Atmos. Meas. Tech. Discuss., 6, C3664–C3672, 2013
www.atmos-meas-tech-discuss.net/6/C3664/2013/

Atmospheric Measurement Techniques
Discussions

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

P. Paatero et al.

norris.gary@epa.gov

Received and published: 13 December 2013

>General comments The manuscript deals with the challenging issue of the rotational uncertainty in multi- variate receptor models introducing a new method that may represent a major break- through in the estimation of model output uncertainty. The manuscript is quite exhaustive and the explanation of methodological aspects and results is appropriate. Nevertheless, restructuring the text and removing redundancy would result in a more synthetic and fluent document.

First a general response to Reviewers 1 and 2: Reviewers 1 and 2 suggest enhancements (e.g. error estimates for G factor elements and more simulations, e.g. varying the number of zeros in assumed true factors) that would strengthen the paper and help

users in applying the presented EE methods. In principle, we agree with the reviewers: the suggested enhancements would be useful. However, there are several reasons why these suggestions cannot be followed. (1) Purpose of simulations. The simulations are not intended, and cannot act as a proof of the usefulness of the methods because real data cause different kinds of unspecified modeling errors. Also, simulations are not a reliable basis for determining the most useful parameter values of the method, such as the maximum allowed change in Q (dQmax) in DISP and BS-DISP. Experience with modeling various types of real data is necessary. Simulations are presented in the hope that they illustrate the EE process and give examples of possible outcomes that may be obtained. (2) Length of paper. This paper is already fairly long, possibly causing that some readers will skip parts of it. (3) Time. Reviewers hope that the EE methods should be available by the end of year 2013. If more simulation studies should be carried out, that would need several months for the simulations. After that, much of the report of simulation results would need to be rewritten, and the whole review cycle, beginning with internal EPA review, would be repeated. The delay might be almost one year. (4)Resources. The entire project of EE development proved to be much more laborious than originally estimated. Much of writing this paper has been achieved based on unpaid (volunteer) work done by authors. More of such unpaid resources are not currently available, and further funding is unclear. If new resources should become available in 2014, it would be more useful (from the viewpoint of EE users) to devote those resources to enhancing the convergence rate of the algorithms applied in EE computations. In theory, an order-of-magnitude acceleration appears possible. If it could be achieved, it would be a real help in practical work.

>As the authors recognize, this study is only a starting point and more experience with synthetic and real-world datasets is necessary to understand the strengths and the limitations of the method. For instance, more replicates would be required to assess the repeatability of the method including also replicates with the same number of zeroes. In addition, comparison with other techniques for overall output uncertainty estimation such as intercomparison exercises or benchmarking tools could be beneficial to vali-

date the method.

Response: In principle, we agree with referee 2: this manuscript does not cover the strengths and the lim- itations of the method when applied to real data sets. Please see the general response, above, for discussion why this manuscript cannot cover all such desirable material. Note that analyses of real data will be published in a companion paper, to be submitted soon.

>The uncertainty estimation with DISP seems to be sensitive to the user's defined maximum allowed change in Q (dQmax) which is in turn linked to the modeling errors and the size of the dataset. In addition, it is unclear what portion of the total uncertainty is represented by the estimated intervals. >Moreover, the selection of active and passive species may impact the evaluation of the uncertainty. For example, the combined effect of many species each of which with small individual mass but high uncertainty could significantly contribute to the overall output uncertainty.

Response: We inserted a paragraph in section 6, discussing the expected behavior of DISP vs. changes of dQmax . Referee writes: "it is unclear what portion of the total uncertainty is represented by the estimated intervals". We are not sure what "total uncertainty" means here. Our viewpoint is about the probability for the computed uncertainty interval to contain the true value. In simulations this is easily observed, and in manycases, the probability was almost 100%. In these cases, the uncertainty intervals represent the total uncertainty for the variables in question. For real data, containing modeling errors, determination of the probability is a major research problem and outside of this work. Some fraction of the total uncertainty caused by modeling errors is then missed by the DISP method, and this fraction may only be determined by analyses of real data.

