MS. However, this cannot be true: in the supplement, they show how the G factor columns (time series for individual factors) are different for AMS and PTR. This would not be possible in the suggested setup where the same G matrix column applies both to AMS and to PTR-MS.

— Terminology

It is unfortunate that this ms continues using such wrong terminology that has already been identified as wrong. Specific fields of chemometrics simply cannot live and prosper if they create their own "mathematical terminology" that deviates from standard math terminology and even from terminology used elsewhere in chemometrics. I emphasize this aspect strongly because it is very frustrating for a reviewer to repeatedly correct the same mistakes. Please understand that you are the tail with a capacity to wag the dog: This ms will become one of those important papers that will be widely read and cited. If there are math errors in this paper, they will be diligently reproduced

by a large number of follow-up papers. Thus the requirement of mathematical and statistical correctness is even higher than in a "standard" paper.

The following is an almost verbatim copy from my review of an earlier AMS manuscript, published in ACPD:

"Using the term "residual" in order to denote Q is wrong. The word residual denotes the (signed) difference (measured-fitted) for any data value x_ij. The symbol Q denotes the sum of squares of scaled residuals, summed over all data values. The sum may be simply called "Q value". Sums of squares of scaled residuals over parts of data matrix may be called "Q contributions". If in doubt, you may include these definitions in your text. But do not call Q a "residual" because this word already has another essential use."

"Equal weighting": the authors call their method of weighting the member matrices "equal weighting". This term is most unfortunate because it can be understood in three or four different ways. A much better term would be "balanced weighting". Even this term does not uniquely suggest what quantity was balanced. However, it guides one to think in the right direction.

— Rotational ambiguity

The authors have correctly noted that there are two problems: (1) selecting a good weighting parameter C_PTR and (2) analyzing the rotational problem (= determining the range of possible rotations of the factors). They notice that variations of weighting have stronger influence on solutions than variations of rotations. Hence, they decide that it is enough to solve the first problem, and they settle for one particular rotation, viz. the rotation obtained with Fpeak=0.0, for both the member matrices as analyzed separately, and also for the combined matrix. Also, they seem to ignore the possibility of multiple solutions. If presence of multiple solutions was in fact excluded by running PMF from a sufficient number of random initial solutions, this fact should be documented!

This decision (of totally ignoring the rotational problem and the possibility of multiple solutions) might be OK if this ms would only be a presentation of the weighted approach to the two-instrument PMF analysis. However, the ms also contains several pages of discussion of the computed factors. This is a contradiction. If valid discussion of sources is attempted, then it is a MUST to either (1) demonstrate that the three solutions are rotationally unique (at least the PTR-MS solution is not!), or (2) to discuss the family of all rotationally possible solutions, both for member matrices analyzed separately and also for the combined matrix.

— Computation of errors of AMS data, interpretation of Q values, etc.

In another earlier AMS manuscript, the error estimates of AMS data values were erroneously computed so that a convolution with a box-car function of width three was ignored in the error computations. Apparently this convolution happens in the "standard" AMS software. The present ms does not explain if this problem was avoided or not. Considering that there is the risk that this box-car convolution still happens in the AMS software, the authors must definitely clarify the situation regarding this problem. If this question remains unanswered, interpretation of AMS Q values is impossible.

There are also problems in the interpretation of PTR-MS Q values. The ms says

"The theoretical value of Q, denoted Q_expected, is therefore equal to the number of elements in the input matrix X."

This definition is based on an approximation that is useful if the number of columns in X is much larger than the number of fitted factors. For the PTR-MS data, this approximation is not satisfactory. A better approximation for the expected Q value is obtained from the expression

Q_expected = (number of elements in X) - ((number of elements in G) + (number of elements in F))

Although this expression is not accurate either, it should nevertheless be used because

it is the best that can be known for the expected value of Q unless quite complicated analysis of scaled residuals is performed.

On third line after Eq(2), the ms says that in an ideal solution, absolute value of the residual e_ij is equal to s_ij. This statement is absolutely wrong. In an ideal solution, these absolute values range from zero to more than 2*s_ij.

The ms hides all the important facts in the long verbose text. For better examples of presenting your results, please examine a few earlier PMF papers. Facts should be collected in tables. This ms does not have even a single table. Why? If I wish to know how many measurements were used, I have to search for this trivial fact by reading through the manuscript. Collect various statistics regarding both the original data and also the computed results in a few tables! Efficient use of tables makes the paper shorter and easier to read.

The obtained Q values are an important diagnostic for assessing the validity of PMF analyses. In this ms, Q values are not properly reported. The insets in figures are not sufficient for assessing Q values. From figure 1, one can estimate that the obtained Q was too large by a factor of 3 ... 5. Why did this fact not alarm the authors?

— Lack of treatment of outliers

Lack of proper treatment of outliers is the major fault in this manuscript. There are samples for which the fit is very bad. Clearly the used PMF model is not valid at all for these samples. This applies both to AMS and also to PTR-MS. The authors have chosen not to use robust mode. A combination of strong outliers and non-robust analysis makes the results so questionable that the whole modeling work must be redone.

