Responses to Reviewers

Manuscript: "*Uncertainty in Aerosol Radiative Forcing Impacts the Simulated Global Monsoon in the 20th Century*"

Please find our responses to the reviewers in this document. Our responses are highlighted in blue, and updated elements of the manuscript are highlighted in yellow and green in the Track Changes document included in the upload, and marked using the comment numbers in this document.

Responses to Reviewer One

This paper aims to contribute to the question of how persisting uncertainty in past emmissions of anthropogenic aerosols (AA) and their radiative forcing have impacted the climate system during the last ~100 years. To this end, they analyse the impact of AA magnitude on planetary-scale characteristics ranging from the global-mean surface temperature to monsoon area and intensity in the SMURPHS ensemble of the coupled HadGEM3-CG3.1. model. The topic is well within the scope of ACP, the paper is well written with good graphics, and it presents a (mostly - see my ITCZ-related comment) clear documentation of how changing the AA magnitude impacts climate. In my view, a weakness of the paper is that it very much remains at this documentary level, and gives little insight into the mechanisms that underlie the AA impact. It also limits its discussion to time-mean changes between 1950 and today, although it would seem that shorter-term changes would be more powerful to understand whether a low or high AA scaling (and thus AA radiative forcing) is more plausible. This is a long-lasting debate, and it would seem that SMURPHS could contribute here. While this somewhat limits the implications of the work, I am still in support of publication in ACP. Below are a number of smaller comments that the authors might want to consider.

> We thank the reviewer for their encouraging overall view of our submitted manuscript. While the point is well taken that AA mechanisms are not discussed in detail in this paper, on the basis that these mechanisms have been discussed extensively in previous literature we argue that their inclusion here would make the present work too long. Among previous literature (much of it cited in our introduction), there is strong evidence that the direct radiative effects and indirect cloud interaction effects of aerosol emissions have contributed at least in part to declining monsoon rainfall. At the scale of an individual region, there is evidence that both local and remotely emitted aerosols contribute to declining rainfall trends.
>
> Please see our responses to the specific questions below.
>
> Note: figure numbers in this response refer to the updated figures in the latest version of the manuscript.

A1. P2, L16: Maybe cite the updated radiative forcing estimate provided by Bellouin et al. (2020, https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019RG000660)?

We have added this reference in the Introduction section along with their estimates. It has also been added in the full reference list at the end.

A2. P2, L20: Is Fig. 1b referring to Fig. 1b of Samset et al.? I assume so, and I think this should be pointed out by writing "their Fig. 1b" (if not then in L19 China is referred to as East Asia in Fig. 1, and consistent wording would be preferrable ... but since the SSP are for the future and Fig. 1 is for the past century I assume that Fig. 1 is that of Samset et al.).

Yes, the Figure 1b referred to here is in Samset et al (2019). This has been clarified.

A3. around P2, L25: I agree that a single model can estimate the impact of AA changes in that particular model. But a model ensemble would be needed to know if the results from the single model are plausible. This needs to be acknowledged here.

In a CMIP-type ensemble, uncertainty due to aerosol forcing strength alone cannot be cleanly separated from uncertainty due to structural model differences (e.g. due to different parameterisations). The strength of the SMURPHS ensemble is that we are able to cleanly quantify one of these uncertainties. We agree that, in an ideal world, a multi-model SMURPHS-type ensemble would be a valuable addition to our datasets. The end of Section 4 already contains a recommendation that other modelling centres should perform similar SMURPHS-type simulations, and we have added more detail here highlighting that such simulations would be valuable to check that the results here are robust. We regard the findings of our manuscript as an important motivator for other modelling centres or the international community to perform a coordinated multi-model experiment. Without suitable "pathfinder" studies published in the literature, as we hope this work will become, it is difficult to launch such major multinational studies.

A4. P2, L24: Really for the first time? I think there were other studies that played with varying aerosol emissions. E.g., the work by Dan Westervelt and colleagues (https://doi.org/10.5194/acp-18-12461-2018). This is just a study that came to my mind immediately, there are more.

The appearance of the phrase "first time" in this case is ambiguous as to what it points to -- indeed, there are many other studies that have scaled aerosols or set them to zero, including Westervelt et al (2018). We have rewritten this sentence to clarify what it is that was done for the first time in the SMURPHS ensemble -- that is, performing a set of historical simulations with time-varying aerosols scaled to sample a large fraction of the uncertainty in historical radiative forcing from IPCC AR5, but in a single model. We have removed the words "first time". See also a response to the other reviewer (comment B5). The reference of Westervelt et al (2018) has also been added.

A5. P4, L20: I don't think standard deviation is a good measure of inter-ensemble spread for a 5-member ensemble. Why not give the range instead?

Standard deviation has been replaced with range for all of the vertical bars on all relevant figures. The corresponding captions have been updated and the text in Section 2 appended accordingly.

