Atmos. Chem. Phys. Discuss., 14, C11038–C11041, 2015
www.atmos-chem-phys-discuss.net/14/C11038/2015/

Atmospheric Chemistry and Physics Discussions

Interactive comment on "A science-based use of ensembles of opportunities for assessment and scenario study: a re-analysis of HTAP-1 ensemble" *by* E. Solazzo and S. Galmarini

Anonymous Referee #3

Received and published: 9 January 2015

General Comment:

The paper aims at demonstrating the possibility of improving the performance of a multi-model ensemble, made of data provided by many (21) participants (using different Eulerian chemical air quality models) at an inter comparison exercise, by an inspection. The inspection allows reducing the number of members of the ensemble to reach a better accuracy and, eventually, a more accurate estimate of the effects of a different emission scenario. Even if inspecting a multi-model ensemble of this kind is essential for many reasons, the methodology proposed is not robust and is not scientific rigorous (see all the major comments). The conclusions of the papers cannot be proven by the

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

methodology proposed. For these reasons, my opinion is that the paper doesn't meet the standard to be published on a scientific journal.

Major comments:

Line 55: The conclusion about the BIAS of two independent models is false. The two models m1 and m2 can be statistically independent but may have biases that don't cancel out, i.e. the sum of the two model biases: mean(m1)+mean(m2)-2*mean(observations) may not be equal to zero. This sentence should be reformulated.

Line 56-59: The definition of spread has not been given. Even if it could be considered straight forward it should be specified (see [1]). I guess that the authors refer to the standard deviation about the ensemble mean. Furthermore, the statistical independence between the members doesn't guarantee that the spread is a reliable measure of the model uncertainty. There are many other factors that may influence this skill of the ensemble, such as the number of members and particularly the ability of the ensemble members in reproducing the PDF of the observations. The statistical consistency of an ensemble is verified if an observation being forecast by a dynamical ensemble is statistically indistinguishable from the ensemble members [2]. Line 148-151: A more formal approach should be followed. The definition of variability should be given.

Line 158-187: The whole section lacks of a rigorous formal approach to allow better understanding the procedure, even without reading the others cited papers. How many data are used to compute the covariance matrix? Just using 12 data of the monthly means? If that were the case, the statistical significance of each element of the covariance matrix would be very low. The bootstrap confidence intervals could help in assessing such significance. How the mentioned projection of the so called "observation anomalies" is used? Because my understanding form the text is that only the explained variance of the first Eigen-vectors is used to draw conclusions on how the ensemble is "wise?

ACPD

14, C11038–C11041, 2015

Line 188-214: The rank histogram, as the authors correctly mention, is meaningful if the number of pairs (forecast, observation) is much larger than the number of the ensemble members. In this paper, the former is 12 the latter is 21. As a consequence any conclusion drawn by the plots in Figure 2 is not statistical significant. If the authors added to each bin a consistency bar computed by using a binomial distribution (see for instance [3,4,5]) they would see that an empty "bin" is still consistent with a perfect flat histogram, because of the small number of data available. Furthermore the rank histograms should be plotted with N+1 bins on the x-axis where N is the number of members. If the members are 21, showing a scale of 25, as in figure 2, is misleading and can give a false "feeling" of over dispersion. (line 206 the bin must be 22 not 21!). A horizontal line indicating the prefect model should also be plotted. It's not correct to state that an ensemble is "ill" just because a few data are analyzed or available. I would invite the authors to repeat this analysis by using more data such as hourly mean or daily means.

Line 270-272: The statement is not very clear; my understanding is that a better precision (as defined by the authors) also implies a lower RMSE.

Table 2 and Figure 2: Which is the statistical significance of the values reported? The "best" models are selected computing the RMSE on 12 data? A bootstrap analysis would probably show several combinations of models exhibiting a RMSE with the same level of statistical significance.

Line 320-326: Considering what mentioned in the previous comment, how the authors can be sure that the "best" combinations of models will provide the best performances also with a new emission scenario? Especially considering that the numerical models haven't a linear response to a change of emission scenario. To prove that the best combinations remain the same in different conditions (meteorological or emissions), the data-set should be divided into two parts. One should be used to find the best combinations, the other to verify that the best combinations remain the same.

ACPD
14, C11038–C11041, 2015

Minor comments: line 187: to me -> to be line 190 and -> an line 318 must be :"described in section 2" line 320 "four monde" ???

[1] V. Fortin, M. Abaza, F. Anctil, and R. Turcotte, 2014: Why Should Ensemble Spread Match the RMSE of the Ensemble Mean?. J. Hydrometeor, 15, 1708–1713.

[2] Wilks DS. Statistical Methods in the Atmospheric Sciences. 2nd Ed. Academic Press, 627 pp. 2006.

[3] Brocker J, Smith LA. 2007. Increasing the reliability of reliability diagrams. Weather Forecast. 22: 651–661

[4] Pinson P, McSharry P, Madsen H. 2009. Reliability diagrams for nonparametric density forecasts of continuous variables: accounting for serial correlation. Q. J. R. Met. Soc. , 00: 1–16

[5] Alessandrini, S., L. Delle Monache, S. Sperati, J. N. Nissen. 2015. Short-term wind power forecasting with an analog ensemble. Renewable Energy 76, 768-781

Interactive comment on Atmos. Chem. Phys. Discuss., 14, 30523, 2014.