Referee writes: "selection of active and passive species may impact the evaluation of the uncertainty". We have added a paragraph in Section 6, containing a more complete explanation of the impact to uncertainty intevals of defining a species as passive or

active.

Response: If high-noise species are not downweighted, they may downgrade the quality of modeling, especially if the number of high-noise full-weight species is large, more than five, say. It is not known how uncertainty estimates would behave in suchsituation (presence of several full-weight high-noise species).

>More effort is needed to provide reference values for the Q drop in DISP aiming at defining standardized thresholds (e.g. based on the dataset size) for rejecting solutions. For users concerned by the resource-intensity of the method, a reasonable choice would be to run only BS-DISP. It combines re-arrangement of samples and factor profiles perturbation to obtain an estimation that represents the rotational and the random components of uncertainty. In addition, it is less impacted by data uncertainties inaccuracies than DISP and provides more satisfactory intervals for downweighted variables.

Response: Reference values of Q drop cannot be specified with current limited experience. Prior experience (Paatero and Hopke 2003, "Discarding or downweighting high-noise values...") has demonstrated that when we specified a tentative expression, clearly indicating that it is tentative, colleagues took this expression as must-use, even in cases where it lead to absurd results. It seems better to not specify anything when the alternative might lead users astray by suggesting a tentative rule that may soon break down in some cases. See also the general response at top of this text.

Section 3.8 was inserted, discussing computational workloads. It is seen that a complete DISP and a very limited BS-DISP (only one element active on each row of F) cause comparable workloads. Such stripped-down BS-DISP does not appear reliable enough. For this reason, it seems that DISP (with all major elements active) is needed as a first step, then perhaps complemented by a BS-DISP using a small number of active elements.

Specific comments >Page 5 line 11. "Classical type of error" is a rather vague expression. Also systematic bias is part of the classical interpretation of error.

Response: The sentence containing “Classical type of error” has been deleted. A mention of bias is included in this paragraph.

>Page 6 line 7. Also wrong decisions about outliers introduce errors that may lead to loss of true environmental information (e.g. temporary sources, advection episodes).

Response: Good point, we had not noticed it. A note about outliers has been inserted.

>Page 7 line 22. Please, specify that you are referring to model output uncertainty estimation.

Response: Subsection heading has been enhanced as suggested

>Page 7 line 23. Also intercomparison exercises and benchmarking tools provide information about models repeatability and bias.

Response: Intercomparison exercises are often useful but they may also be misleading. One of us has participated in an intercomparison where all participants except one agreed with very good precision. Later analysis demonstrated that the single outlying result was the right one. Intercomparisons evaluate both the method and the scientist who applies the method. We would rather not complicate our presentation by also discussing intercomparisons at this stage. Later on, when several groups are (hopefully) familiar with these methods, it will be time to arrange an intercomparison.

>Page 8 line 7. According to the guide for the expression of uncertainty in measurements, also Monte Carlo Methods are suitable for propagation of distributions in uncertainty estimation (Joint Committee for Guides in Metrology 101:2008).

Response: Mention of “Monte Carlo Methods” has been inserted.

>Page 8 line 8 and 19. The concept of noise insertion is not clear. Resampling is associated with observation rearrangement but no modification of observations is made. Perturbation/noise insertion seems a different kind of technique. Please, put some

words here.

Response: The explanation has been expanded with more detail. Hopefully it is better now.

>Page 15 lines 1-2. Please, explain better what is Qopt. Is it the Q of the base case PMF solution? What do you mean by “varying all factor elements”? Explaining this variable to the reader is fundamental for understanding dQ in DISP analysis.

