

Interactive comment on “Methods for estimating uncertainty in factor analytic solutions” by P. Paatero et al.

Anonymous Referee #1

Received and published: 23 September 2013

This is a potentially useful paper that provides a description of a new and interesting approach to estimation of the uncertainties in part of factor analysis solutions. However, it seems to be a strange paper to put in a measurement method journal like AMT. That is an editor's decision, but it would seem more appropriate in a statistical or chemometric journal.

On page 5, there should be some discussion of the potential for unique solutions given a sufficient number of true zero (edge points) in the data set based on the Anderson (An Introduction to Multivariate Statistical Analysis, (2nd edition). Wiley: New York, 1984). People need to be reminded that the ability to resolve sources depends on the availability of edge points. This point can be reinforced as part of the discussion of self-modeling curve resolution in analytical chemistry.

C2669

In the discussion of rotational ambiguity, it would be useful to refer the interested reader to the Paatero and Hopke (2009) paper discussing rotations in more detail. Another question about rotation that should be at least raised is how much of the rotational ambiguity is removed when you force the use of some fixed profiles as has been done by Amato and collaborators.

Another key issue that is not adequately discussed is the variability of profiles in environmental data. There is a true, absolute absorbance spectrum for any given compound. The degree to which the measurements match that spectrum depend on things like slit width, dispersion in the monochromator, signal to noise in the detector, etc. Thus, it is not only the better precision with which AC measurements can be made, but the absolutely fixed shape of the profile. Such fixed profiles do not exist in the environmental receptor modeling problem.

All of the discussion is focused on the estimation of the errors in the F matrix, but the output of the model with policy implications is the G matrix because that points to those sources that contribute significant mass to the samples, particularly those samples that drive the violation of standards for which you are likely to be collecting and analyzing samples. Thus, there has to be some discussion as to why errors in G could not be estimated either in an analogous manner or if one could take the asymmetric intervals in the F matrix and estimate a range for each g value so we have some idea as to the likely accuracy of the contribution estimates. This is where the rubber meets the road with respect to the application of this technology to practical solutions of PM pollution issues. With EPA as an active participant in this paper, it is hard to understand why that perspective is not reflected anywhere in it.

From my experience with beta testing of V5.0, the use of DISP or DISP-BP is EXTREMELY time intensive. Thus, some discussion of the extent of computational resources needed to make the calculations should be provided. People have come to expect relatively instant results and here we are looking at many hours of computer time to produce a DISP solution for even a few input species. To really use it, you

C2670

want to set it running for a weekend on a computer with a UPS to be sure there are no interruptions.

The conclusions section has essentially no conclusions. It is a summary. It would be good to have some clear conclusions as to the value of the DISP and DISP-BP approaches relative to just the BP and some recommendations/guidelines as to when to apply what method. Right now it leaves the reader uninformed as to what this work means. Although it is not possible to provide guidance that covers all situations, there should be some ideas as to how to proceed to use these error estimation tools and how to interpret the results as to what of them are meaningful and what are not. Right now, there is not much clarity in how to apply them. It only says here they are.

Minor Issues Below equation (1), it says “capital bold-face letters denote entire matrices, g_k denotes columns of the factor contribution matrix G ,” g_k should be in bold in parallel with f_k

One hopes that the program will actually be released in 2013 since it has been a long time in beta testing.

Interactive comment on Atmos. Meas. Tech. Discuss., 6, 7593, 2013.