

Reply to RC4 (comments in blue, reply in black)

Summary

This paper presents a novel study which attempts to create better projections by calibrating large ensembles over a calibration period where we have both observations and large ensemble simulations. This study investigates three methods of calibration and finds that while all methods perform well, no method performs substantially better than the others. They then show improvement by using a dynamical decomposition method. They find that the calibration works much better for temperature than precipitation, and attribute this to the lack of clear forced change in the calibration period for precipitation. For temperature they find improvement for both large ensembles over Europe by using this calibration method and find that it reduces warming as compared to the calibrated ensemble. I recommend publication with a few minor points to be addressed.

We thank the reviewer for the positive comments on our study.

Minor points:

Page 3 line 2 should be 'ensembles'

This will be corrected.

Page 3 lines 27/28 MPI-GE is initialized from different years of a long pre-industrial control run, not in the same way as LENS

This is an important distinction and was an oversight on our part. A description to this effect will be added in the revised manuscript.

Page 4 line 22 should be 'projections'

This will be corrected.

Section 2.3.1 Are your results sensitive to the choice of reference period? For the dynamical decomposition can you explain why and how you use SLP?

No the results are not very sensitive to the reference period. We tested from 30-years up to the full 97-year periods and the verification statistics generally improve with the length of the period, which is why we use the full reference period here. Text describing these tests will be added to the revised manuscript.

The SLP is used to estimate what seasonally anomalies can be attribute to large-scale circulation anomalies (assessed in terms of SLP anomalies). This will be clarified in the revised manuscript, including a schematic illustrating how the dynamical decomposition is applied to produce the calibrated projections.

Page 7 lines 7/8. Please explain what you mean by "The raw ensemble is clearly has a positive bias"

This refers to the observations over the reference period – this will be clarified.

Section 3.3 The explanation at the beginning of the section should be in Section 2.4

Agreed, the description in section 2.4 will be expanded in the revised manuscript and will also include a schematic to visualise how this is used to produce calibrated projections.

Additional studies that may be of interested: only cite if you feel appropriate.

<https://www.earth-syst-dynam-discuss.net/esd-2019-69/>

<https://journals.ametsoc.org/doi/full/10.1175/JCLI-D-16-0905.1>

*Deser, C., F. Lehner, K. B. Rodgers, T. Ault, T. L. Delworth, P. N. DiNezio, A. Fiore, C. Frankignoul, J. C. Fyfe, D. E. Horton, J. E. Kay, R. Knutti, N. S. Lovenduski, J. Marotzke, K. A. McKinnon, S. Minobe, J. Randerson, J. A. Screen, I. R. Simpson and M. Ting, 2020: Insights from earth system model initial-condition large ensembles and future prospects. *Nat. Clim. Change*, doi: 10.1038/s41558-020-0731-2.*

Agreed, the description in section 2.4 will be expanded in the revised manuscript.

We thank the reviewer for their very useful comments and suggestions.