A6. The model version is a development version towards the final CMIP6 version of HadGEM3-GC3.1. Can the authors comment on how important they deem the model differences? I.e., can one combine the SMURPS ensemble with simulations with the final model version, or should the two be considered two models?

> The two model versions differ only in their treatment of prescribed ozone concentrations. The issue is described in the Supplementary Information in Dittus et al (2020): “There is a known issue in the model version used here that causes stratospheric ozone concentrations to occur in the upper troposphere as the tropopause rises with warming, causing a small amount of unphysical warming. This issue has been resolved in the UK’s contribution to CMIP6 (Andrews et al., 2020, Hardiman et al., 2019). We can compare the simulations from our ensemble with the updated versions to demonstrate that the effect of this issue is negligible over the historical period, at least for the standard scaling (Figure S1). We cannot rule out an effect for the scalings which produce larger warming levels but we expect it to be small.” We have added a sentence to emphasise this.

A7. Fig. A1: The authors state that the figure demonstrates the fidelity of the SMURPHYS simulations (P4, L6). I find that hard to see from Fig. A1. The figure shows that the simulations capture the magnitude of the variability in time, but I am wondering whether they also capture some of the signals that one believes are driven by AA changes, and whether one would expect the simulations to capture such signals or not. The authors should expand on this point.

> Figure A1 has now been promoted to the main body of the paper, and more detail on the verification has now been added in the final paragraph of Section 2. The figure has been neatened and the bottom four panels removed for conciseness (they showed nothing that is not already clear from the GMA and GMI panels). Using this figure, we demonstrate that the model is capable of capturing aerosol-driven changes by examining the period 1950--1980, in which aerosol forcing rises. We see that the rate of change of global mean temperature and hemispheric temperature contrast vary as expected with scaling factor, with higher scaling factors leading to a more rapid decline in both properties (Figures 2a, 2b).

A8. Fig. 2 and Tab. 1: It would be helpful if the figure and table would include the observational estimates. These are included in Fig. A1, so it should be easy to include them here as well.

> It would be easy, although we feel it would detract from the main point of this part of the study. The emphasis in Figure 3 (also Figures 4, 6 and 7) is the impact of the scaling factor on the variables, rather than a comparison model and observations, which has been attended to already by the inclusion of Figure 2 in the main body of the paper. There are offsets between model and observations, particularly in derived rainfall quantities such as GMI and GMA -- hence observation lines on Figures 3 and 4 may lie outside the range presented. Comparison between model and observations on Figures 3 and 4 would be best represented using some sort of normalised anomaly, although this is already done in Figure 2 and would mask important features such as the sign of HTC and LSTC in Figure 3.

A9. ITCZ definition (p5, L7): I do not understand how the ITCZ is defined, and Shonk et al. (2018) did not help me either because Shonk et al. (2018) indicate a zonal-mean that does not seem to be applied here for the 50% criterion. This should be made more transparent, or, and this would seem preferable to me, the authors should consider using one of the established ITCZ defitions that are based on the precipitation centroid between 20N/S (e.g., Adam et al., 2016, doi: 10.1175/JCLI-D-15-0512.1). I would also prefer to define the ITZC based on the zonal-mean precipitation directly instead of averaging the zonal values of the ITCZ, as I would expect this to be more robust and more closely related to the hemispheric perturbations in the atmospheric energy budget introduced by the AA. This would also explain why the authors diagnose the ITCZ to be at the equator or even in the SH in the zonal-mean, which is at odds with a large body of previous work that has put the ITCZ at several degress north of the equator.

We have repeated the ITCZ location calculation using the “centroid” method of Adam et al (2016), and this has now been incorporated into Figures 2 and 3 and Table 1, replacing the Shonk et al (2018) method. The location of the ITCZ is similar using both methods, which indicates extra robustness. For reference, the Shonk et al centroid approach was applied to the global zonal mean rainfall -- this has been clarified. The location of the ITCZ in HadGEM3-GC3 varies around the equator in both definitions, which is at odds with the findings of Adam et al, but the model tends to put its ITCZ rainfall too far south (Williams et al, 2018). The text describing the ITCZ location has been updated to reflect all of this.

A10. Definition of monsoon metrics: These include ocean areas, as can be seen in Fig. 4. I know there is some debate about whether a monsoon should be thought of to only exist over land or not. It would help if the authors could at least briefly acknowledge this.

We have added a statement about this in the part of the paper that introduces GMA, at the start of Section 4.

A11. P8, L3: typo –> although they lie

This has been amended.

Responses to Reviewer Two

The authors present a high-level analysis of the temperature and precipitation response in monsoon regions to a wide range of aerosol emissions, scaled from the historical CMIP6 dataset. The paper is well written and the figures are clear and of high quality. The topic is squarely within the scope of the journal. However, the paper is very short, and the discussion and analysis are quite shallow. If the authors can add a little more depth to their analysis, I can recommend this for publication.

