Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2018-174-RC1, 2018

Atmospheric Chemistry and Physics Discussions

Interactive comment on "The importance of comprehensive parameter sampling and multiple observations for robust constraint of aerosol radiative forcing" *by* Jill S. Johnson et al.

J. Mülmenstädt (Referee)

johannes.muelmenstaedt@uni-leipzig.de

Received and published: 18 April 2018

Dear Patrick,

I have reviewed "The importance of comprehensive parameter sampling and multiple observations for robust constraint of aerosol radiative forcing" by Jill Johnson et al. The manuscript describes the use of a statistical emulator to explore the model parameter space in the HadGEM3–UKCA aerosol–climate model. It concludes that the parameter space that is consistent with synthetic "observations" of the present-day state of the climate system is large; that this large parameter space leads to a large spread in aerosol effective radiative forcing (ERF) estimates; and that other methods of estimat-

Printer-friendly version

Discussion paper

ing the ERF uncertainty, such as multi-model ensembles and “emergent constraints” applied to such ensembles, give a deceptively small answer.

The manuscript has a clear message and is well written (I particularly enjoyed the succinct discussion of the different meanings of “constraint” in the introduction), which made it very pleasant to read. The issue of reliable uncertainty estimates is an important one, and I believe the exploration of model parameter space is a key contribution toward making progress here. I therefore strongly support the publication of this manuscript. I have one main criticism that I believe needs to be addressed (Section 1). I also have a fairly minor suggestion to the authors which they should feel free to adopt or disregard (Section 2), as well as a litany of minor comments (Section 3).

1 Are the synthetic observations used in this analysis representative of model tuning?

The authors argue that producing one tuned configuration of a climate model underestimates the size of the parameter space (no disagreement there); that the tuning process is often done one parameter at a time, underestimating the non-additive effects of varying multiple parameters at once (only mild disagreement there); and that an analysis based on an ensemble of ERF estimates from tuned models underestimates the true ERF uncertainty. For this last conclusion to be true, the range of ERFs that a single model can plausibly produce (“plausible” meaning consistent with the observed present-day climate) must be non-negligible compared to the intermodel ERF spread, as sketched in Figure 10.

In their analysis, the authors subject the different parameter combinations of their emulated aerosol–climate model to a consistency check against nine observables of the present-day climate. This is the emulator-world equivalent of tuning a single climate model to agree with the present-day climate; the further conclusions made in the

Printer-friendly version

Discussion paper

manuscript about the uncertainty on the model's ERF estimate thus hinge crucially on whether this consistency with observations is indeed equivalent to model tuning.

I am concerned that the answer may be "no", for the following reason. The observational dataset used here is the July mean conditions over Europe. Europe is a tiny portion of the globe, and July only samples one point in the seasonal cycle of aerosol and cloud properties. Since, as the authors point out, the model uncertainty stems from the interaction of aerosol and host model parameters, it seems like one would want to use the widest spectrum of weather and aerosol conditions available to be able to discard observationally excluded parameter combinations.

As a GCM tuning strategy, a European seasonal approach would fail because the constraint on the global-mean climate would be negligible. I expect that many of the ensemble members that are consistent with the European July observations will have outlandish global-mean TOA fluxes, cloud fields, etc., that would get them rejected in a constraint that uses global observations. This, I expect, would narrow the estimate of the single-model uncertainty compared to what the authors find using their constraint strategy.

The interesting question, in my mind, is how much using global constraints would narrow the uncertainty, i.e., whether Figure 10(b) would still look mostly as it does now or start to look more like Figure 10(a).

I realize that Europe was chosen for a reason, namely that global observations do not exist for all of the fields used as constraints (CCN; decadal trends of surface shortwave radiation and AOD). The most accurate equivalence with GCM tuning would probably be to use global fields where they are available and European fields (but for more than a single month) otherwise.

2 Some of the conclusions of the paper could be phrased as recommendations on the future direction of the aerosol forcing community

Apart from the main finding of large single-model uncertainty, there are other interesting findings in the paper that I think are underemphasized:

- The emulator–PPE method is a necessary step toward estimating the model uncertainty correctly. But to narrow the uncertainty range, it will also be necessary to assemble the most powerful (i.e., discriminating) combination of observables. Therefore, I encourage the authors to advocate for both of these necessary steps at once – the more they go hand-in-hand, the more efficient we are likely to be at making progress. (For example, the emulator–PPE method could point to which missing observable would provide the greatest additional constraint if it were added to the existing set of observables; see my comment on Figure 9 in the next section.) This sentiment is kind of there in the last sentence of the abstract and in the text (p. 17, l. 20; p. 23, l. 16), but it would be a missed opportunity not to phrase it more forcefully and more positively in the abstract.

