Atmos. Chem. Phys. Discuss., 6, S6069–S6073, 2007
www.atmos-chem-phys-discuss.net/6/S6069/2007/

Atmospheric
Chemistry
and Physics
Discussions

Interactive comment on "Source apportionment of submicron organic aerosols at an urban site by linear unmixing of aerosol mass spectra" *by* V. A. Lanz et al.

V. A. Lanz et al.

Received and published: 10 January 2007

The helpful comments by P. Paatero are gratefully acknowledged. Consideration of the given items 1-8 will certainly improve this manuscript. We fully agree that the checks mentioned by Pentti Paatero are necessary for a careful data analysis. All of the suggested checks have been carried out by the authors prior to submission. However, the results were considered to be fairly technical for a reader with a main interest in atmospheric sciences. For reasons of clarification and completeness all of the main concerns will be addressed in the revised manuscript.

1. Use and report Q-values The Q-value calculated for the presented PMF solution relative to its expected Q-value (which can be approximated by the number of matrix el-

ements) is clearly lower than a factor of 2. The Q-values from 2- to 6-factorial-solutions relative to its expected Q-value show a rapid decrease as a function of the number of assumed factors. The decrease of Q up to the presented solution is much larger than the sum of the dimensions of the matrix in question and its change -d(Q)/d(number of factors) - is moving towards 0 when the number of factors is increased clearly above 6 factors. In the revised manuscript, these diagnostics will be reported, discussed and included in Figure 2 (absolute Q-values and the dependence of the Q-values relative to the expected Q-values on increasing number of factors). Average Q contributions from each row and column respectively will also be described in the revised manuscript; it can not be concluded that there is a good reason to exclude neither a particular column nor row from the analysis (see also item 2). Only a small fraction of the scaled residuals ($<$1%) exceeded the default outlier limits. Nevertheless, using the robust mode has also been tested for this data set: outliers (in terms of scaled residuals) were downweighted using default settings and variants of it. Even though some PMF diagnostics slightly improved, it was not helpful for interpretation of the results. (In addition, different sets of rows and columns (see item 2) have been discarded entirely in a pre-analysis according to different criteria.) Even though the calculated Q-values do not exceed the expected level by a factor of 2, different error models have been tested (EM=-12, EM=-14, as well as different C-values) - also assuming 5% modelling error that is usually assumed for environmental data sets. In fact, the PMF diagnostics max(rotmat) as well as the final Q-value are lower than with the error model presented here. However, interpretability (e.g. in terms of spectral similarity) is losing significantly. It seem like James Allan's error calculation provide very good uncertainty estimates for PMF modelling of AMS data - at least for the present data set.

2. Check for high-noise columns The signal-to-noise ratios for all organic fragments (or columns) will be plotted together with the average Q contribution for each m/z in the revised manuscript. This will show that the signal-to-noise ratio decreases with increasing masses and the average sum of squares of scaled residuals per column is decreasing, too; there is no indication that any column should be excluded from the

analysis in advance. In contrast, it can be concluded that there is no disproportionate Q contribution from any high mass fragment used in the analysis. Nevertheless, four different thresholds have been imposed on the average signal-to-noise ratio, yielding subsets of 270 (ratio$>$0), 172 (ratio$>$1), 106 (ratio$>$2) and 68 (ratio$>$3) m/z's (noise here is defined as S uncertainty (see manuscript, Eq. 4)). Using those different m/z subsets accordingly in the PMF analyses yields two (ratio$>$3), four (ratio$>$2) and five (ratio$>$1) interpretable sources (as opposed to most PMF studies published so far, there is a scientific basis to interpretability in the present manuscript: its is based on spectral similarities of calculated to measured reference spectra). The same number of sources can be interpreted for both, including all m/z's as well as those that exceed the threshold of ratio = 1. The six-factorial solution based on 172 masses compared to 270 masses yields virtually the same (or even a bit lower) spectral similarities of the F profiles to reference spectra. Obviously, high-mass fragments contain important information that is indispensable for source identification in this data set. Thus, in this case it is not favourable to exclude masses based on the criterion of averaged signal-to-noise ratios. Even if those high masses may have lower signal-to-noise ratios on average, they may be nonetheless significantly above the detection limit temporarily and yield important information for source identification. The actual choice of 270 masses can additionally be supported by PMF diagnostics (such as the Q-values and max(rotmat)).

