**Reply to review #1 of the paper:**

*Two-scale multi-model ensemble: Is a hybrid ensemble of opportunity telling us more?"*

by Stefano Galmarini et al.

Let us thank the reviewer for his understanding of the specific relevance of our work and his comments that indeed have improved the manuscript.

Herewith we will respond in a point-by-point fashion, direct responses are in red.

- The paper structure: The paper has been restructured following the reviewer suggestions, which makes it more coherent and readable. The figures have not been reduced in number but some have been included in the supplemental material. The figure organization has been changed slightly.

- Figure labeling and readability: This aspect has also been improved by making the figure more self explanatory.

- Quantitative conclusions summary in the abstract: More care has been given to the abstract an to adding more quantitative conclusions

- Additional points

- In the conclusions the authors refer to analysis of annual hourly, JJA hourly, and annual daily maximum records. However, until that point there seems to be no discussion of the latter two metrics. The authors should consider elaborating in the previous sections on the analysis they performed using these metrics.

It is not clear to us why the reviewer refers to data frequencies as metrics. The annual and JJA time periods are analysed to consider the all period of the simulation and the period in which ozone production is dominant. The daily maximum is a standard indicator of ozone presence. The text has been modified to make these distinctions clearer.

- It seems like the kzFO is introduced but barely discussed, and should probably be dropped from the main text.

kzFO is an ensemble analysis developed in the past that completes the spectrum of the available treatments. It appears to be neglected because it does not produce any substantial improvement. We'd rather keep it as part of the analysis for the sake of completeness. However we have made clear in the text that due to the above mentioned reasons it won't be much analysed.

- The lack of clarity in the figures is mirrored by the introduction of a large number of confusing acronyms in the text (mme_G, mmeS_GR, mmeW_R, kzFO. . .). These are often unnecessary, and the manuscript would greatly benefit from the use of more complete descriptions of the ensembles being discussed even if it means a small increase in length. I would suggest using the following names in place of the acronyms: o mme_GR: Hybrid ensemble o mme_G: Global ensemble o mme_R: Regional ensemble o mmeS_GR: Optimized hybrid ensemble o mmeS_G: Optimized global ensemble o mmeS_R: Optimized regional ensemble o mmeW_GR: Weighted hybrid ensemble o mmeW_G: Weighted global ensemble o mmeW_R: Weighted regional ensemble Furthermore the "kzFO" ensemble is almost never discussed, is not found to offer an improvement over the other ensemble options, and simply adds to the confusion. I would recommend moving all discussion of the kzFO ensemble to the SI.

We agree with the proposed naming strategy the text has been harmonised accordingly

- Although I understand that this is an ensemble of opportunity and that the authors have no control over what data is available to them, it seems odd to classify hemispheric CMAQ as a "global model".

The reviewer is right in the sense that H-CMAQ is a hemispheric model, however for practical purposes, we do not think this distinction makes any difference for the analysis – there is little cross-hemispheric transport on the time scales that were simulated here and the area of analysis is EU only. We added a statement in the database section that one of the models (H-CMAQ) is not truly global but for the purposes of this analysis is expected to behave the same as global models over the Northern Hemisphere and, therefore, for the rest of the paper we will refer to as "global models. A footnote to the table has been added where it is clear that we acknowledge the fact that H-CMAQ is not strictly global.

- The authors should justify their choice not to include urban monitor data, as better capturing the role of non-linear chemistry in urban environments is stated advantage of regional-scale modeling.

The large difference in model resolution between the global and regional scale models already produces as a result that many more monitoring stations fall in one grid cell of a global model that in a regional one. This produces as an effect that one model value is associated to many monitoring points where high local variability is also measured. Using rural stations, which are in principle less prone to local emissions, should reduce a physiological difference that would only penalise the global models. Based on this argument, it is clear that by adding the large number of urban/sub-urban monitoring station would exacerbate this problem. This point has been clarified in the text.

- The fact that different meteorological fields are being used for the different models should be explicitly mentioned by the authors as this could be a key factor in the differences between the various models.

This point is not hidden but mentioned in section and detailed in the table

- Two of the global models seem to be nearly identical (models 1 and 2), providing only different resolutions. Can these really be considered as giving "original and independent contributions" (1st paragraph)?

Model independence is a nasty beast. Are the other models more independent than these two just because they have different names but may share 75% of modules or methodologies? Is a different name or different resolution sufficient for us to define two models as independent? Obviously not. The one pointed out by the reviewer is a patent case of dependency of which we are aware, it should be pointed out that two of the ensemble methodologies adopted (S and W) are accounting for redundancy and the lack of independence is dealt with in the paper. However excluding 2 models simply because they are two versions of the same one would imply that we can assume that the others are independent when we all know they are not and as the regional scale model redundancy analysis demonstrates.

- The information in table 2 is inconsistent – sometimes degree symbols are used, sometimes they aren't; sometimes pressure units are hPa, sometimes mbar. The table would benefit from being cleaned up.

Units have been harmonised thanks.

- In table 2, the GEOS-Chem model top should be 0.01 hPa and not 0.066 hPa as listed (0.066 hPa is the second-to-last pressure edge: http://wiki.seas.harvard.edu/geoschem/index.php/GEOS-Chem_vertical_grids#47-layer_reduced_vertical_grid). This is simply the model I am most familiar with; I recommend that the authors also re-check the details of the other models in both tables.

We have contacted the model user. He is in agreement with what stated by the reviewer however since the values extracted are generated at the base of the cell, he preferred to give the pressure level of that element which constitutes in fact the 47$^{th}$ level rather than the cell top (48$^{th}$ level as stated in the GEOSS-CHEM specs). You were both right. Thanks