Article:

Fogt, R.L., Goergens, C.A., Jones, M.E. et al. (3 more authors) (2016) Antarctic Station Based Seasonal Pressure Reconstructions Since 1905, Part 1: Reconstruction Evaluation. Journal of Geophysical Research: Atmospheres, 121 (6). pp. 2814-2835. ISSN 2169-897X

https://doi.org/10.1002/2015JD024564

Antarctic station-based seasonal pressure reconstructions since 1905: 1. Reconstruction evaluation

Ryan L. Fogt[1], Chad A. Goergens[1], Megan E. Jones[1], Grant A. Witte[1], Ming Yueng Lee[1], and Julie M. Jones[2]

[1]Department of Geography and Scalia Laboratory for Atmospheric Analysis, Ohio University, Athens, OH, USA
[2]Department of Geography, University of Sheffield, Sheffield, United Kingdom

Corresponding author address: Ryan L. Fogt, 122 Clippinger Laboratories, Department of Geography, Ohio University, Athens, OH, 45701. Email: fogtr@ohio.edu

KEY POINTS:

1. Seasonal station-based pressure reconstructions across Antarctica back to 1905 are possible.
2. The highest reconstruction skill is found in austral summer and along the Antarctic Peninsula in all seasons.
3. Using gridded pressure estimates over ocean basins as additional input significant aids reconstruction skill in austral winter.

Abstract

Seasonal mean Antarctic pressures at 17 stations are reconstructed based on the method of principal component regression, employing midlatitude pressure data as predictors. Several reconstruction methods were performed in order to assess the stability and reliability of the reconstructions obtained, including performing the reconstructions over a shorter 30-year window and withholding the remaining data for an independent validation. Generally, there were small differences between the various approaches, but typically reconstructions conducted on data with the trends still present and over the full period of observations achieved the highest skill. Seasonally, reconstruction skill was high in austral summer across the entire Antarctic continent. Reconstructions that employed gridded pressure data over oceans as well as the observations (here termed ‘pseudo-reconstructions) also performed remarkably well in austral winter. Spatially, the reconstruction skill was highest near the Antarctic Peninsula in all seasons, and weakest in coastal East Antarctica and the Antarctic Interior during austral spring and autumn; the spatial variability of the skill in part reflects the distance to the nearest mid-latitude predictor. Nonetheless, for nearly all seasons and locations the observed trends since 1957 were well captured by the reconstructions, as was the low-frequency decadal scale variability. These results suggest Antarctic pressure observations can be extended throughout the 20th century with high confidence, especially in summer, allowing for a more precise understanding of the role and magnitude of natural atmospheric circulation variability across Antarctica.

1. Introduction

The development of the Reference Antarctic Data for Environmental Research [READER, www.antarctica.ac.uk/met/READER; *Turner et al.*, 2004] archive of Antarctic surface (both staffed and from automatic weather stations) and upper air meteorological observations has been a valuable tool in understanding the extent of atmospheric variations and changes across the Antarctic continent over the last ~60 years. This archive hosts long-term meteorological data (at least 25 years of record) that have been quality controlled, and metadata are provided when available. For the latter, this includes identifying how much of the daily observations were available before calculating the monthly mean (at least 90% and 30% of daily surface and upper air observations, respectively, are needed to calculate an accurate mean). Due to its widespread use, later research archives covering ice core data (iceREADER, www.icereader.org/icereader) and physical observations from the Southern Ocean (OceanREADER, www.antarctica.ac.uk/met/SCAR_ssg/ps/OceanREADER) were later developed. Together, these provide a fairly comprehensive online resource for understanding the wide range of climate variability across Antarctica on monthly and longer timescales.

