International Journal of Climatology
RMetS Royal Meteorological Society

An Assessment of Early 20th Century Antarctic Pressure Reconstructions using Historical Observations

Journal:	*International Journal of Climatology*
Manuscript ID	JOC-19-0857.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	26-May-2020
Complete List of Authors:	Fogt, Ryan; Ohio University, Geography Belak, Connor; Ohio University, Geography; Purdue University, Department of Earth, Atmospheric, and Planetary Sciences Jones, Julie; University of Sheffield, Department of Geography Slivinski, Laura; University of Colorado Boulder, Cooperative Institute for Research in Environmental Sciences; NOAA, Physical Sciences Laboratory Compo, Gilbert; University of Colorado Boulder, Cooperative Institute for Research in Environmental Sciences; NOAA, Physical Sciences Laboratory
Keywords:	Climate < 6. Application/context, Pressure, Data recovery
Country Keywords:	Antarctica

An Assessment of Early 20th Century Antarctic Pressure Reconstructions using Historical Observations

Short title: An assessment of early 20th century Antarctic pressure reconstructions

Ryan L. Fogt[1], Connor P. Belak[1], Julie M. Jones[2], Laura C. Slivinski[3,4], Gilbert P. Compo[3,4]

[1]Department of Geography and Scalia Laboratory for Atmospheric Analysis, Ohio University, Athens, OH
[2]Department of Geography, University of Sheffield, Sheffield, UK
[3] University of Colorado, Cooperative Institute for Research in Environmental Sciences, Boulder, CO
[4]NOAA Physical Sciences Laboratory, Boulder, CO

Corresponding author address: Ryan L. Fogt, Ohio University Department of Geography, 122 Clippinger Laboratories, Athens OH 45701 ph: 740-593-1151 fax: 740-593-1149
email: fogtr@ohio.edu

Keywords: Antarctica, climate, pressure, data recovery

Abstract

While gridded seasonal pressure reconstructions poleward of 60°S extending back to 1905 have been recently completed, their skill has not been assessed prior to 1958. To provide a more thorough evaluation of the skill and performance in the early 20th century, these reconstructions are compared to other gridded datasets, historical data from early Antarctic expeditions, ship records, and temporary bases.

Overall, the comparison confirms that the reconstruction uncertainty of 2-4 hPa (evaluated after 1979) over the Southern Ocean is a valid estimate of the reconstruction error in the early 20th century. Over the interior and near the coast of Antarctica, direct comparisons with historical data are challenged by elevation-based reductions to sea level pressure. In a few cases, a simple linear adjustment of the reconstruction to sea level matches the historical data well, but in other cases, the differences remain greater than 10 hPa. Despite these large errors, comparisons with continuous multi-season observations demonstrate that aspects of the interannual variability are often still captured, suggesting that the reconstructions have skill representing variations on this timescale, even if it is difficult to determine how well they capture the mean pressure at these higher elevations. Additional comparisons with various 20th century reanalysis products demonstrate the value of assimilating the historical observations in these datasets, which acts to substantially reduce the reanalysis ensemble spread, and bring the reanalysis ensemble mean within the reconstruction and observational uncertainty.

1. Introduction

Despite recent dramatic climate-related changes in Antarctica, including warming of West Antarctica (Steig *et al.*, 2009; Bromwich *et al.*, 2012; Jones *et al.*, 2019; Turner *et al.*, 2019), retreat of several marine-based glaciers in the Amundsen Sea embayment (Rignot *et al.*, 2013, 2019; Edwards *et al.*, 2019), and rapid sea ice loss beginning in austral spring 2016 (Stuecker *et al.*, 2017; Turner *et al.*, 2017; Purich and England, 2019), there are significant challenges to understanding how unique these events are in a historical context or how likely they are to change in the future. This is in part due to the high degree of natural climate variability across Antarctica and the fully coupled nature of many of these changes occurring at the atmosphere-ice-ocean interface. When combined with the short observational climate records primarily beginning around the international Geophysical Year (IGY) activities of 1957/1958, interpreting these changes in the Antarctic climate has proven to be very difficult (Jones *et al.*, 2016).

Data from ice cores across Antarctica help to place ongoing change in a longer context (Bracegirdle *et al.*, 2019), especially through coordinated efforts like the International Trans-Antarctic Scientific Expedition (ITASE; Mayewski *et al.*, 2005). Ice core evidence has helped to compare recent Antarctic warming with variability over the past 2000 years (Stenni *et al.*, 2017) as well as changes in West Antarctica snowfall (Thomas *et al.*, 2008), and when combined with a climate model, longer estimates of Antarctic surface mass balance (Agosta *et al.*, 2019), as examples. Nonetheless, dating of specific events and the suppressed sub-annual temporal resolution in many cores pose challenges for these longer-term estimates of Antarctic climate variability.

Other tools to examine historical Antarctic climate variability during the last century include gridded datasets that span the 20th century, such as existing historical reanalyses that only assimilate surface pressure, like the National Oceanic and Atmospheric Administration/Cooperative Institute for Research in Environmental Sciences (NOAA-CIRES) Twentieth Century Reanalysis version 2c (hereafter 20CRv2c, Compo *et al.*, 2011) and the new NOAA-CIRES-DOE version 3 (20CRv3, Slivinski *et al.*, 2019). Similar “sparse-input” reanalyses have been completed by the European Centre for Medium Range Weather Forecasts (ECMWF), including their 20th-century reanalysis (ERA-20C, Poli *et al.*, 2016) and a coupled ocean-atmosphere reanalysis of the 20th century (CERA-20C, Laloyaux *et al.*, 2018), each of which assimilates marine wind observations and surface pressure. As a coupled reanalysis, CERA-20C also assimilates subsurface ocean temperature and salinity observations (Laloyaux *et al.*, 2018). While these are spatially and temporally complete throughout the 20th century, unsurprisingly they are sensitive to the number of assimilated surface pressure observations, and therefore the skill changes considerably over the high southern latitudes throughout the 20th century (Schneider and Fogt, 2018; Fogt *et al.*, 2019; Slivinski *et al.*, 2019).

An alternative approach to address these challenges is to generate seasonal pressure reconstructions, both at individual Antarctic stations (Fogt *et al.*, 2016a, 2016b), as well as spatially complete poleward of 60°S (Fogt *et al.*, 2017a, 2019). When compared to gridded climate datasets, these reconstructions showed far less sensitivity to changes in the number of observations across Antarctica (Schneider and Fogt, 2018), which is perhaps not surprising given that the reconstructions were based on statistical relationships with a spatially and temporally fixed network of stations in the mid and high latitudes of the Southern Hemisphere (Fogt *et al.*, 2016a). However, the skill of the reconstructions was only thoroughly evaluated after the IGY, (1957-1958), when most Antarctic observations began; spatially complete comparisons in reconstruction skill were even more limited and only possible after 1979 through evaluating against the ECMWF Interim reanalysis, ERA-Interim (ERA-Int; Dee *et al.*, 2011). Therefore, the skill of these reconstructions in the early 20th century before the IGY remains unknown, although the underlying relationships governing them are likely stationary (Clark and Fogt, 2019). This work aims to provide a more complete evaluation of these reconstructions poleward of 60°S by using available pressure observations that are independent from the reconstructions in the early 20th century.

2. Data and Methods

We make use of the best performing seasonal spatially-complete Antarctic pressure reconstructions, which are available from 1905-2013 and based on surface pressure anomalies from the 1981-2010 climatology from ERA-Int (Fogt *et al.*, 2019). The reconstructions are based on a kriging interpolation of 18 Antarctic station reconstructions and observations at Orcadas (locations plotted in bottom right panel of Fig. 1) to an 80km x 80 km Cartesian grid centered over the South Pole. For comparison here, this grid has been converted to a 0.75°x0.75° latitude-longitude grid. Importantly, comparisons with ERA-Int after 1979 (Fogt *et al.*, 2017a, 2019) demonstrated that the reconstruction skill varies seasonally, with the highest skill (defined as the best agreement with ERA-Int) in austral summer (December – February, DJF), and the lowest skill in the austral autumn and spring (March-May, MAM, and September-November, SON, respectively). In general, higher reconstruction skill is found over the Antarctic continent (especially near the Antarctic Peninsula), and lower skill at the northern edge of the domain near 60°S, especially in the South Pacific (Fogt *et al.*, 2019).

Historical pressure data prior to 1957 poleward of 60°S were extracted from three primary datasets to further evaluate the reconstruction skill in the early 20th century. The largest source of pressure data comes from the International Surface Pressure Databank version 3.2.9 (ISPD; Cram *et al.*, 2015), which contains subdaily measurements of mean sea level and/or surface pressure from both ships and temporary Antarctic bases (during field expeditions) and early Antarctic stations that began prior to the IGY. Additional sea level and/or surface pressure data were obtained from the International Comprehensive Ocean-Atmosphere Data Set release 3 (ICOADS; Freeman *et al.*, 2017); we only make use of ship records in ICOADS that were not available in ISPD. As many new historical observations from both ships and early Antarctic expeditions are still being recovered and digitized, we also make use of a few newly digitized additional pressure observations stemming from the Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative (Allan *et al.*, 2011), obtained directly from Rob Allan, the lead of the ACRE project. Only those ACRE pressure observations that were not part of the current releases of ISPD and ICOADS are utilized here, particularly from Antarctic expeditions in the first few decades of the 20th century. Most historical observations reported both SLP and surface pressure; however, a few observations only reported SLP, and therefore the reconstruction is compared to historical SLP observations throughout. We directly use the historical observations as reported in ISPD, ICOADS, or in the ACRE data for SLP and surface pressure, and do not reduce any surface pressure observations to SLP.

In making comparisons, it is important to consider that historical observations can have their own error or bias for many reasons, potentially including incorrect reported location (latitude, longitude, or elevation); problems with the barometer; errors associated with adjustments to sea level pressure; or misspecified corrections to pressure measurements from temperature or gravity. Because measurement errors likely vary in time and space, it is difficult to fully quantify how much these errors contribute to the reported pressure observations (Kent and Berry, 2005). Reanalyses generally define instantaneous, random observation errors that depend on platform and/or time (Compo *et al.*, 2011; Poli *et al.*, 2016; Laloyaux *et al.*, 2018; Slivinski *et al.*, 2019), and analyses comparing these expected errors with background statistics suggest estimates of about 1.5-2 hPa in the early 20th century are reasonable (Slivinski *et al.*, 2019). However, Slivinski *et al*. (2019) and more recent comparisons (not shown) also suggest that there is a relatively large systematic error in observations in the Southern Hemisphere in the early 20th century. We therefore assume a conservative estimate of 2 hPa total error in the seasonally-average observation estimates (or an observation error variance of 4 hPa^2).

In order to compare the historical observations to the reconstructions, additional analysis was required. First, all historical pressure data were converted to hPa, the same units as the pressure reconstructions. Monthly means for all the historical observations were calculated if at least 75% of the days within a given month had at least one observation available, to ensure a reliable monthly pressure estimate. Seasonal means (following the traditional seasonal divisions, DJF, MAM, JJA (June – August), and SON) were calculated to compare to the seasonal pressure reconstructions if at least two monthly means existed using our 75% daily data threshold; data for the third month was added to calculate the seasonal mean if 2 months had met the 75% daily data threshold, and the third month had more than 50% of its daily data. This approach allows for comparison with the maximum amount of data possible. Comparisons of a single monthly mean historical observation (for cases when historical data only had one month meeting our threshold) with the seasonal mean pressure reconstruction produced larger differences (not shown), and so this paper only focuses on comparisons of at least two months of historical data to define a seasonal mean. Altogether, these constraints yielded 271 seasonal means from historical pressure observations prior to 1957. The majority (>75%) of these historical observations occur only in DJF (austral summer), although it is noted that there are several observations after 1940 complete for the entire year. For reference, the year of austral summer refers to the year of December throughout this study.

Two more steps were taken prior to making comparisons. First, seasonal mean locations for the historical records were calculated using the coordinates provided in the data archives. These locations varied little for the early Antarctic stations; for observations collected on moving ships, we calculated a seasonal mean location by year for comparison since the reconstruction data are only available seasonally, which limits comparisons at higher temporal and spatial frequency. Of the 130 seasonal means from the ship records, more than 75% of the subdaily pressure observations showed less than a 2 degree standard deviation in the latitude, and half of the ships' standard deviations were less than 15 degrees longitude. In terms of pressure, there was no significant difference between ships that had a standard deviation of more and less than 20 degrees longitude in the standard deviations of sub-daily pressure observations. We therefore suggest the error associated with the mean location is more strongly related to the number of strong storms a ship encountered, which are more unique to a specific location rather than the mean location of the ship. As a rough estimate of this error, the mean standard deviation of pressure from the subdaily ship observations, 3.45 hPa, can be used, although more than half of the ships have a pressure standard deviation below 3.0 hPa. This standard deviation of subdaily pressure within one season is much smaller than the interannual standard deviation of monthly pressure across the South Pacific, which ranges from 3-7 hPa based on ERA5 data (contours in Fig. 1, bottom right panel).

Lastly, to compare the pressure reconstructions to the historical pressure observations, elevation adjustments were needed. The pressure reconstructions, originally constructed as surface pressure anomalies based on the underlying topography of ERA-Int at 0.75°x0.75° latitude / longitude resolution, were adjusted to sea level pressure using a rough linear estimate of 12 hPa pressure decrease per 100 m elevation gain. The 12 hPa approximation is larger than the 10 hPa per 100m assumed in the middle latitudes in the lower troposphere given the colder and drier (and therefore denser) air commonly observed poleward of 60°S. Without simultaneous measurements of temperature and humidity, it is challenging to provide a more accurate adjustment to sea level pressure. Furthermore, reduction to sea level pressure is known to be problematic in nearly all gridded climate datasets (and observations themselves) over Antarctica, including 20CRv3 (Slivinski *et al.*, 2019). This adjustment to sea level was necessary even for historical observations that had both surface and sea level pressure data, as due to the smoothing of the topography of ERA-Int at even the 0.75° resolution, model elevations and observed elevations (where known – this value is not always given in the historical data) are often quite different; these elevation differences similarly make it challenging to adequately compare surface pressure observations with the surface pressure (anomaly) reconstructions. Importantly, elevation differences can also arise due to the smoothing of the steep Antarctic coastline in ERA-Int, and therefore influence comparisons from ships close to the Antarctic continent. This adjustment to sea level introduces the greatest error and uncertainty in our comparison, as will be discussed in detail throughout. Nonetheless, the simple linear adjustment allows us to evaluate interannual (or intra-annual) variability and provide further information on reconstruction performance than otherwise would be possible.

In the comparisons, the reconstruction data are extracted from the gridpoint closest to the seasonal mean location in the reconstruction field. To provide further comparison for select observations that span more than one season, from the reanalyses we also extract both surface and sea level pressure data from the closest gridpoint using the ensemble mean in all but ERA-20C (which has only one estimate). This approach also introduces some error, called the 'error of representativeness', reflecting how well a reanalysis gridpoint represents the point location from the observation; this error is assumed to be one of the largest components of observation errors in the reanalyses, at least in 20CRv3 (Slivinski *et al.*, 2019), and is similar to the elevation-based errors for the reconstruction evaluations described previously. Importantly, these reanalyses are not independent of the observations used for comparison here: ERA-20C and CERA-20C assimilate surface pressure and marine surface wind observations from ISPD version 3.2.6 and ICOADS version 2.5.1 (Poli *et al.*, 2016; Laloyaux *et al.*, 2018); 20CRv2c assimilates pressure from ISPD version 3.2.9 and ICOADS version 2.5.P; and 20CRv3 assimilates pressure from ISPD version 4.7 and ICOADS version 3+v2 (Slivinski *et al.*, 2019). Since the comparisons are made with ISPD v3.2.9 and ICOADS v3, the observations were at least available to be assimilated in both version of 20CR, and likely for ERA-20C and CERA-20C. However, the assimilation combines the instantaneous observations with a background guess from the forecast model, so fields from different reanalyses are expected to differ from each other (as they each use different forecast models and assimilation algorithms), as well as from the seasonally-averaged observation estimates analyzed here. In addition, each reanalysis system has its own quality control algorithm that will blacklist observations deemed unfit for assimilation; thus, even if a given set of observations are available to the assimilation algorithm, they may not all have been assimilated. Finally, the reanalyses each employ their own observation bias correction scheme, which removes significant, consistent differences between the observations and the background fields. For simplicity, we only consider the uncorrected observations here, but note that systematic differences between the reanalyses and observations in our comparison may have been ameliorated by taking into account the observation bias corrections calculated within each reanalysis.

Reconstruction performance compared to the historical observations is evaluated using three primary statistics: overall bias, defined as the mean difference between the seasonal historical observation estimates and reconstructions at sea level (reconstruction minus observations); the mean absolute error (MAE), defined as the mean absolute difference to remove offsetting effects of positive and negative difference; and the root mean square difference (RMSD), which due to the squaring of the differences prior to calculating the mean, gives slightly higher weighting to larger absolute differences than the MAE. For observations that span only one season or have the bias of the same sign across all seasons, the absolute value of the bias and MAE will yield identical results. In all cases, these three statistics were calculated over the full length of each observational record. Further averaging both spatially and temporally is conducted to assess the overall reconstruction performance. Since the reconstruction error determined by Fogt et al. (2019) over 1979-2013 is assumed to be constant in time, it is assumed that there are no (temporally) correlated errors in the reconstruction; the evaluations in this paper help to determine the persistence of reconstruction errors through the early 20th century.

3. Results

3.1. Historical data availability

Prior to estimating the seasonal pressure reconstruction skill in the early 20th century, it is important to recognize the large changes in the number of observations poleward of 60°S, especially before 1957. Figure 1 displays the location and data type for all of the 271 seasonal means that were compared to the reconstruction; open circles represent seasonal mean positions of ship data while filled circles denote base or station records that were stationary or only had minor movements over time. As expected, there are very few observations prior to 1930, although there is a relative increase in the 1910-1919 during the height of the 'Age of Heroes' which includes expeditions to the South Pole, the Australian Antarctic Expeditions, and the British Imperial Trans-Antarctic Expedition led by Ernest Shackleton. The decrease in the number of observations in the 1940s is not only a challenge in the high southern latitudes but also worldwide, associated with the Second World War. Nearly half of the historical observations used for comparison here come from the years just prior to the IGY (1950-1957), with several early Antarctic stations established over the continent (especially along the Antarctic Peninsula), and frequent ships along the East Antarctic coastline. A notable gap in observational coverage is within the Weddell Sea (east of the Antarctic Peninsula) and throughout much of the South Atlantic. In addition to persistent sea ice in the Weddell Sea (Cavalieri and Parkinson, 2008), the South Atlantic area is far from commercial sailing or whaling routes and is likely one of the greatest spatial data voids globally (for example, see Fig. 1 of Allan and Ansell, 2006).

3.2. Overall reconstruction performance

With this temporal evolution of historical data in mind, the reconstruction performance is displayed as a bar chart in Fig. 2; the overall statistics (averaged over all available observations) are given at the bottom: mean bias = -0.79 hPa; mean MAE = 4.05 hPa; mean RMSD = 4.26 hPa. Recalling that the overall comparison is primarily during austral summer, these values are considerably higher than the skill assessed in earlier work in comparison to ERA-Int (Fogt *et al.*, 2017a, 2019), which estimated an MAE over the Antarctic continent generally less than 1.5 hPa and over the Southern Ocean around 2-3 hPa. Indeed, the values seem even higher than the MAE in Fogt *et al.* (2019) in the non-summer months in the South Pacific, the region of highest MAE compared to ERA-Int which ranged from 2-4 hPa.

This quick comparison initially suggests that the seasonal pressure construction performance is of much lower quality prior to 1957 than it is after 1957. However, when considering that the observation estimates have their own errors (assumed to be approximately ~2 hPa in the early 20th century), the higher MAE prior to 1957 is not far from the assumed observational error. Further, when looking at the average skill as a function of latitude (Fig. 2), the bias is near zero in both of the latitude bands that primarily lie over the Southern Ocean (60°-65°S and 60°-70°S), potentially suggesting that the historical observation estimates tend to fluctuate near the reconstruction seasonal mean overall (both above and below, cancelling to near zero). The MAE in these latitude bands, near 3 hPa, better reflects the overall performance as it removes the cancellation between positive and negative differences. The MAE of 3 hPa is consistent with the skill reflected in Fogt *et al.* (2019), and is also close to the assumed observation error of 2 hPa. In contrast, where Fogt *et al.* (2019) demonstrate lower MAE over the Antarctic continent (due to more station observations constraining the reconstruction skill), comparisons with the historical observation estimates clearly show an increase in MAE poleward of 65°S, which is primarily from stations over the Antarctic continent (poleward of 70°S). As noted earlier, this increase in MAE is an artifact of the elevation sensitivity when comparing the historical observations at (often unknown) elevations different from the seasonal pressure reconstruction based on the underlying topography of the ERA-Int 0.75°x0.75° grid. The simple linear offset of 12 hPa per 100 m creates sea level pressures that are consistently too low (evidenced by the negative bias), which gives rise to large MAE and RMSD over the Antarctic continent.

The sensitivity to elevation is also apparent in the comparisons as a function of longitude. There is a strong negative bias in the 330°-30°E range, along the mountainous Antarctic Peninsula. In ERA-Int, this terrain was greatly smoothed but still elevated, while the observations are often along the coast near sea level. In other longitude zones, the mean bias is typically in the range of ±1 hPa, with MAE around 3-4 hPa. Given that Fig. 1 indicates that outside of 330°-30°E, the seasonal mean observation estimates primarily are over the ocean and less influenced by elevation corrections, the comparison by longitude again reflects a similar skill of the reconstruction (over the Southern Ocean at least) as observed in Fogt *et al.* (2019) during 1979-2013. Comparisons by decade typically have MAE also in the 3-4 hPa range, consistent with the previous evaluation. One outlier is the 1910s, which from Fig. 1 has a relatively high percentage of historical observations on or near the Antarctic continent, suggesting that elevation corrections are again giving a misleading impression of relatively low reconstruction skill.

To further visualize the spatial and elevation dependence of reconstruction performance, Fig. 3 displays the decadal mean RMSD for each station by decade. Several key points emerge from Fig. 3. First, there is an indication from ships in the South Pacific north of the Amundsen and Bellingshausen Seas (roughly 60°-70°S, 150°-90°W) that the reconstruction skill is considerably lower, with RMSD values often above 6 hPa. Since these are seasonal pressure estimates from ships (Fig. 1), elevation correction is not an issue. Rather, the larger errors are consistent with the much lower reconstruction performance in this region of high interannual variability (which often exceed observational errors based on interannual monthly pressure standard deviations as large as 7 hPa in Fig. 1), as discussed in previous evaluations (Fogt *et al.*, 2017a, 2019). The larger interannual pressure variability here may also suggest that using the mean ship location could be introducing more error in this sector compared to other regions around Antarctica (Fig. 1). Second, for comparisons in the Southern Ocean for nearly all other regions, the RMSD is generally below 4 hPa, and the majority of the comparisons with the seasonal mean ship observation estimates show RMSD in the 3-4 hPa range. These values are consistent with the slightly lower performance of the reconstruction over the Southern Ocean. However, these values of RMSD still fall within expected error when considering errors in both observation estimates and the reconstruction (taken as the square root of the sum of the observational and reconstruction error variances). Lastly, it is clear that many of the higher errors (exceeding 6 hPa) are found along the coast, the Antarctic Peninsula, or in the interior of the continent. Even on the Ross Ice Shelf, RMSD values are typically larger than 4 hPa. As will be more clearly demonstrated later, these larger differences are due to elevation corrections to the historical observations in or near areas of high or vastly varying terrain. We will show that the reconstruction skill is likely much higher in these areas than suggested by the RMSD values in Fig. 3. Remarkably, in every decade since the 1930s, there are multiple locations where the RMSD values (over the ocean) are less than 1hPa, which is an excellent agreement with the historical observations. Given that we make comparisons with the seasonal mean ship position, that the historical observation estimates are rarely complete for the full season, and that they themselves have errors, this high agreement is noteworthy. It suggests that the reconstruction can be used as a reliable approximation of Antarctic pressure in most regions on and around the continent back to at least the 1930s.

3.3 Reconstruction performance evaluated with selected historical observations

The reconstruction performance is perhaps best evaluated by comparing to individual historical observations, which is the focus for the remainder of the study. To facilitate these assessments, a subset of representative stations and ship seasonal locations has been selected; their average location, name, and mean RMSD values over the entire record length are plotted in Fig. 4. We have purposely selected historical pressure estimates with longer records than a single season, and as many complete records from earlier portions of the 20th century as possible, although the sample size is quite limited (Fig. 1). In general, the reconstruction performance indicated by the mean RMSD in Fig. 4 at these locations is comparable to the overall decadal mean performance in Fig. 3, although we do not closely examine any stations with RMSD less than 2 hPa, and only one station where inadequate elevation corrections challenge the assessment of reconstruction skill (Cape Denison). Recall, all of these data were available for assimilation into 20CRv2c and 20CRv3 (and likely available for CERA-20C and ERA-20C), so the comparisons with the reanalyses are not necessarily always independent as discussed in section 2.

Figure 5 displays three seasonally-averaged historical observation estimates where the mean RMSD values are 2-3 hPa, consistent with the reconstruction skill and uncertainty assessed in Fogt *et al.* (2019), and within the assumed observational error variance of 4 hPa^2. Plotted with the historical observations (in red) are the mean sea level pressure (solid) and surface pressure (dashed, when available) from the nearest gridpoint of 20CRv2c (green), 20CRv3 (orange), ERA-20C (purple), and CERA-20C (blue). To compare with the skill evaluation in Fogt *et al.* (2019), the correlations (if more than 10 values are available) and MAE (in hPa) for each gridded dataset are given at the bottom of each panel: the first number is based on the MSLP, and the second number (where available) is based on surface pressure.

Station 889340 (from the ISPD archive) is situated along the Antarctic Peninsula (Fig. 4) and has a continuous pressure record beginning in 1948 for all seasons. As such, it is one of many such stations that provide a useful evaluation of reconstruction skill, and the overall MAE from the reconstruction is less than 2 hPa. Moreover, the temporal variability is well captured ($r > 0.80$ in all datasets). The various reanalysis products have similar MAE (generally from 1-2 hPa), well within the likely observational uncertainty, with CERA-20C having the lowest for MSLP, and 20CRv2c having the lowest for surface pressure. Despite the low MAE / RMSD at this station, adjustment of the reconstruction to sea level pressure may play a small role in its performance, as there are differences in both the historical MSLP and surface pressure, as well as the MSLP and surface pressure in the reanalyses. Nonetheless, this station shows the viability of the reconstruction shortly before the IGY along the Antarctic Peninsula, a region of relatively higher reconstruction skill compared to ERA-Int after 1979 (Fogt *et al.*, 2017a, 2019).

