In the responses below, the original reviewer reports are in black, while all our comments are in blue. We have also numbered all the reviewer comments and our replies for clarity. We have *quoted text from the manuscript in grey italics*.

Reviewer 1 Comment 1
Dear authors, I appreciate your paper that incorporates and compares a wide variety of different RCMs with different qualities. I would like to contribute to the progress of your RCM inter-comparison project and provide a review on the manuscript submitted to Geoscientific Model Development.

Reviewer 1 Reply 1
Thank you for taking the time to review our paper, it is greatly appreciated by all of the co-authors.

One general comment: We really appreciate the various suggestions for improvement and interesting cross-comparisons of the various modelling group's results. We wholeheartedly agree about the importance of these investigations. However, in response to reviewer 2 and after consulting with the editor, we have had to sharpen the paper's focus and accordingly turned it into a MIP description paper. We do really appreciate the suggestions and hope to be able to respond to them in future work, but we feel we cannot do all of them justice in the confined space we now have.

Reviewer 1 Comment 2
As far as comments on the content are concerned, the question arises how do different RCMs introduce nonlinearities of the temperature response. Your paper is about quantifying the temperature response and does not discuss different concepts that provide conceptual understanding. However, the (equilibrium) temperature response does not always scale linearly with CO2 forcing, and explaining the reader why we have nonlinearities of the temperature response (e.g. explicit feedback temperature dependence, among others) might be helpful for the reader to understand different or common model behavior.

Reviewer 1 Reply 2
Thank you for the comment. We agree that we have not discussed the many different reasons for model differences in any detail. For reasons of scope, we do not feel that we have room to do so in this paper, especially not after the comments of the other reviewers who have asked for further details on the project protocol. We hope to do so in a separate paper and hope that this choice of presenting the manuscript in the style of a MIP description paper is agreeable.

Reviewer 1 Comment 3
Another aspect that is important for an unexperienced reader and related to the former comment is why are different RCMs fitted to different numbers of CMIP models. For instance, some models are likely to runaway in the case of high forcing input, and this runaway can be attributed to different model parameters.

Reviewer 1 Reply 3

Thank you for your comment. In the revised manuscript, we clarify that each model is fitted to a different number of CMIP models due to different calibration choices by different modelling teams. In other words, calibrations depend on each RCM development team's individual capacity. We have also added a clarification of how differences in model parameters have been handled at this very early stage of RCMIP.

Reviewer 1 Comment 4

Further, I can hardly imagine that a parameter which represents feedback temperature dependence is well constrained by the observational record. I wonder how strong model parameters vary between fits to the reference period/observations and abrupt CO2 experiments. Adding brief, explicit paragraphs would be helpful.

Reviewer 1 Reply 4

Thank you for the comment. We agree that we have not discussed the nuances of model constraining at all. For reasons of scope, as in Reply 2 we do not feel that we have room to do so in this MIP description paper. We hope that such work can take place in future research such as https://www.earth-syst-dynam-discuss.net/esd-2019-82/.

Reviewer 1 Comment 5

This also holds true for the discussion on probabilistic projections. You mention very important aspects but how do the different models actually compare?

Reviewer 1 Reply 5

Thank you for the comment. We agree this is an important question but feel it is beyond the scope of the MIP description (see Reply 4).

Reviewer 1 Comment 6

I've a specific comment on the understanding of time-and state-dependent feedback (lines 417-427). It is said that models with time or state-dependent feedback avoid the problem that linear models predict an equally large amplitude to negative radiative forcing as positive radiative forcing. This holds true for state-dependent feedback or the combination of time- and state-dependent feedback but the temperature response of purely time-dependent feedback scales linearly with forcing.

Reviewer 1 Reply 6

Thank you for the comment. We agree this is an important question but have had to remove this more detailed discussion in response to other review comments because it is beyond the scope of the MIP description. We hope this decision is understandable and that future work can consider this question in more detail.

Reviewer 1 Comment 7

As a short technical note, please revise your plotting routines in the supplementary material.

Reviewer 1 Reply 7

Thank you for your comment, we have updated the plots.