Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2019-1205-RC2, 2020

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Effective Radiative forcing from emissions of reactive gases and aerosols – a multimodel comparison" *by* Gillian D. Thornhill et al.

Anonymous Referee #2

Received and published: 11 May 2020

I've read the paper "Effective radiative forcing and adjustments in CMIP6 models" by Smith et al. The paper presents an important analysis of effective radiative forcing from CMIP6 RFMIP experiment. The analysis is sound, but I have a few suggestions for improvement.

The text could use some smoothing overall. In the initial sections in particular, the text is sometimes dis-jointed, with concepts and methods partially introduced, but then not really described until a later section.

Hopefully the results can be updated with model data made available since 12/19/19.

The paper could use a little more discussion of different ERF calculation methods.

This could usefully go in the conclusions section. How might the results presented here change if different methodologies and/or definitions were used. For example, the method and definition of effective forcing excludes "fast" responses related to sea-ice and SST changes. How potentially important are these for the different forcing components?

It would be helpful if the authors also discussed a bit more in the last section the potential limitations of this work and what future work might be warranted to improve forcing estimates.

It would be useful also to comment on how

Need to clarify if in numerical statements such as these: "ERF is -1.04 (±0.23) W m-2 from 12 models." the 1 standard deviation uncertainty provided is (I presume) just due to combining the individual model results, and that no additional uncertainty is included. (e.g., this certainly is an underestimate of the total uncertainty, although by how much it is not clear - as noted in next comment.)

My major quantitative comment is that no uncertainty bars are included for the individual model results. While I realize that inter-model differences are likely do dominate the overall uncertainty estimates, but this may not always be the case and some more analysis should be performed. In particular, it would be useful to better understand which model results are significantly different(statistically) from other models. I consider this a critical issue (thus "major revision"), however I don't believe that addressing this necessarily will entail a huge amount of additional work (assuming the authors current implicit assumption that uncertainty in the individual model results is small enough to be neglected is validated).

It is not sufficient to simply "Using 30 year timeslices generally results in standard absolute errors of less than 0.1 W m-2 (Forster et al., 2016)." How about in cases where the errors are larger than 0.1 W m-2? There are many models here, some are noisier than others, or have weaker signals.

There is some level of uncertainty in the data fits used in this method. This should should be easily quantifiable by the authors and should be reported (perhaps largely in SI). There is also potential biases and uncertainty introduced by using this specific methods - this should at least be discussed (and in some cases quantified, see specific comment on the kernel method below).

For example, if the combined methodological uncertainty of the individual model results was 0.1 W m-2 (as stated, but not shown), then, assuming normal distributions and independence, ± 0.23 becomes ± 0.25. So a fairly small change. However if the methodological uncertainty was larger, this would not necessarily be such a small change.

Some estimate of uncertainty is necessary to better justify the neglect of uncertainty in the individual model results. An example showing the uncertainty in fit for each individual model result, for at least a couple key examples (e.g., CO2 aerosol effects), would be useful to determine if there are some model results where the uncertainty is higher than average.

I would also suspect that for some effects, particularly related to cloud responses, the uncertainty might be larger than for other effects. If so, some of the differences between models might not be statistically signifiant, particularly where forcing is small. This would be important to know.

Also, numerical values for all main results, by model, e.g., in Figue 1 need to be provided in the SI. Results by forcing component and model should also be provided as data files (e.g. .csv file) as part of the supplemental material (GMD does make it obvious that this can be done, but many articles now provide such information). The results from this paper are likely to be widely used, and providing data files will help immensely to avoid transcription errors and save many person-hours in aggregate.

Only data sources (RFMIP and HadGEM3 kernels) are cited in the data availability. sector. While this does not seem to be a requirement of this journal (editors can clarify), it is preferable in this era that the codes used to perform the main calculations

were also made available along with the journal paper and archived in an appropriate archival repository. At a minimum, as mentioned above, the numerical results from the paper need to be made available in numerical form in either the paper supplement or an archival repository (note that an author's or departmental web site is no longer acceptable for this purpose.).

Specific comments:

pg 4, line 77 this text "The experiments and results presented in this study follow on from the assessment" is a bit confusing. This really isn't a follow on study, as such, but the two results are certainly related.

pg 5. The bulleted list here needs an introductory phrase to the effect of "In this work:"

Line 167 " With the exception of stratospheric temperature adjustments to greenhouse-gas forcing, structural differences introduced by using different kernels are well within 0.1 W $m{-}2$" Clarify - does this mean that the kernel method introduces this level of error, or that applying a kernel derived from one model to results from a different model introduces this magnitude of error? There are many models used here, I would think it is quite possible that some might have behavior that differs more widely from the kernel that is used? (Or is there some physical reason why this would not be the case?)

It would be useful to add a couple sentences to section 4.2.4 Kernel masking to walk the reader through what is happening in Equation 5.

line 257 "As shown in fig. 1 and discussed in section 5.1, ERF is approximately equal to RF for CO2, and we apply the Etminan formula to ERF". Reword since this is section 5.1.

line 324 "range of aerosol ERF estimates for 2014 versus" re-word - forcing is always relative to a specified reference point, so this doesn't quite make sense.

line 334 clarify if CNRM-ESM2-1 and UKESM1-0-LL were not in the PDRMIP models.

line 336 was MRI-ESM2-0 in PDRMIP?

line 345 edit to clarify, which experiment this is referring to. e.g., "The increase in cloud albedo leads to a strong negative SW radiative effect in the RFMIP all aerosol experiment"

line 369 something wrong with wording here "It may be the case that aerosol forcing over the whole historical transient aerosol forcing is "

end of Section 5.3.3 Please comment on how the lack of ice cloud interactions in most of the models impacts the multi-model mean forcing results. If, for example, the other models had a response to ice clouds similar to the "four models that include ice cloud interactions", would the forcing shift noticeably one way or the other?

Line 411 " ERFari, the APRP and double-call methods sometimes disagree". For those six models averaged together, is there a net difference between the two methods for ERFari (which might indicate a bias in one or the other methodologies), or are the differences of different sign between the models and cancel?

Line 447 - Is this referring to tropospheric and/or stratospheric ozone? Did all models include tropospheric ozone changes? If not, that would impact the implied ozone forcing values (e.g., actual ozone forcing would be larger). Clarify.

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-1205, 2020.