Deck 215 (Fig. 5b) from ICOADS has adequate observational coverage only during austral summer. The reconstruction skill is comparable to the skill seen in the southern Indian Ocean during much of the 1930s (Fig. 3). The reconstruction MAE is 2.5 hPa, slightly higher than all reanalyses but ERA-20C, which was found to be one of the lower performing reanalyses in the early 20th century near Antarctica (Schneider and Fogt, 2018). Despite the higher MAE for the reconstruction (which, unlike the reanalysis products, is entirely independent from these historical observations), the historical observation values nearly all fall within the reconstruction uncertainty (95% confidence interval, gray shading), and the reconstruction uncertainty would clearly overlap with the observational uncertainty throughout time. Together, the comparison with Deck 215 data suggests the original assessment of the reconstruction skill provides a good approximation of its performance in the early 20th century. Although the interannual variability is not as strongly captured as in Fig. 5a (correlation is not provided since only six years of data exist), it should be noted that the reanalysis datasets also do not reproduce the interannual variability well (perhaps due to observational errors or the spatial averaging). We also observe that the surface pressure in 20CRv2c is notably different than the surface pressure from the other reanalyses, probably due to the spectral effects in the 20CRv2c elevation field noted by Slivinski et al. (2019).

For a longer observational record over the Southern Ocean, Deck 899 (Fig. 4, from ICOADS) also has pressure observations only for austral summer but over multiple decades (Fig. 5c). The location varied during these many voyages, but overall the observational estimate of the seasonal mean falls within or very near the reconstruction uncertainty. Indeed, the reconstruction agrees better with the seasonally-averaged ship-based values than the reanalyses (MAE of 2.23 hPa compared to 2.3 – 3.7 hPa from reanalyses, although almost all datasets would fall within the observational uncertainty at this location). The correlations are much lower for these ship data across all datasets, due to differing fluctuations between a few years (i.e., 1933-1934, 1936-1937, 1945-1946), however the interannual variability is better captured after 1946. In general, the reconstruction also aligns well with the MSLP from the reanalyses, except for DJF 1934 when the ship traversed a large span of the southern Indian Ocean (from 88°E in

December 1934 to 47°E in February 1935). Therefore, the use of the seasonal mean location could be artificially suggesting a lower reconstruction skill for this observational estimate. Nonetheless, as in Figs. 5a and 5b, this comparison further suggests that the uncertainty of the reconstruction is a good estimate of the reconstruction error in the early 20th century. Overall, the reconstruction agrees well with many historical seasonal-mean observation values that were withheld during the reconstruction's development. As in Fig. 5b, we note that there are larger differences in the surface pressure from 20CRv2c which clearly fall outside the uncertainty from using a seasonal mean ship location.

Figure 5 presents several comparisons where elevation corrections did not have a noticeable influence on the evaluation of the reconstruction and are perhaps more representative of the overall reconstruction quality. In many other locations, elevation adjustments appear to play a more important role (especially on or near the Antarctic continent) in comparing the reconstruction and observational estimates. At a few of these locations, a simple linear adjustment of 12 hPa per 100m from the ERA-Int elevation at the reconstruction's gridpoint proved sufficient to readily compare it to the historical values. A subset of these stations is presented in Fig. 6.

Perhaps the most famous early American expeditions to Antarctica were those led by Admiral Richard E. Byrd, who set up a station called Little America in several different field campaigns spanning three decades (Byrd, 2003). Although the mean location of Little America was on the Ross Ice Shelf (it varied negligibly during each campaign in comparison to the resolution of the gridded datasets used in this study; Fig. 4), there were notable differences in observed MSLP and surface pressure at the location (Fig. 6a). When adjusting the reconstruction to sea level, a good match is observed both in overall mean pressure (MAE = 3.63 hPa) and interannual variability. Surprisingly, the reconstruction performs much better than the MSLP from 20CRv2c or 20CRv3, most notably due to the much higher MSLP in these two datasets from the 1940s onward compared to the reconstruction and observed values. However, the 20CRv2c surface pressure nearly perfectly matches the observed surface pressure (an MAE of only 2.39 hPa), while 20CRv3 has very similar values for surface and sea level pressure observations. We note that although there are significant offsets from the observed values for most of the datasets (MAE > 6 hPa for both), the interannual variability is well-captured, with the reconstruction correlation of 0.85, and surface pressure correlations from the reanalyses all above 0.85. The consistent difference between the observations and 20CRv3 could be due to elevation issues or to a detected bias in the observations; this will be examined in more detail later. CERA-20C, as seen earlier, performs with the highest skill for MSLP ($r=0.9$). Also note that the offset from the surface pressure to MSLP is not constant, as this difference, through the hypsometric equation, is influenced by both temperature and humidity. Therefore, some of the larger errors in the reconstruction and the reanalyses, particularly in the 1950s, could be due to too strong of an increase in the surface pressure as it was reduced to sea level, since the gap between surface pressure and sea level pressure is much lower in the historical observations during the 1950s. Observational uncertainties, including the various bias corrections employed by the reanalyses as discussed previously, can also create some of these differences between products when compared to the observations.

Much earlier in Antarctic history, under the leadership of Jean-Baptiste Charcot, the French conducted their second expedition that wintered over on Peterman Island in 1908-1910. The ship associated with this expedition was Porquoi Pas, with a mean location near the Antarctic Peninsula (Fig. 4). Due to the close proximity of the high terrain of the Antarctic

Peninsula, elevation adjustments to the reconstruction were necessary but may not have been sufficient at this location. From historical observations, the difference between surface pressure and sea level pressure ranged from nearly 20 hPa to as little as 5 hPa in DJF 1909 (Fig. 6b). Despite the challenges in correcting for the elevation differences, as seen in Fig. 6b, the adjustment brings the observed MSLP to the far upper-limits of the reconstruction uncertainty, and clearly within the observational uncertainty. Further, the reconstruction agrees well with the observational estimates: the MAE of 3.15 hPa is nearly consistent with time, and the seasonal cycle of the pressure during this portion of the early 20th century is well captured by the reconstruction. It is also encouraging to see high skill in both 20CR datasets, and again in CERA-20C at this location, all within the observational uncertainty; the lower performance of ERA-20C could be related to a cold bias or perhaps resolution issue, as there is a much larger difference between surface pressure (dashed line) and MSLP (solid line) in ERA-20C compared to the observations or CERA-20C. The differences in surface pressure between 20CRv3 and the observations are also noteworthy, with an MAE of 38.60, reflecting differences in model and observed orography.

Similarly situated near the Antarctic Peninsula (Fig. 4), Argentine Island provides a long continuous record much like in Fig. 5a, but with a noticeable influence of elevation-based pressure corrections (Fig. 6c). As with Porquoi Pas, the linear SLP adjustment brings the reconstruction close to the historical estimates, but they fall at the upper-bound of the reconstruction uncertainty, and are much closer to the reanalyses overall. All comparisons are within the expected observational uncertainty. The interannual variability is well-captured by the reconstruction (r= 0.82), although it may be slightly dampened in the early 1950s. Interestingly, ERA-20C surface pressure aligns closely with the historical observations (with an

MAE of 1.63 hPa), but this reanalysis again demonstrates the highest MAE compared to MSLP (4.88 hPa). In contrast, CERA-20C shows the lowest MAE based on MSLP (0.66 hPa), but considerably lower surface pressure than observed (an MAE of 18.29 hPa), though not as low as surface pressure in 20CRv3 (an MAE of 33.42). All reanalyses show very high correlations with observations for both MSLP and surface pressure (r > 0.85).

The long, albeit discontinuous record at the Little America station location (Fig. 4) provides an opportunity to further evaluate the reconstruction in comparison to the historical reanalyses, and importantly, to assess the reconstruction skill in light of the reanalyses' internally estimated uncertainties. For the reanalyses studied here, all but ERA-20 are ensemble reanalyses, and therefore the ensemble spread from the seasonal mean ensemble members can be further employed to assess the quality of these products through time and provide a more complete comparison of their skill relative to that of the reconstruction. The MSLP from the closest gridpoint of 20CRv2c, 20CRv3, and CERA-20C are plotted in Fig. 7, along with 95% confidence intervals calculated as 1.96 times the ensemble standard deviation of seasonally averaged MSLP from the 56, 80, and 10 ensemble members of 20CRv2c, 20CRv3, and CERA-20C, respectively. The performance varies slightly at neighboring gridpoints for MSLP, but varies considerably if using surface pressure, suggesting elevation gradients from the nearby Roosevelt Island or the edge of the Ross Ice Shelf can influence the reanalysis estimate at this location.

Similar to but slightly larger than the reconstruction, Fig. 7 shows that CERA-20C has a pronounced annual cycle of MSLP at Little America, while both 20CRv2C and 20CRv3 have a dampened annual cycle. Nonetheless, Fig. 7 also demonstrates that there is a marked decrease in the reanalyses' uncertainties at the times when the Little America data were assimilated, which effectively constrained the reanalysis estimates. There are also other reductions in the reanalysis uncertainties prior to the establishment of Little America, including the data assimilated from the South Pole race of 1911-1912 (the Norwegian base Framheim was very close to Little America (Fogt *et al.*, 2017b)) and the data from the first British Antarctic Expedition (1908-1909, which established a base near present day McMurdo station west of Little America, at Cape Royds). At these times, the reanalyses and reconstruction show considerable agreement, and the reanalysis uncertainty (colored shading in Fig. 7) overlaps the reconstruction uncertainty (gray shading in Fig. 7). Exceptions are in 1929 for all reanalyses and in 1911 for 20CRv2c and 20CRv3 when the reconstruction and reanalysis uncertainties do not overlap, with the reconstruction seasonal mean MSLP lower than the reanalyses. Nonetheless, the overall agreement at all other times suggests that although there may still be larger MAE in the reanalyses MSLP when compared only during the times of direct observations (as in Figs. 5-6), the observations and reconstruction both fall within the reanalyses uncertainty (estimated by the ensemble spread), and the reanalyses have benefited greatly from assimilating this historical data (moving much closer to, if not within, the observational uncertainty). Clearly, at other periods when data are not available, the reanalysis spread is considerably larger, but the reconstruction nearly always falls within the reanalyses' uncertainties (the high positive pressure values in CERA-20C in 1944 are one exception). We note that the reanalysis uncertainty estimates themselves require further improvement, particularly in the Southern Hemisphere: Slivinski et al (2019) show artificial signals in the uncertainty of 20CRv2c, and demonstrate that the uncertainty in this region still remains too large in 20CRv3. Conversely, Laloyaux *et al*. (2018) acknowledge that the small ensemble of CERA-20C can result in overly-confident estimates of uncertainty. Consistent with the comparison to observations (Fig. 6a), the CERA-20C MSLP and reconstruction (Fig. 7c) show the most similarity throughout the early 20th century at the Little America location (RMSD = 4.32 hPa), and for this location 20CRv3 performs better than 20CRv2c (RMSD of 5.95 hPa compared to 8.22 hPa).

Evaluation of the reconstruction skill in Figs. 2 and 3 often indicates substantial differences, suggesting that the seasonal mean values from the historical observations do not always fall within the reconstruction uncertainty when elevation corrections are applied, particularly near high terrain (even when their uncertainty is accounted for; not shown). To examine this issue further, Figure 8 depicts two cases – a set of Antarctic expedition ship records (Deck 246, Fig. 8a) and a temporary base (Cape Denison, Fig. 8b), where elevation corrections have mixed results; MAE values in Fig. 8 are only calculated for MSLP. Ships included in Deck 246 (from ICOADS) operated discontinuously near present day Dumont d'Urville station in the Ross Sea sector of Antarctica (Fig. 4); the data from ICOADS have two disparate locations in January – February 1912 (the mean of these would be over the continent), so instead of averaging data from both locations, each were treated as separate data points in Fig. 8a and the comparison statistics. The two different values for MAE in Fig. 8a are therefore based on the MSLP at these two different locations, with the data closer to the Ross Ice Shelf (second value) agreeing better with all gridded datasets than the data closer to the East Antarctic plateau, near the location plotted in Fig. 4. Even with this potentially conflicting information, the reconstruction performs very well after adjustment to sea level pressure, with an MAE of 2.32 / 1.85. A closer look shows that the majority of this error is from DJF 1910, when the reconstruction was about 6 hPa lower than the observational estimate, otherwise it is generally within 1.5 hPa. This performance exceeds the reanalyses' performance for Deck 246, as the reanalyses typically have much higher MSLP values.

In contrast, the famous base for the Australian Antarctic Expedition, Cape Denison (Mawson, 1998), situated along the East Antarctic coast south of Australia (Fig. 4) is marked with much lower performance across all datasets (Figure 8b). Elevation corrections and the models' orography have a strong effect here, clearly demonstrated by the large differences in surface and sea level pressure in observations and reanalyses (note, surface pressure is plotted on the right axis, but because of the large differences in surface pressure the MAE is not given). Whereas the seasonal means of the historical observations demonstrate surface pressures around 920 hPa on average, sea level pressure estimates from the station are around 70 hPa higher, at around 990 hPa. Such high differences in the two over the cold ice sheet plateau compromise the quality of the reduced sea level pressure (not only in observations, but across all datasets). The MAE for the reconstruction (13.50 hPa) is one of the highest of all locations examined. Although the reanalyses' errors are nearly half of this (around 5-7 hPa, likely improved by making further use of temperature and humidity calculated within the reanalyses to reduce surface pressure to sea level), the errors are still large, and the reanalyses and reconstructions all likely fall outside the observational uncertainty. Furthermore, it is difficult to assess how well the interannual variability of the MSLP is reproduced, since there are fewer than ten observations and the reduction to sea level varies considerably by season (affected by temperature and humidity), but the reconstruction was adjusted to sea level uniformly (and linearly) across all seasons. Nonetheless, the limited comparison demonstrates that overall there is still good agreement in the surface pressure interannual variability (despite large differences in magnitude) between the reanalyses and the reconstruction, which potentially suggests that the reconstruction is still capturing aspects of the pressure variability at this location in the early 20th century. Importantly, due to the influence of crude elevation adjustments, it is highly likely that the reconstruction is performing better than indicated by the MAE (Fig. 8b) or RMSD (Fig. 4). Unfortunately, the reconstruction skill is difficult to precisely determine at this or any other location (as suggested in reviewing Fig. 2) where large elevation adjustments make the comparison to historical estimates challenging.

4. Discussion and Conclusions

The analysis presented here has compared seasonal Antarctic pressure reconstructions to numerous historical observations and reanalyses throughout the early 20th century. As none of these historical data were included in the reconstruction calibration, they serve as an independent evaluation of the reconstruction skill during a period when relatively little is known about Antarctic climate variability.

The results overall confirm that the reconstruction error and uncertainty assessed in earlier work (Fogt *et al.*, 2017b, 2019), a mean absolute error of around 2-4 hPa across the Southern Ocean, is supported when comparing with ship observations. A few ship records suggest even higher reconstruction skill (MAE less than 2 hPa), while others situated north of the Amundsen and Bellingshausen Seas demonstrate lower skill (MAE above 4 hPa), consistent with earlier work. Furthermore, most comparisons with ship observations that span multiple seasons indicate the reconstruction also captures the interannual variability well (correlations often greater than 0.80). Comparison with historical reanalyses provide further evaluation of the reconstruction's performance, as they assimilate most of the historical observations used here but are independent of the reconstruction.

While the comparisons with observation estimates taken at sea level are relatively straightforward and further validate the reliability of the early portions of the reconstruction, it is far more challenging to make assessments of the reconstruction skill over areas of higher elevation or near the coastline. In many of these locations, reduction of the reconstruction pressure (which was constructed as surface pressure anomalies relative to the ERA-Int model topography) to sea level pressure using a simple linear adjustment does not satisfactorily agree with the historical data, even after considering the potential observational error. In many of these locations, there are also large differences between 20th century reanalysis MSLP and the historical observational estimates, highlighting the reduced reliability of MSLP from all sources over the cold, high Antarctic continent. While the general comparisons conducted here suggest a larger reconstruction MAE of nearly 6 hPa or more over high elevation, a closer examination at select locations reveals that the reconstruction uncertainty is likely much lower than this and perhaps as low as 1-3 hPa, as indicated in Fogt *et al.* (2019).

This work has demonstrated the value of digitizing historical observations from ships and temporary bases for both understanding long term change across the high southern latitudes and evaluating gridded datasets. While their temporary nature may make them difficult to use for assessing long-term variability and change, when coupled with gridded climate datasets like the seasonal Antarctic pressure reconstructions evaluated here, they serve as an independent and valuable tool of documenting historical climate. As one recent example, the use of newly digitized historical observations and pressure reconstructions shed new light on exceptional conditions during the South Pole race of 1911-1912 (Fogt *et al.*, 2017c, 2018; Sienicki, 2018). Future work will hopefully continue to unlock the power of these and other historical observations, so that the ongoing change across the high southern latitudes can be placed in a much-needed longer historical context (Jones *et al.*, 2016).

Acknowledgments

Data from the Antarctic pressure reconstructions are available from figshare (https://doi.org/10.6084/m9.figshare.5325541). RLF and CPB acknowledge support from the National Science Foundation (NSF), grant PLR-1341621 and ANT-1744998. JMJ acknowledges support from the Leverhulme Trust through a research Fellowship (RF-2018-183). Support for the Twentieth Century Reanalysis Project is provided by the U.S. Department of Energy, Office of Science Biological and Environmental Research, by the National Oceanic and Atmospheric Administration Climate Program Office, and by the NOAA Physical Sciences Laboratory.

References

Agosta C, Amory C, Kittel C, Orsi A, Favier V, Gallée H, van den Broeke MR, Lenaerts JTM, van Wessem JM, van de Berg WJ, Fettweis X. 2019. Estimation of the Antarctic surface mass balance using the regional climate model MAR (1979–2015) and identification of dominant processes. *The Cryosphere*, 13(1): 281–296. https://doi.org/10.5194/tc-13-281-2019.

Allan R, Ansell T. 2006. A new globally complete monthly historical gridded mean sea level pressure dataset (HadSLP2): 1850–2004. *Journal of Climate*, 19(22): 5816–5842.

Allan R, Brohan P, Compo GP, Stone R, Luterbacher J, Brönnimann S. 2011. The International Atmospheric Circulation Reconstructions over the Earth (ACRE) Initiative. *Bulletin of the American Meteorological Society*, 92(11): 1421–1425. https://doi.org/10.1175/2011BAMS3218.1.

Bracegirdle TJ, Colleoni F, Abram NJ, Bertler NAN, Dixon DA, England M, Favier V, Fogwill CJ, Fyfe JC, Goodwin I, Goosse H, Hobbs W, Jones JM, Keller ED, Khan AL, Phipps SJ, Raphael MN, Russell J, Sime L, Thomas ER, van den Broeke MR, Wainer I. 2019. Back to the Future: Using Long-Term Observational and Paleo-Proxy Reconstructions to Improve Model Projections of Antarctic Climate. *Geosciences*, 9(6): 255. https://doi.org/10.3390/geosciences9060255.

Bromwich DH, Nicolas JP, Monaghan AJ, Lazzara MA, Keller LM, Weidner GA, Wilson AB. 2012. Central West Antarctica among the most rapidly warming regions on Earth. *Nature Geoscience*, 6(2): 139–145. https://doi.org/10.1038/ngeo1671.

Byrd RE. 2003. *Alone: the classic polar adventure*. Island Press/Shearwater Books: Washington, DC.

Cavalieri DJ, Parkinson CL. 2008. Antarctic sea ice variability and trends, 1979–2006. *Journal of Geophysical Research*, 113(C7). https://doi.org/10.1029/2007JC004564.

Clark L, Fogt R. 2019. Southern Hemisphere Pressure Relationships during the 20th Century—Implications for Climate Reconstructions and Model Evaluation. *Geosciences*, 9(10): 413. https://doi.org/10.3390/geosciences9100413.

Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N, Allan RJ, Yin X, Gleason BE, Vose RS, Rutledge G, Bessemoulin P, Brönnimann S, Brunet M, Crouthamel RI, Grant AN, Groisman PY, Jones PD, Kruk MC, Kruger AC, Marshall GJ, Maugeri M, Mok HY, Nordli ø., Ross TF, Trigo RM, Wang XL, Woodruff SD, Worley SJ. 2011. The Twentieth Century Reanalysis Project. *Quarterly Journal of the Royal Meteorological Society*, 137(654): 1–28. https://doi.org/10.1002/qj.776.

Cram TA, Compo GP, Yin X, Allan RJ, McColl C, Vose RS, Whitaker JS, Matsui N, Ashcroft L, Auchmann R, Bessemoulin P, Brandsma T, Brohan P, Brunet M, Comeaux J, Crouthamel R, Gleason BE, Groisman PY, Hersbach H, Jones PD, Jonsson T, Jourdain S, Kelly G, Knapp KR, Kruger A, Kubota H, Lentini G, Lorrey A, Lott N, Lubker SJ, Luterbacher J, Marshall GJ, Maugeri

M, Mock CJ, Mok HY, Nordli O, Rodwell MJ, Ross TF, Schuster D, Srnec L, Valente MA, Vizi Z, Wang XL, Westcott N, Woollen JS, Worley SJ. 2015. The International Surface Pressure Databank version 2. *Geoscience Data Journal*, 2(1): 31–46. https://doi.org/10.1002/gdj3.25.

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette J-J, Park B-K, Peubey C, de Rosnay P, Tavolato C, Thépaut J-N, Vitart F. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656): 553–597. https://doi.org/10.1002/qj.828.

Edwards TL, Brandon MA, Durand G, Edwards NR, Golledge NR, Holden PB, Nias IJ, Payne AJ, Ritz C, Wernecke A. 2019. Revisiting Antarctic ice loss due to marine ice-cliff instability. *Nature*, 566(7742): 58–64. https://doi.org/10.1038/s41586-019-0901-4.

Fogt RL, Goergens CA, Jones JM, Schneider DP, Nicolas JP, Bromwich DH, Dusselier HE. 2017a. A twentieth century perspective on summer Antarctic pressure change and variability and contributions from tropical SSTs and ozone depletion. *Geophysical Research Letters*, 44(19): 9918–9927. https://doi.org/10.1002/2017GL075079.

Fogt RL, Goergens CA, Jones ME, Witte GA, Lee MY, Jones JM. 2016a. Antarctic station-based seasonal pressure reconstructions since 1905: 1. Reconstruction evaluation: Antarctic Pressure Evaluation. *Journal of Geophysical Research: Atmospheres*, 121(6): 2814–2835. https://doi.org/10.1002/2015JD024564.

Fogt RL, Jones JM, Goergens CA, Jones ME, Witte GA, Lee MY. 2016b. Antarctic station-based seasonal pressure reconstructions since 1905: 2. Variability and trends during the twentieth century. *Journal of Geophysical Research: Atmospheres*, 121(6): 2836–2856. https://doi.org/10.1002/2015JD024565.

Fogt RL, Jones ME, Goergens CA, Solomon S, Jones JM. 2018. Reply to “Comment on ‘An Exceptional Summer during the South Pole Race of 1911/12.’” *Bulletin of the American Meteorological Society*, 99(10): 2143–2145. https://doi.org/10.1175/BAMS-D-18-0088.1.

Fogt RL, Jones ME, Solomon S, Jones JM, Goergens CA. 2017b. An Exceptional Summer during the South Pole Race of 1911-1912. *Bulletin of the American Meteorological Society*. https://doi.org/10.1175/BAMS-D-17-0013.1.

Fogt RL, Jones ME, Solomon S, Jones JM, Goergens CA. 2017c. An Exceptional Summer during the South Pole Race of 1911-1912. *Bulletin of the American Meteorological Society*, 98. https://doi.org/10.1175/BAMS-D-17-0013.1.

Fogt RL, Schneider DP, Goergens CA, Jones JM, Clark LN, Garberoglio MJ. 2019. Seasonal Antarctic pressure variability during the twentieth century from spatially complete

reconstructions and CAM5 simulations. *Climate Dynamics*. https://doi.org/10.1007/s00382-019-04674-8.

Freeman E, Woodruff SD, Worley SJ, Lubker SJ, Kent EC, Angel WE, Berry DI, Brohan P, Eastman R, Gates L, Gloeden W, Ji Z, Lawrimore J, Rayner NA, Rosenhagen G, Smith SR. 2017. ICOADS Release 3.0: a major update to the historical marine climate record. *International Journal of Climatology*, 37(5): 2211–2232. https://doi.org/10.1002/joc.4775.

Jones JM, Gille ST, Goosse H, Abram NJ, Canziani PO, Charman DJ, Clem KR, Crosta X, de Lavergne C, Eisenman I, England MH, Fogt RL, Frankcombe LM, Marshall GJ, Masson-Delmotte V, Morrison AK, Orsi AJ, Raphael MN, Renwick JA, Schneider DP, Simpkins GR, Steig EJ, Stenni B, Swingedouw D, Vance TR. 2016. Assessing recent trends in high-latitude Southern Hemisphere surface climate. *Nature Climate Change*, 6(10): 917–926. https://doi.org/10.1038/nclimate3103.

Jones ME, Bromwich DH, Nicolas JP, Carrasco J, Plavcová E, Zou X, Wang S-H. 2019. Sixty Years of Widespread Warming in the Southern Middle and High Latitudes (1957–2016). *Journal of Climate*, 32(20): 6875–6898. https://doi.org/10.1175/JCLI-D-18-0565.1.

Kent EC, Berry DI. 2005. Quantifying random measurement errors in Voluntary Observing Ships' meteorological observations. *International Journal of Climatology*, 25(7): 843–856. https://doi.org/10.1002/joc.1167.

Laloyaux P, de Boisseson E, Balmaseda M, Bidlot J-R, Broennimann S, Buizza R, Dalhgren P, Dee D, Haimberger L, Hersbach H, Kosaka Y, Martin M, Poli P, Rayner N, Rustemeier E, Schepers D. 2018. CERA-20C: A Coupled Reanalysis of the Twentieth Century. *Journal of Advances in Modeling Earth Systems*, 10(5): 1172–1195. https://doi.org/10.1029/2018MS001273.

Mawson D. 1998. *The home of the blizzard: a true story of Antarctic survival*. St. Martin's Press: New York.

Mayewski PA, Frezzotti M, Bertler N, Ommen TV, Hamilton G, Jacka TH, Welch B, Frey M, Dahe Q, Jiawen R, Simões J, Fily M, Oerter H, Nishio F, Isaksson E, Mulvaney R, Holmund P, Lipenkov V, Goodwin I. 2005. The International Trans-Antarctic Scientific Expedition (ITASE): an overview. *Annals of Glaciology*, 41: 180–185. https://doi.org/10.3189/172756405781813159.

Poli P, Hersbach H, Dee DP, Berrisford P, Simmons AJ, Vitart F, Laloyaux P, Tan DGH, Peubey C, Thépaut J-N, Trémolet Y, Hólm EV, Bonavita M, Isaksen L, Fisher M. 2016. ERA-20C: An Atmospheric Reanalysis of the Twentieth Century. *Journal of Climate*, 29(11): 4083–4097. https://doi.org/10.1175/JCLI-D-15-0556.1.

