

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

Anonymous Referee #2

Received and published: 5 March 2018

Galmarini et al. show a statistical analysis of the performance of a multi-model ensemble of regional and global-scale chemistry-transport models in reproducing ozone surface concentrations over Europe. The simulations have been conducted as part of phase 3 of the AQMEII-HTAP initiative. A number of methods are introduced to construct optimal sub-ensembles for only global, only regional, and hybrid ensembles, and the performance of these is pitted against each other. In general, they show a small improvement of the optimally-selected hybrid ensemble over the pure global and pure regional ensembles. A cautious conclusion is drawn that the multi-scale nature of the hybrid ensemble actually proves to be beneficial.

General comments:

My general suggestion to the editor is to publish this manuscript after a number of
C1

points have been addressed, as the analysis seems to be sound, the methods are reasonably well explained, most of the discussion is reasonable, and the conclusions drawn are sufficiently cautious. I have given it "major revisions" in the hope to see it again before it is published, but in general the manuscript is very close to "minor revisions" for me.

It should be noted that surface ozone concentrations (and their observations) are a notoriously difficult measure to analyse and compare against models due to the strong effects of local emissions and deposition, so care should be taken to limit the scope of this manuscript to the performance of the ensemble in terms of surface ozone. This should be discussed in the introduction.

Before going into details I would finally urge the authors to reconsider putting 14 (!) figures into this manuscript and rather move several of them to a supplement in order to improve readability.

Specific comments:

P2, L70 define which "spectra" you refer to here.

P2, L74-75 is it a good thing that you use reg-glob equally at only 15 % of stations?

P5, L60-66 this is a slightly confusing amalgamation of arguments. Limitations in space are combined with different representations of gas-phase chemistry. I suggest rephrasing this paragraph and simply discussing gas-phase chemistry mechanism diversity.

P7, L221: regional models biased towards C-IFS?

P7, L229: "can take stock" seems inappropriately used in this context, revisit expression.

Figure 2: please use the same axis (Period in days is fine) for 2a as for 2b and c.

P8, L249ff: why did you average the spectrum? how did you average the spectrum? Are 24h peaks as pronounced in the model as in the measurements?

C2

P8, L249ff: Spectrum analysis also results in mangling daily maxima and daily minima. I am not sure if I like the fact that this means you are evaluating both photochemistry (daily maxima) and representation of boundary layer parameterization (nighttime titration)... Could be worth discussing in the text.

P8, L255: "The time series of the rural monitoring ...", these are simply your "observations", right? Not a subset or something? I suggest replacing "rural monitoring stations" with observations so as not to confuse readers.

P8, L268: Interesting! Do you dare to speculate as to why this is?

P8, L269-271: "A weak sec..." this sentence seems repetitive of what comes afterwards, remove or merge with remainder of paragraph.

P9, L298-300: "An element of surprise...". Surprising sentence - it reads like the introduction to the paragraph before. Why is it at the end? Is it a conclusion? Might want to rephrase.

P11, L337-338: if it is important, please tell us what you think about GLO vs REG now. Is the fact that GLO have a higher POD but also a higher FAR a good thing? Not really, no?

P12, L371: "... transport in the case of a global model, " are you not talking about regional models here? Consider cleaning up the paragraph.

P13, L401ff: I disagree with this assessment of the combined histogram (Talagrand diagram). The regional models (6b) actually came quite close to the ideal Talagrand diagram (equal distribution). Combining them and increasing bin number does not (necessarily) increase the value range. Now you still have an overdispersed model system, which just happens to be more correctly distributed amongst values closer to the mean (bins 5-18). I suggest to rephrase this paragraph.

Fig 9: very small panels, please improve this to make it readable.

C3

P17, L532: Figure 11 does not have panels a, b and c...

P17, L539ff: calling this improvement "systematic" is an exaggeration - there is no estimate of the uncertainty of these numbers, hence you have no idea whether the difference ("improvement") is statistically significant. This should be rephrased and written more cautious.

Fig 13, and also P18, L572-573: it would be very helpful if the optimal point would be marked in these plots. If I understand this plot correctly, for a 6 member ensemble the optimal point is ($x=16,667$, $y=1$; description in text says in line 573: ($x=100/(\text{Number of Models}), y=1$) with number of models = 6). All points are pretty far off of this optimal point in all the plots!

P18, L579: typo "form"

P18, L580ff: y scales 0-1, hence the "y values" cited here are for x, probably.

P18, L583: I guess you are talking about Figure 13d (mmeS6) as being superior here - rephrase to make this clear.

Fig 14: legends miss "obs" description (if that is what the thick red line is supposed to show)

P19, 587-600: this paragraph deserves its own (sub)caption, as it concerns a quite different and important point than the previous text

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

C4