Referee #1: **Matthieu. Plu**

We thank Dr Matthieu Plu for the positive and helpful comments that have improved the manuscript. They have all been taken on board and addressed in the revised version of our manuscript.

General Comments:

The manuscript entitled "Improving the deterministic skill of air quality ensembles" reports about the properties and the scores of different ensembles applied on the model outputs of the two phases of AQMEII (2006 and 2010) over Europe. The presentation of the manuscript and of figures is good. To my view, the study has several merits and the results that are presented are original enough to be published. However, the manuscript should be improved along the following recommendations in order to go over the Discussion step.

Specific Comments:

1. There is a lack of focus of the manuscript on the main relevant original ideas that are demonstrated. Many interesting results are presented, and the manuscript needs to focussed on one or two main new scientific questions that the manuscript addresses. These lines should be followed from the abstract to the conclusion. To make my argument more understandable, I would like to point out the following:

- the results that are reported in the first paragraph of the abstract - from lines 6 to 10 - are not new as they were demonstrated in past research articles. These lines mislead the reader on the purpose of the article;

- many times (page2-line12, page4-line4, page7-line12), "two ensembles" appear in the text, but it is not clear whether it refers to two different ensemble methods (and actually, the article compares more than two methods) or to the two AQMEII phases.

A suggestion would be to present the article as a comparison of different ensemble methods (mmW, mmS, KZ, ...), applied on two different datasets (both phases of AQMEII). Actually, the manuscript does not present new methods, but it compares the performance of existing ensemble methods on different pollutants and on different periods. What I consider also to be original are the diagnostics (such as Figure 5) that have been developed and that are used to analyse the ensemble properties. The objectives written page 4 (lines 20-23) may also be the relevant lines to follow, which is not fully obvious in the present manuscript.

Response: Thank you for the valuable suggestion. The manuscript has been rewritten as an analysis of the performance of different ensemble techniques rather than a comparison of the results from the two phases of the AQMEII activity, focusing on the originality of the study that includes: (a) the comparison of several ensemble methods on pollutants of different skill using different datasets, (b) the introduction of an approach based on high-dimension spectral optimization, (c) the introduction of innovative charts for the interpretation of the error of the unconditional ensemble mean with respect to indicators reflecting the skill difference and error dependence of the models as well as the effective number of models.

2. In the same line of thoughts, the title of the manuscript is too general and it should be more specific. A general title such as the one that appears now could apply to many papers that have already been published.

Response: Done as suggested. The title has been changed to "Insights in the deterministic skill of air quality ensembles from the analysis of AQMEII data"

3. Comparing the ensemble performance between the two AQMEII phases does not bring much to the study and can be misleading, since:

- the observation dataset changes (no PM10 observations available from UK nor France in Phase II, page8-lines7-8),

- the period (meteorological regimes, types of pollution, etc) change,

- the individual models change in depth.

The differences in ensemble performances between phases I and II (page9 for instance) are subject to all these differences. The attribution of differences of ensemble performances between the two phases should be done cautiously, making only one variable change at each time for any interpretation. If it is not possible, I suggest then to remove the discussions about the differences between Phase I and Phase II of AQMEII.

Response: Done. The new presentation is after an analysis of the performance of different ensemble techniques rather than a comparison of the results from the two phases of the AQMEII activity.

4. There is a lack of description of the experimental setup of the two phases of AQMEII, that would help the reader to understand some of the conclusions that are drawn, such as the arguments at page9-line19, page16-line, among others. The manuscript as it is written now is not self-consistent. To improve this, I would suggest to add in section 3.2 the key facts of both AQMEII phases: general experimental setup (domain, periods, common input data and setup for all models) and the different models that partici- pate (name, chemical and aerosol schemes, resolution, meteorological model, etc). At least the key facts that are needed to understand the discussions should appear in the manuscript. For the rest, the manuscript should cite some AQMEII reference articles.

Response: Done as suggested. A new section 3 with the experimental setup has been added.

Minor comments:

- page 3, if the "Recent results" (line 7) refer to the citation (Eskes, 2002) (line 9), then the word "Recent" does not apply; if they refer to an actual recent other work, please cite it,

Response: Done as suggested. The word 'recent' has been removed.

- the manuscript would gain in clarity if the KZ methods (page6) were described more in depth; for instance the page12-lines(7-10) sentence is somehow enigmatic.

Response: Done as suggested. Two paragraphs have been introduced in section 2. The first describes the rationale behind spectral optimization (spectral decomposition equations are provided in the Appendix). The second presents the examined ensemble estimators with reference to their theoretical basis.

- Is the quarter (September-October-November) chosen for NO2 the most relevant one? Do not the December-January-February quarters show higher NO2 concentration levels?

Response: We choose a continuous seasonal time series for each pollutant.

- The sentence page7-line22 would better fit in section 3.1.

Response: Done as suggested.

- page8-lines5-8: the sentence "the decline ... due to .. sampling stations." should be proven by some diagnoses or adequate citation.

Response: The sentence has been removed from the text.

- page10-line 2: the sentence "the benefits of ensemble ... members)." is not fully clear and maybe not true: what happens if we take the 6 "worse performing" models?

Response: We have rephrased the sentence to emphasize that we do not mean particular models.

- page10-line17: reference to the relevant figure is needed.

Response: The sentence has been removed from the text.

- page17-line28: remove '*'

Response: Done as suggested.