Purich A, England MH. 2019. Tropical teleconnections to Antarctic sea ice during austral spring 2016 in coupled pacemaker experiments. *Geophysical Research Letters*. https://doi.org/10.1029/2019GL082671.

Rignot E, Jacobs S, Mouginot J, Scheuchl B. 2013. Ice-Shelf Melting Around Antarctica. *Science*, 341(6143): 266–270. https://doi.org/10.1126/science.1235798.

Rignot E, Mouginot J, Scheuchl B, van den Broeke M, van Wessem MJ, Morlighem M. 2019. Four decades of Antarctic Ice Sheet mass balance from 1979–2017. *Proceedings of the National Academy of Sciences*, 116(4): 1095–1103. https://doi.org/10.1073/pnas.1812883116.

Schneider DP, Fogt RL. 2018. Artifacts in Century-Length Atmospheric and Coupled Reanalyses Over Antarctica Due to Historical Data Availability. *Geophysical Research Letters*, 45(2): 964–973. https://doi.org/10.1002/2017GL076226.

Sienicki K. 2018. Comments on "An Exceptional Summer during the South Pole Race of 1911/12." *Bulletin of the American Meteorological Society*, 99(10): 2139–2143. https://doi.org/10.1175/BAMS-D-17-0282.1.

Slivinski LC, Compo GP, Whitaker JS, Sardeshmukh PD, Giese BS, McColl C, Allan R, Yin X, Vose R, Titchner H, Kennedy J, Spencer LJ, Ashcroft L, Brönnimann S, Brunet M, Camuffo D, Cornes R, Cram TA, Crouthamel R, Domínguez-Castro F, Freeman JE, Gergis J, Hawkins E, Jones PD, Jourdain S, Kaplan A, Kubota H, Le Blancq F, Lee T, Lorrey A, Luterbacher J, Maugeri M, Mock CJ, Moore GWK, Przybylak R, Pudmenzky C, Reason C, Slonosky VC, Smith C, Tinz B, Trewin B, Valente MA, Wang XL, Wilkinson C, Wood K, Wyszyn'ski P. 2019. Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. *Quarterly Journal of the Royal Meteorological Society*. https://doi.org/10.1002/qj.3598.

Steig EJ, Schneider DP, Rutherford SD, Mann ME, Comiso JC, Shindell DT. 2009. Warming of the Antarctic ice-sheet surface since the 1957 International Geophysical Year. *Nature*, 457(7228): 459–462. https://doi.org/10.1038/nature07669.

Stenni B, Curran MAJ, Abram NJ, Orsi A, Goursaud S, Masson-Delmotte V, Neukom R, Goosse H, Divine D, van Ommen T, Steig EJ, Dixon DA, Thomas ER, Bertler NAN, Isaksson E, Ekaykin A, Werner M, Frezzotti M. 2017. Antarctic climate variability on regional and continental scales over the last 2000 years. *Climate of the Past*, 13(11): 1609–1634. https://doi.org/10.5194/cp-13-1609-2017.

Stuecker MF, Bitz CM, Armour KC. 2017. Conditions leading to the unprecedented low Antarctic sea ice extent during the 2016 austral spring season. *Geophysical Research Letters*, 44(17): 9008–9019. https://doi.org/10.1002/2017GL074691.

Thomas ER, Marshall GJ, McConnell JR. 2008. A doubling in snow accumulation in the western Antarctic Peninsula since 1850. *Geophysical Research Letters*, 35(1). https://doi.org/10.1029/2007GL032529.

Turner J, Marshall GJ, Clem K, Colwell S, Phillips T, Lu H. 2019. Antarctic temperature variability and change from station data. *International Journal of Climatology*. https://doi.org/10.1002/joc.6378.

Turner J, Phillips T, Marshall GJ, Hosking JS, Pope JO, Bracegirdle TJ, Deb P. 2017. Unprecedented springtime retreat of Antarctic sea ice in 2016: The 2016 Antarctic Sea Ice Retreat. *Geophysical Research Letters*, 44(13): 6868–6875. https://doi.org/10.1002/2017GL073656.

Figure Captions

Figure 1. Maps of seasonal mean data location grouped by decade. Open circles represent seasonal mean locations from ship records, and filled circles are for temporary bases on the continent. The bottom left plot shows the location of all 271 seasonal mean observations compared, while the bottom right plot shows the locations of the 18 station reconstructions (brown) and observations from Orcadas (grey) used to generate the spatially complete pressure reconstruction. Contours on the bottom right panel are the standard deviations of monthly ERA5 surface pressure anomalies for reference, contoured every 0.5 hPa.

Figure 2. Mean reconstruction skill statistics (columns; bias, MAE, and RMSD) compared to historical observations, and averaged over various latitudes (top row), longitudes (middle row), and decades (bottom row). The statistics calculated over all observations are listed at the bottom of the figure.

Figure 3. Decadal mean RMSD plotted by decade.

Figure 4. Map showing observational mean (over full length of record) RMSD for select representative locations examined in more detail.

Figure 5. Time series of historical observations, reconstruction (with 95% confidence interval in grey shading), and gridded reanalysis problems for observations representative of the lowest RMSD (solid lines for MSLP, dashed lines for surface pressure). The name is the record identifier provided in ISPD or ICOADS. In a), the x-axis varies by season, and the labels represent the DJF seasons for each year; in b) and c) only DJF data are plotted and the label represents the DJF season. The white spaces in c) represent discontinuities in the observations. The values at the bottom of each panel are the correlations (if more than 10 data points are available) and MAE values (first numbers based on MSLP, second (where available) based on surface pressure) for each dataset. The data for station 889340 were from ISPD, while Deck 215 and Deck 899 were from ICOADS.

Figure 6. As in Fig. 5, but representative of stations where elevation corrections (reduction to sea level pressure) play an important aspect of the reconstruction performance evaluation. All data in this figure were obtained from ISPD.

Figure 7. Time series of seasonal mean MSLP for all four seasons at Little America (from ISPD) on the northern edge of the Ross Ice Shelf for the reconstruction along with values from a) 20CRv2c; b) 20CRv3; c) CERA-20C. The gray shading in each panel represents the 95% confidence interval for the reconstruction, while the colored shading represents 95% confidence intervals for each of the reanalyses (calculated as the 1.96 times the standard deviation across the seasonal mean ensemble members). The overall RMSD compared to the reconstruction is given in the upper right for each dataset.

Figure 8. As in Fig. 5, but for a) Deck 246, which operated near the East Antarctic coast discontinuously between 1910-1930, and b) observations at Cape Denison during the Australian Antarctic Expedition of 1911-1914. Note in b) that surface pressure is plotted on the right axis.

MAE values are based only on MSLP since there is a wide range of surface pressure values (>30 hPa), all primarily reflecting elevation differences in the underlying models. For a) the MAE values (both based on MSLP) are calculated using the two different locations in DJF 1911. Cape Denison data were obtained from ISPD, while data for Deck 246 were obtained from ICOADS version 3.

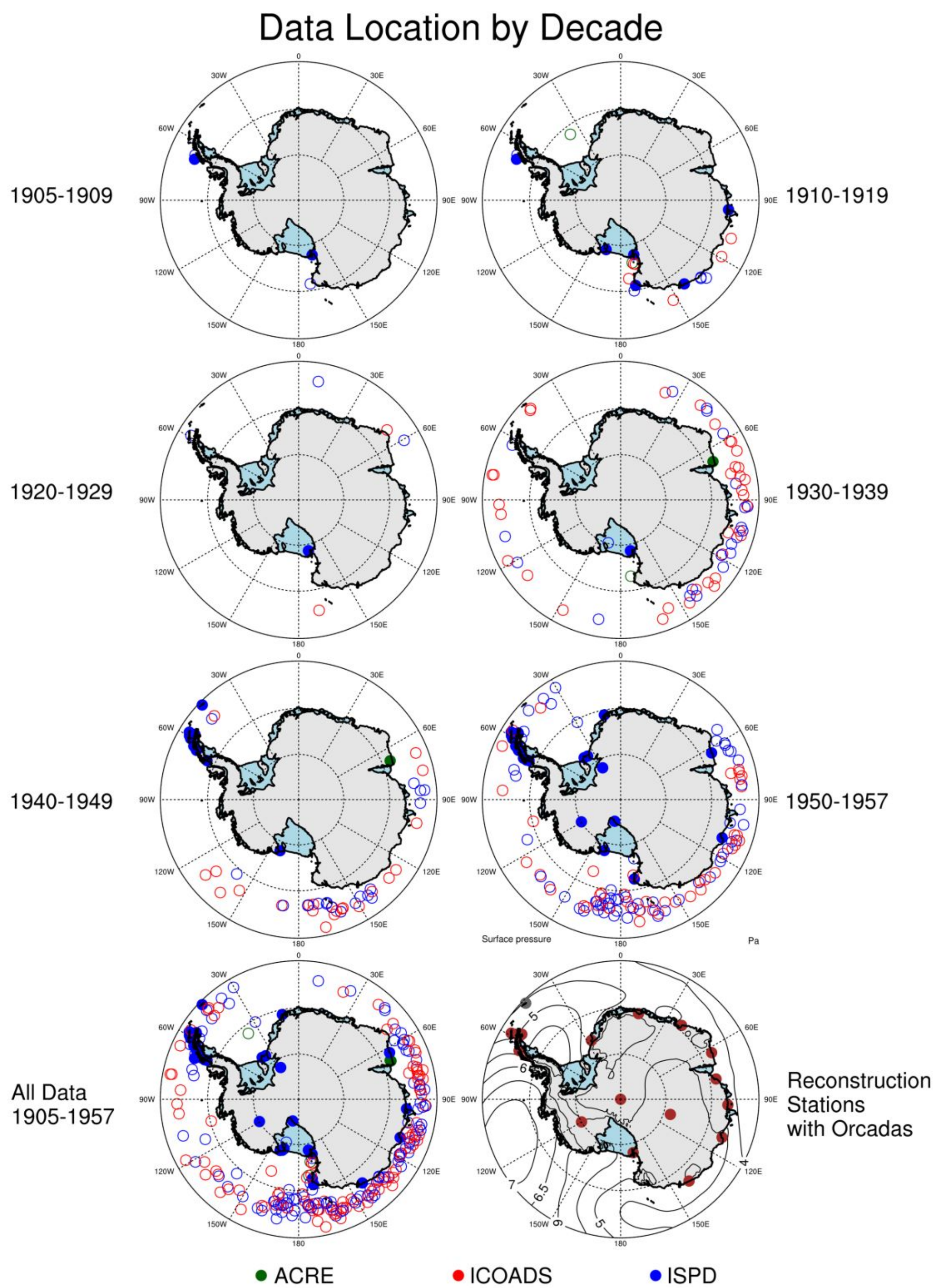

Figure 1. Maps of seasonal mean data location grouped by decade. Open circles represent seasonal mean locations from ship records, and filled circles are for temporary bases on the continent. The bottom left plot shows the location of all 271 seasonal mean observations compared, while the bottom right plot shows the locations of the 18 station reconstructions (brown) and observations from Orcadas (grey) used to generate the spatially complete pressure reconstruction. Contours on the bottom right panel are the standard deviations of monthly ERA5 surface pressure anomalies for reference, contoured every 0.5 hPa.

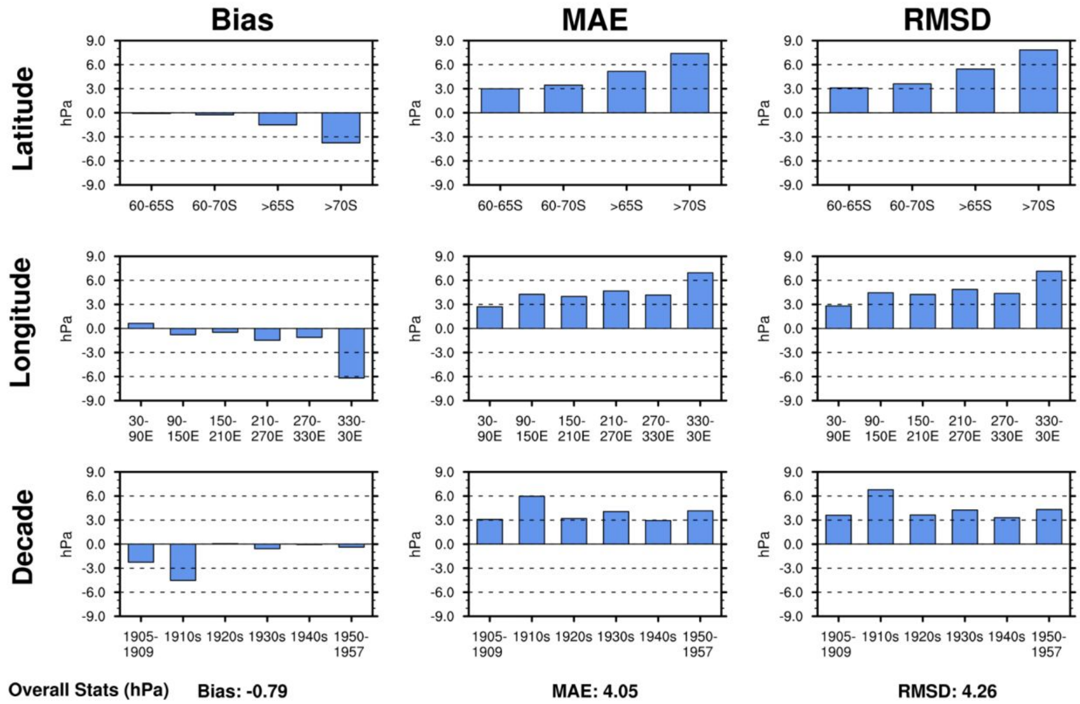

Figure 2. Mean reconstruction skill statistics (columns; bias, MAE, and RMSD) compared to historical observations, and averaged over various latitude bands (top row), longitude bands (middle row), and decades (bottom row). The statistics calculated over all observations are listed at the bottom of the figure.

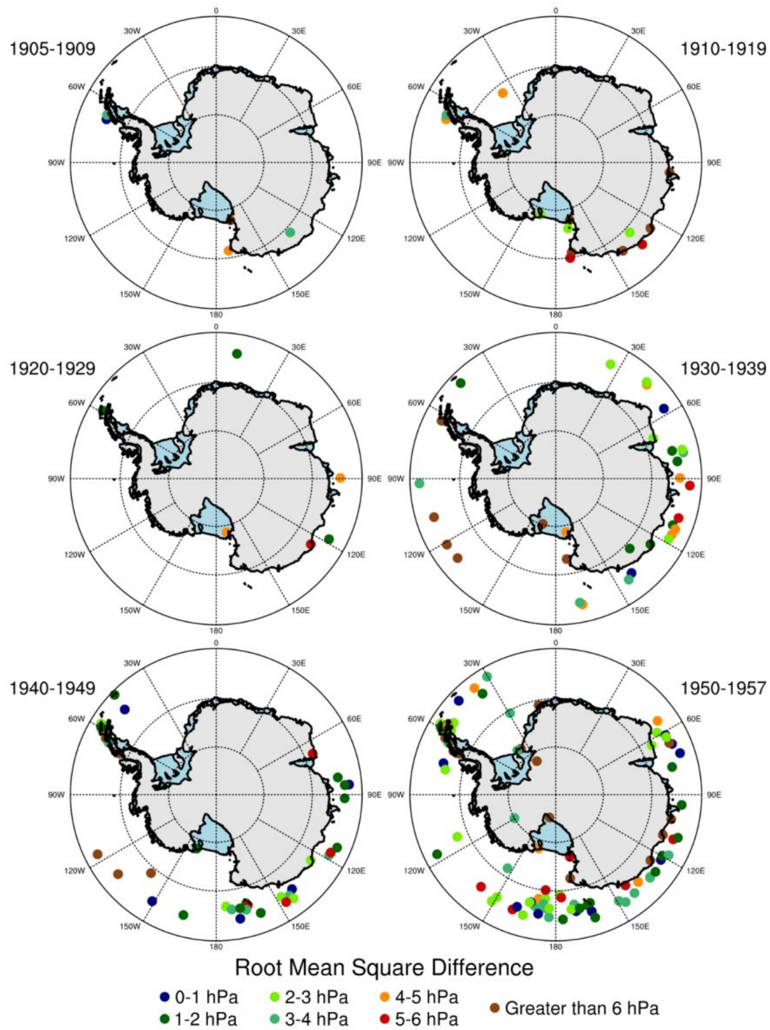

Figure 3. Decadal mean RMSD plotted by decade.

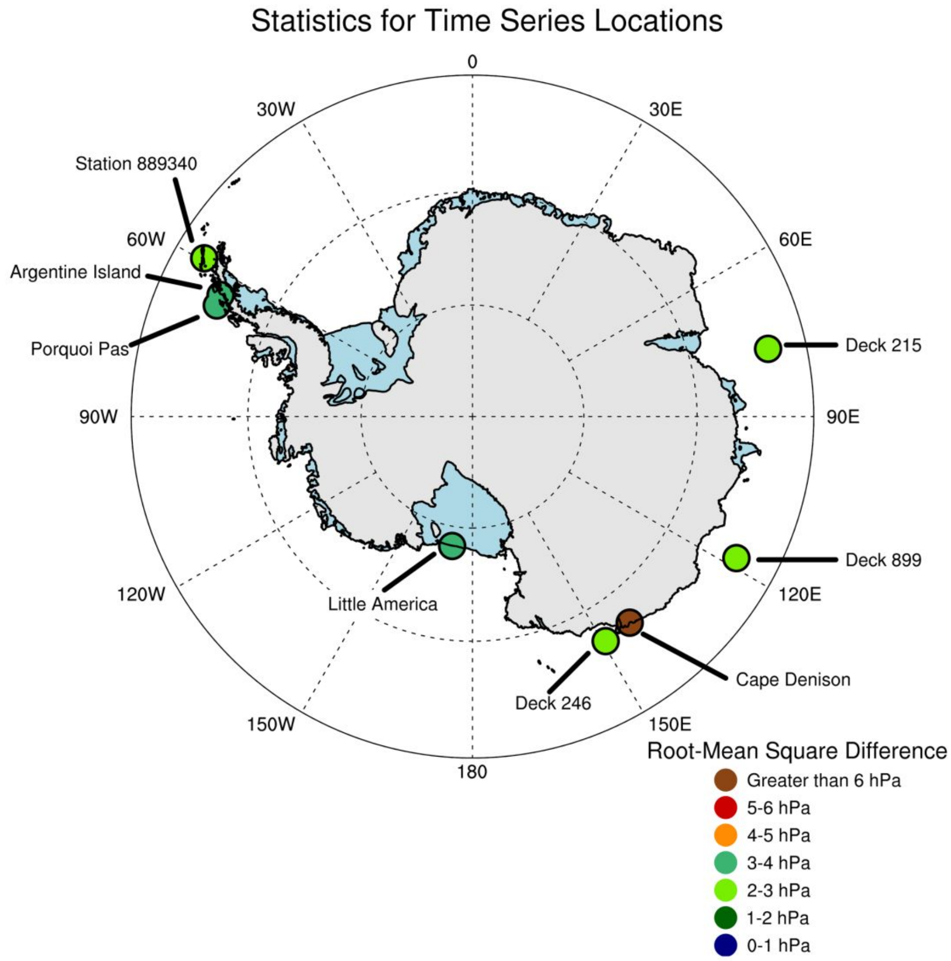

Figure 4. Map showing observational mean (over full length of record) RMSD for select representative locations examined in more detail.

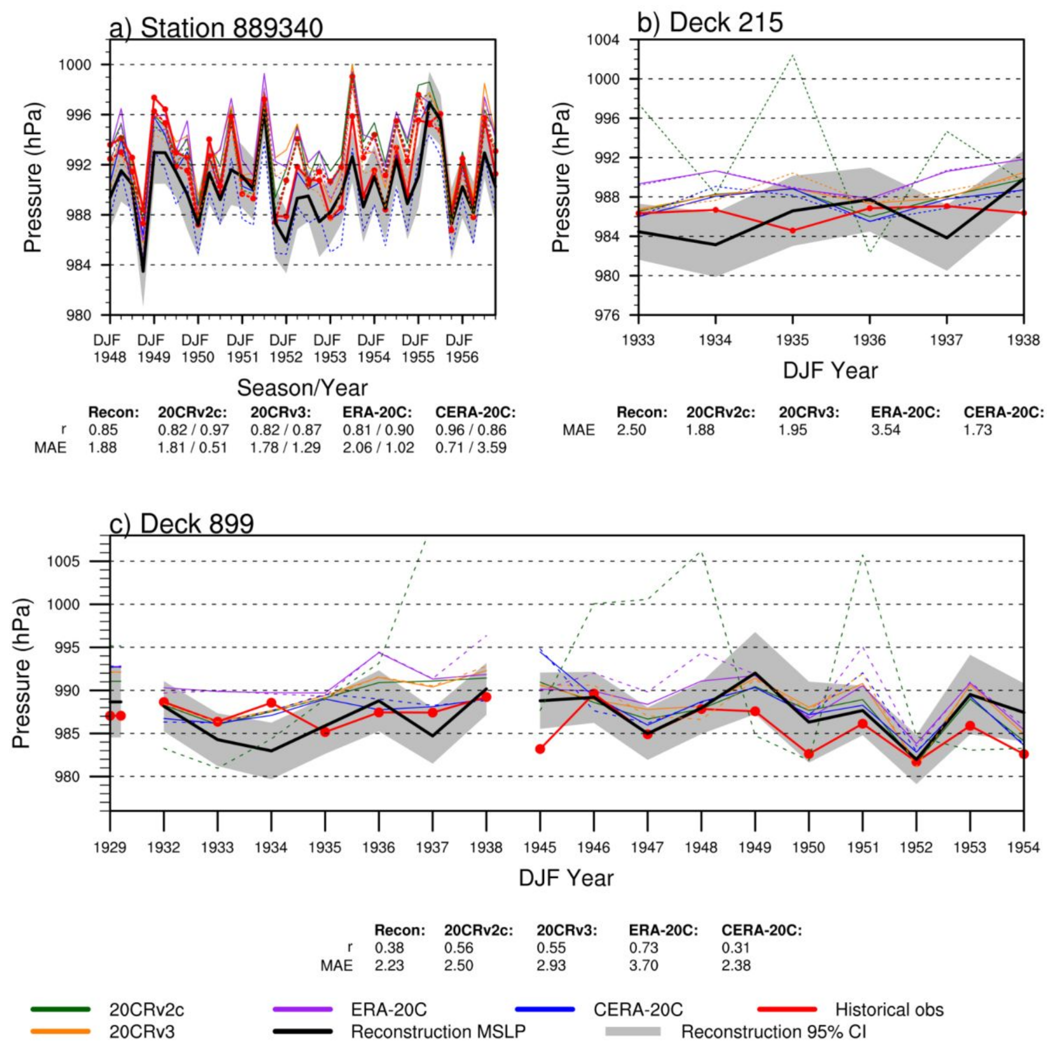

Figure 5. Time series of historical observations, reconstruction (with 95% confidence interval in grey shading), and gridded reanalysis products for observations representative of the lowest RMSD (solid lines for MSLP, dashed lines for surface pressure). The name is the record identifier provided in ISPD or ICOADS. In a), the x-axis varies by season, and the labels represent the DJF seasons for each year; in b) and c) only DJF data are plotted and the label represents the DJF season. The white spaces in c) represent discontinuities in the observations. The values at the bottom of each panel are the correlations (if more than 10 data points are available) and MAE values (first numbers based on MSLP, second (where available) based on surface pressure) for each dataset. The data for station 889340 are from ISPD, while Deck 215 and Deck 899 are from ICOADS.

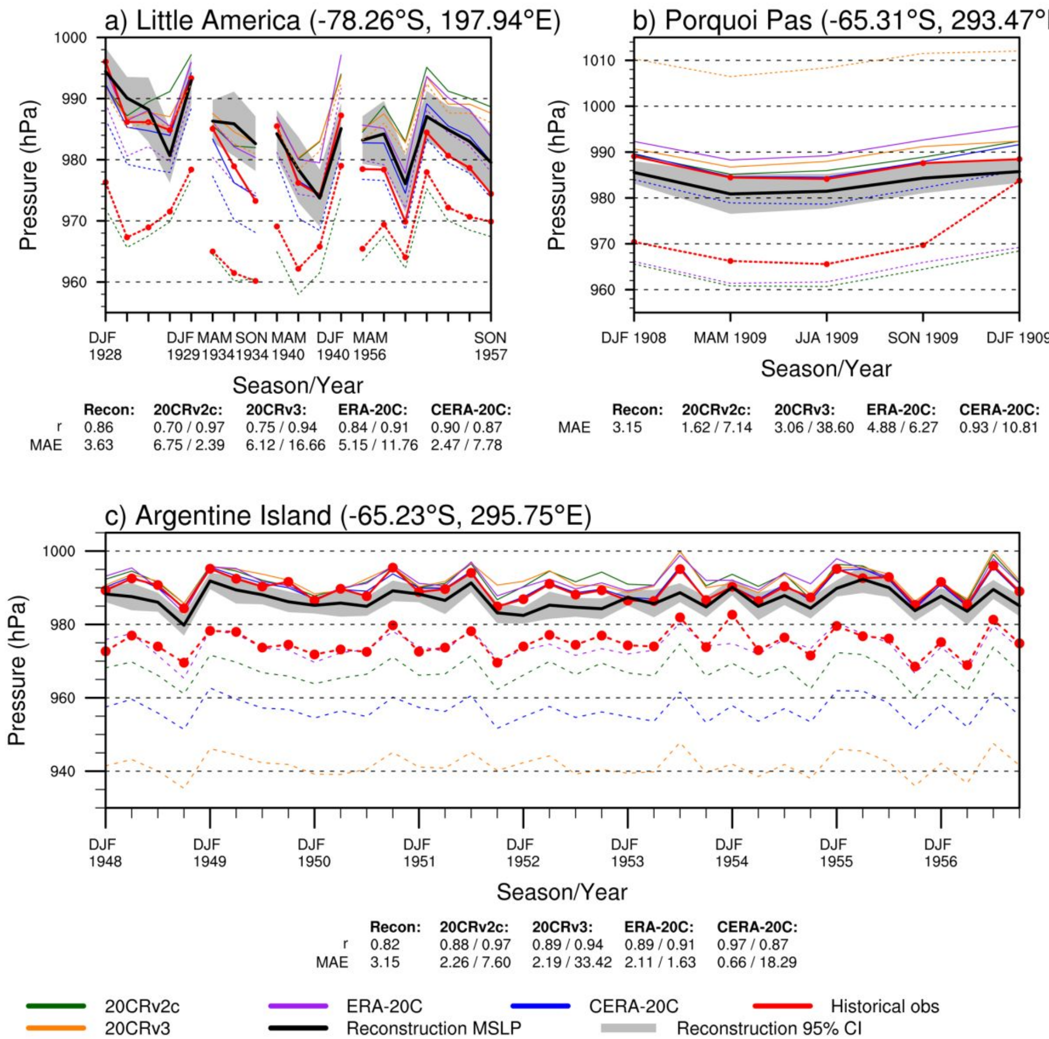

Figure 6. As in Fig. 5, but representative of stations where elevation corrections (reduction to sea level pressure) play an important aspect of the reconstruction performance evaluation. In all panels, the x-axis varies by season. All data in this figure were obtained from ISPD version 3.

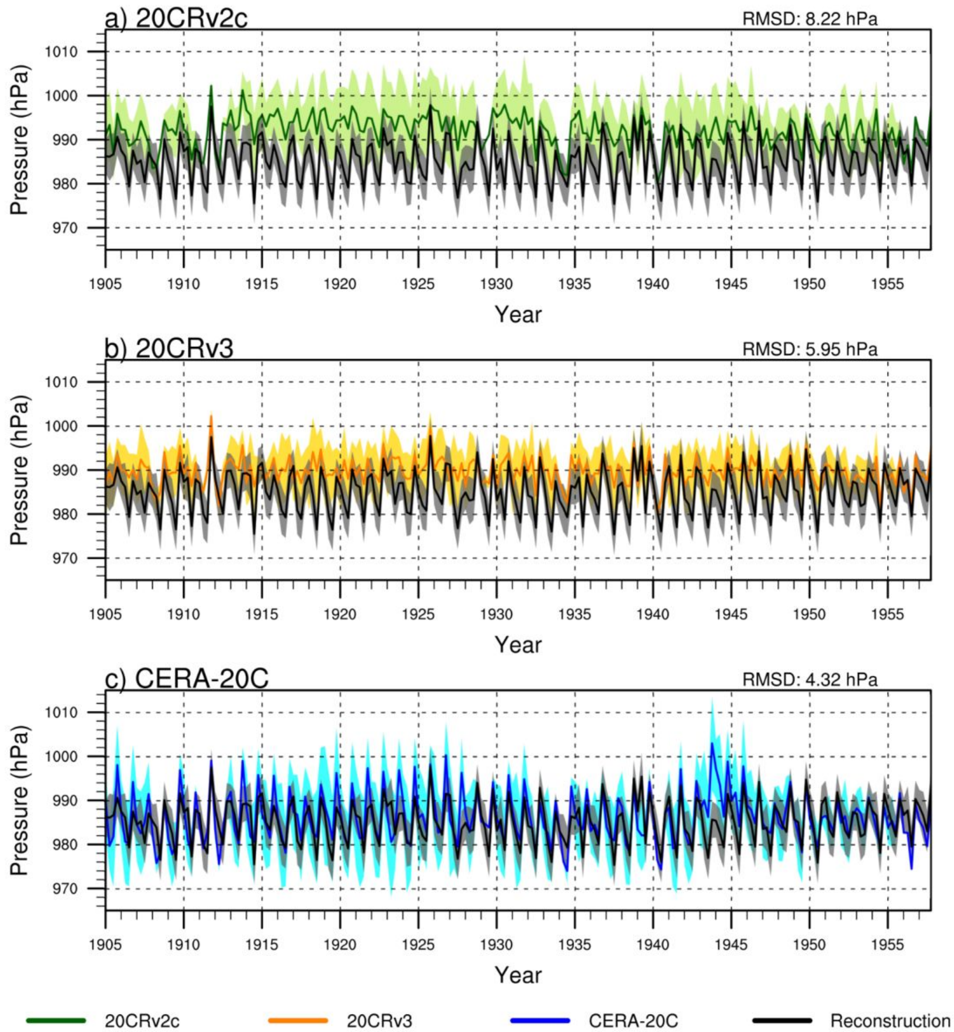

Figure 7. Time series of seasonal mean MSLP for all four seasons at Little America (from ISPD) on the northern edge of the Ross Ice Shelf for the reconstruction along with values from a) 20CRv2c; b) 20CRv3; c) CERA-20C. The gray shading in each panel represents the 95% confidence interval for the reconstruction, while the colored shading represents 95% confidence intervals for each of the reanalyses (calculated as the 1.96 times the standard deviation across the seasonal mean ensemble members). The overall RMSD compared to the reconstruction is given in the upper right for each dataset.

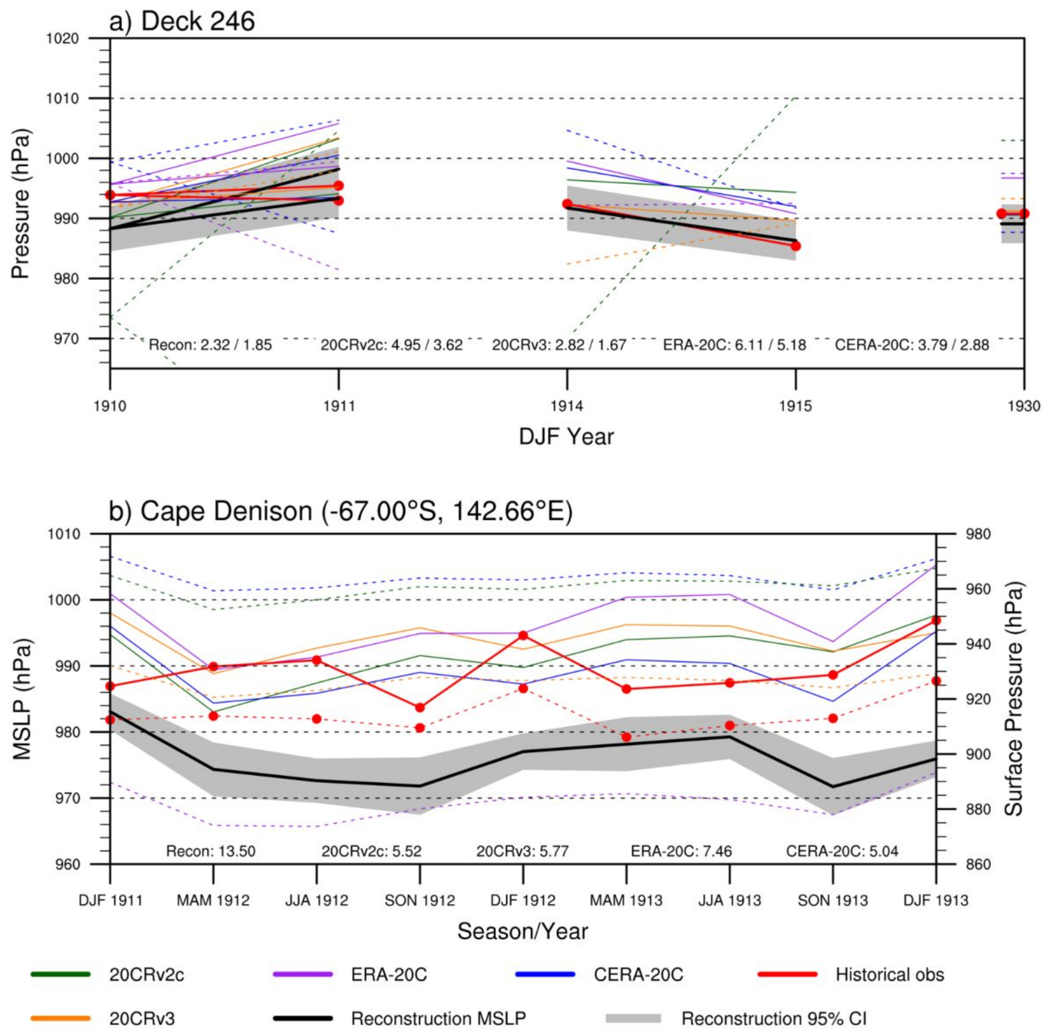

Figure 8. As in Fig. 5, but for a) Deck 246, which operated near the East Antarctic coast discontinuously between 1910-1930, and b) observations at Cape Denison during the Australian Antarctic Expedition of 1911-1914. Note in b) that surface pressure is plotted on the right axis. MAE values are based only on MSLP since there is a wide range of surface pressure values (>30 hPa), all primarily reflecting elevation differences in the underlying models. For a) the MAE values (both based on MSLP) are calculated using the two different locations in DJF 1911. Cape Denison data were obtained from ISPD, while data for Deck 246 were obtained from ICOADS version 3.

An Assessment of Early 20th Century Antarctic Pressure Reconstructions using Historical Observations

Short title: An assessment of early 20th century Antarctic pressure reconstructions

Ryan L. Fogt[1], Connor P. Belak[1], Julie M. Jones[2], Laura C. Slivinski[3,4], Gilbert P. Compo[3,4]

[1]Department of Geography and Scalia Laboratory for Atmospheric Analysis, Ohio University, Athens, OH
[2]Department of Geography, University of Sheffield, Sheffield, UK
[3] University of Colorado, Cooperative Institute for Research in Environmental Sciences, Boulder, CO
[4]NOAA ~~Earth System Research~~Physical Sciences Laboratory~~, Physical Sciences Division~~, Boulder, CO

Corresponding author address: Ryan L. Fogt, Ohio University Department of Geography, 122 Clippinger Laboratories, Athens OH 45701 ph: 740-593-1151 fax: 740-593-1149
email: fogtr@ohio.edu

Keywords: Antarctica, climate, pressure, data recovery

Abstract

While gridded seasonal pressure reconstructions poleward of 60°S extending back to 1905 have been recently completed, their skill has not been assessed prior to 1958. To provide a more thorough evaluation of the skill and performance in the early 20th century, these reconstructions are compared to other gridded datasets, historical data from early Antarctic expeditions, ship records, and temporary bases ~~to further evaluate their performance~~.

Overall, the comparison confirms that the reconstruction uncertainty of 2-4 hPa (evaluated after 1979) over the Southern Ocean is a valid estimate of the reconstruction error in the early 20th century. Over the interior and near the coast of Antarctica, direct comparisons with historical data are challenged by elevation-based reductions to sea level pressure. In a few cases, a simple linear adjustment of the reconstruction to sea level matches the historical data well, but in other cases, the differences remain greater than 10 hPa. Despite these large errors, comparisons with continuous multi-season observations demonstrate that aspects of the interannual variability are often still captured, suggesting that the reconstructions have skill representing variations on this timescale, even if it is difficult to determine how well they capture the mean pressure at these higher elevations. Additional comparisons with various 20th century reanalysis products demonstrate the value of assimilating the historical observations in these datasets, which acts to substantially reduce the reanalysis ensemble spread, and bring the reanalysis ensemble mean within the reconstruction and observational uncertainty.

1. Introduction

Despite recent dramatic climate-related changes in Antarctica, including warming of West Antarctica (Steig *et al.*, 2009; Bromwich *et al.*, 2012; Jones *et al.*, 2019; Turner *et al.*, 2019), retreat of several marine-based glaciers in the Amundsen Sea embayment (Rignot *et al.*, 2013, 2019; Edwards *et al.*, 2019), and rapid sea ice loss beginning in austral spring 2016 (Stuecker *et al.*, 2017; Turner *et al.*, 2017; Purich and England, 2019), there are significant challenges to understanding how unique these events are in a historical context or how likely they are to change in the future. This is in part due to the high degree of natural climate variability across Antarctica and the fully coupled nature of many of these changes occurring at the atmosphere-ice-ocean interface. When combined with the short observational climate records primarily beginning around the international Geophysical Year (IGY) activities of 1957/1958, interpreting these changes in the Antarctic climate has proven to be very difficult (Jones *et al.*, 2016).

Data from ice cores across Antarctica help to place ongoing change in a longer context (Bracegirdle *et al.*, 2019), especially through coordinated efforts like the International Trans-Antarctic Scientific Expedition (ITASE; Mayewski *et al.*, 2005). Ice core evidence has helped to compare recent Antarctic warming with variability over the past 2000 years (Stenni *et al.*, 2017) as well as changes in West Antarctica snowfall (Thomas *et al.*, 2008), and when combined with a climate model, longer estimates of Antarctic surface mass balance (Agosta *et al.*, 2019), as examples. Nonetheless, dating of specific events and the suppressed sub-annual temporal resolution in many cores pose challenges for these longer-term estimates of Antarctic climate variability.

Other tools to examine historical Antarctic climate variability during the last century include gridded datasets that span the 20th century, such as existing historical reanalyses that only assimilate surface pressure, like the National Oceanic and Atmospheric Administration/Cooperative Institute for Research in Environmental Sciences (NOAA-CIRES) Twentieth Century Reanalysis version 2c (hereafter 20CRv2c, Compo *et al.*, 2011) and the new NOAA-CIRES-DOE version 3 (20CRv3, Slivinski *et al.*, 2019). Similar "sparse-input" reanalyses have been completed by the European Centre for Medium Range Weather Forecasts (ECMWF), including their 20th-century reanalysis (ERA-20C, Poli *et al.*, 2016) and a coupled ocean-atmosphere reanalysis of the 20th century (CERA-20C, Laloyaux *et al.*, 2018), each of which assimilates marine wind observations and surface pressure. As a coupled reanalysis, CERA-20C also assimilates subsurface ocean temperature and salinity observations (Laloyaux *et al.*, 2018). While these are spatially and temporally complete throughout the 20th century, unsurprisingly they are sensitive to the number of assimilated surface pressure observations, and therefore the skill changes considerably over the high southern latitudes throughout the 20th century (Schneider and Fogt, 2018; Fogt *et al.*, 2019; Slivinski *et al.*, 2019).

An alternative approach to address these challenges is to generate seasonal pressure reconstructions ~~poleward of 60°S~~, both at individual Antarctic stations (Fogt *et al.*, 2016a, 2016b)~~(Fogt *et al.*, 2016a, 2016b, 2017a, 2019)~~, as well as spatially complete poleward of 60°S (Fogt *et al.*, 2017a, 2019). When compared to gridded climate datasets, these reconstructions showed far less sensitivity to changes in the number of observations across Antarctica (Schneider and Fogt, 2018), which is perhaps not surprising given that the reconstructions were based on statistical relationships with a spatially and temporally fixed network of stations in the mid and high latitudes of the Southern Hemisphere (Fogt *et al.*, 2016a). However, the skill of the reconstructions was only thoroughly evaluated after the ~~International Geophysical Year (~~IGY, (1957-1958), when most Antarctic observations began; spatially complete comparisons in reconstruction skill were even more limited and only possible after 1979 through evaluating against the ECMWF Interim reanalysis, ERA-Interim (ERA-Int; Dee *et al.*, 2011). Therefore, the skill of these reconstructions in the early 20th century before the IGY remains unknown, although the underlying relationships governing them are likely stationary (Clark and Fogt, 2019). This work aims to provide a more complete evaluation of these reconstructions poleward of 60°S by using available pressure observations that are independent from the reconstructions in the early 20th century.

2. Data and Methods

We make use of the best performing seasonal spatially-complete Antarctic pressure reconstructions, which are available from 1905-2013 and based on surface pressure anomalies from the 1981-2010 climatology from ERA-Int (Fogt *et al.*, 2019). The reconstructions are~~were~~ based on a kriging interpolation of 18 Antarctic station reconstructions and observations at Orcadas (locations plotted in bottom right panel of Fig. 1) to ~~originally created on~~ an 80km x 80 km Cartesian grid centered over the South Pole. For comparison here, this grid has been converted to a 0.75°x0.75° latitude-longitude grid. Importantly, comparisons with ERA-Int after 1979 (Fogt *et al.*, 2017a, 2019) demonstrated that the reconstruction skill varies seasonally, with the highest skill (defined as the best agreement with ERA-Int) in austral summer (December – February, DJF), and the lowest skill in the austral autumn and spring (March-May, MAM, and September-November, SON, respectively). In general, higher reconstruction skill is found over the Antarctic continent (especially near the Antarctic Peninsula), and lower skill at the northern edge of the domain near 60°S, especially in the South Pacific (Fogt *et al.*, 2019).

Historical pressure data prior to 1957 poleward of 60°S were extracted from three primary datasets to further evaluate the reconstruction skill in the early 20th century. The largest source of pressure data comes from the International Surface Pressure Databank version 3.2.9 (ISPD; Cram *et al.*, 2015), which contains subdaily measurements of mean sea level and/or surface pressure from both ships and temporary Antarctic bases (during field expeditions) and early Antarctic stations that began prior to the IGY. Additional sea level and/or surface pressure data were obtained from the International Comprehensive Ocean-Atmosphere Data Set release 3 (ICOADS; Freeman *et al.*, 2017); we only make use of ship records in ICOADS that were not available in ISPD. As many new historical observations from both ships and early Antarctic expeditions are still being recovered and digitized, we also make use of a few newly digitized additional pressure observations stemming from the Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative (Allan *et al.*, 2011), obtained directly from Rob Allan, the lead of the ACRE project. Only those ACRE pressure observations that were not part of the current releases of ISPD and ICOADS are utilized here, particularly from Antarctic expeditions in the first few decades of the 20th century. Most historical observations reported both SLP and surface pressure; however, a few observations only reported SLP, and therefore the reconstruction is compared to historical SLP observations throughout. We directly use the historical observations as reported in ISPD, ICOADS, or in the ACRE data for SLP and surface pressure, and do not reduce any surface pressure observations to SLP.

In making comparisons, it is important to consider that historical observations can have their own error or bias for many reasons, potentially including incorrect reported location (latitude, longitude, or elevation); problems with the barometer; errors associated with adjustments to sea level pressure; or misspecified corrections to pressure measurements from temperature or gravity. Because measurement errors likely vary in time and space, it is difficult to fully quantify how much these errors contribute to the reported pressure observations (Kent and Berry, 2005). Reanalyses generally define instantaneous, random observation errors that depend on platform and/or time (Compo *et al.*, 2011; Poli *et al.*, 2016; Laloyaux *et al.*, 2018; Slivinski *et al.*, 2019), and analyses comparing these expected errors with background statistics suggest estimates of about 1.5-2 hPa in the early 20th century are reasonable (Slivinski *et al.*, 2019). However, Slivinski *et al.* (2019) and more recent comparisons (not shown) also suggest that there is a relatively large systematic error in observations in the Southern Hemisphere in the early 20th century. We therefore assume a conservative estimate of 2 hPa total error in the seasonally-average observation estimates (or an observation error variance of 4 hPa^2).

In order to compare the historical observations to the reconstructions, additional analysis was required. First, all historical pressure data were converted to hPa, the same units as the pressure reconstructions. Monthly means for all the historical observations were calculated if at least 75% of the days within a given month had at least one observation available, to ensure a reliable monthly pressure estimate. Seasonal means (following the traditional seasonal divisions, DJF, MAM, JJA (June – August), and SON) were calculated to compare to the seasonal pressure reconstructions if at least two monthly means existed using our 75% daily data threshold; data for the third month was added to calculate the seasonal mean if 2 months had met the 75% daily data threshold, and the third month had more than 50% of its daily data. This approach allows for comparison with the maximum amount of data possible. Comparisons of a single monthly mean historical observation (for cases when historical data only had one month meeting our threshold) with the seasonal mean pressure reconstruction produced larger differences (not shown), and so this paper only focuses on comparisons of at least two months of historical data to define a seasonal mean. Altogether, these constraints yielded 271 seasonal means from historical pressure observations prior to 1957. The majority (>75%) of these historical observations occur only in DJF (austral summer), although it is noted that there are several observations after 1940 complete for the entire year. For reference, the year of austral summer refers to the year of December throughout this study.

Two more steps were taken prior to making comparisons. First, seasonal mean locations for the historical records were calculated using the coordinates provided in the data archives. These locations varied little for the early Antarctic stations; for observations collected on moving ships, we calculated a seasonal mean location by year for comparison since the reconstruction data are only available seasonally, which limits comparisons at higher temporal and spatial frequency. Of the 130 seasonal means from the ship records, more than 75% of the subdaily pressure observations showed less than a 2 degree standard deviation in the latitude, and half of the ships' standard deviations were less than 15 degrees longitude. In terms of pressure, there was no significant difference between ships that had a standard deviation of more and less than 20 degrees longitude in the standard deviations of sub-daily pressure observations. We therefore suggest the error associated with the mean location is more strongly related to the number of strong storms a ship encountered, which are more unique to a specific location rather than the mean location of the ship. As a rough estimate of this error, the mean standard deviation of pressure from the subdaily ship observations, 3.45 hPa, can be used, although more than half of the ships have a pressure standard deviation below 3.0 hPa. This standard deviation of subdaily pressure within one season is much smaller than the interannual standard deviation of monthly pressure across the South Pacific, which ranges from 3-7 hPa based on ERA5 data (contours in Fig. 1, bottom right panel).

Lastly, to compare the pressure reconstructions to the historical pressure observations, elevation adjustments were needed. The pressure reconstructions, originally constructed as surface pressure anomalies based on the underlying topography of ERA-Int at 0.75°x0.75° latitude / longitude resolution, were adjusted to sea level pressure using a rough linear estimate of 12 hPa pressure decrease per 100 m elevation gain. The 12 hPa approximation is larger than the 10 hPa per 100m assumed in the middle latitudes in the lower troposphere given the colder and drier (and therefore denser) air commonly observed poleward of 60°S. Without simultaneous measurements of temperature and humidity, it is challenging to provide a more accurate adjustment to sea level pressure. Furthermore, reduction to sea level pressure is known to be problematic in nearly all gridded climate datasets (and observations themselves) over Antarctica, including 20CRv3 (Slivinski *et al.*, 2019). This adjustment to sea level was necessary even for historical observations that had both surface and sea level pressure data, as due to the smoothing of the topography of ERA-Int at even the 0.75° resolution, model elevations and observed elevations (where known – this value is not always given in the historical data) are often quite different; these elevation differences similarly make it challenging to adequately compare surface pressure observations with the surface pressure (anomaly) reconstructions. Importantly, elevation differences can also arise due to the smoothing of the steep Antarctic coastline in ERA-Int, and therefore influence comparisons from ships close to the Antarctic continent. This adjustment to sea level introduces the greatest error and uncertainty in our comparison, as will be discussed in detail throughout. Nonetheless, the simple linear adjustment allows us to evaluate interannual (or intra-annual) variability and provide further information on reconstruction performance than otherwise would be possible.

In the comparisons, the reconstruction data are extracted from the gridpoint closest to the seasonal mean location in the reconstruction field. To provide further comparison for select observations that span more than one season, from the reanalyses we also extract both surface and sea level pressure data from the closest gridpoint using the ensemble mean in all but ERA-20C (which has only one estimate). This approach also introduces some error, called the 'error of representativeness', reflecting how well a reanalysis gridpoint represents the point location from the observation; this error is assumed to be one of the largest components of observation errors in the reanalyses, at least in 20CRv3 (Slivinski *et al.*, 2019), and is similar to the elevation-based errors for the reconstruction evaluations described previously. Importantly~~However~~, these reanalyses are not independent of the observations used for comparison here: ERA-20C and CERA-20C assimilate surface pressure and marine surface wind observations from ISPD version 3.2.6 and ICOADS version 2.5.1 (Poli *et al.*, 2016; Laloyaux *et al.*, 2018); 20CRv2c assimilates pressure from ISPD version 3.2.9 and ICOADS version 2.5.P; and 20CRv3 assimilates pressure from ISPD version 4.7 and ICOADS version 3+v2 (Slivinski *et al.*, 2019). Since the comparisons are made with ISPD v3.2.9 and ICOADS v3, the observations were at least available to be assimilated in both version of 20CR, and likely for ERA-20C and CERA-20C. However, t~~T~~he assimilation combines the instantaneous observations with a background guess from the forecast model, so fields from different reanalyses are expected to differ from each other (as they each use different forecast models and assimilation algorithms), as well as from the seasonally-averaged observation estimates analyzed here. In addition, each reanalysis system has its own quality control algorithm that will blacklist observations deemed unfit for assimilation; thus, even if a given set of observations are available to the assimilation algorithm, they may not all have been assimilated. Finally, the reanalyses each employ their own observation bias correction scheme, which removes significant, consistent differences between the observations and the background fields. For simplicity, we only consider the uncorrected observations here, but note that systematic differences between the reanalyses and observations in our comparison may have been ameliorated by taking into account the observation bias corrections calculated within each reanalysis.

Reconstruction performance compared to the historical observations is evaluated using three primary statistics: overall bias, defined as the mean difference between the seasonal historical observation estimates and reconstructions at sea level (reconstruction minus observations); the mean absolute error (MAE), defined as the mean absolute difference to remove offsetting effects of positive and negative difference; and the root mean square difference (RMSD), which due to the squaring of the differences prior to calculating the mean, gives slightly higher weighting to larger absolute differences than the MAE. For observations that span only one season or have the bias of the same sign across all seasons, the absolute value of the bias and MAE will yield identical results. In all cases, these three statistics were calculated over the full length of each observational record. Further averaging both spatially and temporally is conducted to assess the overall reconstruction performance. Since the reconstruction error determined by Fogt et al. (2019) over 1979-2013 is assumed to be constant in time, it is assumed that there are no (temporally) correlated errors in the reconstruction; the evaluations in this paper help to determine the persistence of reconstruction errors through the early 20th century.

3. Results

3.1. Historical data availability

Prior to estimating the seasonal pressure reconstruction skill in the early 20th century, it is important to recognize the large changes in the number of observations poleward of 60°S, especially before 1957. Figure 1 displays the location and data type for all of the 271 seasonal means that were compared to the reconstruction; open circles represent seasonal mean positions of ship data while filled circles denote base or station records that were stationary or only had minor movements over time. As expected, there are very few observations prior to 1930, although there is a relative increase in the 1910-1919 during the height of the 'Age of Heroes' which includes expeditions to the South Pole, the Australian Antarctic Expeditions, and the British Imperial Trans-Antarctic Expedition led by Ernest Shackleton. The decrease in the number of observations in the 1940s is not only a challenge in the high southern latitudes but also worldwide, associated with the Second World War. Nearly half of the historical observations used for comparison here come from the years just prior to the IGY (1950-1957), with several early Antarctic stations established over the continent (especially along the Antarctic Peninsula), and frequent ships along the East Antarctic coastline. A notable gap in observational coverage is within the Weddell Sea (east of the Antarctic Peninsula) and throughout much of the South Atlantic. In addition to persistent sea ice in the Weddell Sea (Cavalieri and Parkinson, 2008), the South Atlantic area is far from commercial sailing or whaling routes and is likely one of the greatest spatial data voids globally (for example, see Fig. 1 of Allan and Ansell, 2006).

3.2. Overall reconstruction performance

With this temporal evolution of historical data in mind, the reconstruction performance is displayed as a bar chart in Fig. 2; the overall statistics (averaged over all available observations) are given at the bottom: mean bias = -0.79 hPa; mean MAE = 4.05 hPa; mean RMSD = 4.26 hPa. Recalling that the overall comparison is primarily during austral summer, these values are considerably higher than the skill assessed in earlier work in comparison to ERA-Int (Fogt *et al.*, 2017a, 2019), which estimated an MAE over the Antarctic continent generally less than 1.5 hPa and over the Southern Ocean around 2-3 hPa. Indeed, the values seem even higher than the MAE in Fogt *et al.* (2019) in the non-summer months in the South Pacific, the region of highest MAE compared to ERA-Int which ranged from 2-4 hPa.

