Review of "An Updated Reconstruction of Antarctic Near-Surface Air Temperatures at Monthly Intervals Since 1958" by David Bromwich et al.

Overall Assessment

The study describes a monthly reconstruction product of Antarctic near-surface air temperature from 1958 to near present-day, which the authors refer to as 'RECON'. It is not entirely novel as it builds upon a similar product developed in an earlier study (Nicolas and Bromwich, 2014), although it appears to be a clear advance on this previous iteration. Broadly, it leverages the few (15) long-term, reliable records that exist (with infilling of data gaps first), in conjunction with spatial extrapolation according to a kriging method, using weights derived from the ERA5 reanalysis. A key difference is that CFSR reanalysis was used instead in Nicolas and Bromwich (2014). These weights are essentially the square of the correlation coefficients between the anomaly of monthly temperatures at each station (after linear detrending) and the monthly anomalies according to ERA5 at each grid point across the continent over the 1981-2020 period. Validation of the algorithm is performed at the locations of the long-term stations, revealing the very high correspondence and close agreement compared to the observations, with moderately strong correlations according to a series of other shorter record stations and automatic weather stations that are completely independent.

The importance of RECON is clear as long-term, reliable records across Antarctica from in situ observations are sparse and reanalyses such as ERA5 are not considered particularly reliable prior to 1979 (which marks the start of the satellite era). Even thereafter, artefacts or discontinuities exist due to several reasons (e.g., temporal irregularities in satellite products assimilated), so the comprehensive coverage of ERA5 cannot always be relied upon for accurate monitoring of temperature trends and variability. The RECON dataset convincingly provides a more accurate picture of long-term monthly air temperature trends, particularly on a more regional scale. However, it is likely that RECON does not necessarily improve our understanding at more local scales, as evidence by closer agreement between observations and ERA5 in some cases, highlighting an important caveat that warrants further attention (perhaps in a follow-up study or using other complementary approaches).

I judge that this paper warrants prompt publication in ESSD and only have a few comments that the authors may wish to consider before acceptance, which I hope they find helpful.

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?

- 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 wellrounded 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?

- 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.

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.

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?

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)?

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.

L152-154: Is it maybe a little surprising that the average R^2 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').

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.

Technical Corrections

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

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

L193: Double spacing after '2000.'

L235: 'longtime scales' → 'long timescales'.