All least-squares-based data analysis methods (including the simple process of computing averages) are by definition "not robust". This means that outliers (= infrequent data values that cause large residuals in the least squares (LS) fit) are able to com-

pletely distort the results of LS methods. see e.g.

Pentti Paatero, Least squares formulation of robust non-negative factor analysis, Chemometrics and Intelligent Laboratory Systems 37 (1997) 23-35.

The PMF model (as all factor analytic models) is intended for describing the recurrent or regular features in multivariate data. It is not suitable for modeling one-off events or events that have different compositions at different times. Such data are "outliers". This does not mean that they must be erroneous data, they simply cannot be fitted with PMF because they deviate from the assumptions underlying PMF. Such unfitted data values have a tendency to severely distort PMF results.

When large residuals have been noticed, it is necessary to perform a preprocessing of the data set, especially in cases where a robustized LS method cannot be used. The simplest form of preprocessing is to omit from the data set the samples with extra large residuals. This should be done already when AMS or PTR-MS is fitted separately. Omitting the samples should be done in stages: first omit only samples having the largest residuals. Perform a new fit and again delete the samples with largest remaining residuals, until a satisfactory fit is obtained. When a sample is omitted, omit from both data sets so that the same set of acceptable samples remains in both member matrices. – It is important that this preprocessing is documented in the publication. Useful statistics should be reported about how many values were rejected, did they display understandable patterns, and about possible sources causing these extra large (or extra small?) concentrations. N.B. an outlier may also be caused by equipment failure, e.g. so that some concentrations are erroneously reported as zero.

It may be useful to also study the rejected residuals by suitable statistical methods: the unusual values may well contain useful scientific information. A PMF fit separates the usual from the unusual, yet both may be scientifically valuable. I recommend that the statistical study of residuals be published in a companion or follow-up paper, not in the present ms.

The authors mention that the outlier behavior is already seen in the 1-minute data, before averaging has been performed. Thus rejection of outliers can be started already within the 1-minute data values: if one 1-min period differs dramatically from the remaining 14, reject that period and compute the 15 min average by using the remaining 14 periods. Alternatively, perform the initial "averaging" by using "trimmed means" instead of "arithmetic means". (See any statistical textbook for the definition of "trimmed mean".) The trimmed means approach is the quickest because there are ready-made programs for doing it. No visual inspection of all data values is needed, the trimmed means should be computed systematically for all 15-minute periods. Also, use of trimmed means preserves information better than omitting entire 15-min periods or omitting individual hand-picked 1-minute intervals. However, outlier scrutiny must also be performed based on PMF residuals, even if trimmed averaging was performed, because some of the outliers may be of a longer duration so that they cannot be reliably detected when averaging for one 15-minute aggregate sample.

Instead of rejecting outliers, there is an alternative method: the outliers may be strongly downweighted "manually", by a factor of 10 or more (i.e. the s_ij values of outlying samples are multiplied by a downweighting factor of 10 or more). This approach has the merit that the residuals of the outlier values are obtained as a by-product of the analysis. This is useful if a follow-up statistical analysis of the outliers is planned in order to complement the factor analytic modeling of the regular or recurrent behavior.

There is a slight probability that the results of this original ms will not change when the outliers are eliminated. If this happens, the authors should not think that they have done wasted work. In any case, correct handling of outliers is a necessary quality control step. Without this step, there is no credibility in the results. – On the other hand, it is also possible that the number of meaningful factors may increase when outliers are eliminated. The authors should be open for this possibility – do not take the present results, such as the number of factors, as granted.

It should be noted that the presence of significant outliers may strongly diminish the

amount of rotational ambiguity. This might sound a desirable effect but it is not. The outliers may force the solution to one specific unique rotation, so that the scientist thinks that this is THE solution. However, this is a fallacy. The apparent uniqueness of rotation would be achieved because the model attempts to fit such data values (outliers) that are in inherent conflict with the model. Only after removing the outliers, the rotational ambiguity of the model-conforming data can be observed.

— Discussion of results

As there is a significant possibility that the numerical results may change, I will not discuss the present results and their interpretation at all. Such discussion might well turn out to be wasted effort.

It would be good if the authors could discuss the basic assumptions underlying the work: when do the aerosol and gas concentrations vary together so that they form one factor in this model, and when do they vary so that they cannot enter in the same factor. If one source type emits both aerosol and gas, then covariation might be expected. However, if a source emits gas that then transforms into aerosols in the atmosphere, then the situation is more complicated. – If you can discuss this, please do. If not, that is OK, too.

The results of the current modeling seem to indicate that when relative weighting of AMS vs. PTR-MS is changed by a "small" amount (such as a factor of two) the results change a lot. This behavior means that there is a conflict between AMS and PTR-MS. The weights act as a "partial" referee in this contest. If there would be no conflict, the adjustments of weighting would only have minor influence on the results. This aspect of the results should be discussed in the paper. However, at the current stage, the conflict can also come (I might even say, probably comes) from the outliers: the outliers in one member are not in harmony with the other member of the unified dataset. In order that the possible "real" conflicts can be identified, it is necessary to get rid of the conflict caused by the outliers. This is yet another reason why it is vital to process the outliers

in a preprocessing step, as already discussed.

Interactive comment on Atmos. Chem. Phys. Discuss., 9, 6739, 2009.