3. Terminology We will change the terminology in our manuscript as proposed by P. Paatero. Also the title will be changed to “Source apportionment of submicron organic aerosols by factor analytic modelling of aerosol mass spectra” as suggested. We however suggest to keep the terminology ‘Algorithm 2’ because this is the expression used by Zhang et al. (2005) and already used in literature. In the present manuscript it is not claimed that neither PMF2 nor PMF3 nor ME-2 is based on alternating regression (ALS). The sentence on p. 11684, line 8 probably has caused this confusion. This issue will be clarified in the revised version.

4. Columns scaling The manuscript will be adjusted according to the suggestions on

columns scaling. It is correct, columns scaling is not necessary here.

5. Interpretation of the sixth factor It is most important to note that we did not equal the sixth factor to food cooking. It has been labelled “minor source (that may be influenced by food cooking)”. Howsoever, we will restate this even more carefully throughout the revised manuscript. Misunderstanding the manuscript probably has lead to the interpretation that the “main purpose of the sixth factor is to model the average signal in high-mass columns”. The intensity of high-mass signals is not exceptionally high in the sixth factor: it should be noted that the range of the y-axis in each plot is in the order of the maximum intensity m/z (normalized intensities). As an example, this value equals about 0.16 for OOA, type I (Fig. 3). It should also be noted that the maximum intensity of a single m/z is in the range of 0.06 for the sixth factor. If this latter value were used to scale the y-axes of all other profiles, high m/z patterns could be detected in other factors as well. On the other hand, setting 0.06 as the maximum for all factors in Fig. 3 would distort real ratios of the most important m/z’s. This misunderstanding can be avoided by publishing more readable figures (see item 8). Visual inspection of the time series of the sixth factor is not conclusive to judge whether a single local source could have caused such signal. The y-axis of the sixth factor is still to be rescaled to the level of the highest peak. After that rescale, the time series of the sixth factor looks less different from the other primary aerosol contributions in time. This point will be clarified by sticking to the referee’s suggestions given in item 8.

6. Handling isolated peaks There are indeed spikes in this highly time-resolved data set. Therefore, several different criteria (e.g. concentration thresholds) were set to eliminate isolated measurement peaks and it was checked whether this would cause significant changes in source apportionments. No significant changes in source apportionments could be observed. ‘No significant changes’ here means that other choices introduce more uncertainties (e.g. choosing the right number of factor as reported in Tab. 2) to the modelled aerosol source contributions. Further, one aim of the study is to model a real-world ambient situation. For this purpose, it would not be helpful to

arbitrarily discard samples with highest concentrations from the data set, especially as some peaks are non-randomly and can be observed repeatedly: they are representative for an urban summer situation at that site (e.g. charbroiling) (and their highly aged fraction can be described by OOA, type I). In addition, the extracted factor profiles for this data set are very similar to measured reference spectra from many different independent campaigns/experiments, supporting their representativeness of typical aerosol sources. Mean and median for the overall aerosol contributions are nearly identical: even if there were peaks that possibly may have been caused by a sporadic, non-representative local source, it would not affect the main results. This study deals with a very large data matrix (about 15'000 rows). Between each tick mark in Fig. 6, one day or 720 values in time are represented. Thus, the time span of e.g. one day lies within a few mm in the printed version. Many calculated aerosol peaks appear isochronal at first sight but they are not, in fact. To reduce this kind of misunderstandings, we will enlarge Figures 3 and 6 as much as possible (see item 8). In any case, it is possible that one activity in time causes emissions of more than one mass spectral signature and also that emission activities of two different sources coincide.

7. Error estimates in figures 7 and 9 The symbols in the box-whisker plots will be explained. We will check the theory behind the notches and will probably omit them and use ‘normal’ box-whisker plots. Using the bootstrap method to estimate uncertainties of the medians is an interesting approach, however we feel that this information is not too important here. We therefore suggest not to put too much emphasis on it. The diurnal variations given in Figs. 7 and 9 intend to check the variation of the estimated source impacts with a priori knowledge of the temporal variation of the corresponding source activities.

8. Figures The manuscript will be changed according to the suggestions by the referee.

Interactive comment on Atmos. Chem. Phys. Discuss., 6, 11681, 2006.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper