



EGUsphere, referee comment RC2
<https://doi.org/10.5194/egusphere-2022-976-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on egusphere-2022-976

Tim Andrews (Referee)

Referee comment on "Better-constrained climate sensitivity when accounting for dataset dependency on pattern effect estimates" by Angshuman Modak and Thorsten Mauritsen, EGU sphere, <https://doi.org/10.5194/egusphere-2022-976-RC2>, 2022

Review of "Better constrained climate sensitivity when accounting for dataset dependency on pattern effect estimates" by Modak and Mauritsen.

This manuscript uses the MPI-ESM1.2-LR model to investigate the dependence of the pattern effect on underlying SST datasets used to force AGCMs with. By forcing the AGCM with 8 different reconstructions of historical SST patterns they find a substantial spread in the diagnosed pattern effect. By combining these results with published results on the inter-model spread the authors produce a new constraint on ECS from historical observations of past climate changes.

I found the manuscript well written and presented, and applaud the authors for tackling the important question of SST dataset dependency in methods that quantify the pattern effect. I think this manuscript will make a really useful contribution to the literature.

While there are a quite a comments here, all ought to be surmountable and are intended to be constructive.

Major Comments:

1. Presentation of SST patterns: the authors present a useful analysis of the geographical patterns of SST trends across the different datasets, and show how they differ (e.g. Fig 3), but are unable to find a geographical region or principal explanation for the variation in feedback seen in their experiments (e.g. Fig 4 and Section 4.3). This is done by calculating the SST trend as a function of time (i.e. K /century). However I wonder if this is the most appropriate method to tackle the question being investigated. I think a more appropriate calculation (of more relevance to feedback and patterns effects) is the SST

change per global-mean dT (analogous to how the feedback is calculated as radiative response per global-mean dT), rather than a trend in time. This normalised pattern has the additional advantage of removing any global-mean dT differences between the datasets (as shown in Figure 2).

Hence I recommend the authors repeat the analysis of Figure 3, Figure 5, Figure 6 and discussion in Section 4.3 but this time calculating the SST patterns per global-mean dT (ideally calculated like the feedback parameter, i.e. with linear regression, in this case regressing $dT(\text{lat}, \text{lon}, t)$ against global-mean $dT(t)$). It might turn out to make little difference, in which that is fine – but it would be good to know and comment on if so. On the other hand, as it ought to relate better to the feedback and pattern effect it might improve the relationships in Fig 6. If so, I would recommend simply dropping the K/century trends and replacing all this with $K(\text{lat}, \text{lon}) / \overline{dT}$ where \overline{dT} denotes global-mean in this case.

2. Calculation of λ_{4xCO_2} : the authors quote the 4xCO₂ λ for MPI-ESM1.2-LR as $-1.48 \pm 0.03 \text{ Wm}^{-2} \text{ K}^{-1}$ from regression in 150yrs of abrupt-4xCO₂. However two other published papers (Andrews et al., 2022; <https://doi.org/10.1029/2022JD036675>) and Zelinka et al. (2020, GRL, updated to more models here <https://zenodo.org/badge/latestdoi/259409949>; see specifically https://github.com/mzelnka/cmip56_forcing_feedback_ecs/blob/master/CMIP6_ECS_ERF_fbks.txt) both report $-1.39 \text{ Wm}^{-2} \text{ K}^{-1}$ for this model using similar linear regression on 150yrs of abrupt-4xCO₂. This difference of $\sim 0.1 \text{ Wm}^{-2} \text{ K}^{-1}$ will feed into all estimates of the pattern effect, and is a significant % of the reported pattern effect, and could even change the sign in some cases. I wonder if the difference is because this manuscript calculates the 4xCO₂ changes relative to the “mean of the last 500 years of piControl simulations”, whereas both Andrews et al. and Zelinka et al. use the corresponding section of piControl and account for control drift? The authors ought to check, and I would recommend using the corresponding section and control drift method. If this is the cause of the difference, then this sensitivity to methodological choices ought to be discussed and acknowledged as a large source of uncertainty and potential bias in the pattern effect estimate.

3. Discussion of the limitations/assumptions in the approach: when coming up with a combined estimate of the pattern effect, I think the manuscript might be making an assumption that isn't explicitly discussed. Specifically, if I've understood correctly, using the mean estimate and distribution from the observed-piForcing simulations is implicitly assuming that not only are all the various SST reconstructions independent (as the manuscript acknowledges, but clearly isn't true), but also that they are equally plausible reconstructions of the truth. We do not know which dataset is 'best', right? The real world could have looked like any or none of them. If the authors have any thoughts on this, and how it could be taken forward – it might be useful to explicitly discuss. Related to this, I hesitate at statements such as line 223: “... AMIP II dataset is an outlier suggesting that earlier studies may have overestimated the pattern effect...”. It gives the impression (perhaps not intended) that the AMIP II results are not trustworthy, whereas we have no idea if AMIP II SSTs are any less likely than any other SST reconstruction. They might be fine. For example, just as a counter argument, there are other situations where simulations with AMIPII SSTs have been shown to perform better than simulations with HadISST. As Andrews et al. (2022) discuss: “...Zhou et al. (2021; <https://doi.org/10.1029/2022JD036675>) showed that TOA radiative fluxes simulated by CAM5.3 correlated better with CERES observations when forced with AMIP II SSTs rather

