Atmos. Meas. Tech. Discuss., 4, C1708–C1712, 2011
www.atmos-meas-tech-discuss.net/4/C1708/2011/

Atmospheric Measurement Techniques Discussions

Interactive comment on "Three-dimensional factorization of size-resolved organic aerosol mass spectra from Mexico City" *by* I. M. Ulbrich et al.

P. Paatero (Referee)

Pentti.Paatero@helsinki.fi

Received and published: 5 October 2011

This manuscript first describes different methods that are available for analyzing atmospheric three-way measurements, i.e. data that have been measured in a three-way grid of variables, such as chemical characterization, time, and particle size.

Second, the ms applies two methods to one large 3-way data set, size-segregated AMS measurements from Mexico City. The results obtained by applying the methods in different ways are carefuly analyzed and a recommendation is formulated about the best way to analyze this data set.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

This manuscript is of high quality. Different aspects of the methods are highlighted in the first part. Thus the first part can also be read as an advanced introduction to factor analytic methods applicable to environmental data. The second part illustrates the results by using a rich variety of different summaries, such as contributions to error function Q by different parts of the data set.

The manuscript is rather long. However, all of its contents are well motivated and help building an understanding of the methods. Thus I firmly recommend that nothing should be omitted from the text. Also, splitting the manuscript in two publications would be counter-productive. It would be more difficult for the reader to study the presentation if it is split in two.

This manuscript will most likely become one of those "classical" texts, to be referenced over and over again, also by scientists in fields other than AMS. For this reason, I do not agree with the recommendations by J Allan: I would like to keep Figs 1 & 2, and also the first sections of Supplementary info, as they are now. This will help other scientists in applying this work.

Recommendation: I recommend that the manuscript should be published in AMT essentially in its present form. However, there are some details that need to be enhanced or corrected, as listed below.

The ms does not discuss the nature of rotational ambiguity. It is clear that thorough discussion of this topic goes outside of this paper. However, rotational ambiguity is sometimes the main reason for obtaining inaccurate or "wrong" results. Avoiding rotational ambiguity is often the main incentive for using 3-vector models in science. Thus it would be good to offer a few sentences about the nature of rotational ambiguity in vector-matrix and 3-vector models, considering that most readers have no prior information about rotations in 3-D models. Something like the following sentences should be included: "In vector-matrix models, there is in principle always some amount of rotational ambiguity. The extent of rotations is limited only by non-negativity constraints of

AMTD

4, C1708–C1712, 2011

factor elements. Factors with few or no zero values tend to suffer more from rotational ambiguity. In 3-vector models, there is in the most general case no rotational ambiguity at all. Thus the solutions are well defined even without the help from non-negativity constraints. (There may be multiple solutions, however). In practical cases, one dimension (particle size profile, say) may be (almost) identical for two factors. Then rotational ambiguity appears between such two factors. Similarly as in vector-matrix models, such ambiguity is constrained only by non-negativity constraints."

The ms should refer to the following early papers describing 3-way analysis of aerosol data:

Y.-L. Xie, P. K. Hopke, P. Paatero, L. A. Barrie and S.-M. Li, Identification of Source Nature and Seasonal Variations of Arctic Aerosol by the Multilinear Engine, Atmospheric Environment 33 (1999) 2549-2562 and Philip K. Hopke, YuLong Xie, and Pentti Paatero, Mixed multiway analysis of airborne particle composition data. Journal of Chemometrics (1999) Vol 13, 343-352.

These two papers analyze a 3-way data set of aerosol concentrations. The dimensions are years, months, and chemical species. Three methods are applied: 2-way PMF (vector-matrix model), PARAFAC (3-vector model), and a "mixed model" where some factors are modeled using vector-matrix models and some using 3-vector models.

Section 3.4.3, "Guidelines for choosing a solution" is excellent. There is one detail that might perhaps be included here: sometimes it may happen that there are no objective criteria for choosing between two solutions that have different physical interpretations. Then one should preferably report both solutions in the report, instead of picking one alternative by using questionable criteria.

Section 3.4.4, uncertainties in the solution, p. 4585. Here seems to be a misunderstanding, possibly due to unclear wording in PMF handbook. The standard deviations reported by PMF2 represent uncertainties of individual factor matrices (G or F) while considering the other matrix fixed. In contrast, PMF3 derives the uncertainties of A, B, and C from the diagonal of the joint covariance matrix of all factor elements (i.e. of all elements of A, B, and C). Thus the computed std-deviations of A, B, and C will also contain uncertainty due to rotational ambiguity, if such ambiguity happens to be present in the 3-vector model in question. – The remark that "the estimates appear to be too small in our experience" (line 28) is not surprising. Too small estimates may well be caused by the presence of systematic errors (e.g. variation of chemical or size profiles) in data. Even if such "modeling errors" might be small in comparison to random errors, uncertainties of results due to systematic errors might be larger than those due to random data errors. Intuitively this may be understood as follows: Least squares fit resembles averaging. Random errors tend to cancel each other in averaging, while systematic errors may pass through with much less cancellation. In PMF3, the covariance matrix is computed by assuming that data errors are statistically independent, i.e. not systematic. Hence, PMF3 error estimates underestimate the uncertainty caused by systematic errors. I agree with the authors: deeper examination of error estimation methods is outside of the scope of this paper. However, please change the erroneous sentence on lines 24, 25, 26 so that the reader is not mislead.

SNR: Signal to noise ratio may be defined in different ways. It would be good to show the equation that was used in this work.

Supporting info line 488: "... the fail ..." should be "... fail the ..."

The first reviewer J Allan asked about smoothing, as mentioned e.g. in Figs 7 & 8. This is a relevant question. In a linear process, it would not matter if linear smoothing is performed before or after the analysis. However, PMF is bilinear. Then order of computations does matter. It can be demonstrated (the derivation is difficult, see Paatero & Hopke, 2003) that linear smoothing before PMF may enable the recovery of weak factor(s) that would not be detectable without such smoothing. – Hopefully smoothing was applied -before- factor analytic computations in this work.

Problems in Figures ——————- Please expand the horizontal scale of Fig. 4 and

AMTD

4, C1708–C1712, 2011

Interactive Comment

similar figures so that the entire page width is used. This will be helpful in situations where a paper copy is used. (When examining the .pdf file on computer screen, it is possible to expand the figure so that all details become visible. Then the original width of the figure does not matter.)

It appears that the same set of curves is shown both for HOA and for LOA in Fig. 8c. Please check and correct if needed.

Fig 11. Is this for one chosen day (which day) or is it average over all days? Please specify in figure caption.

Fig S1. The unit on x axis is given as "ms". It seems that "us" (microsecond) is intended.

Figs S7 and S8. Delete last sentence of both figure captions, beginning with "The scale for LOA ...".

Interactive comment on Atmos. Meas. Tech. Discuss., 4, 4561, 2011.