Response: Thanks for pointing out this unclear detail. Qopt stands for the Q value of the PMF model that is about to be DISPed, thus first the Qopt is the Q of the base case, for pure DISP, and later, it is the Q of the resampled data set for BS-DISP. This explanation has been inserted in the text. – “varying all factor elements” meant that Q was obtained in a minimization over all elements of G and F, as in equation (5). In the updated text, the expression was changed to the correct one: “a minimum with respect to all elements of factor matrices G and F”

>Page 15 lines 16 and following. It seems that the length of the intervals is quite dependent on dQmax and that little is known about the portion of total uncertainty represented by the intervals. Are intervals obtained with different dQmax comparable?

Response: Discussion of different dQmax has been inserted in Section 6. In brief, we say that for rotation-free cases, intervals are proportional to square root of dQmax. For cases with ample rotation, intervals are almost independent of dQmax. In simulations, computed intervals tend to contain the known true values: this indicates that the computed intervals are valid confidence intervals so that “they represent total uncertainty”.

>Page 17 line 21. Why do you consider F (factor profile identity) rather than G (factor time trend) in DISP? Is it a way to complement BS that deals only with samples?

Response: No. The reason is that we determine confidence intervals of F factor elements. Both BS and DISP and BS-DISP do the same. We interpolate in order to determine the “critical” profile shape where dQ is equal to dQmax. In future, if rotational uncertainty of G elements is desired, we would also interpolate G.

>Page 20 line 6: Pearson is widely used in chemistry and engineering as well. It is also used in EPA-PMF 3.

Response: The sentence is updated to include chemistry and engineering.

>Page 20 lines 9-10. This is true but cosine similarity is sensitive to scale and location changes.

Response: True. We specifically need sensitivity to location changes. This can also be expressed by saying that for concentrations, zero has a meaning, in contrast to e.g. test scores where the scale is arbitrary. Location invariant methods would lose some information. Sensitivity to scale changes is not a problem because all G columns are normalized to unit average.

>Page22 lines 11-15. In the three cases the relative uncertainties for values close to zero may be too small.

Response: True. In this respect, the simulation does not mimic atmospheric real data as closely as might be desired. Nevertheless, the simulation is valid research and properly reported.

>Page 23 line 11. Why did you set the correlation threshold to 0.8? In EPA PMF 3 the default value for comparing bootstrap results is 0.6.

Response: Here we chose a more stringent threshold than the default in EPA PMF; this value in EPA PMF can easily be changed.

>Page 23 line 25. Is there any relationship between the choice of dQmax and the drop of Q?

Response: Probably not. The drop of Q observed in DISP indicates that the base case solution was not the global minimum of Q. The base case solution is not connected with dQmax in any way.

>Page 26 line 26. I suggest to mention “Table 3” here to help the reader following your explanation.

Response: Table 3 is now mentioned here.

>Page 27 line 5. Two replicates seems not enough and in any case an odd number is recommended.

Response: Please see the general response regarding timing and available resources.

>Page 28 line 3. I suggest testing synthetic replicates with the same number of zeroes. Please see the general response regarding timing and available resources.

Response: The suggestion will be considered again if/when new research is begun regarding error estimation of the PMF model.

>Page 29 line 21. It is also recommended to report the variations of Q as absolute and relative values (percentage) for a better interpretation and comparison between different cases.

Response: done

>Page 31 lines 9 -14. You should also consider the uncertainty of the source profile.

Response: The formulation “non-Diesel values” implies that a possible non-diesel source profile has some leeway or “uncertainty” in its composition.

>Page 31 lines 15-18. I don’t think factor uncertainty should be used for estimating source uncertainty.

Response: “diesel source profile” was changed to “diesel factor”. In addition, how can be defined what is “small enough”. In the example discussed in this paragraph, “small enough” would mean that the uncertainty of EC is so small that it does not allow that EC reaches down to non-Diesel values, hence confirming the identification as a Diesel source.

Technical corrections >Page 8 lines 1-5. there is a repetition of two sentences.

Response: This was not evident in the version posted on the AMTD website, no change done

>Page10 line 20. Use only first name initial

Response: done

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