This quick comparison initially suggests that the seasonal pressure construction performance is of much lower quality prior to 1957 than it is after 1957. However, when considering that the observation estimates have their own errors (assumed to be approximately ~2 hPa in the early 20th century), the higher MAE prior to 1957 is not far from the assumed observational error. Further, when looking at the average skill as a function of latitude (Fig. 2), the bias is near zero in both of the latitude bands that primarily lie over the Southern Ocean (60°-65°S and 60°-70°S), potentially suggesting that the historical observation estimates tend to fluctuate near the reconstruction seasonal mean overall (both above and below, cancelling to near zero). The MAE in these latitude bands, near 3 hPa, better reflects the overall performance as it removes the cancellation between positive and negative differences. The MAE of 3 hPa is consistent with the skill reflected in Fogt *et al.* (2019), and is also close to the assumed observation error of 2 hPa. In contrast, where Fogt *et al.* (2019) demonstrate lower MAE over the Antarctic continent (due to more station observations constraining the reconstruction skill), comparisons with the historical observation estimates clearly show an increase in MAE poleward of 65°S, which is primarily from stations over the Antarctic continent (poleward of 70°S). As noted earlier, this increase in MAE is an artifact of the elevation sensitivity when comparing the historical observations at (often unknown) elevations different from the seasonal pressure reconstruction based on the underlying topography of the ERA-Int 0.75°x0.75° grid. The simple linear offset of 12 hPa per 100 m creates sea level pressures that are consistently too low (evidenced by the negative bias), which gives rise to large MAE and RMSD over the Antarctic continent.

The sensitivity to elevation is also apparent in the comparisons as a function of longitude. There is a strong negative bias in the 330°-30°E range, along the mountainous Antarctic Peninsula. In ERA-Int, this terrain was greatly smoothed but still elevated, while the observations are often along the coast near sea level. In other longitude zones, the mean bias is typically in the range of ±1 hPa, with MAE around 3-4 hPa. Given that Fig. 1 indicates that outside of 330°-30°E, the seasonal mean observation estimates primarily are over the ocean and less influenced by elevation corrections, the comparison by longitude again reflects a similar skill of the reconstruction (over the Southern Ocean at least) as observed in Fogt *et al.* (2019) during 1979-2013. Comparisons by decade typically have MAE also in the 3-4 hPa range, consistent with the previous evaluation. One outlier is the 1910s, which from Fig. 1 has a relatively high percentage of historical observations on or near the Antarctic continent, suggesting that elevation corrections are again giving a misleading impression of relatively low reconstruction skill.

To further visualize the spatial and elevation dependence of reconstruction performance, Fig. 3 displays the decadal mean RMSD for each station by decade. Several key points emerge from Fig. 3. First, there is an indication from ships in the South Pacific north of the Amundsen and Bellingshausen Seas (roughly 60°-70°S, 150°-90°W) that the reconstruction skill is considerably lower, with RMSD values often above 6 hPa. Since these are seasonal pressure estimates from ships (Fig. 1), elevation correction is not an issue. Rather, the larger errors are consistent with the much lower reconstruction performance in this region of high interannual variability (which often exceed observational errors based on interannual monthly pressure standard ~~deviaitons~~deviations as large as 7 hPa in Fig. 1~~likely exceed observational errors~~), as discussed in previous evaluations (Fogt *et al.*, 2017a, 2019). The larger interannual pressure variability here may also suggest that using the mean ship location could be introducing more error in this sector compared to other regions around Antarctica (Fig. 1). Second, for comparisons in the Southern Ocean for nearly all other regions, the RMSD is generally below 4 hPa, and the majority of the comparisons with the seasonal mean ship observation estimates show RMSD in the 3-4 hPa range. These values are consistent with the slightly lower performance of the reconstruction over the Southern Ocean. However, these values of RMSD still fall within expected error when considering errors in both observation estimates and the reconstruction (taken as the square root of the sum of the observational and reconstruction error variances). Lastly, it is clear that many of the higher errors (exceeding 6 hPa) are found along the coast, the Antarctic Peninsula, or in the interior of the continent. Even on the Ross Ice Shelf, RMSD values are typically larger than 4 hPa. As will be more clearly demonstrated later, these larger differences are due to elevation corrections to the historical observations in or near areas of high or vastly varying terrain. We will show that the reconstruction skill is likely much higher in these areas than suggested by the RMSD values in Fig. 3. Remarkably, in every decade since the 1930s, there are multiple locations where the RMSD values (over the ocean) are less than

1hPa, which is an excellent agreement with the historical observations. Given that we make comparisons with the seasonal mean ship position, that the historical observation estimates are rarely complete for the full season, and that they themselves have errors, this high agreement is noteworthy. It suggests that the reconstruction can be used as a reliable approximation of Antarctic pressure in most regions on and around the continent back to at least the 1930s.

3.3 Reconstruction performance evaluated with selected historical observations

The reconstruction performance is perhaps best evaluated by comparing to individual historical observations, which is the focus for the remainder of the study. To facilitate these assessments, a subset of representative stations and ship seasonal locations has been selected; their average location, name, and mean RMSD values over the entire record length are plotted in Fig. 4. We have purposely selected historical pressure estimates with longer records than a single season, and as many complete records from earlier portions of the 20th century as possible, although the sample size is quite limited (Fig. 1). In general, the reconstruction performance indicated by the mean RMSD in Fig. 4 at these locations is comparable to the overall decadal mean performance in Fig. 3, although we do not closely examine any stations with RMSD less than 2 hPa, and only one station where inadequate elevation corrections challenge the assessment of reconstruction skill (Cape Denison). Recall, all of these data were available for assimilation into 20CRv2c and 20CRv3 (and likely available for CERA-20C and ERA-20C), so the comparisons with the reanalyses are not necessarily always independent as discussed in section 2.

Figure 5 displays three seasonally-averaged historical observation estimates where the mean RMSD values are 2-3 hPa, consistent with the reconstruction skill and uncertainty assessed in Fogt *et al.* (2019), and within the assumed observational error variance of 4 hPa^2. Plotted with the historical observations (in red) are the mean sea level pressure (solid) and surface pressure (dashed, when available) from the nearest gridpoint of 20CRv2c (green), 20CRv3 (orange), ERA-20C (purple), and CERA-20C (blue). To compare with the skill evaluation in Fogt *et al.* (2019), the correlations (if more than 10 values are available) and MAE (in hPa) for each gridded dataset are given at the bottom of each panel: the first number is based on the MSLP, and the second number (where available) is based on surface pressure.

Station 889340 (from the ISPD archive) is situated along the Antarctic Peninsula (Fig. 4) and has a continuous pressure record beginning in 1948 for all seasons. As such, it is one of many such stations that provide a useful evaluation of reconstruction skill, and the overall MAE from the reconstruction is less than 2 hPa. Moreover, the temporal variability is well captured (r > 0.80 in all datasets). The various reanalysis products have similar MAE (generally from 1-2 hPa), well within the likely observational uncertainty, with CERA-20C having the lowest for MSLP, and 20CRv2c having the lowest for surface pressure. Despite the low MAE / RMSD at this station, adjustment of the reconstruction to sea level pressure may play a small role in its performance, as there are differences in both the historical MSLP and surface pressure, as well as the MSLP and surface pressure in the reanalyses. Nonetheless, this station shows the viability of the reconstruction shortly before the IGY along the Antarctic Peninsula, a region of relatively higher reconstruction skill compared to ERA-Int after 1979 (Fogt *et al.*, 2017a, 2019).

Deck 215 (Fig. 5b) from ICOADS has adequate observational coverage only during austral summer. The reconstruction skill is comparable to the skill seen in the southern Indian Ocean during much of the 1930s (Fig. 3). The reconstruction MAE is 2.5 hPa, slightly higher than all reanalyses but ERA-20C, which was found to be one of the lower performing reanalyses in the early 20th century near Antarctica (Schneider and Fogt, 2018). Despite the higher MAE for the reconstruction (which, unlike the reanalysis products, is entirely independent from these historical observations), the historical observation values nearly all fall within the reconstruction uncertainty (95% confidence interval, gray shading), and the reconstruction uncertainty would clearly overlap with the observational uncertainty throughout time. Together, the comparison with Deck 215 data suggests the original assessment of the reconstruction skill provides a good approximation of its performance in the early 20th century. Although the interannual variability is not as strongly captured as in Fig. 5a (correlation is not provided since only six ~~-~~years of data exist), it should be noted that the reanalysis datasets also do not reproduce the interannual variability well (perhaps due to observational errors or the spatial averaging). We also observe that the surface pressure in 20CRv2c is notably different than the surface pressure from the other reanalyses, probably due to the spectral effects in the 20CRv2c elevation field noted by Slivinski et al. (2019).

For a longer observational record over the Southern Ocean, Deck 899 (Fig. 4, from ICOADS) also has pressure observations only for austral summer but over multiple decades (Fig. 5c). The location varied during these many voyages, but overall the observational estimate of the seasonal mean falls within or very near the reconstruction uncertainty. Indeed, the reconstruction agrees better with the seasonally-averaged ship-based values than the reanalyses (MAE of 2.23 hPa compared to 2.3 – 3.7 hPa from reanalyses, although almost all datasets would fall within the observational uncertainty at this location). The correlations are much lower for ~~this~~ these ship data across all datasets, due to differing fluctuations between a few years (i.e., 1933-1934, 1936-1937, 1945-1946), however the interannual variability is better captured after 1946. In general, the reconstruction also aligns well with the MSLP from the reanalyses, except for DJF 1934 when the ship traversed a large span of the southern Indian Ocean (from 88°E in December 1934 to 47°E in February 1935). Therefore, the use of the seasonal mean location could be artificially suggesting a lower reconstruction skill for this observational estimate. Nonetheless, as in Figs. 5a and 5b, this comparison further suggests that the uncertainty of the reconstruction is a good estimate of the reconstruction error in the early 20th century. Overall, the reconstruction agrees well with many historical seasonal-mean observation values that were withheld during the reconstruction's development. As in Fig. 5b, we note that there are larger differences in the surface pressure from 20CRv2c which clearly fall outside the uncertainty from using a seasonal mean ship location.

Figure 5 presents several comparisons where elevation corrections did not have a noticeable influence on the evaluation of the reconstruction and are perhaps more representative of the overall reconstruction quality. In many other locations, elevation adjustments appear to play a more important role (especially on or near the Antarctic continent) in comparing the reconstruction and observational estimates. At a few of these locations, a simple linear adjustment of 12 hPa per 100m from the ERA-Int elevation at the reconstruction's gridpoint proved sufficient to readily compare it to the historical values. A subset of these stations is presented in Fig. 6.

Perhaps the most famous early American expeditions to Antarctica were those led by Admiral Richard E. Byrd, who set up a station called Little America in several different field campaigns spanning three decades (Byrd, 2003). Although the mean location of Little America was on the Ross Ice Shelf (it varied negligibly during each campaign in comparison to the resolution of the gridded datasets used in this study; Fig. 4), there were notable differences in observed MSLP and surface pressure at the location (Fig. 6a). When adjusting the reconstruction to sea level, a good match is observed both in overall mean pressure (MAE = 3.63 hPa) and interannual variability. Surprisingly, the reconstruction performs much better than the MSLP from 20CRv2c or 20CRv3, most notably due to the much higher MSLP in these two datasets from the 1940s onward compared to the reconstruction and observed values. However, the 20CRv2c surface pressure nearly perfectly matches the observed surface pressure (an MAE of only 2.39 hPa), while 20CRv3 has very similar values for surface and sea level pressure observations. We note that although there are significant offsets from the observed values for most of the datasets (MAE > 6 hPa for both), the interannual variability is well-captured, with the reconstruction correlation of 0.85, and surface pressure correlations from the reanalyses all above 0.85. The consistent difference between the observations and 20CRv3 could be due to elevation issues or to a detected bias in the observations; this will be examined in more detail later. CERA-20C, as seen earlier, performs with the highest skill for MSLP (r=0.9). Also note that the offset from the surface pressure to MSLP is not constant, as this difference, through the hypsometric equation, is influenced by both temperature and humidity. Therefore, some of the larger errors in the reconstruction and the reanalyses, particularly in the 1950s, could be due to too strong of an increase in the surface pressure as it was reduced to sea level, since the gap between surface pressure and sea level pressure is much lower in the historical observations during the 1950s. Observational uncertainties, including the various bias corrections employed by the reanalyses as discussed previously, can also create some of these differences between products when compared to the observations.

Much earlier in Antarctic history, under the leadership of Jean-Baptiste Charcot, the French conducted their second expedition that wintered over on Peterman Island in 1908-1910. The ship associated with this expedition was Porquoi Pas, with a mean location near the

Antarctic Peninsula (Fig. 4). Due to the close proximity of the high terrain of the Antarctic Peninsula, elevation adjustments to the reconstruction were necessary but may not have been sufficient at this location. From historical observations, the difference between surface pressure and sea level pressure ranged from nearly 20 hPa to as little as 5 hPa in DJF 1909 (Fig. 6b). Despite the challenges in correcting for the elevation differences, as seen in Fig. 6b, the adjustment brings the observed MSLP to the far upper-limits of the reconstruction uncertainty, and clearly within the observational uncertainty. Further, the reconstruction agrees well with the observational estimates: the MAE of 3.15 hPa is nearly consistent with time, and the seasonal cycle of the pressure during this portion of the early 20th century is well captured by the reconstruction. It is also encouraging to see high skill in both 20CR datasets, and again in CERA-20C at this location, all within the observational uncertainty; the lower performance of ERA-20C could be related to a cold bias or perhaps resolution issue, as there is a much larger difference between surface pressure (dashed line) and MSLP (solid line) in ERA-20C compared to the observations or CERA-20C. The differences in surface pressure between 20CRv3 and the observations are also noteworthy, with an MAE of 38.60, reflecting differences in model and observed orography.

Similarly situated near the Antarctic Peninsula (Fig. 4), Argentine Island provides a long continuous record much like in Fig. 5a, but with a noticeable influence of elevation-based pressure corrections (Fig. 6c). As with Porquoi Pas, the linear SLP adjustment brings the reconstruction close to the historical estimates, but they fall at the upper-bound of the reconstruction uncertainty, and are much closer to the reanalyses overall. All comparisons are within the expected observational uncertainty. The interannual variability is well-captured by the reconstruction (r= 0.82), although it may be slightly dampened in the early 1950s.

Interestingly, ERA-20C surface pressure aligns closely with the historical observations (with an MAE of 1.63 hPa), but this reanalysis again demonstrates the highest MAE compared to MSLP (4.88 hPa). In contrast, CERA-20C shows the lowest MAE based on MSLP (0.66 hPa), but considerably lower surface pressure than observed (an MAE of 18.29 hPa), though not as low as surface pressure in 20CRv3 (an MAE of 33.42). All reanalyses show very high correlations with observations for both MSLP and surface pressure ($r > 0.85$).

The long, albeit discontinuous record at the Little America station location (Fig. 4) provides an opportunity to further evaluate the reconstruction in comparison to the historical reanalyses, and importantly, to assess the reconstruction skill in light of the reanalyses' internally estimated uncertainties. For the reanalyses studied here, all but ERA-20 are ensemble reanalyses, and therefore the ensemble spread from the seasonal mean ensemble members can be further employed to assess the quality of these products through time and provide a more complete comparison of their skill relative to that of the reconstruction. The MSLP from the closest gridpoint of 20CRv2c, 20CRv3, and CERA-20C are plotted in Fig. 7, along with 95% confidence intervals calculated as 1.96 times the ensemble standard deviation of seasonally averaged MSLP from the 56, 80, and 10 ensemble members of 20CRv2c, 20CRv3, and CERA-20C, respectively. The performance varies slightly at neighboring gridpoints for MSLP, but varies considerably if using surface pressure, suggesting elevation gradients from the nearby Roosevelt Island or the edge of the Ross Ice Shelf can influence the reanalysis ~~solution~~ estimate at this location.

Similar to but slightly larger than the reconstruction, Fig. 7 shows that CERA-20C has a pronounced annual cycle of MSLP at Little America, while both 20CRv2C and 20CRv3 have a dampened annual cycle. Nonetheless, Fig. 7 also demonstrates that there is a marked decrease in the reanalyses' uncertainties at the times when the Little America data were assimilated, which effectively constrained the reanalysis ~~solutions~~estimates. There are also other reductions in the reanalysis uncertainties prior to the establishment of Little America, including the data assimilated from the South Pole race of 1911-1912 (the Norwegian base Framheim was very close to Little America (Fogt *et al.*, 2017b)) and the data from the first British Antarctic Expedition (1908-1909, which established a base near present day McMurdo station west of Little America, at Cape Royds). At these times, the reanalyses and reconstruction show considerable agreement, and the reanalysis uncertainty (colored shading in Fig. 7) overlaps the reconstruction uncertainty (gray shading in Fig. 7). Exceptions are in 1929 for all reanalyses and in 1911 for 20CRv2c and 20CRv3 when the reconstruction and reanalysis uncertainties do not overlap, with the reconstruction seasonal mean MSLP lower than the reanalyses. Nonetheless, the overall agreement at all other times suggests that although there may still be larger MAE in the reanalyses MSLP when compared only during the times of direct observations (as in Figs. 5-6), the observations and reconstruction both fall within the reanalyses uncertainty (estimated by the ensemble spread), and the reanalyses have benefited greatly from assimilating this historical data (moving much closer to, if not within, the observational uncertainty). Clearly, at other periods when data are not available, the reanalysis spread is considerably larger, but the reconstruction nearly always falls within the reanalyses' uncertainties (the high positive pressure values in CERA-20C in 1944 are one exception). We note that the reanalysis uncertainty estimates themselves require further improvement, particularly in the Southern Hemisphere: Slivinski et al (2019) show artificial signals in the uncertainty of 20CRv2c, and demonstrate that the uncertainty in this region still remains too large in 20CRv3. Conversely, Laloyaux *et al.* (2018) acknowledge that the small ensemble of CERA-20C can result in overly-confident estimates of uncertainty. Consistent with the comparison to observations (Fig. 6a), the CERA-20C MSLP and reconstruction (Fig. 7c) show the most similarity throughout the early 20th century at the Little America location (RMSD = 4.32 hPa), and for this location 20CRv3 performs better than 20CRv2c (RMSD of 5.95 hPa compared to 8.22 hPa).

Evaluation of the reconstruction skill in Figs. 2 and 3 often indicates substantial differences, suggesting that the seasonal mean values from the historical observations do not always fall within the reconstruction uncertainty when elevation corrections are applied, particularly near high terrain (even when their uncertainty is accounted for; not shown). To examine this issue further, Figure 8 depicts two cases – a set of Antarctic expedition ~~a~~ ship records (Deck 246, Fig. 8a) and a temporary base (Cape Denison, Fig. 8b), where elevation corrections have mixed results; MAE values in Fig. 8 are only calculated for MSLP. Ships included in Deck 246 (from ICOADS) operated discontinuously near present day Dumont d'Urville station in the Ross Sea sector of Antarctica (Fig. 4); the data from ICOADS have two disparate locations in January – February 1912 (the mean of these would be over the continent), so instead of averaging data from both locations, each were treated as separate data points in Fig. 8a and the comparison statistics. The two different values for MAE in Fig. 8a are therefore based on the MSLP at these two different locations, with the data closer to the Ross Ice Shelf (second value) agreeing better with all gridded datasets than the data closer to the East Antarctic plateau, near the location plotted in Fig. 4. Even with this potentially conflicting information, the reconstruction performs very well after adjustment to sea level pressure, with an MAE of 2.32 / 1.85. A closer look shows that the majority of this error is from DJF 1910, when the reconstruction was about 6 hPa lower than the observational estimate, otherwise it is generally within 1.5 hPa. This performance exceeds the reanalyses' performance for Deck 246, as the reanalyses typically have much higher MSLP values.

In contrast, the famous base for the Australian Antarctic Expedition, Cape Denison (Mawson, 1998), situated along the East Antarctic coast south of Australia (Fig. 4) is marked with much lower performance across all datasets (Figure 8b). Elevation corrections and the models' orography have a strong effect here, clearly demonstrated by the large differences in surface and sea level pressure in observations and reanalyses (note, surface pressure is plotted on the right axis, but because of the large differences in surface pressure the MAE is not given). Whereas the seasonal means of the historical observations demonstrate surface pressures around 920 hPa on average, sea level pressure estimates from the station are around 70 hPa higher, at around 990 hPa. Such high differences in the two over the cold ice sheet plateau compromise the quality of the reduced sea level pressure (not only in observations, but across all datasets). The MAE for the reconstruction (13.50 hPa) is one of the highest of all locations examined. Although the reanalyses' errors are nearly half of this (around 5-7 hPa, likely improved by making further use of temperature and humidity calculated within the reanalyses to reduce surface pressure to sea level), the errors are still large, and the reanalyses and reconstructions all likely fall outside the observational uncertainty. Furthermore, it is difficult to assess how well the interannual variability of the MSLP is reproduced, since there are fewer than ten observations and the reduction to sea level varies considerably by season (affected by temperature and humidity), but the reconstruction was adjusted to sea level uniformly (and linearly) across all seasons. Nonetheless, the limited comparison demonstrates that overall there is still good agreement in the surface pressure interannual variability (despite large differences in magnitude) between the reanalyses and the reconstruction, which potentially suggests that the reconstruction is still capturing aspects of the pressure variability at this location in the early 20th century. Importantly, due to the influence of crude elevation adjustments, it is highly likely that the reconstruction is performing better than indicated by the MAE (Fig. 8b) or RMSD (Fig. 4). Unfortunately, the reconstruction skill is difficult to precisely determine at this or any other location (as suggested in reviewing Fig. 2) where large elevation adjustments make the comparison to historical estimates challenging.

4. Discussion and Conclusions

The analysis presented here has compared seasonal Antarctic pressure reconstructions to numerous historical observations and reanalyses throughout the early 20th century. As none of these historical data were included in the reconstruction calibration, they serve as an independent evaluation of the reconstruction skill during a period when relatively little is known about Antarctic climate variability.

The results overall confirm that the reconstruction error and uncertainty assessed in earlier work (Fogt *et al.*, 2017b, 2019), a mean absolute error of around 2-4 hPa across the Southern Ocean, is supported when comparing with ship observations. A few ship records suggest even higher reconstruction skill (MAE less than 2 hPa), while others situated north of the Amundsen and Bellingshausen Seas demonstrate lower skill (MAE above 4 hPa), consistent with earlier work. Furthermore, most comparisons with ship observations that span multiple seasons indicate the reconstruction also captures the interannual variability well (correlations often greater than 0.80). Comparison with historical reanalyses provide further evaluation of the reconstruction's performance, as they assimilate most of the historical observations used here but are independent of the reconstruction.

While the comparisons with observation estimates taken at sea level are relatively straightforward and further validate the reliability of the early portions of the reconstruction, it is far more challenging to make assessments of the reconstruction skill over areas of higher elevation or near the coastline. In many of these locations, reduction of the reconstruction pressure (which was constructed as surface pressure anomalies relative to the ERA-Int model topography) to sea level pressure using a simple linear adjustment does not satisfactorily agree with the historical data, even after considering the potential observational error. In many of these locations, there are also large differences between 20th century reanalysis MSLP and the historical observational estimates, highlighting the reduced reliability of MSLP from all sources over the cold, high Antarctic continent. While the general comparisons conducted here suggest a larger reconstruction MAE of nearly 6 hPa or more over high elevation, a closer examination at select locations reveals that the reconstruction uncertainty is likely much lower than this and perhaps as low as 1-3 hPa, as indicated in Fogt *et al.* (2019).

This work has demonstrated the value of digitizing historical observations from ships and temporary bases for both understanding long term change across the high southern latitudes and evaluating gridded datasets. While their temporary nature may make them difficult to use for assessing long-term variability and change, when coupled with gridded climate datasets like the seasonal Antarctic pressure reconstructions evaluated here, they serve as an independent and valuable tool of documenting historical climate. As one recent example, the use of newly digitized historical observations and pressure reconstructions shed new light on exceptional conditions during the South Pole race of 1911-1912 (Fogt *et al.*, 2017c, 2018; Sienicki, 2018). Future work will hopefully continue to unlock the power of these and other historical observations, so that the ongoing change across the high southern latitudes can be placed in a much-needed longer historical context (Jones *et al.*, 2016).

Acknowledgments

Data from the Antarctic pressure reconstructions are available from figshare (https://doi.org/10.6084/m9.figshare.5325541). RLF and CPB acknowledge support from the National Science Foundation (NSF), grant PLR-1341621 and ANT-1744998. JMJ acknowledges support from the Leverhulme Trust through a research Fellowship (RF-2018-183). Support for the Twentieth Century Reanalysis Project is provided by the U.S. Department of Energy, Office of Science Biological and Environmental Research, by the National Oceanic and Atmospheric Administration Climate Program Office, and by the NOAA ~~Earth System Research Laboratory~~ Physical Sciences ~~Division~~Laboratory.

References

Agosta C, Amory C, Kittel C, Orsi A, Favier V, Gallée H, van den Broeke MR, Lenaerts JTM, van Wessem JM, van de Berg WJ, Fettweis X. 2019. Estimation of the Antarctic surface mass balance using the regional climate model MAR (1979–2015) and identification of dominant processes. *The Cryosphere*, 13(1): 281–296. https://doi.org/10.5194/tc-13-281-2019.

Allan R, Ansell T. 2006. A new globally complete monthly historical gridded mean sea level pressure dataset (HadSLP2): 1850–2004. *Journal of Climate*, 19(22): 5816–5842.

Allan R, Brohan P, Compo GP, Stone R, Luterbacher J, Brönnimann S. 2011. The International Atmospheric Circulation Reconstructions over the Earth (ACRE) Initiative. *Bulletin of the American Meteorological Society*, 92(11): 1421–1425. https://doi.org/10.1175/2011BAMS3218.1.

Bracegirdle TJ, Colleoni F, Abram NJ, Bertler NAN, Dixon DA, England M, Favier V, Fogwill CJ, Fyfe JC, Goodwin I, Goosse H, Hobbs W, Jones JM, Keller ED, Khan AL, Phipps SJ, Raphael MN, Russell J, Sime L, Thomas ER, van den Broeke MR, Wainer I. 2019. Back to the Future: Using Long-Term Observational and Paleo-Proxy Reconstructions to Improve Model Projections of Antarctic Climate. *Geosciences*, 9(6): 255. https://doi.org/10.3390/geosciences9060255.

Bromwich DH, Nicolas JP, Monaghan AJ, Lazzara MA, Keller LM, Weidner GA, Wilson AB. 2012. Central West Antarctica among the most rapidly warming regions on Earth. *Nature Geoscience*, 6(2): 139–145. https://doi.org/10.1038/ngeo1671.

Byrd RE. 2003. *Alone: the classic polar adventure*. Island Press/Shearwater Books: Washington, DC.

Cavalieri DJ, Parkinson CL. 2008. Antarctic sea ice variability and trends, 1979–2006. *Journal of Geophysical Research*, 113(C7). https://doi.org/10.1029/2007JC004564.

Clark L, Fogt R. 2019. Southern Hemisphere Pressure Relationships during the 20th Century—Implications for Climate Reconstructions and Model Evaluation. *Geosciences*, 9(10): 413. https://doi.org/10.3390/geosciences9100413.

Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N, Allan RJ, Yin X, Gleason BE, Vose RS, Rutledge G, Bessemoulin P, Brönnimann S, Brunet M, Crouthamel RI, Grant AN, Groisman PY, Jones PD, Kruk MC, Kruger AC, Marshall GJ, Maugeri M, Mok HY, Nordli ø., Ross TF, Trigo RM, Wang XL, Woodruff SD, Worley SJ. 2011. The Twentieth Century Reanalysis Project. *Quarterly Journal of the Royal Meteorological Society*, 137(654): 1–28. https://doi.org/10.1002/qj.776.

Cram TA, Compo GP, Yin X, Allan RJ, McColl C, Vose RS, Whitaker JS, Matsui N, Ashcroft L, Auchmann R, Bessemoulin P, Brandsma T, Brohan P, Brunet M, Comeaux J, Crouthamel R, Gleason BE, Groisman PY, Hersbach H, Jones PD, Jonsson T, Jourdain S, Kelly G, Knapp KR, Kruger A, Kubota H, Lentini G, Lorrey A, Lott N, Lubker SJ, Luterbacher J, Marshall GJ, Maugeri

M, Mock CJ, Mok HY, Nordli O, Rodwell MJ, Ross TF, Schuster D, Srnec L, Valente MA, Vizi Z, Wang XL, Westcott N, Woollen JS, Worley SJ. 2015. The International Surface Pressure Databank version 2. *Geoscience Data Journal*, 2(1): 31–46. https://doi.org/10.1002/gdj3.25.

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette J-J, Park B-K, Peubey C, de Rosnay P, Tavolato C, Thépaut J-N, Vitart F. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656): 553–597. https://doi.org/10.1002/qj.828.

Edwards TL, Brandon MA, Durand G, Edwards NR, Golledge NR, Holden PB, Nias IJ, Payne AJ, Ritz C, Wernecke A. 2019. Revisiting Antarctic ice loss due to marine ice-cliff instability. *Nature*, 566(7742): 58–64. https://doi.org/10.1038/s41586-019-0901-4.

Fogt RL, Goergens CA, Jones JM, Schneider DP, Nicolas JP, Bromwich DH, Dusselier HE. 2017a. A twentieth century perspective on summer Antarctic pressure change and variability and contributions from tropical SSTs and ozone depletion. *Geophysical Research Letters*, 44(19): 9918–9927. https://doi.org/10.1002/2017GL075079.

Fogt RL, Goergens CA, Jones ME, Witte GA, Lee MY, Jones JM. 2016a. Antarctic station-based seasonal pressure reconstructions since 1905: 1. Reconstruction evaluation: Antarctic Pressure Evaluation. *Journal of Geophysical Research: Atmospheres*, 121(6): 2814–2835. https://doi.org/10.1002/2015JD024564.

Fogt RL, Jones JM, Goergens CA, Jones ME, Witte GA, Lee MY. 2016b. Antarctic station-based seasonal pressure reconstructions since 1905: 2. Variability and trends during the twentieth century. *Journal of Geophysical Research: Atmospheres*, 121(6): 2836–2856. https://doi.org/10.1002/2015JD024565.

Fogt RL, Jones ME, Goergens CA, Solomon S, Jones JM. 2018. Reply to "Comment on 'An Exceptional Summer during the South Pole Race of 1911/12.'" *Bulletin of the American Meteorological Society*, 99(10): 2143–2145. https://doi.org/10.1175/BAMS-D-18-0088.1.

Fogt RL, Jones ME, Solomon S, Jones JM, Goergens CA. 2017b. An Exceptional Summer during the South Pole Race of 1911-1912. *Bulletin of the American Meteorological Society*. https://doi.org/10.1175/BAMS-D-17-0013.1.

Fogt RL, Jones ME, Solomon S, Jones JM, Goergens CA. 2017c. An Exceptional Summer during the South Pole Race of 1911-1912. *Bulletin of the American Meteorological Society*, 98. https://doi.org/10.1175/BAMS-D-17-0013.1.

Fogt RL, Schneider DP, Goergens CA, Jones JM, Clark LN, Garberoglio MJ. 2019. Seasonal Antarctic pressure variability during the twentieth century from spatially complete

reconstructions and CAM5 simulations. *Climate Dynamics*. https://doi.org/10.1007/s00382-019-04674-8.

Freeman E, Woodruff SD, Worley SJ, Lubker SJ, Kent EC, Angel WE, Berry DI, Brohan P, Eastman R, Gates L, Gloeden W, Ji Z, Lawrimore J, Rayner NA, Rosenhagen G, Smith SR. 2017. ICOADS Release 3.0: a major update to the historical marine climate record. *International Journal of Climatology*, 37(5): 2211–2232. https://doi.org/10.1002/joc.4775.

Jones JM, Gille ST, Goosse H, Abram NJ, Canziani PO, Charman DJ, Clem KR, Crosta X, de Lavergne C, Eisenman I, England MH, Fogt RL, Frankcombe LM, Marshall GJ, Masson-Delmotte V, Morrison AK, Orsi AJ, Raphael MN, Renwick JA, Schneider DP, Simpkins GR, Steig EJ, Stenni B, Swingedouw D, Vance TR. 2016. Assessing recent trends in high-latitude Southern Hemisphere surface climate. *Nature Climate Change*, 6(10): 917–926. https://doi.org/10.1038/nclimate3103.

Jones ME, Bromwich DH, Nicolas JP, Carrasco J, Plavcová E, Zou X, Wang S-H. 2019. Sixty Years of Widespread Warming in the Southern Middle and High Latitudes (1957–2016). *Journal of Climate*, 32(20): 6875–6898. https://doi.org/10.1175/JCLI-D-18-0565.1.

Kent EC, Berry DI. 2005. Quantifying random measurement errors in Voluntary Observing Ships' meteorological observations. *International Journal of Climatology*, 25(7): 843–856. https://doi.org/10.1002/joc.1167.

Laloyaux P, de Boisseson E, Balmaseda M, Bidlot J-R, Broennimann S, Buizza R, Dalhgren P, Dee D, Haimberger L, Hersbach H, Kosaka Y, Martin M, Poli P, Rayner N, Rustemeier E, Schepers D. 2018. CERA-20C: A Coupled Reanalysis of the Twentieth Century. *Journal of Advances in Modeling Earth Systems*, 10(5): 1172–1195. https://doi.org/10.1029/2018MS001273.

Mawson D. 1998. *The home of the blizzard: a true story of Antarctic survival*. St. Martin's Press: New York.

Mayewski PA, Frezzotti M, Bertler N, Ommen TV, Hamilton G, Jacka TH, Welch B, Frey M, Dahe Q, Jiawen R, Simões J, Fily M, Oerter H, Nishio F, Isaksson E, Mulvaney R, Holmund P, Lipenkov V, Goodwin I. 2005. The International Trans-Antarctic Scientific Expedition (ITASE): an overview. *Annals of Glaciology*, 41: 180–185. https://doi.org/10.3189/172756405781813159.

Poli P, Hersbach H, Dee DP, Berrisford P, Simmons AJ, Vitart F, Laloyaux P, Tan DGH, Peubey C, Thépaut J-N, Trémolet Y, Hólm EV, Bonavita M, Isaksen L, Fisher M. 2016. ERA-20C: An Atmospheric Reanalysis of the Twentieth Century. *Journal of Climate*, 29(11): 4083–4097. https://doi.org/10.1175/JCLI-D-15-0556.1.

Purich A, England MH. 2019. Tropical teleconnections to Antarctic sea ice during austral spring 2016 in coupled pacemaker experiments. *Geophysical Research Letters*. https://doi.org/10.1029/2019GL082671.

Rignot E, Jacobs S, Mouginot J, Scheuchl B. 2013. Ice-Shelf Melting Around Antarctica. *Science*, 341(6143): 266–270. https://doi.org/10.1126/science.1235798.

Rignot E, Mouginot J, Scheuchl B, van den Broeke M, van Wessem MJ, Morlighem M. 2019. Four decades of Antarctic Ice Sheet mass balance from 1979–2017. *Proceedings of the National Academy of Sciences*, 116(4): 1095–1103. https://doi.org/10.1073/pnas.1812883116.

Schneider DP, Fogt RL. 2018. Artifacts in Century-Length Atmospheric and Coupled Reanalyses Over Antarctica Due to Historical Data Availability. *Geophysical Research Letters*, 45(2): 964–973. https://doi.org/10.1002/2017GL076226.

Sienicki K. 2018. Comments on "An Exceptional Summer during the South Pole Race of 1911/12." *Bulletin of the American Meteorological Society*, 99(10): 2139–2143. https://doi.org/10.1175/BAMS-D-17-0282.1.

Slivinski LC, Compo GP, Whitaker JS, Sardeshmukh PD, Giese BS, McColl C, Allan R, Yin X, Vose R, Titchner H, Kennedy J, Spencer LJ, Ashcroft L, Brönnimann S, Brunet M, Camuffo D, Cornes R, Cram TA, Crouthamel R, Domínguez-Castro F, Freeman JE, Gergis J, Hawkins E, Jones PD, Jourdain S, Kaplan A, Kubota H, Le Blancq F, Lee T, Lorrey A, Luterbacher J, Maugeri M, Mock CJ, Moore GWK, Przybylak R, Pudmenzky C, Reason C, Slonosky VC, Smith C, Tinz B, Trewin B, Valente MA, Wang XL, Wilkinson C, Wood K, Wyszyn'ski P. 2019. Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. *Quarterly Journal of the Royal Meteorological Society*. https://doi.org/10.1002/qj.3598.

Steig EJ, Schneider DP, Rutherford SD, Mann ME, Comiso JC, Shindell DT. 2009. Warming of the Antarctic ice-sheet surface since the 1957 International Geophysical Year. *Nature*, 457(7228): 459–462. https://doi.org/10.1038/nature07669.

Stenni B, Curran MAJ, Abram NJ, Orsi A, Goursaud S, Masson-Delmotte V, Neukom R, Goosse H, Divine D, van Ommen T, Steig EJ, Dixon DA, Thomas ER, Bertler NAN, Isaksson E, Ekaykin A, Werner M, Frezzotti M. 2017. Antarctic climate variability on regional and continental scales over the last 2000 years. *Climate of the Past*, 13(11): 1609–1634. https://doi.org/10.5194/cp-13-1609-2017.

Stuecker MF, Bitz CM, Armour KC. 2017. Conditions leading to the unprecedented low Antarctic sea ice extent during the 2016 austral spring season. *Geophysical Research Letters*, 44(17): 9008–9019. https://doi.org/10.1002/2017GL074691.

Thomas ER, Marshall GJ, McConnell JR. 2008. A doubling in snow accumulation in the western Antarctic Peninsula since 1850. *Geophysical Research Letters*, 35(1). https://doi.org/10.1029/2007GL032529.

Turner J, Marshall GJ, Clem K, Colwell S, Phillips T, Lu H. 2019. Antarctic temperature variability and change from station data. *International Journal of Climatology*. https://doi.org/10.1002/joc.6378.

Turner J, Phillips T, Marshall GJ, Hosking JS, Pope JO, Bracegirdle TJ, Deb P. 2017. Unprecedented springtime retreat of Antarctic sea ice in 2016: The 2016 Antarctic Sea Ice Retreat. *Geophysical Research Letters*, 44(13): 6868–6875. https://doi.org/10.1002/2017GL073656.

Agosta C, Amory C, Kittel C, Orsi A, Favier V, Gallée H, van den Broeke MR, Lenaerts JTM, van Wessem JM, van de Berg WJ, Fettweis X. 2019. Estimation of the Antarctic surface mass balance using the regional climate model MAR (1979–2015) and identification of dominant processes. *The Cryosphere*, 13(1): 281–296. https://doi.org/10.5194/tc-13-281-2019.

Allan R, Ansell T. 2006. A new globally complete monthly historical gridded mean sea level pressure dataset (HadSLP2): 1850–2004. *Journal of Climate*, 19(22): 5816–5842.

Allan R, Brohan P, Compo GP, Stone R, Luterbacher J, Brönnimann S. 2011. The International Atmospheric Circulation Reconstructions over the Earth (ACRE) Initiative. *Bulletin of the American Meteorological Society*, 92(11): 1421–1425. https://doi.org/10.1175/2011BAMS3218.1.

Bracegirdle TJ, Colleoni F, Abram NJ, Bertler NAN, Dixon DA, England M, Favier V, Fogwill CJ, Fyfe JC, Goodwin I, Goosse H, Hobbs W, Jones JM, Keller ED, Khan AL, Phipps SJ, Raphael MN, Russell J, Sime L, Thomas ER, van den Broeke MR, Wainer I. 2019. Back to the Future: Using Long-Term Observational and Paleo-Proxy Reconstructions to Improve Model Projections of Antarctic Climate. *Geosciences*, 9(6): 255. https://doi.org/10.3390/geosciences9060255.

Bromwich DH, Nicolas JP, Monaghan AJ, Lazzara MA, Keller LM, Weidner GA, Wilson AB. 2012. Central West Antarctica among the most rapidly warming regions on Earth. *Nature Geoscience*, 6(2): 139–145. https://doi.org/10.1038/ngeo1671.

Byrd RE. 2003. *Alone: the classic polar adventure*. Island Press/Shearwater Books: Washington, DC.

Cavalieri DJ, Parkinson CL. 2008. Antarctic sea ice variability and trends, 1979–2006. *Journal of Geophysical Research*, 113(C7). https://doi.org/10.1029/2007JC004564.

Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N, Allan RJ, Yin X, Gleason BE, Vose RS, Rutledge G, Bessemoulin P, Brönnimann S, Brunet M, Crouthamel RI, Grant AN, Groisman PY, Jones PD, Kruk MC, Kruger AC, Marshall GJ, Maugeri M, Mok HY, Nordli ø., Ross TF, Trigo RM, Wang XL, Woodruff SD, Worley SJ. 2011. The Twentieth Century Reanalysis Project. *Quarterly Journal of the Royal Meteorological Society*, 137(654): 1–28. https://doi.org/10.1002/qj.776.

Cram TA, Compo GP, Yin X, Allan RJ, McColl C, Vose RS, Whitaker JS, Matsui N, Ashcroft L, Auchmann R, Bessemoulin P, Brandsma T, Brohan P, Brunet M, Comeaux J, Crouthamel R, Gleason BE, Groisman PY, Hersbach H, Jones PD, Jonsson T, Jourdain S, Kelly G, Knapp KR, Kruger A, Kubota H, Lentini G, Lorrey A, Lott N, Lubker SJ, Luterbacher J, Marshall GJ, Maugeri M, Mock CJ, Mok HY, Nordli O, Rodwell MJ, Ross TF, Schuster D, Srnec L, Valente MA, Vizi Z, Wang XL, Westcott N, Woollen JS, Worley SJ. 2015. The International Surface Pressure Databank version 2. *Geoscience Data Journal*, 2(1): 31–46. https://doi.org/10.1002/gdj3.25.

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette J-J, Park B-K, Peubey C, de Rosnay P, Tavolato C, Thépaut J-N, Vitart F. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656): 553–597. https://doi.org/10.1002/qj.828.

Edwards TL, Brandon MA, Durand G, Edwards NR, Golledge NR, Holden PB, Nias IJ, Payne AJ, Ritz C, Wernecke A. 2019. Revisiting Antarctic ice loss due to marine ice-cliff instability. *Nature*, 566(7742): 58–64. https://doi.org/10.1038/s41586-019-0901-4.

Fogt RL, Goergens CA, Jones JM, Schneider DP, Nicolas JP, Bromwich DH, Dusselier HE. 2017a. A twentieth century perspective on summer Antarctic pressure change and variability and contributions from tropical SSTs and ozone depletion. *Geophysical Research Letters*, 44(19): 9918–9927. https://doi.org/10.1002/2017GL075079.

Fogt RL, Goergens CA, Jones ME, Witte GA, Lee MY, Jones JM. 2016a. Antarctic station-based seasonal pressure reconstructions since 1905: 1. Reconstruction evaluation: Antarctic Pressure Evaluation. *Journal of Geophysical Research: Atmospheres*, 121(6): 2814–2835. https://doi.org/10.1002/2015JD024564.

Fogt RL, Jones JM, Goergens CA, Jones ME, Witte GA, Lee MY. 2016b. Antarctic station-based seasonal pressure reconstructions since 1905: 2. Variability and trends during the twentieth century. *Journal of Geophysical Research: Atmospheres*, 121(6): 2836–2856. https://doi.org/10.1002/2015JD024565.

Fogt RL, Jones ME, Goergens CA, Solomon S, Jones JM. 2018. Reply to "Comment on 'An Exceptional Summer during the South Pole Race of 1911/12.'" *Bulletin of the American Meteorological Society*, 99(10): 2143–2145. https://doi.org/10.1175/BAMS-D-18-0088.1.

Fogt RL, Jones ME, Solomon S, Jones JM, Goergens CA. 2017b. An Exceptional Summer during the South Pole Race of 1911-1912. *Bulletin of the American Meteorological Society*. https://doi.org/10.1175/BAMS-D-17-0013.1.

Fogt RL, Jones ME, Solomon S, Jones JM, Goergens CA. 2017c. An Exceptional Summer during the South Pole Race of 1911-1912. *Bulletin of the American Meteorological Society*, 98. https://doi.org/10.1175/BAMS-D-17-0013.1.

Fogt RL, Schneider DP, Goergens CA, Jones JM, Clark LN, Garberoglio MJ. 2019. Seasonal Antarctic pressure variability during the twentieth century from spatially complete

reconstructions and CAM5 simulations. *Climate Dynamics*. https://doi.org/10.1007/s00382-019-04674-8.

Freeman E, Woodruff SD, Worley SJ, Lubker SJ, Kent EC, Angel WE, Berry DI, Brohan P, Eastman R, Gates L, Gloeden W, Ji Z, Lawrimore J, Rayner NA, Rosenhagen G, Smith SR. 2017. ICOADS Release 3.0: a major update to the historical marine climate record. *International Journal of Climatology*, 37(5): 2211–2232. https://doi.org/10.1002/joc.4775.

Jones JM, Gille ST, Goosse H, Abram NJ, Canziani PO, Charman DJ, Clem KR, Crosta X, de Lavergne C, Eisenman I, England MH, Fogt RL, Frankcombe LM, Marshall GJ, Masson-Delmotte V, Morrison AK, Orsi AJ, Raphael MN, Renwick JA, Schneider DP, Simpkins GR, Steig EJ, Stenni B, Swingedouw D, Vance TR. 2016. Assessing recent trends in high-latitude Southern Hemisphere surface climate. *Nature Climate Change*, 6(10): 917–926. https://doi.org/10.1038/nclimate3103.

Jones ME, Bromwich DH, Nicolas JP, Carrasco J, Plavcová E, Zou X, Wang S-H. 2019. Sixty Years of Widespread Warming in the Southern Middle and High Latitudes (1957–2016). *Journal of Climate*, 32(20): 6875–6898. https://doi.org/10.1175/JCLI-D-18-0565.1.

Laloyaux P, de Boisseson E, Balmaseda M, Bidlot J-R, Broennimann S, Buizza R, Dalhgren P, Dee D, Haimberger L, Hersbach H, Kosaka Y, Martin M, Poli P, Rayner N, Rustemeier E, Schepers D. 2018. CERA-20C: A Coupled Reanalysis of the Twentieth Century. *Journal of Advances in Modeling Earth Systems*, 10(5): 1172–1195. https://doi.org/10.1029/2018MS001273.

Mawson D. 1998. *The home of the blizzard: a true story of Antarctic survival*. St. Martin's Press: New York.

Mayewski PA, Frezzotti M, Bertler N, Ommen TV, Hamilton G, Jacka TH, Welch B, Frey M, Dahe Q, Jiawen R, Simões J, Fily M, Oerter H, Nishio F, Isaksson E, Mulvaney R, Holmund P, Lipenkov V, Goodwin I. 2005. The International Trans-Antarctic Scientific Expedition (ITASE): an overview. *Annals of Glaciology*, 41: 180–185. https://doi.org/10.3189/172756405781813159.

Poli P, Hersbach H, Dee DP, Berrisford P, Simmons AJ, Vitart F, Laloyaux P, Tan DGH, Peubey C, Thépaut J-N, Trémolet Y, Hólm EV, Bonavita M, Isaksen L, Fisher M. 2016. ERA-20C: An Atmospheric Reanalysis of the Twentieth Century. *Journal of Climate*, 29(11): 4083–4097. https://doi.org/10.1175/JCLI-D-15-0556.1.

Purich A, England MH. 2019. Tropical teleconnections to Antarctic sea ice during austral spring 2016 in coupled pacemaker experiments. *Geophysical Research Letters*. https://doi.org/10.1029/2019GL082671.

Rignot E, Jacobs S, Mouginot J, Scheuchl B. 2013. Ice-Shelf Melting Around Antarctica. *Science*, 341(6143): 266–270. https://doi.org/10.1126/science.1235798.

Rignot E, Mouginot J, Scheuchl B, van den Broeke M, van Wessem MJ, Morlighem M. 2019. Four decades of Antarctic Ice Sheet mass balance from 1979–2017. *Proceedings of the National Academy of Sciences*, 116(4): 1095–1103. https://doi.org/10.1073/pnas.1812883116.

Schneider DP, Fogt RL. 2018. Artifacts in Century-Length Atmospheric and Coupled Reanalyses Over Antarctica Due to Historical Data Availability. *Geophysical Research Letters*, 45(2): 964–973. https://doi.org/10.1002/2017GL076226.

Sienicki K. 2018. Comments on "An Exceptional Summer during the South Pole Race of 1911/12." *Bulletin of the American Meteorological Society*, 99(10): 2139–2143. https://doi.org/10.1175/BAMS-D-17-0282.1.

Slivinski LC, Compo GP, Whitaker JS, Sardeshmukh PD, Giese BS, McColl C, Allan R, Yin X, Vose R, Titchner H, Kennedy J, Spencer LJ, Ashcroft L, Brönnimann S, Brunet M, Camuffo D, Cornes R, Cram TA, Crouthamel R, Domínguez-Castro F, Freeman JE, Gergis J, Hawkins E, Jones PD, Jourdain S, Kaplan A, Kubota H, Le Blancq F, Lee T, Lorrey A, Luterbacher J, Maugeri M, Mock CJ, Moore GWK, Przybylak R, Pudmenzky C, Reason C, Slonosky VC, Smith C, Tinz B, Trewin B, Valente MA, Wang XL, Wilkinson C, Wood K, Wyszyn'ski P. 2019. Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. *Quarterly Journal of the Royal Meteorological Society*. https://doi.org/10.1002/qj.3598.

Steig EJ, Schneider DP, Rutherford SD, Mann ME, Comiso JC, Shindell DT. 2009. Warming of the Antarctic ice-sheet surface since the 1957 International Geophysical Year. *Nature*, 457(7228): 459–462. https://doi.org/10.1038/nature07669.

Stenni B, Curran MAJ, Abram NJ, Orsi A, Goursaud S, Masson-Delmotte V, Neukom R, Goosse H, Divine D, van Ommen T, Steig EJ, Dixon DA, Thomas ER, Bertler NAN, Isaksson E, Ekaykin A, Werner M, Frezzotti M. 2017. Antarctic climate variability on regional and continental scales over the last 2000 years. *Climate of the Past*, 13(11): 1609–1634. https://doi.org/10.5194/cp-13-1609-2017.

Stuecker MF, Bitz CM, Armour KC. 2017. Conditions leading to the unprecedented low Antarctic sea ice extent during the 2016 austral spring season. *Geophysical Research Letters*, 44(17): 9008–9019. https://doi.org/10.1002/2017GL074691.

Thomas ER, Marshall GJ, McConnell JR. 2008. A doubling in snow accumulation in the western Antarctic Peninsula since 1850. *Geophysical Research Letters*, 35(1). https://doi.org/10.1029/2007GL032529.

Turner J, Marshall GJ, Clem K, Colwell S, Phillips T, Lu H. 2019. Antarctic temperature variability and change from station data. *International Journal of Climatology*. https://doi.org/10.1002/joc.6378.

Turner J, Phillips T, Marshall GJ, Hosking JS, Pope JO, Bracegirdle TJ, Deb P. 2017. Unprecedented springtime retreat of Antarctic sea ice in 2016: The 2016 Antarctic Sea Ice

~~Retreat. *Geophysical Research Letters*, 44(13): 6868–6875.
https://doi.org/10.1002/2017GL073656.~~

Figure Captions

Figure 1. Maps of seasonal mean data location grouped by decade. Open circles represent seasonal mean locations from ship records, and filled circles are for temporary bases on the continent. The bottom left plot shows the location of all 271 seasonal mean observations compared, while the bottom right plot shows the locations of the 18 station reconstructions (brown) and observations from Orcadas (grey) used to generate the spatially complete pressure reconstruction. Contours on the bottom right panel are the standard deviations of monthly ERA5 surface pressure anomalies for reference, contoured every 0.5 hPa.~~Map of seasonal mean data location grouped by decade, with the bottom plot showing the location of all 271 seasonal mean observations compared. Open circles represent seasonal mean locations from ship records.~~

Figure 2. Mean reconstruction skill statistics (columns; bias, MAE, and RMSD) compared to historical observations, and averaged over various latitudes (top row), longitudes (middle row), and decades (bottom row). The statistics calculated over all observations are listed at the bottom of the figure.

Figure 3. Decadal mean RMSD plotted by decade.

Figure 4. Map showing observational mean (over full length of record) RMSD for select representative locations examined in more detail.

Figure 5. Time series of historical observations, reconstruction (with 95% confidence interval in grey shading), and gridded reanalysis problems for observations representative of the lowest RMSD (solid lines for MSLP, dashed lines for surface pressure). The name is the record identifier provided in ISPD or ICOADS. In a), the x-axis varies by season, and the labels represent the DJF seasons for each year; in b) and c) only DJF data are plotted and the label represents the DJF season. The white spaces in c) represent discontinuities in the observations. The values at the bottom of each panel are the correlations (if more than 10 data points are available) and MAE values (first numbers based on MSLP, second (where available) based on surface pressure) for each dataset. The data for station 889340 were from ISPD, while Deck 215 and Deck 899 were from ICOADS.

Figure 6. As in Fig. 5, but representative of stations where elevation corrections (reduction to sea level pressure) play an important aspect of the reconstruction performance evaluation. All data in this figure were obtained from ISPD.

Figure 7. Time series of seasonal mean MSLP for all four seasons at Little America (from ISPD) on the northern edge of the Ross Ice Shelf for the reconstruction along with values from a) 20CRv2c; b) 20CRv3; c) CERA-20C. The gray shading in each panel represents the 95% confidence interval for the reconstruction, while the colored shading represents 95% confidence intervals for each of the reanalyses (calculated as the 1.96 times the standard deviation across the seasonal mean ensemble members). The overall RMSD compared to the reconstruction is given in the upper right for each dataset.

