I had my first read of this paper before the other reviewers posted their comments, but it looks like I am the last to finally organize my thoughts. I am not going to repeat things requested by the other reviewers (including a summary of results), and rather focus on those aspects not already covered.

We want to thank the reviewer for the detailed comments, which helped improve the quality of the manuscript. Following these suggestions, our plan to address it in a revised manuscript is the following:

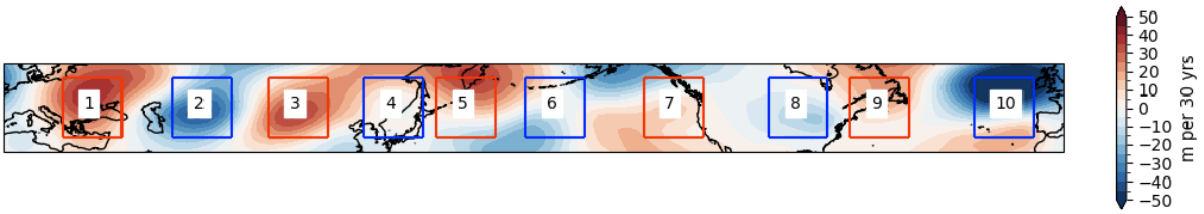
My main general comment is that the authors seem to prematurely dismiss volcanos as being important, even though figure 3 indicates a pattern correlation as strong in the multi-model mean as the other runs. The authors dismiss this high correlation because the amplitude of the signal is too-weak, but then again all of the amplitudes are too weak (though not quite as much as hist-volc) and issues with signal-to-noise are raised as a possible factor behind generically too-weak signals. While I can accept that the models likely do have S2N issues, this just begs the question of why they then rule out hist-volc as being important. If we are to inflate variances or bias correct, then perhaps hist-volc will turn out to be important after all?

Thank you for this comment. We agree that high correlation from volcanoes has been overlooked in the current manuscript despite its high pattern correlation (and based on new analysis also high temporal correlation). In the revised manuscript (and following new results) we will make sure volcanoes are not excluded from the results discussion.

I think the piece of analysis that is most missing from the paper in its current form isn't anything that reviewer 2 requested, but rather a proper quantification of whether these models have S2N issues in summer. Most of the S2N paradox literature concerns winter, though there is some on summer too (Dunstone et al 2023) but with a focus on the NAO and not on the wave-5 pattern. Please do a formal analysis of the RPC in these models! I suggest following the methodology of Hardimann et al 2022 for the RPC definition, though there are others out there that would also work ok.

Thank you for pointing this out. We agree that a formal S2N analysis would add context to our findings and give answer to a knowledge gap in current literature. In order to quantify the S2N issues we plan to do the following (we are already working on it):

1. Make an index for the wave-5 pattern to reduce the dimensionality and be able to apply current methods in the literature, which use a single time series to do the analysis. We plan to design the index as follows: Add the Z200_az values around the positive centers of the trends and subtract the Z200_az values around the negative centers for each season. The specific locations of these centers are based on the 30-60°N meridional mean local maximas and minimas along the longitude dimension, with the specific areas being a 20°x20° box centered at each of the centers in longitude and 45°N in latitude.



2. Compute the index for the entire period (1950-2014) after applying a 9yr moving average as a low pass filter to filter out interannual variability (similar methodology is employed in existing literature such as Smith et al 2020 or Smith et al. 2025), as LESFMIP are free-running simulations and we do not expect them to predict the specific interannual variability but rather decadal change.
3. Compute correlation and ratio of predictable components to assess if there is a signal-to-noise issue in the forced responses of the simulations for each of the different forcings.

Smith, Doug M., Adam A. Scaife, Rosie Eade, P. Athanasiadis, Alessio Bellucci, I. Bethke, Roberto Bilbao et al. "North Atlantic climate far more predictable than models imply." Nature 583, no. 7818 (2020): 796-800.

Smith, D. M., Dunstone, N. J., Eade, R., Hardiman, S. C., Hermanson, L., Scaife, A. A., & Seabrook, M. (2025). Mitigation needed to avoid unprecedented multi-decadal North Atlantic Oscillation magnitude. Nature Climate Change, 15(4), 403–410. https://doi.org/10.1038/s41558-025-02277-2

Assuming there are RPC issues, the next step would be to pick individual ensemble members that most closely the ensemble mean response in ~10-20 years subsamples of the record, and then plot the Z200asym maps for those individual ensemble members. [Note that the chosen members will change in each 10 year subset, and so you will be computing a trend over a varying set of members (See Smith et al 2020). Assuming the models have a reasonable amount of variance [which isn't guaranteed, but at least the bias likely won't be as big as possible issues with the signal], the signal in this timeseries and the accompanying trend should more closely match the amplitude of that in observations. This signal can be computed for both hist-vol and hist-aer, and I think this is the proper way to assess the relatively importance of the two forcings for the observed signal while also clarifying whether S2N issues are causing a too-weak response in the grand ensemble mean. Because you have large ensembles the "matching method" of Smith et al 2020 **should** work.

Thank you for the comment. Following the previous remark, we are working on computing the RPC to assess whether there are signal-to-noise issues. If there are, our next step will be to apply the correction described by the reviewer (and by Smith et al 2020) to the affected forcings to assess the relative importance of each of them.

Smith, Doug M., Adam A. Scaife, Rosie Eade, P. Athanasiadis, Alessio Bellucci, I. Bethke, Roberto Bilbao et al. "North Atlantic climate far more predictable than models imply." Nature 583, no. 7818 (2020): 796-800.