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

Note- Our responses to the reviewer's comments are in red color text.

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, EGUsphere, https://doi.org/10.5194/egusphere-2022-976-RC2, 2022

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.

We thank the reviewer for appreciating the study and for the constructive comments that helped improve the manuscript. The point by point responses are provided below.

1 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(lat,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(lat,lon)/ were <> denotes global-mean in this case.

Thanks. In the revised manuscript, we have updated the analysis of Figure 3, Figure 5 and Figure 6 by regressing dT(lat,lon,t) against global-mean dT(t) instead of temperature trends as a function of time. We find that the overall results remains similar.

2. Calculation of lambda4xCO2: the authors quote the 4xCO2 lambda for MPI-ESM1.2-LR as -1.48 +- 0.03 Wm-2 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/mzelinka/cmip56_forcing_feedback_ecs/blob/master/CMIP6_ECS_ERF_fbks.txt)` both report -1.39 Wm-2 K-1 for this model using similar linear regression on 150yrs of abrupt-4xCO2. This difference of 0.1 Wm-2 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.

   We find that the difference is because of our use of surface temperature (ts) instead of surface air temperature as in Andrews et al. 2022 and Zelinka et al. 2020. IN the revised manuscript we write "Note, since we use surface temperature in our analysis, our estimate of λ_{4xCO2} for MPI-ESM1.2-LR is larger than the estimate from Zelinka et al. (2020) and Andrews et al. (2022) which uses surface air temperature instead".

   The difference in trends under global warming in surface and surface air temperature was investigated as part of IPCC AR6. Generally, models warm their surface air temperature by 5-10 percent more than the surface. However, observations of night time marine air temperatures co-located with surface temperatures show the opposite of similar magnitude. Therefore IPCC AR6 assessed the difference to be zero in all cases, however, with the issue that this causes some risk of confusion.

   In this case we use the same definition throughout, so there is no comparing apples to bananas. There is one exception, though, when using Andrews et al. (2022) pattern effects, these are calculated based on surface air temperature. Therefore, the pattern effects and corresponding correction to ECS over that of AR6 is a bit higher by a few percent. The downward correction is then a bit smaller, in turn.

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…". If 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.

Thanks, yes we do not intend to discuss the validity of the reconstructions in this study, and so by outlier we mean in a statistical sense. In the revised manuscript, we now rewrite "...it turns out that the pattern effect estimated from the AMIPII dataset is by far the largest suggesting that earlier studies may have overestimated the pattern effect."

Also, we add "However, we reiterate that any of the SST dataset could be a possible path Earth could have taken."

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 +- 0.47 Wm-2 K-1 in Andrews et al. (2022), compared to 0.64 +- 0.40 Wm-2 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 0.2 Wm-2 K-1 in hadSSTpiForcing, whereas the model mean is 0.48 Wm-2 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 Wm-2 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.

Thanks. In the revised manuscript, we have applied Andrews et al. 2022 and updated the calculations. We now write "We find the combined pattern effect estimate to be 0.37 $\mathrm{Wm^{-2}K^{-1}}$ [-0.14 to 0.88 $\mathrm{Wm^{-2}K^{-1}}$]. Our calculations for the combined estimate considers only the AMIPII based pattern effect estimates from Andrews et al. (2022). The combined estimate is obtained by subtracting the mean of ECHAM6.3 and MPI-ESM1.2-LR estimate in Andrews et al. (2022) from the sum of our estimate and their multi-model mean estimate. The uncertainty range in the combined estimate is deduced by adding in quadrature the standard deviations from our dataset dependent estimate and their multi-model estimate.

Thus, our calculation also accounts for the weaker pattern effect in ECHAM6.3 and MPI-ESM1.2-LR than in the multi-model mean. While accounting for the weaker pattern effect the assumption is that ECHAM6.3/MPI-ESM1.2LR is different from all the other models in all datasets as in AMIPII. However, one can argue this assumption. Inferred from Andrews et al. (2022), Figure A4 shows that not only ECHAM6.3/MPI-ESM1.2LR but also other models though produce larger pattern effects has a large difference in pattern effect estimates based on AMIPII and HadISST datasets."

It is certainly possible that MPI-ESM is more sensitive to changes in datasets than other models, however, this cannot be determined based on our study. We are therefore currently conducting a model intercomparison project to investigate this possibility further.