Figure 8. As in Fig. 5, but for a) Deck 246, which operated near the East Antarctic coast discontinuously between 1910-1930, and b) observations at Cape Denison during the Australian Antarctic Expedition of 1911-1914. Note in b) that surface pressure is plotted on the right axis. MAE values are based only on MSLP since there is a wide range of surface pressure values (>30 hPa), all primarily reflecting elevation differences in the underlying models. For a) the MAE values (both based on MSLP) are calculated using the two different locations in DJF 1911. Cape Denison data were obtained from ISPD, while data for Deck 246 were obtained from ICOADS version 3.

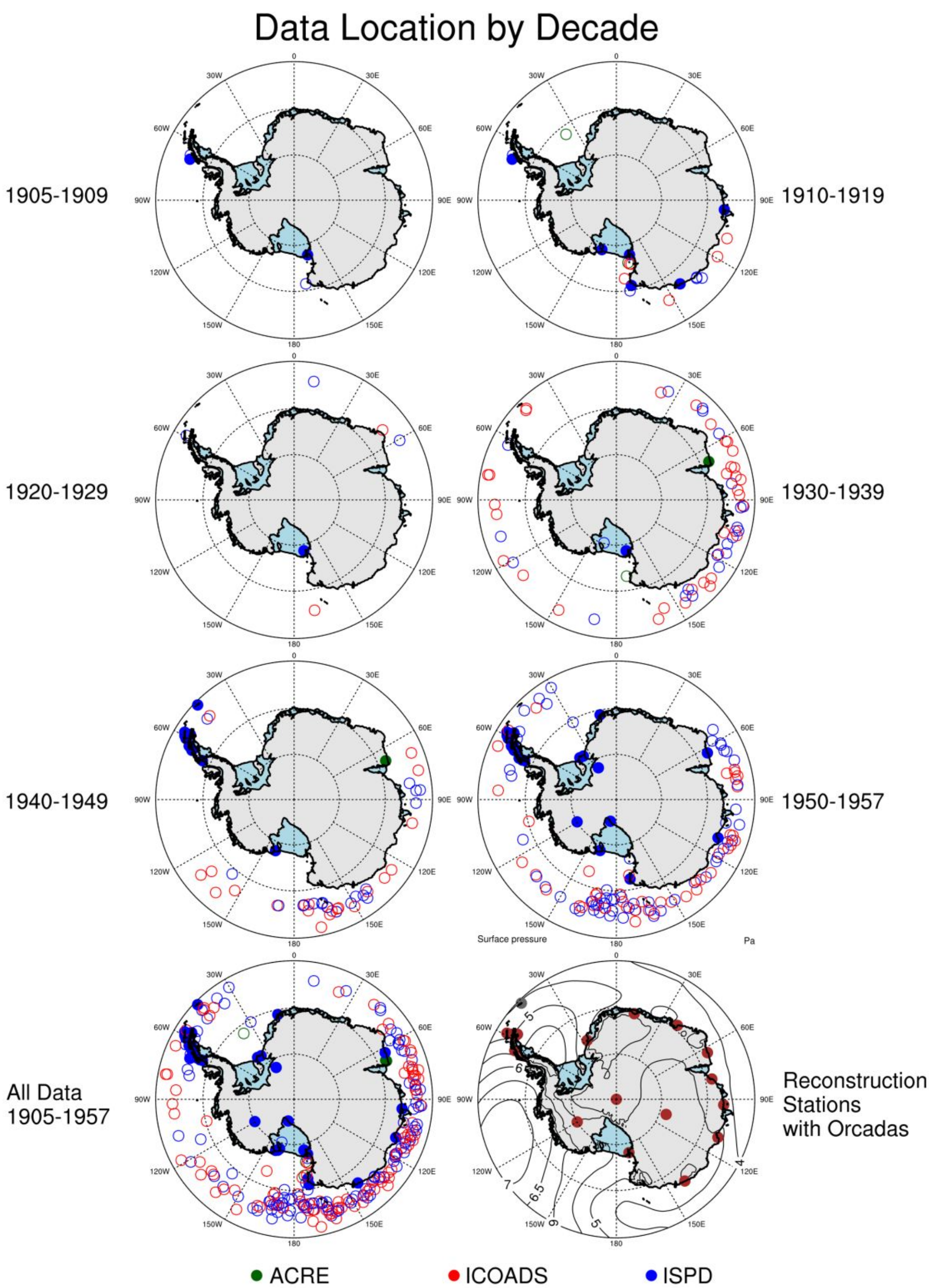

Figure 1. Maps of seasonal mean data location grouped by decade. ~~, with the bottom plot showing the location of all 271 seasonal mean observations compared.~~ Open circles represent seasonal mean locations from ship records, and filled circles are for temporary bases on the continent. The bottom left plot shows the location of all 271 seasonal mean observations compared, while the bottom right plot shows the locations of the 18 station reconstructions (brown) and observations from Orcadas (grey) used to generate the spatially complete pressure reconstruction. Contours on the bottom right panel are the standard deviations of monthly ERA5 surface pressure anomalies for reference, contoured every 0.5 hPa.

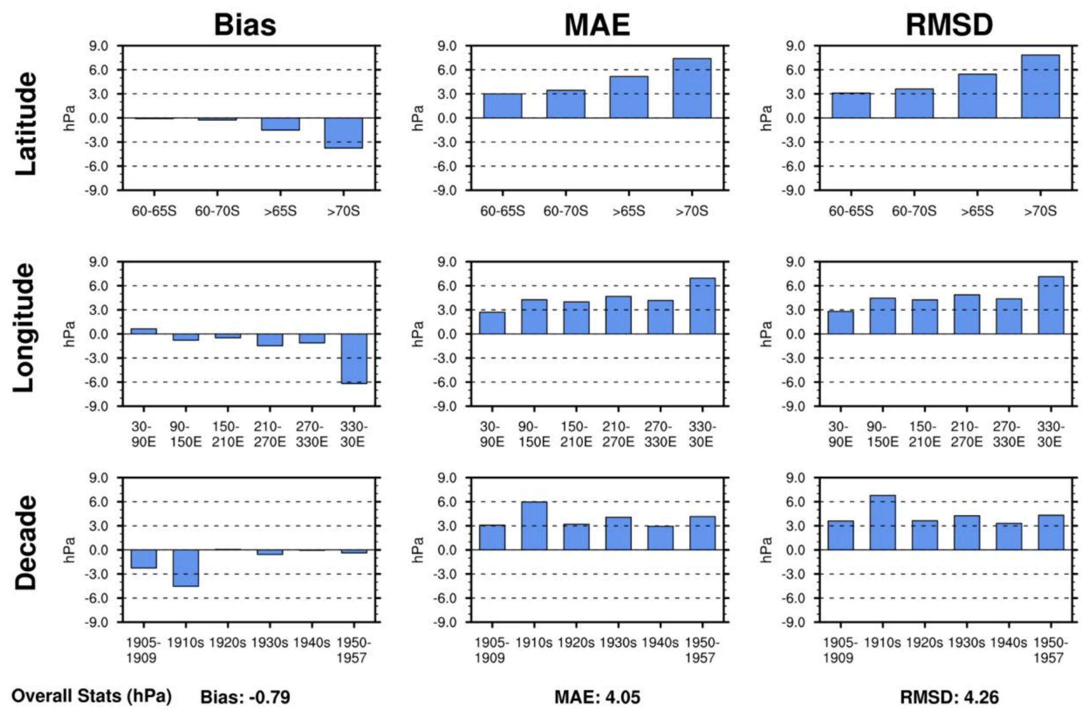

Figure 2. Mean reconstruction skill statistics (columns; bias, MAE, and RMSD) compared to historical observations, and averaged over various latitude bands (top row), longitude bands (middle row), and decades (bottom row). The statistics calculated over all observations are listed at the bottom of the figure.

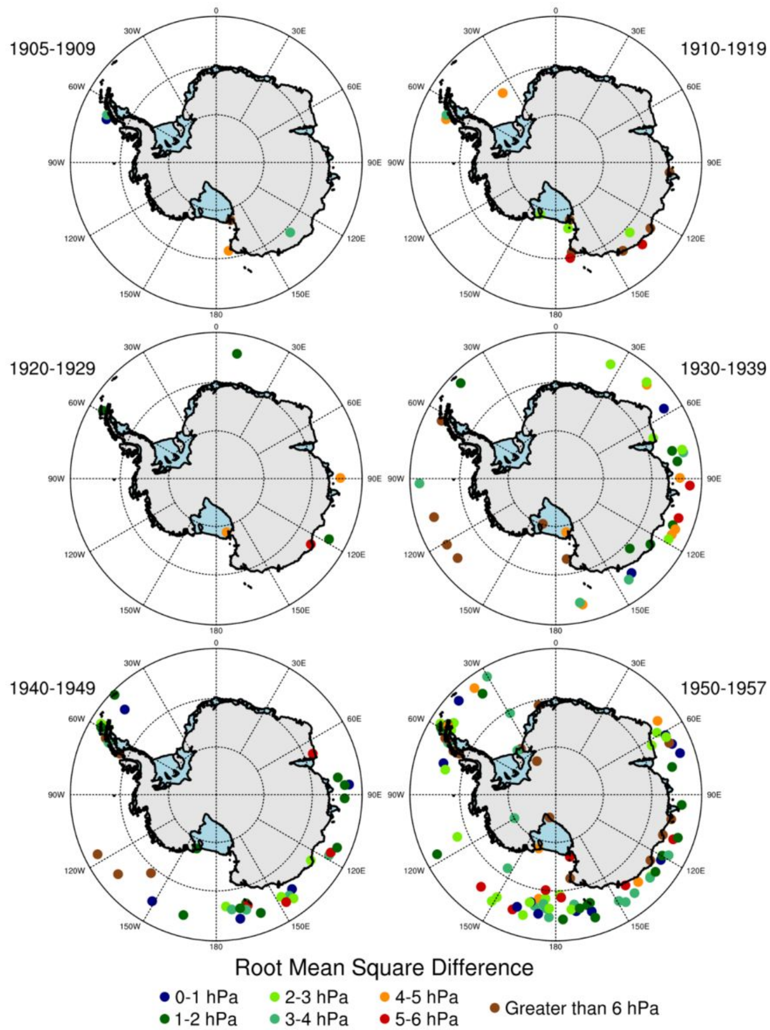

Figure 3. Decadal mean RMSD plotted by decade.

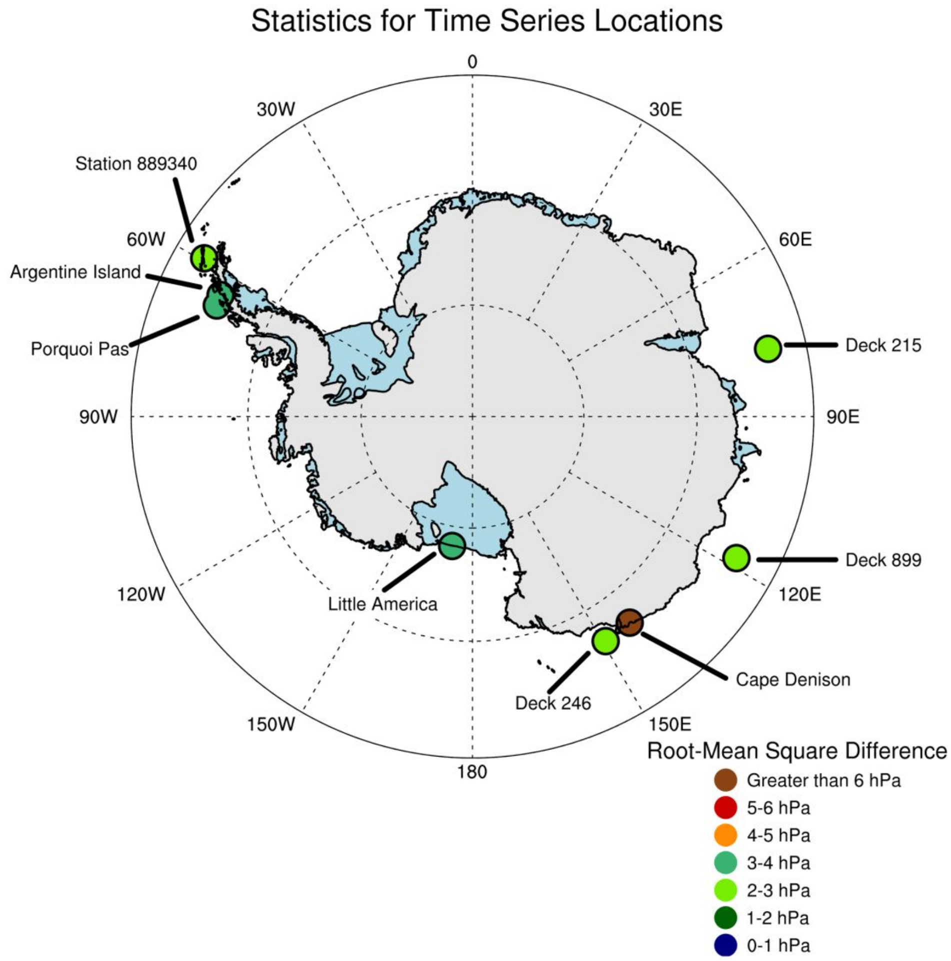

Figure 4. Map showing observational mean (over full length of record) RMSD for select representative locations examined in more detail.

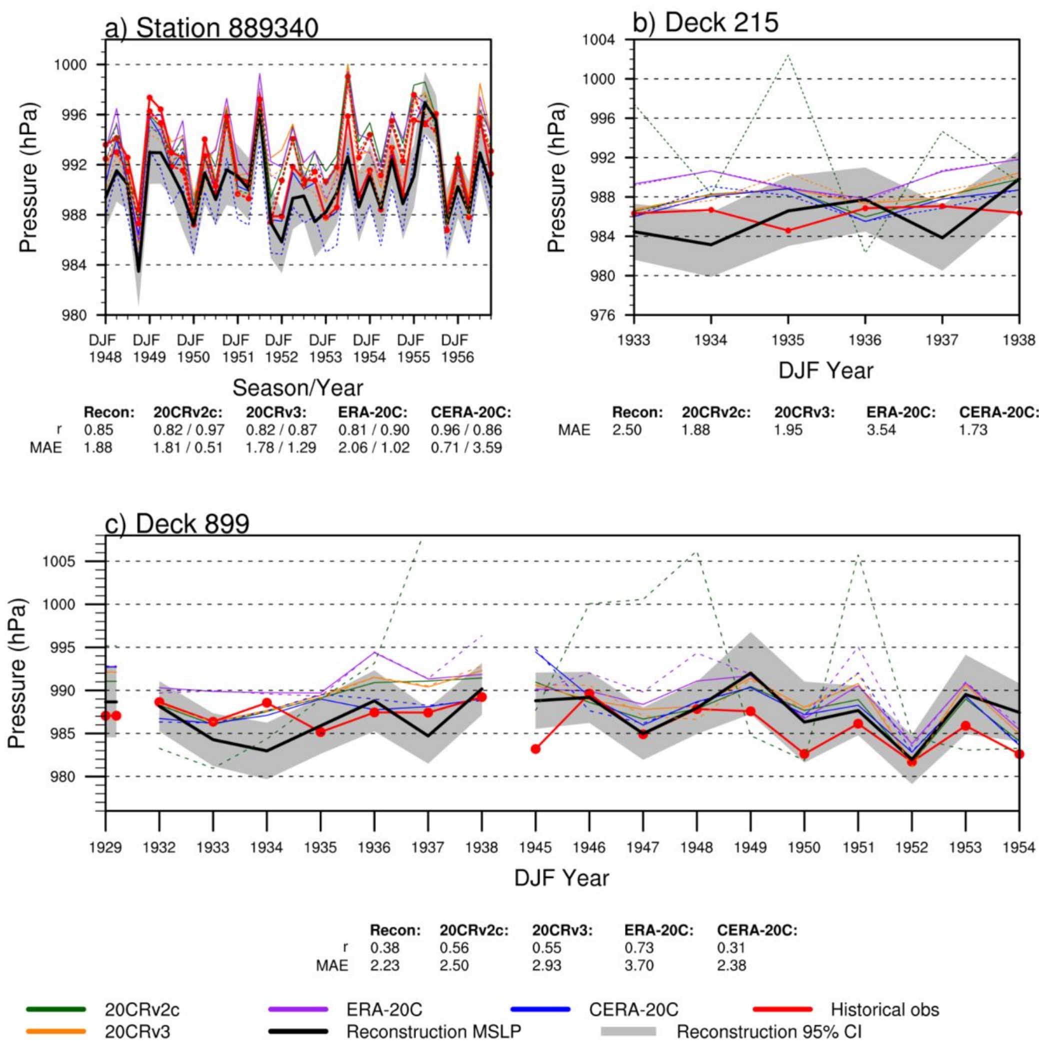

Figure 5. Time series of historical observations, reconstruction (with 95% confidence interval in grey shading), and gridded reanalysis products for observations representative of the lowest RMSD (solid lines for MSLP, dashed lines for surface pressure). The name is the record identifier provided in ISPD or ICOADS. In a), the x-axis varies by season, and the labels represent the DJF seasons for each year; in b) and c) only DJF data are plotted and the label represents the DJF season. The white spaces in c) represent discontinuities in the observations. The values at the bottom of each panel are the correlations (if more than 10 data points are available) and MAE values (first numbers based on MSLP, second (where available) based on surface pressure) for each dataset. The data for station 889340 are from ISPD, while Deck 215 and Deck 899 are from ICOADS.

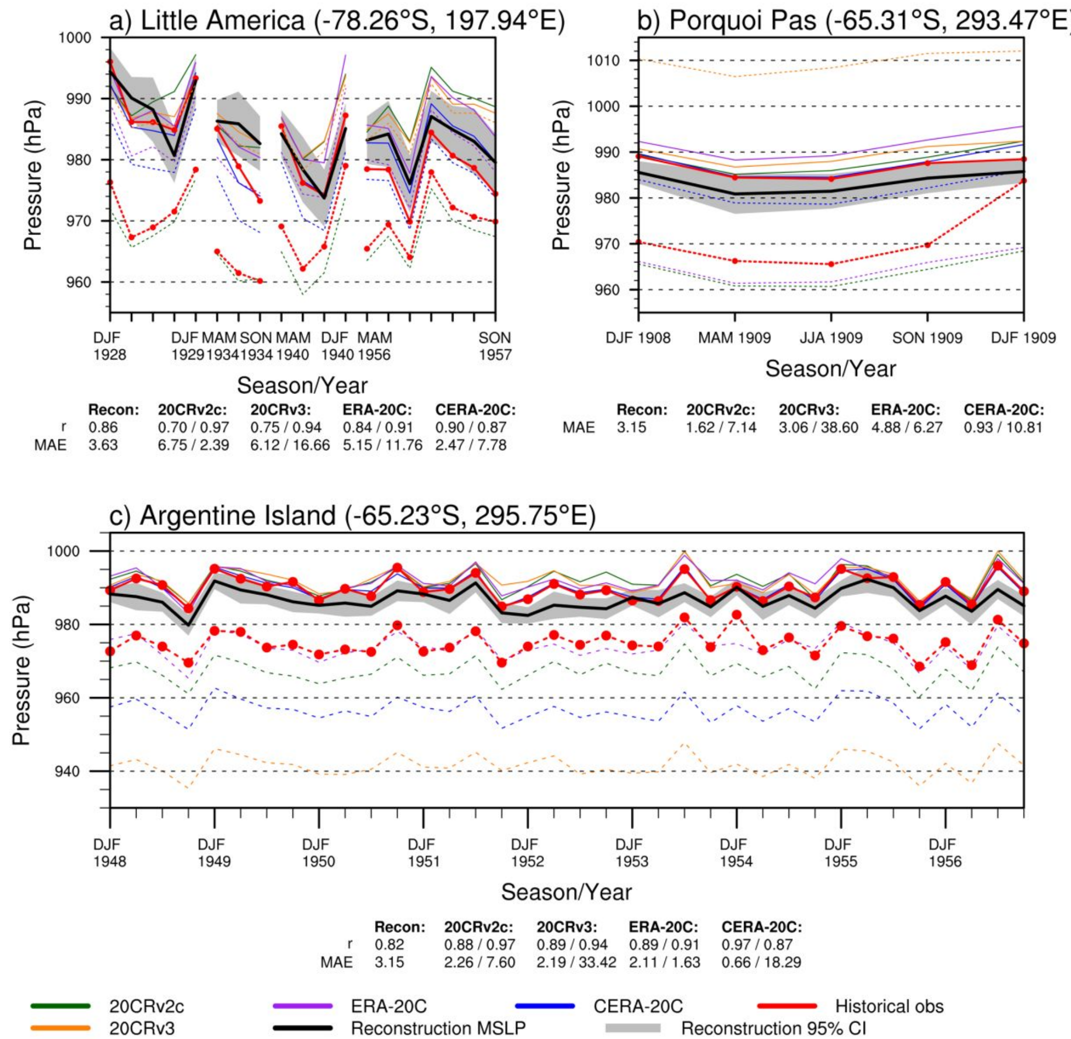

Figure 6. As in Fig. 5, but representative of stations where elevation corrections (reduction to sea level pressure) play an important aspect of the reconstruction performance evaluation. In all panels, the x-axis varies by season. All data in this figure were obtained from ISPD version 3.

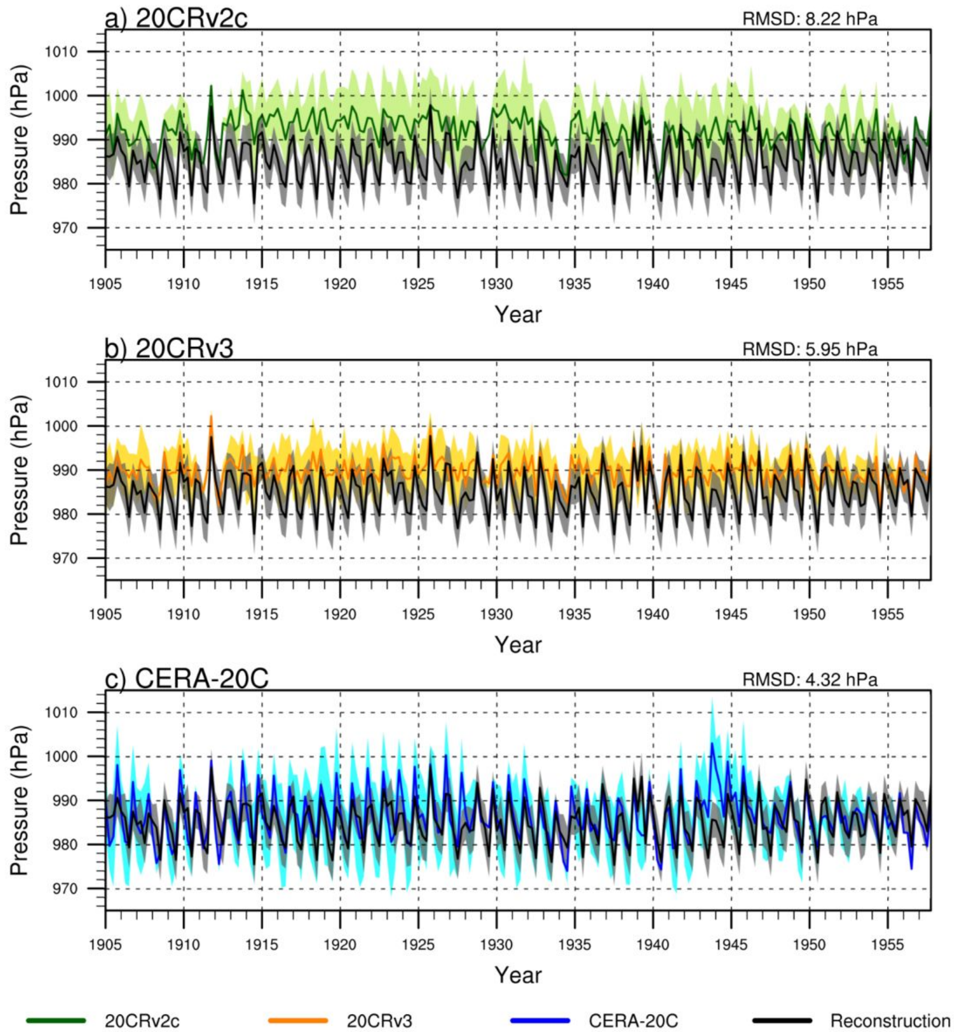

Figure 7. Time series of seasonal mean MSLP for all four seasons at Little America (from ISPD) on the northern edge of the Ross Ice Shelf for the reconstruction along with values from a) 20CRv2c; b) 20CRv3; c) CERA-20C. The gray shading in each panel represents the 95% confidence interval for the reconstruction, while the colored shading represents 95% confidence intervals for each of the reanalyses (calculated as the 1.96 times the standard deviation across the seasonal mean ensemble members). The overall RMSD compared to the reconstruction is given in the upper right for each dataset.

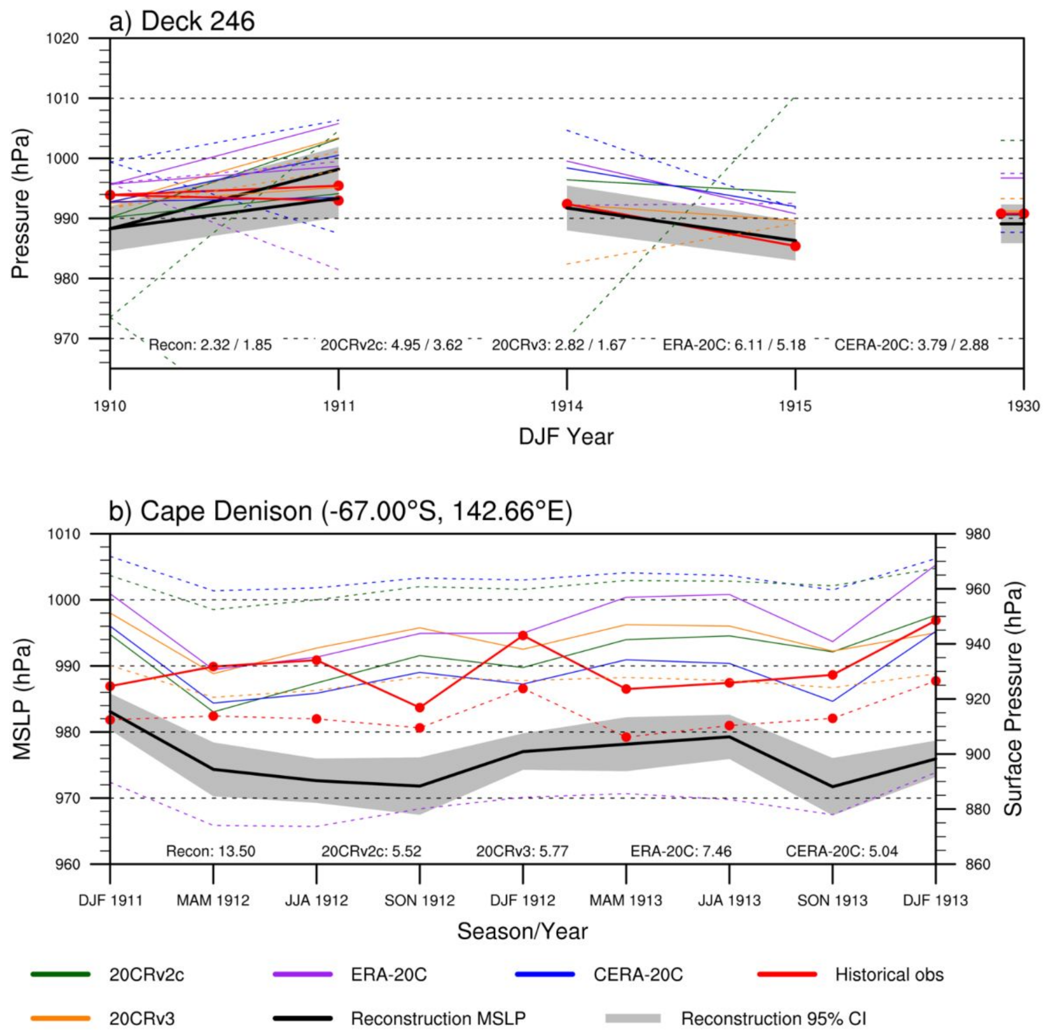

Figure 8. As in Fig. 5, but for a) Deck 246, which operated near the East Antarctic coast discontinuously between 1910-1930, and b) observations at Cape Denison during the Australian Antarctic Expedition of 1911-1914. Note in b) that surface pressure is plotted on the right axis. MAE values are based only on MSLP since there is a wide range of surface pressure values (>30 hPa), all primarily reflecting elevation differences in the underlying models. For a) the MAE values (both based on MSLP) are calculated using the two different locations in DJF 1911. Cape Denison data were obtained from ISPD, while data for Deck 246 were obtained from ICOADS version 3.

