Reviewer comment= black
Reply= blue

Referee comment on "Evaluating Uncertainty in Aerosol Forcing of Tropical Precipitation Shifts" by Amy H. Peace et al., Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2022-11-RC2, 2022

The authors use a single model ensemble to estimate the uncertainty in attributing tropical precipitation shifts to aerosols. The topic of aerosol-forced ITCZ shifts has been reported on repeatedly in the literature, but the authors here show no relationship between aerosol ERF and tropical precipitation shift, and instead argue for shifts associated with volcanic eruptions and modulated by internal variability. Following exactly what the authors did in their simulations in this paper is quite difficult. I therefore recommend major revisions.

My main comment is that the simulation description/setup is extremely difficult to follow. It seems like 13 ensemble members were eventually chosen from an initial 2800 (where does the 2800 come from?). I'm not sure at all how the 47 model parameters in the PPE map into the final 13 simulations chosen. There is a mention of a filtering process and then an assessment of diversity based on ERF from aerosols, ERF due to 4xCO2, and some other CMIP-type simulations. I feel like some kind of table or better a schematic is needed here to explicity describe what exactly the simulations are that the authors are running.

This and subsequent comments by both reviewers highlight that the experimental design could be clearer. Echoing our reply to Referee 1, we have worked to improve and clarify the design process of the PPE, so that those readers who want more details can get a sense of these without reference to the existing papers that document the design of the PPE.

Specifically, we have added more text in the methods to describe how the 13 ensemble members used in our analysis were chosen from an initial pool of 2800 model variants (parameter combinations). We have added more detailed descriptions for each of the experiments used in the filtering process, their purpose and how many model variants were retained in each stage. We have added a schematic (below) to visual this process and show which experiments form this filtering process we used in our analyses.

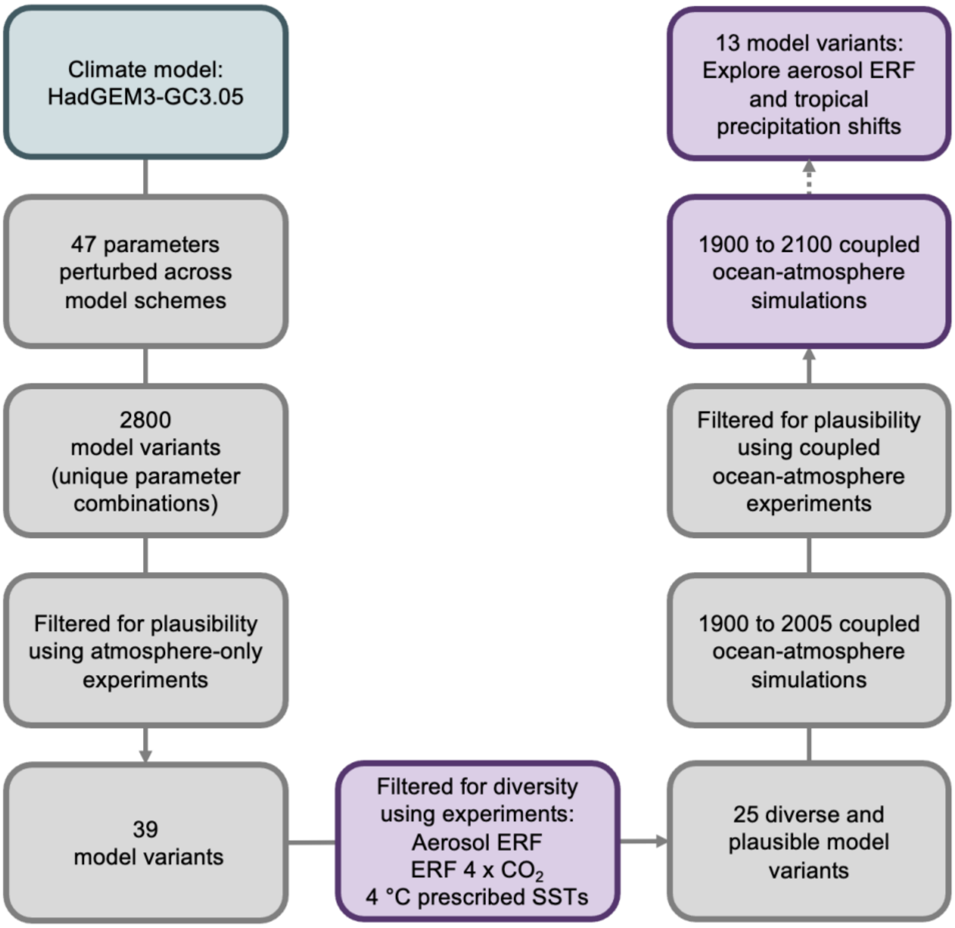

Figure 2: Schematic showing the stages in the design process used in UKCP18 to provide a small subset of model variants that sample a diverse climate response and are plausible when evaluated against historical climate. In this study, we use 13 PPE members of the aerosol ERF and transient coupled ocean-atmosphere experiments which are highlighted in the purple boxes.

