

Reply to reviewer #2

This paper is intended as a documentation paper for a new Large Ensemble of climate model simulations with MIROC6. This is an extensive set of new simulations, including historical & scenario simulations, as well as single forcing experiments. It presents the experimental design and overview of simulations, as well as presents some high-level analyses of global climate characteristics of this ensemble, in particular testing the linear additivity of the single-forcing experiments to make up the all-forcing simulations, and projected changes in ENSO variability.

The paper introduces a valuable new resource for the climate community, is well written, and presents some interesting first analyses of climate characteristics in this ensemble. This is a valuable contribution, and I would recommend publication subject to minor revisions. I have included my comments below. Comments 6) and 12) are slightly more substantial than the others but should be easy to address.

Thank you very much for your useful comments.

- 1) Do the 50 historical ensemble members of the LE include the CMIP6 historical ensemble members, or are they separate? Is the model version identical to CMIP6 (i.e., can both ensembles be merged)?

The LE includes the CMIP6 simulations, and the model version is identical to CMIP6. We will clarify it as “We increased the ensemble sizes of the CMIP6 simulations and performed an experiment recently proposed.”

- 2) How were the initial conditions years selected / spaced out in the piControl run?

We will involve the list of initial conditions as a table.

- 3) Timeseries figures (1, 2, 6 & 7): it would be very helpful to draw a line at 0

We will draw a line at 0.

4) L105: It might be worth noting that the effect of hist-stratO3, hist-totalO3 and hist-lu are small but not zero especially in the 2nd half of the 20th century

We will rewrite this sentence as “Changes in stratospheric and tropospheric ozone (hist-stratO3 and hist-totalO3) and land-use-land-cover (hist-lu) have small effects on the global mean T and Tx: changes in stratospheric ozone and land-use land-cover have small cooling effects in the 2nd half of 20th century and changes in total ozone have small warming effects.”.

5) L108: do you mean to say small trend or no trend? I would suggest either using “the P of the historical runs **only** has a small trend” or “has little trend” depending on the intended meaning

We will use “only has a small trend”.

6) L117-126: The median of the blue bars (corresponding to the linear sum of single forcing experiments) is in some cases not near the middle of the orange bars, suggesting some non-linearities in the medians / ensemble means. There could still be important non-linearities here that would be evident in the ensemble means (or medians); overlap of the blue / orange ranges is not evidence of the absence of non-linearities in the ensemble means. This should be commented on here

Thank you for the useful advice. For the revised version, we have computed the max-min ranges of 1000 ensemble average values of randomly sampled ensemble members with replacement. The range of the ensemble mean of historical runs overlap with the range of the sum of the ensemble average of individual forcing experiments. Therefore we have not found any statistical evidence of non-linearities in the ensemble mean values.

7) L139 typo: faction -> fraction

We will correct it.

8) Figure 4: I would suggest using a title for each row e.g. deltaTx(hist)-deltaTx(hist-nat) to help the reader

We will add titles on Fig. 4.

9) L153-155: I don't think 'nearly disappear' for the aerosol signal by end of 21st century in T is quite correct: substantially reduced yes, nearly disappear I'm not sure – it looks like -0.2 degrees C by the end of the century. That's still about half of the strongest aerosol response in the late 20th century (approx. -0.4C by eye)?

We will revise those lines as “Because aerosol emissions gradually decrease under SSP2-4.5 (Rao 2017, Lund 2019), the negative responses of T, Tx and Px significantly decrease. Although negative responses of P to aerosol forcing also decrease, apparent differences of P changes between ssp245 and ssp245-GHG remain until 2100 mainly due to the large sensitivity of P to aerosol forcing (Shiogama et al. 2010a, b).”

10) L 187: Internal variability (=variabilities) in plural sounds a bit unusual, I think you could use Internal variability singular here and in other places

We will use “Internal variability” in the revised manuscript.

11) L 210: single-ensemble – don't you mean single-member?

We will correct it.

12) The methods and terms used to describe the two methods are somewhat ambiguous (L190ff): I think it should be made clear that the single-member / multi-member name refers to the method for removing the forced response only, since the circle and crosses in Figures 10, 11 and 12 are both estimated from averages across multiple ensemble members. The authors could try using different names: multi-member-mean removed and single-member trend removed, for example

The authors also need to explicitly state how they calculate the variability once the forced response is removed in both methods: standard deviation across time for each member, and then

averaged across all ensemble members? Are you averaging standard deviations or variances?

L 198 is especially ambiguous: does the “best estimate” ensemble averages apply to both methods?

Thank you for the suggestion. We will change the wording to describe the two methods. After your suggestion, we will use “single-member-trend removed (SMTR) estimate” and “multi-member-mean removed (MMMR) estimate” through the manuscript.

The method for calculating the variability will be clarified as follows: “The standard deviation is derived across time for each member, and then averaged across ensemble members. We term the 50 member averages of the standard deviations derived by the MMMR estimate ‘the best estimate’ in this study.”

13) L291: It might be worth stating explicitly that AGCMs cannot simulate changes in coupled modes of variability and SST patterns, and hence can only inform projections and attribution statements conditionally with respect to prescribed SST patterns

We will explain it as “while AGCM simulations can only inform projections and attribution statements conditionally with respect to prescribed SST patterns”

14) L297: perhaps worth inserting “future changes in ENSO variability” or “ENSO amplitude”

We will correct it.

It is somewhat obvious, but the authors state in a number of places that larger ensembles are needed to detect changes between ssp119 and ssp126. This is certainly true, but it might be worth saying somewhere something to the effect of “unsurprisingly, the smaller the difference in forcing, the larger the ensemble needs to be to detect differences in the forced response. This is just a suggestion, the authors can choose to take it or leave it!

We will add “It is expected that larger ensembles are necessary to identify differences in climate

response for smaller differences of forcing.” in the text.