Review of "Does the pace of carbon emissions matter in an atmospheric general circulation model?"

Summary and comments:

In this work, the authors assess whether the mean state and day-to-day variability of global near-surface temperature and precipitation depend on CO_2 concentration increase rates using a small ensemble of GFDL-AM2 simulations, specifically focussing on the atmosphere's role and excluding incluences of a reactive and dynamic ocean.

As the authors reason, theory of the governing atmospheric processes indicates a priori that such path-dependence should not exist when slow ocean processes are eliminated artificially. However, I agree with the authors, that it is worth a quantified assessment and support the trial to sort this out, even if this yields the apparent negative finding: the mean state and day-to-day variability of global near-surface temperature and precipitation in a purely atmospheric setting likely do not depend on the CO_2 concentration change rate.

The basic idea to compare inter- and intra-scenario differences as a robustness criterion is good. I believe the analysis presented in this manuscript illustrates the argument, however is not sufficient to prove the point (as the authors state themselves). While I appreciate the effort of the analysis and promote the analysis for and publication of negative findings, I feel this analysis could do much better at this, even when restricted to the limited data produced by the authors.

My main suggestions are:

- go beyond one decade and test multiple time ranges up to 30 years. I see little inhibiting this in de-trended space.
- Aggregate the grid-cell level results to different higher levels
- Compute signal-to-noise ratios at all spatial levels.
- Consider analysing existing AMIP-style simulations (data ocean instead of slab ocean) from the CMIP6. There are different scenarios with different CO2 concentration rates and several models with multiple ensemble members.

Overall, in my opinion, this work needs major revisions to serve its purpose.

Line-by-line comments:

Title: replace "emissions" with "concentration changes". As far as I understand, you only prescribe CO2 concentrations, not emissions. You are not testing the paces of emissions in a GCM with online carbon cycle modeling. Likewise replace emissions in any equivalent occurrence in the text.

Lines 10-11: this is a good method. In my opinion it additionally needs more quantitative aggregation. Surface area of significantly different variability, magnitude of the difference field, etc.

Lines 37-38: make clear, whether you refer to temperature or precipitation here.

Lines 48-49: Why not a 0.5% scenario?

Lines 70-71: Is this enough for your study and which kind of potential miss-estimation do you expect from such a simple land model?

Line 78: How did you determine N = 5 as adequate ensemble size for the task? Later on, you try to argue further analysis is inhibited by lacking data. Please indicate shortly why you designed a macroensemble (vs. small perturbation / microensemble), why is a one year IC steps a good choice?

Lines 79-80: replace with “too far equatorwards on both hemispheres.”. Does the sea ice modeling in “slab mode” mean that the behavior you mention for GFDL-AM2 is prescribed or modeled and emergent? And if prescribed, why can you not prescribe the sea ice cover more realistically?

Line: 83: Annual or decadal mean GMT?

Line 84: Why don’t you subtract the 10 year mean around initialisation in the parent simulation? This would help to exclude variability from the IC. Additionally, for avoiding influences of model GMT drift, a comparison against an unforced reference simulation (ensemble) would make this more robust.

Figure 1: enumerate the panels please.

please unify the y-axes of the CO2 concentrations for comparability.

There is very little year-to-year variability in the annual mean GMT time series compared to the historical record or Earth System Model simulations. Is this because ocean variability and its feedback are excluded or are there additional other players at work?

Line 86: on “small deviations”: Please quantify how small.

Lines 86-87: Why is it important to have values centered around zero in their distribution in space? As far as I understand you do not consider distribution in space.

Lines 89-91: With which parameters did you smooth the timeseries? Did you check whether such smoothening gives results close to results of averaging over more years?

Lines 92-94: Again, please give details on the parameters and other methodical choices here. Any part of variability you do or do not filter out in this step will not be evaluable downstream in your pipeline and can potentially influence your results of variability estimates. So the reader needs to understand, how you determine your trendline and which frequency ranges of temporal variability you are eliminating beyond a linear trend.

Figure 3: Please enumerate the panels.

Panel a): Why not center this line around 2°C? The phenomena you are trying to show would emerge equivalently. And please highlight the global mean change as vertical line.