- The aerosol forcing change over the coming decades is a much easier target than the forcing relative to preindustrial conditions, assuming the emissions changes are known. Perhaps it is time to move away from ERF as the holy grail of the field and instead focus on future aerosol emissions scenarios.

- AOD is a terrible variable if your aim is to understand aerosol–cloud interactions. We should figure out something better. Ed Gryspeerdt's work seems to show that aerosol index is a much better proxy for CCN. If the authors' model diagnoses AI, it would be easy for them to refute or corroborate this result.

These results have a direct bearing on where the aerosol forcing community would best invest its efforts in order to reduce the forcing uncertainty. I understand that the authors

want to keep the focus on the main result (the large single-model uncertainty), but they might think about ways to give these other findings a prominent place as well. For example, in our recent review article on radiative forcing by aerosol–cloud interactions (https://doi.org/10.1007/s40641-018-0089-y), we eschewed the traditional re-listing of conclusions at the end of the paper in favor of making recommendations to the forcing community. Mentioning these recommendations in the abstract may be appropriate as well.

3 Other minor comments

- Table 1: A bit more detail on parameters 8 and 9 would be nice. (Threshold in what variable? Rate of change with respect to what?)
- p. 10, l. 8: Do we know that the validation carries over to this new emulator? What is the rationale for not simply using the emulator that is definitely validated?
- p. 10, l. 14: Do the two approaches (elicited vs uniform PDF) give very different results?
- p. 11, l. 10: This does not make intuitive sense to me; if the "truth" is chosen to lie at one the edge of parameter space, shouldn't ensemble members from the opposite direction of that edge be penalized much more strongly than they would be if the "truth" were chosen to lie in the center?
- p. 14, l. 2: Much as I hate linear correlation coefficients, perhaps it would be useful to tabulate them to make it easier to follow this discussion. I am having trouble reading them off Figure 2.
- Speaking of Figure 2:

  - Nice to run into fellow R users.
  - All the labels are tiny! `par(cex)` may help.
  - Point clouds are hard to interpret. Perhaps the relationships between variables would be easier to interpret as color maps of the 2D densities of the emulator results?
  - Raster graphics are an abomination. R supports PDF output; this is the best option in general, and in particular for an information-rich figure like this one, where the reader will want to zoom in on interesting features.
  - These comments apply to Figure 3 as well, where additionally I see funny boxes around some of the panels at certain magnifications.

- Figure 4: In this model, ERF variability is dominated by ERFaci variability (see Figure 2). Yet the model does not care one bit about aerosol–precipitation interactions, at least not wet scavenging. I believe this is quite different from other GCMs. Any idea why? (Not necessarily something that needs to go into the paper.)
- p. 25, l. 15: In GCM tuning, this would routinely be done; see my main point of criticism.
- Figure 9: Does the converse also hold (that observations of decoupling would be a useful constraint on the model)?
- p. 27, l. 5: But Cherian et al. do this using models tuned to reproduce the global climate; see my main point of criticism.
- p. 28, l. 10: AOD multidecadal change appears to be double-counted.
- p. 29, l.1: Would AI or fine-mode fraction work better? I think the authors have the opportunity to make a significant statement here about whether there is a way

forward from AOD, which is known to be a poor CCN proxy, via other proxies. See my point in the recommendations section above.

- p. 31, l. 6: I don't understand the point about cancellation of correlated errors, but I would like to. Perhaps the authors could elaborate.
- Craig 1997 has a bunch of cryptic initials instead of editor names.
- Gryspeerdt 2017 a and b are the same publication.
- Stier ACPD 2015 has been superseded by Stier 2016 ACP (https://www.atmos-chem-phys.net/16/6595/2016/acp-16-6595-2016.html).
- Penner 2011: the DOI looks strange.
- Pujol 2008: check that this is still up to date with `citation("sensitivity")`.
- Zhang 2016: Toshi Takemura's name is misspelled.

I hope that these comments will be useful to the authors. I look forward to seeing their updated manuscript.

Best regards,

Johannes

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2018-174, 2018.

Printer-friendly version

Discussion paper