acpd-2014-775 Submitted on 30 Sep 2014

A Science-Based Use of Ensembles of Opportunities for Assessment and Scenario study

E. Solazzo, S. Galmarini
Referee full report – S. Potempski

General Comments

The paper deals with the problem of application of multi-model ensembles for air quality problems basing on the results of Hemisphere Transport of Air Pollution (HTAP) project. The question raised in the paper concerns a quite important issue on the applicability and reliability of the results based on multi-model ensemble analysis and the conclusions drawn from such an analysis.

The authors have proposed to include screening methodology into ensemble practise, based on the techniques used for the reduction of multi-model ensemble. This seems to be reasonable if one has to deal with model results only. Deeper approach could be based on more detailed characteristics of the models, but this is strictly related to model validation as one should get to know weak and strong points of the models.

The choice of the ensemble presented by Fiore et al (2009) is interesting as it was originally done for the sensitivity analysis of emission reduction options. In this respect the authors have shown that the emission can change essentially for various ensemble sets of model's selected, which indicates that the sensitivity analysis prepared by using multi-model ensembles should be performed very carefully. In consequence this shows that there is still a problem of defining good practices in treating multi-model ensembles (which to a certain degree is due to the lack of robust theory of multi-model systems). Hence the paper can be treated as a vote towards further research in this direction.

The paper meets requirements for including it into ACP with some minor corrections included into specific comments.

Specific comments

Lines 41-42: The statement: "An inspected ensemble should always produce a result that is more accurate than the simple average of the multi model results" seems to me as a bit too strong. I can imagine the situation (for example when the models are independent and accurate) that each new model in the ensemble improves, at least slightly, the accuracy.

Lines 54-55: "Under this condition, biases of opposite signs cancel out …". In fact independent models can have various biases – all of them can be positive or negative or partially positive/negative. Hence the statement above is not necessarily true.

Lines 199-202: As the authors indicated the Talagrand diagram is created by sorting the ensemble results to define bins and counting the number of measurement within each bin. Then in order to have any reasonable statistics the number of measurement should be much greater than the number of ensemble members. Otherwise rank histogram is simply not a proper tool for the analysis and should not be used at all. I suggest to put clearly such a statement.

Lines 268-269: In principle measurement errors should be also taken into account in the procedure for reducing the ensemble, but in case where they are significantly smaller than the model ones, RMSE is sufficient measure.

Lines 295-297: I agree with the conclusion on the importance of the inspection of the available results prior to their use in further analysis. However, it would be very nice to make this conclusion more practical, for example, by proposing an algorithm for such screening process. In fact the authors described it (lines 139-144) but I suggest to include a diagram that could in clear way show all the steps that should be done in analyzing any ensemble results. Inspection would be a part of this procedure. One of the aspects is that prior to any analysis it is seldom when one knows from scratch which models should be selected for the ensemble. This means that it is better to start with more models, and then to reduce the ensemble basing on the comparison with measurements. This process, however also depends on what kind of analysis is supposed to be performed i.e. for which purpose the ensemble is created, and which measures or indicators should be applied. That's why I suggest to include a kind of diagram presenting all these elements. The diagram could serve as a starting point for defining good practices in using ensemble methodology in air quality problems.

Typographical errors:
Line 62: Potempsky -> Potempski
Line 187: to me -> to be