The authors present a new estimate of how much the “pattern effect” affects equilibrium climate sensitivity (ECS) inferred from 1871—2017 climate change. Many pattern-effect studies use the AMIPII sea-surface temperature (SST) dataset but there are unresolved inter-dataset differences in SST evolution. The submission shows how, in a single climate model, different SST datasets affect estimated ECS_hist. The paper is in scope and takes a sensible approach. Despite my review’s length, I picked minor revisions because I expect the main conclusions to be robust – with the caveat that all estimates of ECS_hist are expected to have issues.

Before supporting publication, I ask that the authors address a series of comments on several themes below, then end with minor comments.

I strongly suggest that panel Fig. A1(a) goes into the main manuscript. It allows identification of which periods matter for ECS. I was interested in WWII when ERSST and HadSST-based datasets have very different corrections, plus recent decades for which we have satellite-based constraints. Different communities will care about such results which are hard to interpret from Fig. 7(a).

1. Discussion and results to contextualise this paper

I think your discussion of the causes of historical pattern effects is insufficient (e.g. “could be either externally forced or could be internal variability driven”). As I understand it, your ECS_hist correction is only valid if the pattern effect is internal variability and/or non-CO2 forced. If it’s CO2 forced, as proposed in Seager et al. (2019, doi: 10.1038/s41558-019-0505-x), then your correction would be invalid. Please explicitly discuss.

I believe you could also derive time-varying λ_{hist} as in Figs. 7(a) and A1(a) from the historical runs. Does that suggest a role for a time-varying forced pattern effect in this model?

It also seems obvious to compare to prior work. Can you reproduce Fig. 7(a) and Fig. A1(a) with the Lewis & Mauritsen (2021) CAM5.3 Green’s function results for the in-common SST datasets?

Both tests could go in the appendix with brief main-text commentary, unless results are particularly interesting.

2. Method clarifications

Some methodology details confused me.

P5L108 paragraph. Did you remove piControl drift? Was preindustrial volcanism included in the piControl runs? Do these factors change *N* and therefore ECS_hist?

For the standard error from SST datasets, is that just the Gaussian standard error on the mean, treating each SST result as independent?

How do you combine the errors in Sec. 4.5? I followed AR6 Fig. 7.6 to Table 7.SM.14 and the cited notebook (https://github.com/chrisroadmap/ar6/blob/main/notebooks/100_chapter7_fig7.6.ipynb). I don’t see where error combination method is stated. Please clarify.

3. Robustness tests and some extra detail

How linear are things?

Can you show (perhaps appendix?) annual *N* versus *T* by SST dataset? Can you check your lambda regression calculations with something more robust than OLS to outliers (e.g. Theil-Sen)? The results would help say something about robustness.

Long-term linear fits to SST(t) could be misleading. What if Fig. 3, 5 & 6 analysis were repeated with e.g. LOWESS or differences between start and end periods?

Are there periods (e.g. from the 30-year regressions) where the Fig. 5/6 regions seem more predictive of λ? Only interesting results need to go into the paper or appendix. You might find times or locations that are particularly impactful.

For the Figs. 5/6 regional analysis – did you try any combinations e.g. a Pacific west-east difference? Is there no additional info from a larger Sc/trade Cu region? I understand you should limit paper content, but it would be nice to know you checked at least Pacific ascent minus descent region differences given the number of papers that have highlighted the east-west gradient.

You report narrowed ECS uncertainty but to some it may be counterintuitive that adding SST source uncertainty shrinks the final errors. Your title emphasises the reduced uncertainty – but somewhere it should be stated that this just comes from the smaller ECS thanks to how the division-by-lambda works. Spelling this out would be helpful, as would mentioning some extra caveats that I did not see:

- Are SST dataset lambdas treated as independent in the standard error calculation? If yes, are your standard errors too small since they're not actually independent?
- Do you use model spread as an error? If yes, please add the usual comments about ensemble of opportunity etc and how that affects interpretation.
- You said that you "*had to assume that variations in the pattern effect as estimated among the models and across the datasets are independent of each other*". It seems likely there is a correlation though, justifying your suggested next steps in the final sentence. You could do error propagation with assumed correlations between model and SST errors and a sentence or two could then say "strong correlation (r=whatever) does/doesn't greatly affect these conclusions…".

Basically your headline is "reduced uncertainty" and I'd like extra clarity on exactly what you're claiming.

