We thank the editor for their very helpful comments which will lead to improvements in the presentation of our results. We detail how we have responded to each of these comments below (in bold).

Figure 2:
* The small boxes with "NAO", "ENSO" and "SAM" are distracting in the way that the appear on the time line. Suggest instead moving these labels to the left side of each of the time line bars.

* for proxy SAM the December "D" indicator is missing off of the months given.

* I find the way that the black lines marking the transition from proxy to observed data continue into the white space to be distracting. Could these please be restricted just to within the time line box that they refer to.

We have made the changes as suggested and agree that this has improved the figure. Thank you.

Figure 3:
"Reconstructions (orange)" should be "Reconstructions (red)"

We have changed the caption accordingly (although note the reconstructions are actually plotted in brown).

* To what extent might the larger magnitude of reconstructed SAM changes compared to observations (as evident in the 3rd row of figure 3) be related to methodology in the way the observed annual average SAM index is calculated? The Marshall SAM index page gives an annual (Jan-Dec) SAM index that is calculated by first calculating annual mean zonal MSLP and then constructing the SAM. Because you are using a different set of months to calculate the annual SAM I suspect that your observed annual SAM is instead calculated by averaging across the monthly SAM index. This results in a large difference in the magnitude of the SAM index. The SAM index calculated by Marshall by first taking zonal mean annuals has a much larger (though theoretically dimensionless because of the normalisation step in calculating the SAM index) magnitude of variability than an annual SAM value calculated across the monthly SAM index. The Abram et al 2014 SAM index used that former version of annual SAM index to scale the proxy data, hence resulting in a large magnitude of variability in the proxy, which I don't think is comparable to the annual SAM index you use for the observations. See Nicky Wrights paper for details: https://cp.copernicus.org/articles/18/1509/2022/cp-18-1509-2022.html

Thank you for this very thoughtful suggestion. We have double checked the analysis and are confident that it is OK as it is. To be consistent we are actually plotting the annual SAM index (Jan-Dec) for everything (model simulations, observations and reconstructions). We appreciate this is unclear in the figure since this is not what we are assimilating during the final period and have updated the caption to make this clear.

We also have realised that the description of what SAM index was assimilated was also not clear. The metric initially assimilated was the annual mean SAM as

defined in Abram et al. After 1957 the SAM assimilated was the annual mean of the monthly values of the Marshall index (April - March). We have added to the main text so it now reads.

Given that monthly data exists, the mean of the period April through to March is used to coincide with the assimilation time-step, and the metric for this period is calculated as the mean of the monthly values.

We hope this is all clear now.

(let me know if you would like the rescaled version of the Abram 2014 SAM Index using annual average instrumental target calculated from monthly SAM data, as presented in Nicky's paper)

We do not think this is needed but we are grateful for your offer of help.

Figure 5:
On the SAM figure for "observed v actual index" (lower left panel) it seems odd to have Antarctica shaded grey with no data but SAT data showing over the whole of the Southern Ocean. Could ERA5 be used instead to provide coverage over Antarctica? Or could you combine HadCRUT5 with the spatial temperature dataset for Antarctica from Steig et al., 2009, Nature (doi:10.1038/nature07669)?

Note, I then also am confused as to how Figure 6 shows data using HadCRUT5 temperature correlations over Antarctica, whereas the Antarctic continent was shaded grey for missing data in Figure 5?

Thank you again, for an excellent observation. The difference is due to an inconsistency in our analysis. The figures used different versions of the HadCRUT5 dataset – Fig 5 used the infilled dataset which has more coverage and Fig 6 the un-infilled version of the dataset. We have updated figure 6 to use the infilled dataset and have updated all relevant figure captions to make it clear which version was used in each.

Figure 6:
I think that it would be helpful to add % signs after each of the numbers given at the top right of each panel.

Agreed – we have updated this figure to include this suggestion, as well as all relevant figures in the supplement.

Figure 7 b-f:
It would be clearer if the zonal mean labels gave degree symbols and N or S (e.g. "75°S to 45°S" rather than "-75 to -45")
Figure 8:
y-axis labels: remove negative signs as you indicate hemisphere already with N and S.

Agreed – we have made these changes. Thank you.