Thanks to both reviewers for valuable input that significantly improved the manuscript.

Reviewer 1:

General Comments

• The introduction is notably short and does not give any indication of how the rest of the paper is structured. The authors may wish to consider adding this? **Added.**

• I think the authors could consider discussing their results in the context of what other complimentary approaches would yield, which would make for a more well-rounded paper. Examples could include statistical and dynamical downscaling (using RCMs driven by reanalysis, such as available from CORDEX). Particularly, the limitation of RECON at a localized scale could be addressed via these avenues? **A paragraph has been added to the introduction to highlight the contrasts between RECON and global analyses and a comment about the need for statistical downscaling for regions of complex terrain has been added to the last section.**

• I do note a major omission that there is no code provided in generating the NetCDF file of the RECON output. This would be good to provide in ensuring the dataset has been generated robustly and consistently with that specified in the paper. **Now provided at https://github.com/shwang-met/Antarctic_Recon**

Specific Comments

L13-15: "It is based on monthly mean 2-m temperatures at 15 fixed stations that are spatially extrapolated to the entire continent using weights derived from the European Centre for Medium-Range Weather Forecasts 5th generation reanalysis (ERA5)". → What is the grid resolution of the derived RECON product (equal to ERA5 at 0.25°?). I found this detail lacking and think such detail should be added here explicitly. **Added.**

L23: "For those regions of Earth that are remote and sparsely populated, establishing their temperature history from direct observations can be a major challenge". → See general comment. Are there other examples from the literature where other similar approaches have been used to help overcome this. It would be good I think for the reader to have a sense of where and on what scale such approaches have been executed before and the relative degree of success, measured in terms of independent, observation-based validation.

What were the limitations, and have they been factored into the choices made in the authors' study? **See General Comments number 2.**

Table 1: I find some of this information on data sources used for infilling to be hard to follow. For instance, the GHCN QCF is considered under 'Other Observations' and GHCN QFE is provided as a separate column. Also, how is data made available from the other sources that is not present in MET READER (I thought this data source includes at least some of these such as University of Wisconsin-Madison)? **Some more reliable data sets than provided by READER are used. We have clarified the contrast between GHCN QCF and GHCN QFE in the text.**

L144: "Overall, the skill statistics for the current reconstruction dominantly for 1979-2022…"

→ I am not clear what this sentence is conveying. I find the use the word 'dominantly' troubling. **Reworded.**

L152-154: Is it maybe a little surprising that the average R2 for independent is a little lower versus Nicolas and Bromwich (2014)? Perhaps an additional sentence could suggest why this is (is it just a longer timeframe considered or inclusion of a few more 'problematic sites'). **Material added indicating that the longer period may be responsible because of greater recent atmospheric variability.**

Figure 2 Caption: The ERA5 and RECON panels for the annual trends are reversed with respect to the figure caption. The ordering of the seasonal trends is not mentioned for ERA5 and RECON, so this also needs adding. **Corrected.**

Technical Corrections

L59: Double spacing after '(ERA5; Hersbach et al., 2020)'. **Fixed.**

L139: Full stop missing after 'automatic weather stations'. **Fixed.**

L193: Double spacing after '2000.' **Fixed.**

L235: 'longtime scales' → 'long timescales'. **Changed to "long-time" because noun is "scales".**

Reviewer 2:

Major comments

- The paper does not discuss the source data in any detail, except for the newly-added Belgrano station. A reader may assume that the remaining 14 stations are taken directly from station data but this is not the case (for example, the 2014 paper discusses the Byrd reconstruction at some

length). The paper should be clear about the input station data sets and discuss any potential homogeneity issues with them. (To give one example with which I am familiar, the 'Casey' record is presumably a composite of Casey with pre-1969 Wilkes and it is unclear whether any potential inhomogeneities with that site move have been considered). **More details about the station records have been added. All time series have been checked for evidence of discontinuities, but none have been identified.**

- From the results in Table 3 and Figure 2, it appears that recent warming in ERA5 is substantially greater than in the reconstruction. This is an interesting result and I think is worth more discussion that it gets. It may also be of interest to compare warming rates in the reconstruction with the Antarctic component of major global temperature data sets (e.g. HadCRUT5, NOAAGlobalTemp, GISTEMP, Berkeley Earth). **ESSD is a data publication. This topic will be explored elsewhere in a research publication.**

Other comments

- L29 – 'collected initially for weather forecasting purposes' – this is common for historical climate records everywhere, is there anything specific to the Antarctic which requires elaboration here?

  Generally, weather record keeping in Antarctica have not been a high priority with many examples of less care taken than in more temperate climates. The reference by Lazzara et al. (2012) for South Pole has been provided as one example.

- L60-61 – although Gossart et al (2019) found that ERA5 was in general the best-performing reanalysis for temperature over Antarctica, they did find that it did have a warm bias in the cold season over the interior (although less than CFSR). This should be mentioned somewhere; does it have any implications for the results presented here? **Added. This ERA5 winter shortcoming does not seem to impact the trends it depicts.**

- L73-75 – this implies that the pre-1980 'Belgrano' data are in fact a reconstruction from elsewhere – is this correct? This could be made clearer, and it would be useful to get an indication of how far away the data being used in the reconstruction are. **Added.**

- Figure 1 – I think it would be useful to show the location of the Halley site (perhaps in a different colour) so readers can be aware of how Belgrano replaces it. **Provided.**

- L129 - ‘The significantly smaller correlation for Orcadas’ – presumably the fact that it’s an island (and the surrounding oceans are free of sea ice for a significant part of the year) is also relevant here? It’s also surprising to me that the R2 metric in Table 2 is very high for Orcadas when it performs less well on the other metrics, is this worth comment? **Error found in the calculations. Now Orcadas performs like the other anchor sites.**
- Figure 2 – the caption says ERA5 is on the left and RECON on the right but the label on the figures themselves is the other way round. **Fixed.**
- L237-247 – this paragraph is more of a discussion than a conclusion, perhaps the section header could be changed? **Changed.**

ESSD Chief Editor (Ice)

There are some usability issues with your NetCDF files.

Resolved.