Interactive comment on "De praeceptis ferendis: good practice in multi-model ensembles" *by* I. Kioutsioukis and S. Galmarini

1.   Summary

This work provides a detailed (and somehow unnecessarily extensive) mathematical analysis of the main properties of an ensemble being focused on air quality issues utilizing AQMEII data sets. Constructing an optimal ensemble is the main target of this work and a variety of error decomposition methods are being used for this purpose.

Results based on (a) the ensemble mean of all ensemble members, (b) the ensemble mean of certain subsets and (c) the weighted ensemble mean of the total population of the ensemble, are presented and well documented. A variety of different cluster methods are being utilized for this purpose.

For the final assessment a set of different indices and skill scores are being utilized although some additional - quite helpful indices in building an optimal ensemble - are missing (as the Talagrand bin score for example).

2.   Assessment

This is a well-written and well-documented work and I trust it should be published although some major and quite a few more minor issues should be taken care beforehand.

3.   Points of Major Importance

(a)   The paper seems to be quite long. By skipping some "unnecessary" details the paper could be abbreviated and become easier to be comprehended by non-specialist readers as well.

(b)   It has to be documented why the specific data set being used for this study (namely AQMEII) has been an appropriate data set since its time span covers only a year. This becomes even more demanding since there is a clear tendency to generalize the results of this study beyond this "limited" data set.

(c)   Certain clarifications for the selection and utilization of the training (training set) and predictability modes are necessary for better understanding final results.

(d)   There should be a clear distinction between results that are true for any ensemble and what has found to be different for any special data-set ensemble as the one being used.

(e).   The effect of bias correction on the ensemble characteristics (over- or under-prediction) could be easily (and clearly) shown by utilizing a set of Talagrand bin diagrams. It has not been clear also where exactly (i.e., on which data sets or sub-sets) this bias correction has been applied.

4.   Points of Minor Importance

An extensive range of minor grammatical or spelling errors can be found in the document. These errors could be easily spotted and corrected by a native English speaker.