My other main concern has to do with ensemble size. 13 member (and 4 initial condition members) do not constitute a large ensemble that can robustly estimate internal variability and uncertainty. I would like to see the authors better justify the ensemble size when work from other groups doing large ensembles (e.g. NCAR) to estimate uncertainty are using ~40 ensemble members.

We understand that the ensemble size of the PPE and initial condition ensemble is limited. However, our 13 ensemble members of the PPE were carefully selected to be both diverse and observationally plausible. Additional members with reduced plausibility would add uncertainty, but would not necessarily improve our analysis. Whereas the 4 initial condition ensemble members were those submitted to CMIP6. We have worked to address this comment in the two ways below.

Firstly, in the former version of the paper, we used the 4 members of the HadGEM3-GC3.1-LL (low resolution) initial condition ensemble to estimate the role of internal variability in tropical precipitation shifts. In the revised version, we have added the 4 members of the HadGEM3-GC3.1-MM initial condition ensemble to expand the number of ensemble members in our estimates of internal variability from 4 to 8.

Secondly, in the revised version, we have made an effort in the introduction and methods to emphasize the PPE is an ‘ensemble of opportunity’, in a similar fashion to the CMIP multi-model ensembles. The PPE was primarily designed to support a range of impact assessments as part of UK Climate Projections 2018, rather than address our specific research question. However, as the small sample size of the PPE has been designed to sample a broad range of climate responses, we believe it provides a unique viewpoint for assessing the relationship between aerosol forcing and tropical precipitation shifts. We have added a new figure (below and now Figure 1) that shows the PPE spans a similar range of aerosol ERF to the AR6 likely range, but a comparably smaller range of climate sensitivity. Hence, we suggest that despite the small sample size, the PPE may be a useful tool for exploring the impact of the former, rather than the latter, on ITCZ shifts, in a model with consistent physics and where differences in responses can be linked back to underlying parameters/processes.

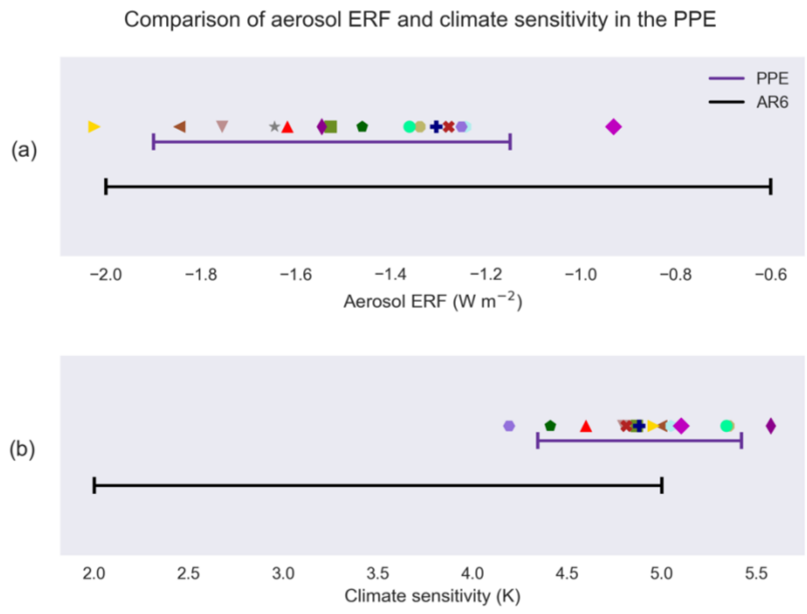

Figure 1: Range of aerosol ERF (a) and equilibrium climate sensitivity (b) across the PPE and AR6. The PPE error bars shows the 90% range across 15 PPE members compared with the AR6 90% ‘very likely’ ranges from (Forster et al., 2021). We use 13 PPE members in this study, excluding two shown in this figure due to model drifts.

Finally, how generalizable are these findings? Though the authors span a range of the parameter space of certain variables, this all occurs within a single model.

The range of aerosol ERF and tropical precipitation shifts spanned within the PPE of our single model are comparable to the range spanned in AR6 and multi-model studies respectively (new Figure 1 and Text S1). Hence, we expect that in other climate models that represent similar parameters and processes, a PPE would span a similar uncertainty range. Furthermore, we compare our results to multi-model ensemble analyses in section 3 & 4 and show the benefits of evaluating uncertainty within a single model. For example, assumed relationships that emerge from multi-model analyses are not always evident when uncertainty in processes is accounted for. This suggests multi-model ensembles would benefit from a perturbed parameter component – a conclusion we think some readers will reach.