2 Minor comments:

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

  Thanks. We considered the suggestion but decided to stay with the original title.

- * 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.

  Thanks. Now it is included.

- * Lines 110: "-1.48 +- 0.03 Wm-2 K-1", what is the level error being presented here and throughout, 5-95% or something?

  We write in section *3. Model, datasets and experiments* "Unless otherwise specified, throughout the text, the displayed results are based on the mean of the 5 ensemble simulations while the uncertainty denotes the standard error from the ordinary least square (OLS) regression and the ranges are the 5-95th percentiles."
  Also, in section *4.5 An updated estimate of ECS* we add, "In this section, note that we consider the best estimate while the uncertainty (standard deviation) and ranges denotes the 5-95% confidence intervals."

- * 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.

  In the revised manuscript we clarify this and write "To calculate F, we account for the land surface warming: we subtract the product of λ_{4xCO2} and the land surface warming from N simulated by historical simulation in fixed-SST configuration (Hansen et al., 2005; Modak et al., 2018)."

- * 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?

  Thanks. We now add "The uncertainty in the pattern effect estimates are calculated by adding the errors from λ_{4xCO2} and respective λ from *observedSST-piForcing* simulations in quadrature".

- * 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.

  We now write "The mean estimate of the pattern effect derived from different dataset is less than the multi-model mean estimate of Andrews et al. (2022) and the mean estimate considered in IPCC AR6 (IPCC, 2021)."

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

  We now write "When we apply our dataset dependent pattern effect estimate - mean of $\Delta\lambda$ deduced from each dataset, 0.20 $Wm^{-2}K^{-1}$ [0.01 to 0.39 $Wm^{-2}K^{-1}$]), the ECS is adjusted to 2.9 K [1.7 to 6.7 K] (Figure 8)."

- * Line 204: "bt" typo.

  Done.

- * 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?

  We now clarify and write "The uncertainty range in the combined estimate is deduced by adding in quadrature the standard deviations from our dataset dependent estimate and their multi-model estimate."

- * 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.

  To update ECS estimates based on the historical warming, the entire long-term period is generally applied, in this case 1871-2017. In Figure 7 we show how the long-term slope evolves when we increase the regression length by a year derived from different SST datasets. Hence, we find it relevant for this study and decide to retain Figure 7 in the main text. We agree that the sliding 30-year evolution of climate feedback is useful and interesting as it helps to identify how the 30-year feedback varies and can be used to compare the datasets. However, we do not use the 30-year periods to update ECS, and so is not central to our work. Hence, we decide to keep it in the supplementary.

I have signed the review.

Tim Andrews.

References

Andrews, T., Gregory, J. M., Dong, Y., Armour, K., Paynter, D., Lin, P., Modak, A., Mauritsen, T., Cole, J., Medeiros, B., and et al. (2022). On the effect of historical sst patterns on radiative feedback. *Earth and Space Science Open Archive*, page 48.

Hansen, J., Sato, M., Ruedy, R., Nazarenko, L., Lacis, A., Schmidt, G. A., Russell, G., Aleinov, I., Bauer, M., Bauer, S., Bell, N., Cairns, B., Canuto, V., Chandler, M., Cheng, Y., Del Genio, A., Faluvegi, G., Fleming, E., Friend, A., Hall, T., Jackman, C., Kelley, M., Kiang, N., Koch, D., Lean, J., Lerner, J., Lo, K., Menon, S., Miller, R., Minnis, P., Novakov, T., Oinas, V., Perlwitz, J., Perlwitz, J., Rind, D., Romanou, A., Shindell, D., Stone, P., Sun, S., Tausnev, N., Thresher, D., Wielicki, B., Wong, T., Yao, M., and Zhang, S. (2005). Efficacy of climate forcings. *Journal of Geophysical Research D: Atmospheres*, 110(18):1–45.

IPCC (2021). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, volume In Press. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Modak, A., Bala, G., Caldeira, K., and Cao, L. (2018). Does shortwave absorption by methane influence its effectiveness? *Climate Dynamics*, 0(0):0.

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., Klein, S. A., and Taylor, K. E. (2020). Causes of Higher Climate Sensitivity in CMIP6 Models. *Geophysical Research Letters*, 47(1):e2019GL085782.