Minor/typos

Suggest extending subscripts to be absolutely clear, specifically "ECS_hist" (or "ECS_amip") to discriminate as you do with your λ subscripts.

P1L11 – 95 percentile -> 95th percentile

P1L19—20: "*To know what is required to meet the Paris Agreement goal it is imperative to better quantify and understand the rate of global warming*"
You aren't really studying warming rates. Maybe something like "…it is imperative to better quantify and understand the ultimate amount of warming in response to a given forcing."

P2 lines messed up but between 25 and 30:
"*It can be framed as N = F +λT, where N is the planetary energy imbalance which is generally measured as the net downward radiative flux at the top-of-the-atmosphere (TOA), F is the external radiative forcing, measured as the effective radiative forcing (ERF)*"
I think you mean "defined" rather than "measured", both for TOA flux (even CERES is pinned to Argo etc for the mean) and ERF is clearly not measured either.

P2 lines and equation:
The equation uses ΔT and the text uses *T* both to denote changes in *T*. The same for *N* and *F*. I suggest using ΔT, ΔF and ΔN consistently to denote changes.

P2L32—33:
"*To reconcile this discrepancy the community looked into the concept of pattern effect which is not taken into account in the traditional energy balance framework.*"
This misses how changes in *T* data have reduced the apparent differences. Both Lewis & Curry and Otto et al. used HadCRUT4. Pick some relevant citation – a recent e.g. is Clarke & Richardson (2021, doi: 10.1029/2020EA001082) whose Table 2 shows how 1850—2019 ECS_hist estimates would increase relative to HadCRUT4. Cowtan & Way infilling is worth about +10 %, while Berkeley

Earth's more substantial improvements (more stations, infilling, improved sea ice treatment) give +17 % to ECS_hist. The authors could persuade me otherwise, but updated global temperature datasets seem both important context for a study based on inter-dataset issues.

P2L42: "*Atmosphere-Ocean General Circulation Model (AOGCM) simulations show that the long term climate response (response to abrupt4xCO2) resembles a temperature pattern similar to 'El-Nino Southern Oscillations' (ENSO)"*
This could be interpreted to mean that models have shown/proven the "real" long-term pattern and should be rephrased to something like: AOGCMs' *simulated* pattern resembles ENSO. I think you wanted to make this point in P3L47 when you said "assumed correct", but I wasn't sure. Change from:
"*…leads to more stabilizing feedback while an assumed correct opposite future temperature distribution would lead to a less stabilizing feedback"*
To something like:
"…leads to more stabilizing feedback while the simulated future temperature distribution would lead to a less stabilizing feedback"

Figure 1 caption: are your fixed-SST simulations really "AOGCM" which you've defined as "atmosphere ocean GCMs"? I interpreted AOGCM as including coupling to an ocean model.

P3L57—58: "*Such estimates are model dependent, nevertheless most models yield a dampening pattern effect based on the AMIPII dataset."*
Alone, "dampening effect" could simply mean negative feedback in general. I assume you mean dampening relative to the long-term abrupt4xCO2 pattern? If yes, please correct.

Fig. 3 – second row first column figure is labelled "HadSST" rather than "HadISST"

P5L113—115: "*For λhist, we take the changes in the global annual mean N and T from 1851-2014 of historical simulation relative to the mean of 1851-1900 of the same. Since, there are 10 ensemble members present in the historical simulation, we use the mean of the ensembles in our calculation"*
Can you rephrase? I think this is saying you took the mean of 1851-2014 minus the mean of 1841-1900 but am not certain. Also, did you mean "ensemble mean" instead of "mean of ensembles"?

P8L142: is this the mean and standard error of the fits to each dataset, calculated using the standard Gaussian assumption, or something else?

P8L145: rogue decimal point in error: 0.0.25

P8L151: typo "expectced" -> "expected"

P8L161: typo gap in 0 .59

P10Fig6: Are *y* axis units W m^{-2} K^{-1}? Please add axis labels or explain why not.

P11L195—196: "*while the uncertainty is deduced from Figure 7.6 of this report." -* "this report" could mean your paper, whereas I infer you mean "AR6 Figure 7.6"

P12L204: typo "bt" -> "by"

P14L245: Table A1 says that AMIPII = HadISST until Nov 1981, is that correct? If yes, then what are the slight differences in Fig. 1? If so, then is it surprising that there are such big, consistent differences in feedbacks in Fig. 7(a) when you've got multiple runs with only land/atmosphere variability?