Panel b): Please plot the field showing actual grid cell values. The contour shading you use hides the models resolution and its effects unnecessarily. This is the case for all maps in this work.

The caption and colorbar label do not indicate clearly what you are showing. Is it the change in mean state in a 2°C warming world vs. a 0°C warming world and is re-centered around 0 instead of 2°C or something else? If this is it, the presentation with temperature changes centered around 0 instead of 2°C is highly confusing. Please reconsider.

Also I do not understand the centrality of this figure to the paper you seem to suggest with its size and prominence. If this is just a sanity check that the mean state behaves alright for the following comparative analysis, please present it accordingly.

Your plot seems to have deficiencies at lon = 0 (no color shown). Please fix this.

Finally, please give the reader some help on which values differ significantly from a 0°C world in your model. I recommend grid-cell wise significance testing with a correction for false discovery.

Line 112: Add “on grid cell level” for clarity.

Lines 114-115: This seems quite normal when only averaging over 10 years and 5 ensemble members (Compare e.g. AMIP simulations spread).

Figure 4: enumerate panels, please.

At this level of detail, I am not interested in single ensemble member comparison. like in panels b) and c). Please aggregate the information graphically across all possible inter- and intraensemble differences.

Lines 117-118: Can you really tell this from visual inspection only. It gives the impression yes, but I would like to see more rigorous analysis here. Please not make the reader aggregate your findings, aggregate them for the reader. Use measures of difference field strength, regional, hemispheric, latitude band areal mean values and show whether your estimates (this holds for all your indicators for both mean state and variability as well as temperature and precipitation) emerge from noise (contrast between members of same ensemble).

Lines 120-123: So why are you using 10yr mean values instead of longer averaging? Especially after de-trending, what inhibits an analysis of a range differently long averaging windows around a common warming level?

Figure 5: So compared to Fig. 3, are these maps now re-centered likewise? As far as I understand, no. Please handle Fig. 3 accordingly. I strongly suggest showing 4% together with 1% also there.

I strongly suggest to use another color scheme for precipitation than for temperature to avoid confusion of the reader with reduction vs. increase of temp vs. precip.

Lines 129-130: Please give the intra-ensemble uncertainty range in mm/d as well.

Lines 133-136: Please elaborate. What is your analysis worth when the reader is left with speculation that all the signal should probably be noise.

Lines 137-138: Valuable context, correct comparison in my opinion. However, when the signal is of same order of magnitude as the noise, there can still be a signal. How about a quantification of signal-to-noise ratios at different aggregation levels? Grid cell, region, hemisphere, global? Absolute value means vs. means would help here for an exact argument.

Lines 145-146: So indeed the sea-ice biases have massive influence on the temperature variability. Please argue more precisely why the GFDL-AM2 estimates are still valuable for an assessment of variability changes.

Lines 159-164: I disagree, quantitative statistical analysis can still help to illustrate how likely it is from the available data that differences in mean states and variability between forcing rates are insignificant.

Lines 180-181: Again, please show this graphically for different levels of aggregation.

Lines 194-196: I agree, this analysis is not sufficient to prove the point. While I appreciate the effort of the analysis and promote the analysis for and publication of negative findings, I feel this analysis could do much better at this, even when restricted to the limited data produced by the authors.

Lines 214: So, I suggest you provide the actual information to the reader, what are known biases of GFDL-AM2 concerning mean state changes and intra-decadal variability? And for indicators, where observations do not give a meaningful reference, how does GFDL behavior compare to other atmospheric models?

Lines 216-217: This is trivial, can be omitted in my opinion. However, there are ensembles of AMIP style simulations (data ocean, but otherwise similar) you could try to use as additional source of data to check other models. Multiple models with 3 and more ensemble members for different rates of concentration increase (e.g. amip-hist vs. amip-4xCO2).
https://esgf-node.ipsl.upmc.fr/search/cmip6-ipsl/ search query: amip, tas, day

Lines 221-223: Please also comment on the impact of the choice of land model on near-surface temperature and precipitation variability. Please also reflect on internal variability coming from carbon feedbacks in the Earth System.