We thank the reviewer for their encouraging view of our submission. We have opted to deepen some of the discussions in the main body of the manuscript, including incorporation of some of the supplementary/appendix material into the main body, and expanding detail in the results sections. Please also see our responses to specific comments below.

B1. The abstract is really brief. There needs to be some detail there. The name of the model isn’t even mentioned in the abstract. It appears the whole article is written more in "Letter" format.

We have added more detail to the Abstract, including a mention of the name of the SMURPHS project and the model used for the simulations. We have added a little more detail on some of the key results, and mentioned the monotonic, roughly linear relationship as requested in comment B8 below. To balance this, we have also added more detail to the Conclusions.

B2. Page 1 Line 24: Not all aerosols cause negative radiative forcing.

The statement of negative radiative forcing here was intended to apply to the net effect across all aerosols rather than individual components -- this has now been clarified.

B3. Page 2, Line 7-8: See also and consider citing Westervelt et al. (2018) Westervelt, D.M., A.J. Conley, A.M. Fiore, J.-F. Lamarque, D.T. Shindell, M. Previdi, N.R. Mascioli, G. Faluvegi, G. Correa, and L.W. Horowitz, 2018: Connecting regional aerosol emissions reductions to local and remote precipitation responses. Atmos. Chem. Phys., 18, 12461-12475.

The reference has been added, although slightly below the suggested location.

B4. Page 2, Line 11: Regarding AA emissions look likely to decrease. This is probably true, but nonetheless it is dependent on projections/IAMs and at the very least a citation is needed here (i.e. one of the RCP or SSP papers).

Lund et al (2019) shows that aerosol emissions from three of the SSPs reduce globally over the period 2015 to 2100. We have added this reference in this sentence.

B5. Page 2, Line 24: I don’t believe this is the first time someone has investigated climate response to a variety of forcing levels (or emissions). Perhaps you mean the first time in this particular model.

The use of “first time” here was intended to highlight the novel features of our study, but the wording did not make that clear. The novel aspect here is that we cover the historical time period since 1850 and systematically sample a large fraction of the IPCC AR5 range of aerosol forcing uncertainty (i.e., ‘plausible’ range of aerosol forcing). As the reviewer rightly points out, there have been many studies scaling emissions, but these have typically been idealised simulations focussed on a specific time period/region/aerosol species, so differ quite substantially in the experimental design, and have in some cases applied unrealistically large aerosol perturbations to better identify the forced response. To our knowledge, the only study to vary historical aerosol forcing through time in a similar manner is Jimenez-de-la-Cuesta and Mauritsen, 2019. However, they did not change aerosol forcing via emissions, so again it is a different experiment design. However, the wording ‘for the first time’ has been replaced with a new sentence to clarify the novel features. See response to comment 4 from the other reviewer.

B6. Section 2. Monsoon regions (especially in Africa) may be strongly impacted by natural aerosols (dust mostly). The reader needs to know what the model is doing for dust.

Mineral dust is simulated interactively in this model version using the CLASSIC aerosol module (Woodward, 2001). Changes in dust emission may arise in these simulations, associated with changes in near-surface winds and soil moisture induced by the differences in anthropogenic aerosol. This means there is the potential for a dust feedback in these simulations, due to an induced change in the dust radiative forcing. However, the dust does not mix with the anthropogenic aerosol. This has been clarified.

B7. Section 3 and Figure 2. Why not present GMST as an anomaly as in commonly done? This will make the results more comparable to the many other studies looking at temperature response to aerosols, since models may have different baseline temperatures.

Global mean surface temperature has been expressed in Figure 2 and Table 1 now as anomalies with respect to the 1900--1929 mean value in the 1.0 scaling experiment.

B8. Page 5 Line 20. "...climate responses vary monotonically and roughly linearly across the 0.2 - 1.5 scalings." I find this to be pretty interesting given the complexity and nonlinearity of the aerosol-climate system. This also may be one of the more novel findings and one that SMURPHS is uniquely positioned to answer. Perhaps this could be a sentence added to the abstract.

This result has now been included in the Abstract, and also added to the Conclusions.

B9. Page 8, line 8, final paragraph. Sorry but I don't see the point of just parachuting in a bunch of appendix figures/tables for the other regions. Referring specifically to Figures A2, A3, and Tables A1 and A2. Any figure in the paper should be discussed and contribute to the narrative, or else it shouldn't be included. It seems that there is a wealth of interesting analysis that could be written about these two tables and two paragraphs.

We have shifted Figures A2 and A3 into the text (as Figures 6 and 7), and merged the tables into a single new Table 2. We have added two more paragraphs highlighting the main results and conclusions that can be drawn from these tables and figures -- primarily that the Asian monsoon regions show a stronger sensitivity of GMA to aerosol forcing than most other regions. We demonstrate that this increased sensitivity of the Asian monsoons to the scaling echoes results from CMIP5 studies, in which a warmer climate led to increases in monsoon area over Asia, yet little change in area elsewhere.