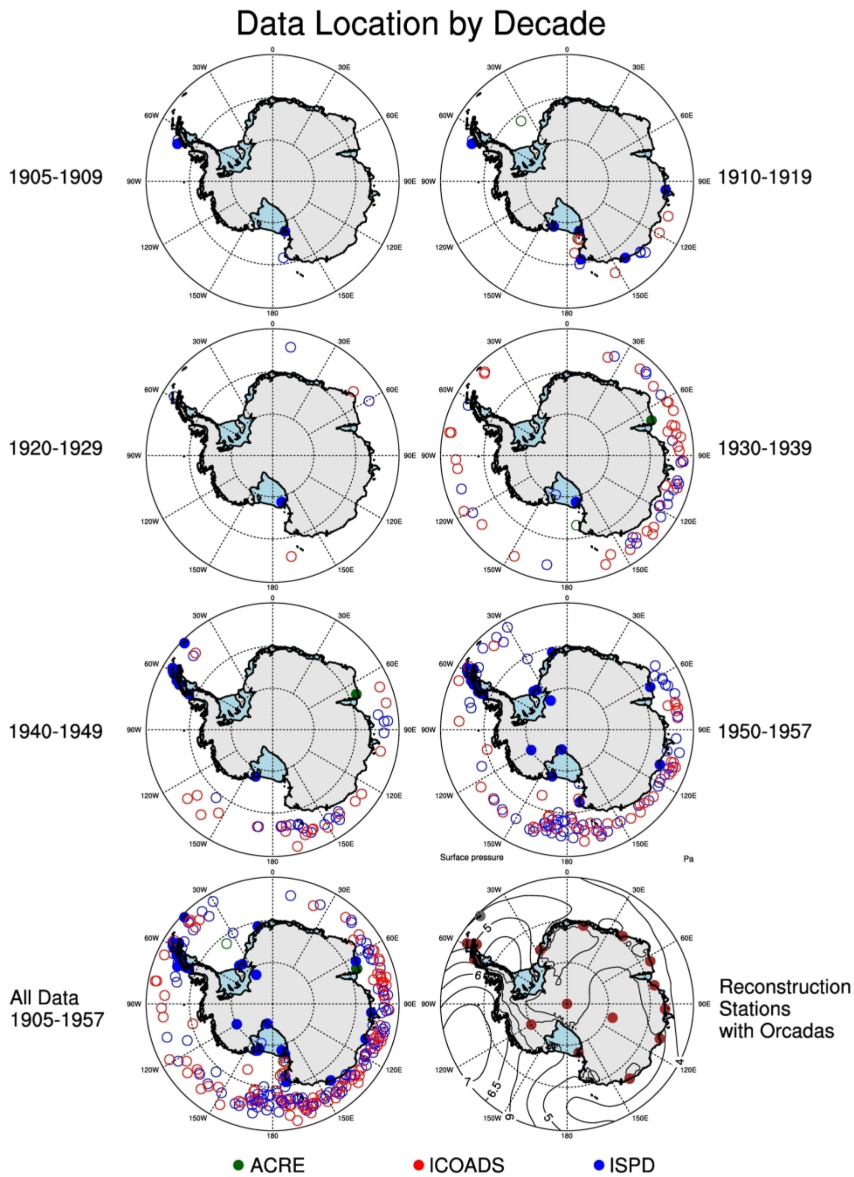

Figure 1. Maps of seasonal mean data location grouped by decade. Open circles represent seasonal mean locations from ship records, and filled circles are for temporary bases on the continent. The bottom left plot shows the location of all 271 seasonal mean observations compared, while the bottom right plot shows the locations of the 18 station reconstructions (brown) and observations from Orcadas (grey) used to generate the spatially complete pressure reconstruction. Contours on the bottom right panel are the standard deviations of monthly ERA5 surface pressure anomalies for reference, contoured every 0.5 hPa.

215x266mm (300 x 300 DPI)

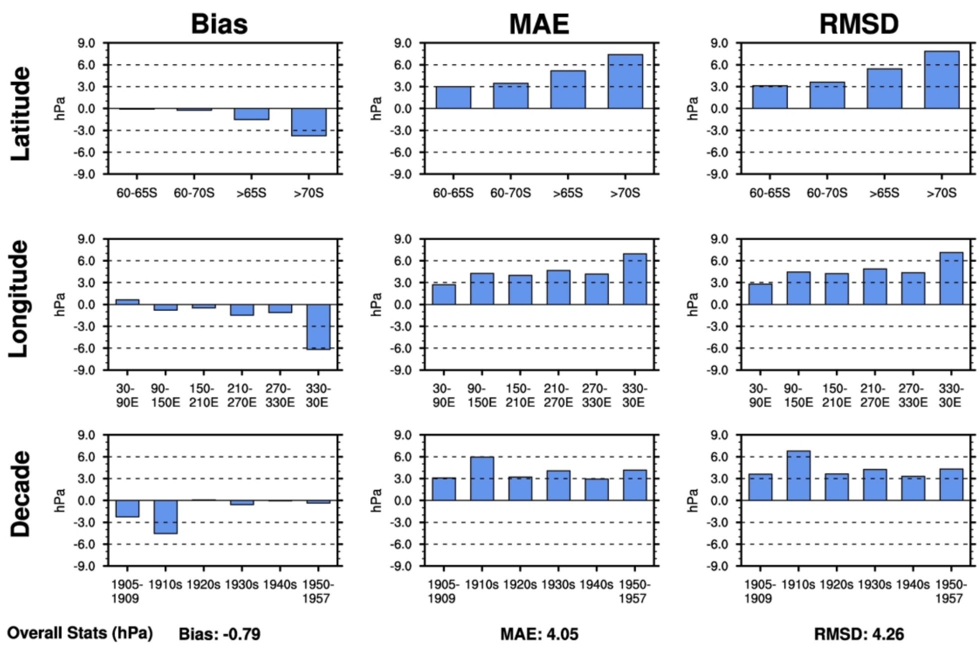

Figure 2. Mean reconstruction skill statistics (columns; bias, MAE, and RMSD) compared to historical observations, and averaged over various latitudes (top row), longitudes (middle row), and decades (bottom row). The statistics calculated over all observations are listed at the bottom of the figure.

190x190mm (300 x 300 DPI)

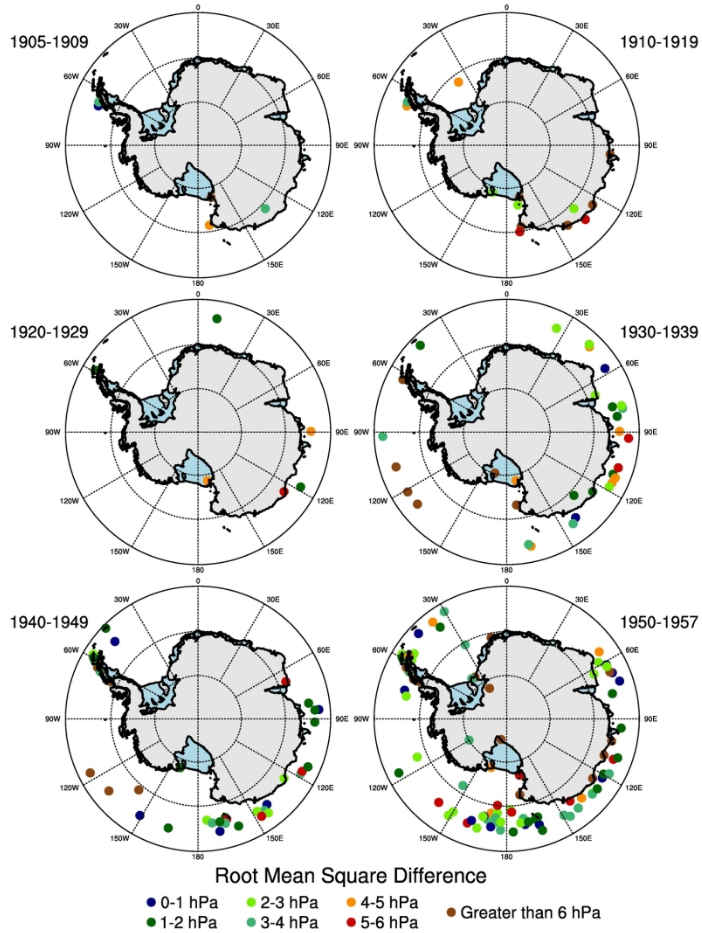

Figure 3. Decadal mean RMSD plotted by decade.

190x190mm (300 x 300 DPI)

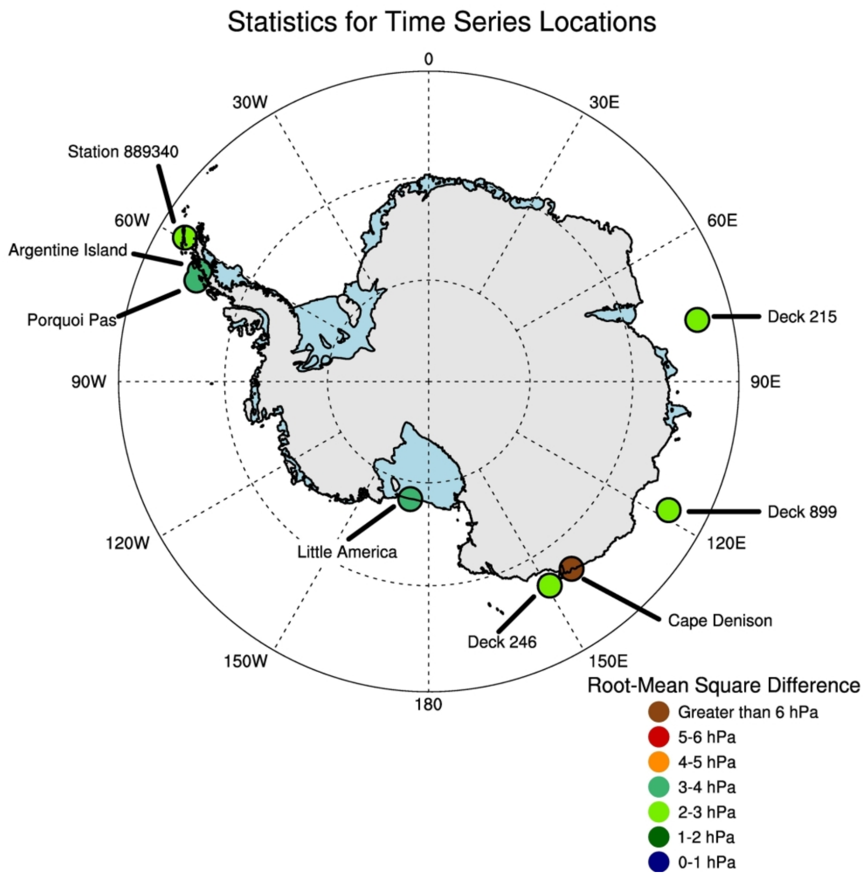

Figure 4. Map showing observational mean (over full length of record) RMSD for select representative locations examined in more detail.

190x190mm (300 x 300 DPI)

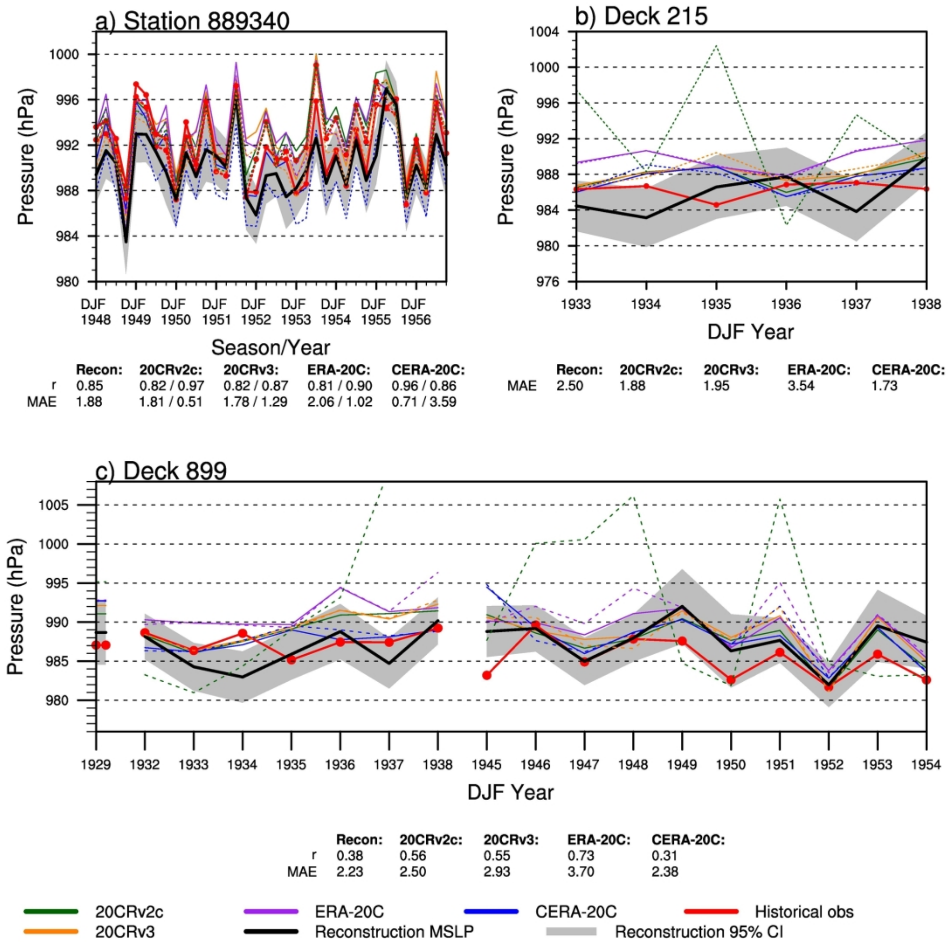

Figure 5. Time series of historical observations, reconstruction (with 95% confidence interval in grey shading), and gridded reanalysis problems for observations representative of the lowest RMSD (solid lines for MSLP, dashed lines for surface pressure). The name is the record identifier provided in ISPD or ICOADS. In a), the x-axis varies by season, and the labels represent the DJF seasons for each year; in b) and c) only DJF data are plotted and the label represents the DJF season. The white spaces in c) represent discontinuities in the observations. The values at the bottom of each panel are the correlations (if more than 10 data points are available) and MAE values (first numbers based on MSLP, second (where available) based on surface pressure) for each dataset. The data for station 889340 were from ISPD, while Deck 215 and Deck 899 were from ICOADS.

190x190mm (300 x 300 DPI)

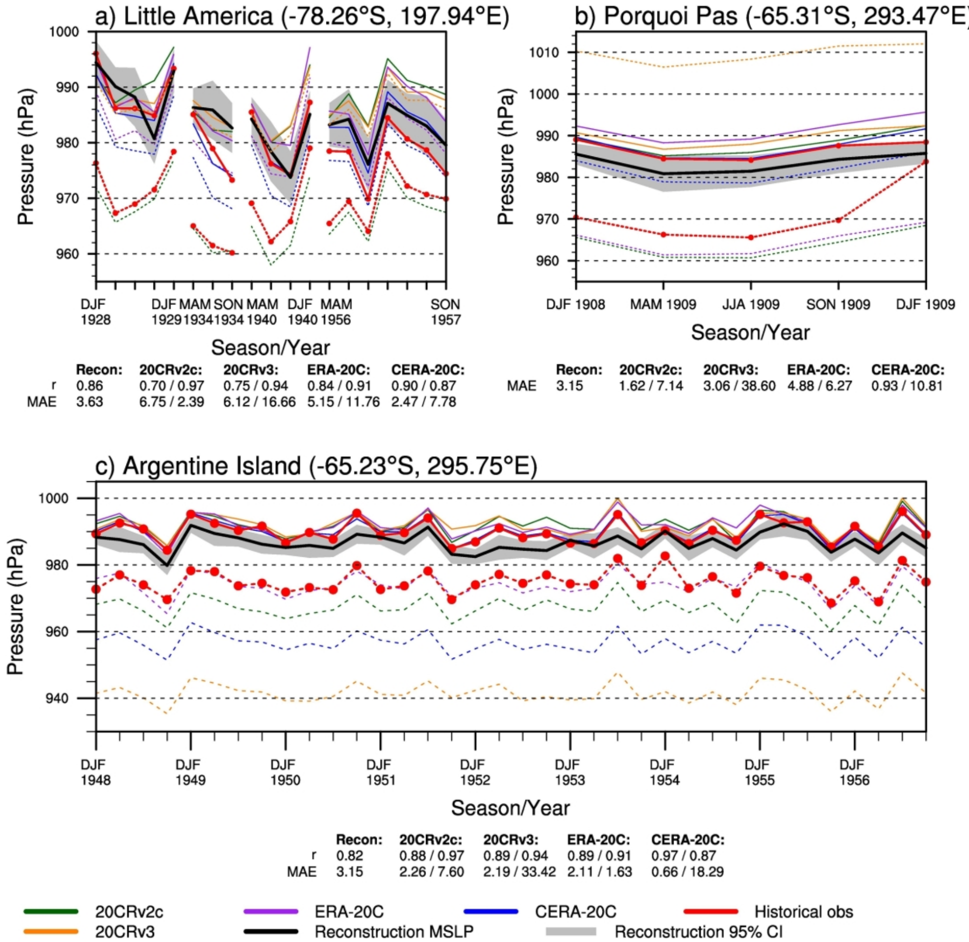

Figure 6. As in Fig. 5, but representative of stations where elevation corrections (reduction to sea level pressure) play an important aspect of the reconstruction performance evaluation. All data in this figure were obtained from ISPD.

190x190mm (300 x 300 DPI)

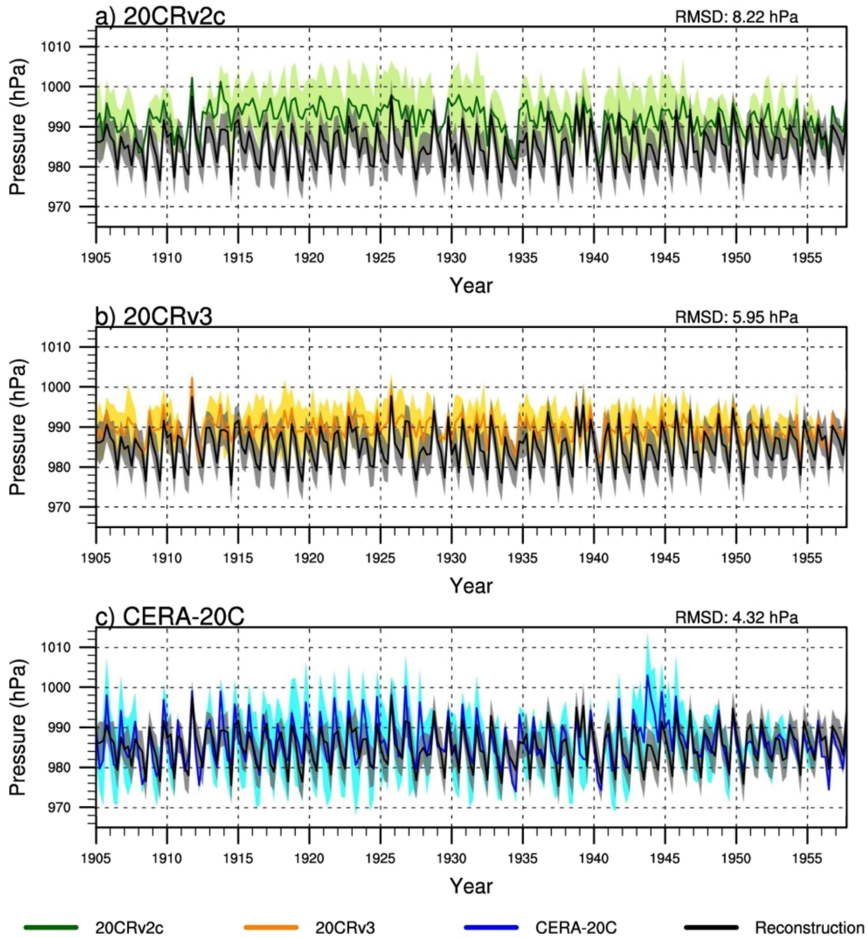

Figure 7. Time series of seasonal mean MSLP for all four seasons at Little America (from ISPD) on the northern edge of the Ross Ice Shelf for the reconstruction along with values from a) 20CRv2c; b) 20CRv3; c) CERA-20C. The gray shading in each panel represents the 95% confidence interval for the reconstruction, while the colored shading represents 95% confidence intervals for each of the reanalyses (calculated as the 1.96 times the standard deviation across the seasonal mean ensemble members). The overall RMSD compared to the reconstruction is given in the upper right for each dataset.

190x190mm (300 x 300 DPI)

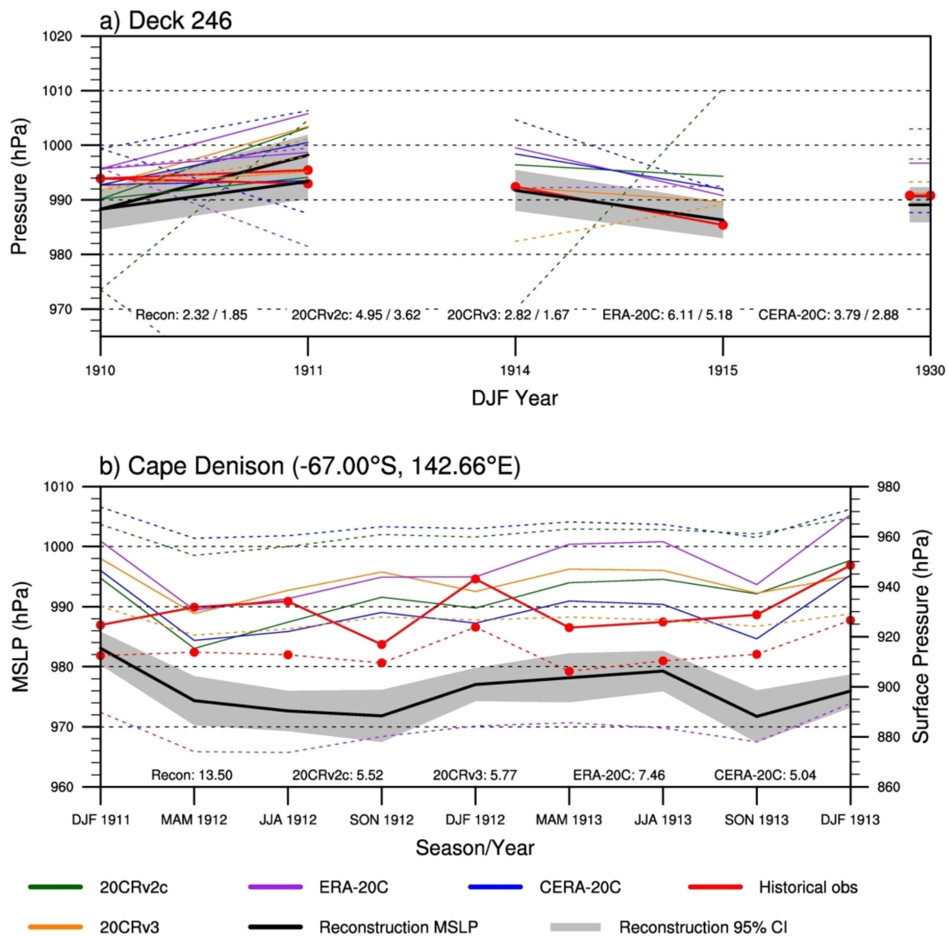

Figure 8. As in Fig. 5, but for a) Deck 246, which operated near the East Antarctic coast discontinuously between 1910-1930, and b) observations at Cape Denison during the Australian Antarctic Expedition of 1911-1914. Note in b) that surface pressure is plotted on the right axis. MAE values are based only on MSLP since there is a wide range of surface pressure values (>30 hPa), all primarily reflecting elevation differences in the underlying models. For a) the MAE values (both based on MSLP) are calculated using the two different locations in DJF 1911. Cape Denison data were obtained from ISPD, while data for Deck 246 were obtained from ICOADS version 3.

190x190mm (300 x 300 DPI)

An Assessment of Early 20th Century Antarctic Pressure Reconstructions using Historical Observations

Ryan L. Fogt[1], Connor P. Belak[1], Julie M. Jones[2], Laura C. Slivinski[3,4], Gilbert P. Compo[3,4]

[1]Department of Geography and Scalia Laboratory for Atmospheric Analysis, Ohio University, Athens, OH
[2]Department of Geography, University of Sheffield, Sheffield, UK
[3] University of Colorado, Cooperative Institute for Research in Environmental Sciences, Boulder, CO
[4]NOAA, Physical Sciences Laboratory, Boulder, CO

Graphical Abstract
While gridded seasonal pressure reconstructions poleward of 60°S extending back to 1905 have been recently completed, their skill has not been assessed prior to 1958. To provide a more thorough evaluation of the skill in the early 20th century, these reconstructions are compared to other gridded datasets, historical data from early Antarctic expeditions, ship records, and temporary bases, such as the Little America base shown in the image, to further evaluate their performance.