Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2018-86-RC2, 2018

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Two-scale multi-model ensemble: Is a hybrid ensemble of opportunity telling us more?" *by* Stefano Galmarini et al.

Anonymous Referee #1

Received and published: 13 March 2018

The researchers aim to evaluate whether a hybrid model ensemble, using a combination of both regional and global models, can better reproduce observed variations in surface ozone than an ensemble made up of only global or only regional models. In order to do so, they take advantage of existing, global-scale model output produced during the second Hemispheric Transport of Air Pollution modeling experiment (HTAP2), while using regional model output produced from the third phase of the Air Quality Model Evaluation International Initiative (AQMEII3). Model output is compared to a full year of hourly ozone data from multiple monitoring stations across Europe. Applying a variety of interesting and appropriate analysis techniques, the authors find that the use of a hybrid rather than single-scale ensemble can yield an improvement in performance on all three metrics.

Printer-friendly version

Discussion paper

The question being addressed is both interesting and important, and the methods used are appropriate. The magnitude of the advance is limited, in that this is a methodological advance which has specific relevance only to ensemble modelers, but in that field presents a significant finding. However, the paper is significantly hampered by its presentation. Although the content of the paper seems generally of high quality, the problem of presentation is sufficient that I recommend major revisions be made before the paper be considered for publication. I have given major issues below in paragraph form, followed by an itemized list of minor issues.

The first issue is the structure of the paper, in particular the use of figures. Although the manuscript clearly has a narrative, it is lost in the volume of analysis: 14 figures, usually with multiple sub-panels, and presented without any subsections to help structure the paper. I would urge the authors to move a significant fraction of both the figures and the analysis into the supplemental information, and retain only the most interesting and relevant findings in the main text. The paper would also benefit from prudent use of headings and sub-headings. Sections 3-4 should really be subsections 3.1 and 3.2, discussing the individual model performance without considering ensembles. Sections 5-7 then cover performance at the ensemble level as sections 4.1 through 4.3, with section 8 becoming the paper's conclusions. Even if the authors do not take up this suggestion, it would help the readability of the manuscript if they included a brief discussion at the end of what is currently section 4 to summarize the ways in which global models are generally providing better or worse results compared to regional models.

The second issue is the presentation of the figures. I would usually consider this to be only a minor issue, but in this case the figures are so minimally labeled or polished that it compromises the readability, and therefore the quality, of the manuscript of a whole. Figure 2 provides a good example. First, the upper panel undergoes no serious analysis, and the information presented in it is redundant as it is presented again with smoothing in the next two panels. Second, the data in the remaining two panels have

no legend, making it difficult to distinguish at first glance between the model results and those from the observations. Third, the vertical axis is labeled only as "|Y(f)|", which is both obscure and incomplete, lacking any units. Beyond just incomplete or redundant information, there are also several stylistic choices which make the figures difficult to interpret. The panels are labeled only as GLO and REG; it seems like it would have taken minimal effort to label these more naturally as "Global" and "Regional". I would also suggest using more natural tick choices on the x-axis (e.g. 1 hour, 12 hours, 1 day, 1 week, 30 days, 90 days, 1 year) and removing the clutter of the grid lines.

Each of the other figures has similarly strange choices. Figure 3 shows tick marks on the color bars which do not correspond to the color bands, the limits on the standard deviation change between sub panels, and the sub panels are completely unlabeled. Figure 4 has ample space for a full vertical label (i.e. "Probability of detection (POD)" and "False alarm rate (FAR)") but instead uses the acronyms POD and FAR, while still using the terse bar labels GLO and REG. Figure 4 also has two lower sub-panels with much smaller fonts than the upper panels, an X-label "Ozone" which lacks either units or temporal period, and a bar series with the name "##%". Similarly, figure 5 uses the bar labels ID, DU, SY, LT, and un; although four of these are explained in the caption, there is plenty of space in the figure itself to include the full labels ("inter-diurnal", "diurnal", "synoptic", "long-term", "residual"), which would make the abbreviations redundant as they are never used in the text. Figure 6 shows Talagrand diagrams, which are difficult to interpret for unfamiliar readers and deserve special care. Unfortunately, as presented the figure has no obvious meaning, with 3 unlabeled panels and vertical axes labeled only as "Frequency", while the X-axis label "Model id" is misleading. "Model id" implies parity with the "Model id" label in other figures, whereas it really refers to "(n-1)th to nth lowest model ozone concentration". Similar criticisms are applicable for the remaining figures.

The third issue is that the abstract lacks any quantitative conclusions regarding the paper. I recommend that the authors consider rewriting the abstract to include more of the

information provided in the paper's conclusions, with a particular focus on quantitative outcomes (e.g. the 1-5% improvement observed relative to single-scale ensembles when considering a specific metric).

The remaining issues are relatively minor, and are listed below:

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

- 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.

- 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 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 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.

- 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)?

- 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.

- 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/geos-chem/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.

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2018-86, 2018.