

I have reviewed the revised version of Fiedler et al. Since I was not part of the first round of review, I have mostly restricted myself to determining whether the authors have addressed the comments raised by the other reviewers. However, I am a bit disappointed that those reviewers did not bring up the most poignant criticism of the “simple plumes” (SP) parameterization of aerosol–cloud interactions (ACI), which is that the sporadic transport of anthropogenic pollution into usually very clean regions is underrepresented.

My biggest concern with the manuscript is one raised by both of the original reviewers, which is that the manuscript is difficult to read. This is in part because vague phrasing abounds. One of the reviewers explicitly asked the authors to go through the text line by line to rephrase and reduce potential for confusion. That the authors have made only cursory improvements in the revision is disrespectful of the reviewers (and the manuscript’s future readers). It also made me feel antagonistic enough that I seriously considered recommending rejection before deciding on major revisions; after all, what assurance do I have that the authors will consider my recommendations any more than the original reviewers’?

1 Major organizational issues

That being said, I also disagree with multiple comments made by the original reviewer who recommended rejection. While the manuscript currently reads like a grab bag of resultlets, none of which is fully developed into an interesting conclusion (as the original reviews pointed out), a good-faith effort at reorganizing the content, as well as rewriting to remove platitudes and vagueness, would lead to a manuscript well worth reading. I have collected a few suggestions for results to emphasize and discuss in greater depth:

- On the subject of the averaging required for reliable forcing estimates, have the authors considered the “nudging” approach, where the large-scale dynamics of the model is constrained to reanalysis (e.g., Zhang et al., 2014, ACP)? Can nudging shorten the integration periods or reduce the averaging required to make reliable estimates of transient forcing time evolution? What are the drawbacks or trade-offs that need to be considered?
- The point that it may be more fruitful to calculate forcing differences between different levels of anthropogenic pollution rather than a forcing relative to a poorly characterized “preindustrial” state is well taken. I should note that I do not understand why this point appears in the section it appears in, “Benchmarking RF” (but I also do not understand what “benchmarking” means as used by the authors).
- I believe the original reviewer’s comment about the spatial shift in pollution from Europe/North America to Asia is wrong;¹ contrary to that reviewer’s opinion, the fact that the AOD is similar in both time periods does not imply that the ERF or RF should necessarily be similar.

¹Parenthetically, I also think the reviewer comment about ensemble means is incorrect; the sample mean is the optimal estimate of the population mean regardless of the caveats the reviewer lists. There is absolutely no guarantee that it will converge on the true forcing, but that is an inherent problem in modeling and unrelated to the number of models that participate.

However, for this finding to be illuminating, the authors should endeavor to explain why this is the case rather than simply stating the fact in Sec. 3.4. (As a side note, the numbers on p. 8 l. 26 are slightly different from the original manuscript; why is that?) The same holds for the discussion of the differences in efficacy. Why is the efficacy different between the time periods? Why do the efficacies increase in the less polluted regions? I would assume this is because the ACI sensitivity saturates. If so, what does that imply for the reliability of the SP method, where the model sees essentially the same average concentration of anthropogenic pollution, while in model configuration where the model is allowed to do its own transport, anthropogenic aerosol can sporadically intrude into clean, and therefore highly sensitive, regions? (I would say it indicates that the SP method will lead to a significant underestimate of the RF_{aci}, but my point here is that the authors need to discuss their findings.)

- Continuing on the previous point, in the introduction, the authors say that one of their research questions is how differences in surface albedo, insolation, and cloud regimes affect ERF over time. However, they do not return to this question in the manuscript. If they follow my suggestion in my previous point, it will have the cobenefit of making their introduction more reflective of the paper.

2 Clarity of writing

Regarding clarity of the writing, one of the imprecisions that was a constant irritant was the definition of F_{aci} : I think it might mean ERF_{aci} for EC-Earth and RF_{aci} for the other models, but I still haven't been able to figure it out for sure. As for F_{ari} , I'm pretty sure that is the ERF_{ari}, from the description on page 4. But if F_x refers to ERF for $x = ari$ and RF for $x = aci$, that is extremely confusing.

The original reviewers went to some trouble to identify other particularly unclear passages. I am somewhat deterred by the lack of response by the authors, so I will not expend effort on listing further instances. Just by way of example, in the first paragraph of the conclusions:

- What does “the” in “the five state of the art models” mean? It makes it sound like this is an exhaustive list, so no other models are state of the art. I know that this is not what the authors mean, but the sloppy writing is doing them a disservice.
- “reflecting both natural variability and model differences affecting ERF” is such an unclear way of restating the previous clause in that sentence that it took me forever to figure out that it was meant as a restatement. Writing something like “reflecting that natural variability and model differences both contribute to the model diversity in ERF” would have made it clear immediately. I know that it is hard to identify unclear passages in something that one has written oneself, but this paper has 12 authors, so there was no shortage of opportunity for someone to approach the text in the role of an uninitiated reader.
- What are the “best” model-mean estimates? (For that matter, what does “model-mean” mean in that sentence?)

I do implore the authors to follow the suggestion of having someone (who will receive better compensation than a reviewer) go through the manuscript line by line to improve the writing.

3 Minor comments

- p. 5, l. 32 onward: how can you tell this is not just coincidence?
- p. 6, l. 28: I do not understand this sentence; what does “more than one model ensemble” refer to? More than an ensemble of runs from one model? More than the multi-model ensemble in this study?
- p. 7, l. 1: “are” → “is”
- The word “herein” appears frequently, and I don’t think I understood what it was supposed to mean once
- I like the appendices, and I do not understand why the original reviewer complained about “too much detail” in the model description (the “too many notes, Mr. Mozart” line from *Amadeus* comes to mind). I believe the long form of “DMS” is two words (p. 10, l. 29). While “vertically integrated liquid water content” (p. 12, l. 25) is correct, why not call it by its better known name, liquid water path?
- In the author contributions, “lead” → “led”
- In Tab. 3, what does it mean when a clear-sky ERF is more negative than a cloudy-sky ERF? Positive forcing by ACI?