

Second Review of 'Practice and philosophy of climate model tuning across six U.S. modeling centers' by Schmidt et al

I have reviewed the author's response to my original review and still feel that many of the points I raised have not really been addressed. Hence I am reluctant to accept the paper for publication without some further attention to the key points I raised notably;

1. **The shortening of Section 2.** The authors claim that Section 2 has been shortened but I think it is actually slightly longer in the revised version and I can only see one reference highlighting the differences with Hourdin et al (2016) and how this paper extends its work. I would urge the authors to look at this again
2. **Section 3.** The author's claim they have endeavored to make the discussion of the tuning practices of each modeling center more consistent and removed vague language. However there are very few changes to section 3 overall and specifically none of the three examples of vague language that I highlighted in my original review have been changed. The authors agreed with my points on this, so I feel that they must make changes to reflect this at least.

Specific points (in *italics*) not addressed adequately

P7, l4: Cess climate sensitivity is evaluated using idealized SST +4K simulations'. How is this then used? Are models thrown away if this is outside of some range (e.g. CMIP5)?

The authors responded to this comment with some explanatory text but this is not included in the manuscript. I think it should be.

P7, l16: '...to monitor the combined impact of anthropogenic forcings and climate sensitivity' Again what does 'monitoring' mean? Is action if its deemed to be 'unacceptable'?

Again the authors expanded in their response but not all of it was included in the revised manuscript. Specifically I think Section 2 needs to include the statement 'Thus action is often taken to go into details more closely and see what has happened'

P8, l29: Why were the new model versions constrained to have a ratio of RFP to Cess sensitivity the same as the old model? Presumably so that the evolution of historic temperature will not differ substantially from that achieved by the old model – although it sounds like it didn't work very well. The target for this tuning need to be said more explicitly.

The authors explain that GFDL use this to predict how the model would react but not the basis for the constraint. To me, they are forcing the new model to maintain this ratio but on what physical or observation basis is this valid?

P10, l6: What happens if the coupled model drifts are not 'relatively small'? Do you go back to the start (i.e. component level tuning?)

Again the authors have given a response but not made any changes to the text. I think their answer to my question is yes, once longer runs have exposed the true size of the drift. Please say so in the text.

P11, l26: What does 'bring together' mean.

This now reads 'bring to bear' but the sentence doesn't make sense to me and it remains very vague.

P12, second paragraph: Are the higher resolution simulations tuned independently from the lower resolution ones, even for parameters with no obvious resolution dependence? How does this fit with the seamless idea or is this an explicit recognition of the specific requirements if the different uses/customers?

Once again the authors response is not reflected in any text changes. It would be good to say their views that there is a continued need for tuning with changing resolution.

P13, l17: What is the basis for constraining the net aerosol forcing to be less than -1.5Wm-2?

The authors state in their response that NCAR used the guidance from IPCC AR5 for this constraint. This, at least, needs to be included in the text.

P14, l 15: What tuning to the historic record does happen – no tuning or no fine tuning?

The authors respond that no tuning was done – this needs to be explicitly stated in the revised manuscript.