than HadISST SSTs, suggesting the results from *amip-piForcing* may be more reliable...". I do not think any major changes are required to address this comment, maybe just a simple change of wording or explicit acknowledgement of this point in the manuscript ought to suffice.

4. Structural (model) dependence on SST dataset sensitivity: the authors acknowledge that their results may depend on the model used (MPI-ESM1.2), and I appreciate the effort in trying to combine the model spread and adjusting the mean from previous intercomparisons to account for this. However the authors should update the analysis to use numbers from larger model intercomparison of Andrews et al. (2022), rather than the Andrews et al. (2018) study. For *amip-piForcing* the pattern effect across 14 models is $0.70 \pm 0.47 \text{ Wm}^{-2} \text{ K}^{-1}$ in Andrews et al. (2022), compared to $0.64 \pm 0.40 \text{ Wm}^{-2} \text{ K}^{-1}$ in Andrews et al. (2018). So the difference in the mean to ECHAM6.3 ought to be slightly larger than the authors used here, and the spread slightly larger too. But moreover, the method applied assumes that the ECHAM/MPI-ESM model is similarly different to other model in ALL the datasets as it is in the *amip-piForcing* simulation. But we know from Andrews et al. (2022) that this is not so, the ECHAM and MPI-ESM pattern effects under *hadSST-piForcing* are much further from the mean than under *amip-piForcing*. For example, the pattern effect for ECHAM6.3 and MPI-ESM2.1-LL is $\sim 0.2 \text{ Wm}^{-2} \text{ K}^{-1}$ in *hadSST-piForcing*, whereas the model mean is $0.48 \text{ Wm}^{-2} \text{ K}^{-1}$ (Andrews et al. 2022; Table 2). I'm not sure exactly what the solution is here, but it seems adjusting for the mean difference to ECHAM6.3 of just $0.1 \text{ Wm}^{-2} \text{ K}^{-1}$ based on Andrews et al. (2018) is insufficient. It ought to be larger based on the larger ensemble of Andrews et al. (2022) and larger again given the *hadSST-piForcing* results where ECHAM6.3 pattern effect is further away from the rest of the models. At a minimum this potential structural dependence of the results on the ECHAM6.3/MPI-ESM model ought to be explicitly discussed.

Minor comments:

* Title: "Better constrained" does not particularly read well to me, how about "Improved constraints on.." ? But I'll let the authors decide.

* Lines 45-49: I think it would be appropriate to include Andrews and Webb (2018; JCLIM, <https://doi.org/10.1175/JCLI-D-17-0087.1>) in the discussion of the mechanisms of the pattern effect.

* Lines 110: " $-1.48 \pm 0.03 \text{ Wm}^{-2} \text{ K}^{-1}$ ", what is the level error being presented here and throughout, 5-95% or something?

* Line 116: "... account for land surface warming in the fixed-SST simulation to calculate the forcing.." – the literature has various ways of doing this, it would be good to briefly clarify which approach was used.

* Lines 126: "... pattern effect ranges from -0.01 +- 0.09 Wm⁻² K⁻¹..." – please clarify how the uncertainty is calculated here. Since it is difference between lambda_4xCO2 and lambda_piForcing have you added the errors in quadrature or something?

* Line 129: "... the dataset mean pattern effect is lower than both Andrews et al. (2018) and the values considered in IPCC AR6..." – this reads as if the value is outside the range. So it is not quite right that the value is lower than those "considered" by Andrews et al. and IPCC since it is within their uncertainty range.

* Line 200: "... pattern effect estimate in MPI... averaged over all decades..." – I'm not exactly sure what this means, please clarify.

* Line 204: "bt" typo.

* Line 210-11: This sentence explains how the mean combined pattern effect estimate is arrived at, but it doesn't explain how the uncertainty is combined between the pattern effect dataset dependence in this manuscript and the multi-model spread, i.e. how is the -0.14 to 0.74 Wm⁻² K⁻¹ on line 210 arrived at?

* Appendix: I found the Table summarising the datasets and Figure A1 showing the 30yr moving window lambda useful, and would recommend integrating them into the main text. One option would be to replace Figure 7 in the main text with Fig A1, since I find this figure more useful and interesting. However I do appreciate this is somewhat personal preference, so I do not demand it and leave the authors to choose.

I have signed the review.

Tim Andrews.