Using the READER archive, *Turner et al.* [2005] described the changes in the Antarctic atmosphere, and were among the first to discuss the regional differences between warming in East and West Antarctica. During the period 1971-2001, when most stations had data, they noted a statistically significant warming (p<0.10) across much of the northern Antarctic Peninsula, ranging from 0.2-0.7 °C decade^{-1}. In contrast, statistically insignificant negative temperature trends were observed across much of coastal East Antarctica during the same time period; a few stations (Halley, Amundsen-Scott, and Casey) displayed significant (p<0.10) negative temperature trends in austral autumn. In terms of the atmospheric circulation, *Turner et al.* [2005] found that pressure was decreasing more uniformly across the continent, with the most significant decreases during 1971-2010 across East Antarctica in austral summer and autumn. They related these changes to the increase in the circumpolar westerlies and the positive trend in the Southern Hemisphere Annular Mode (SAM) index over the same period.

Since atmospheric reanalysis datasets have been deemed unreliable prior to the start of the modern satellite era [1979; *Bromwich and Fogt,* 2004; *Bromwich et al.*, 2007], several studies have worked to reconstruct the Antarctic temperature field back to 1957, when most surface observations began with the start of the International Geophysical Year [*Chapman and Walsh*, 2007; *Monaghan et al*., 2008; *Steig et al.*, 2009; *O'Donnell et al.*, 2011; *Nicolas and Bromwich*, 2014]. While there are differences in the magnitude of the warming among the reconstructions, the largest warming is confined to the Antarctic Peninsula and West Antarctica, particularly in austral winter and spring. *Nicolas and Bromwich* [2014] suggest that only in austral spring is the warming significant throughout time; at other locations or seasons the warming becomes insignificant or is marked with weak (i.e., statistically insignificant) cooling trends.

Additional work has tried to understand the cause of the changes, and has linked the changes to variations in the sea ice cover (especially in the Ross and Amundsen / Bellingshausen Seas) and changes in the atmospheric circulation, forced from the tropics or from Antarctic stratospheric ozone depletion. A study by *Holland and Kwok* [2012] determined that much of the trends in the sea ice extent and motion in the Ross, Amundsen, and Bellingshausen Seas was tied to the changes in the wind pattern, manifested in the deepening of the Amundsen Sea Low off the coast of West Antarctica [*Fogt et al.*, 2012a; *Turner et al.*, 2013; *Hosking et al.*, 2013; *Raphael et al.*, 2015].

Despite how these studies have improved the understanding of the ongoing Antarctic climate change, all of them are plagued with working with very short observational records. Since many of the changes are related to the atmospheric circulation (as indicated by pressure trends), the primary goals of these two companion papers are to reconstruct, evaluate, and analyze seasonal station-based pressure changes across Antarctica during the 20th century. We use principal component regression, a proven reconstruction technique detailed further in section 2, and station observations of pressure across the midlatitudes of the Southern Hemisphere (from the major continents). The remainder of this paper is laid out as follows: section 2 provides a discussion of the data employed while section 3 gives an overview of the various reconstruction methodologies performed. Section 4 evaluates the reconstructions across the Antarctic continent. Finally, 'pseudo' reconstructions employing mean sea level pressure (MSLP) estimates over ocean basins from various long-term gridded datasets, in addition to the direct observations, will be presented and evaluated in section 5. A summary and conclusions is offered in section 6; the companion paper [*Fogt et al.*, 2016] delves deeper into the reconstructions themselves and highlights Antarctic atmospheric circulation changes over the last century.

2. Data Used

a) Observational pressure records

The reconstructions were based on selected Antarctic stations from the READER archive, which have been quality-controlled prior to publishing online [*Turner et al.,* 2004]. Although small errors may exist in these data, many of the uncertainties in the underlying daily data are reduced when focusing on seasonal means used in our study. We chose the 17 stations with the longest and most complete records, and a map of the stations is provided in Fig. 1; Byrd station in West Antarctica was also reconstructed. However, as discussed for the temperature record [*Bromwich et al.*, 2013, 2014], most of the data during the 1970s is missing for this station, until the automatic weather station (AWS) data became available in 1980. The missing AWS data at Byrd were patched using bi-linearly interpolated surface pressure data from the European Centre for Medium Range Weather Forecasts (ECMWF) Interim-reanalysis (ERA-Int) through monthly linear regression. Unfortunately, other data sources used to patch many missing temperature records for Byrd during the 1970s [*Bromwich et al.*, 2014] were not available for surface pressure, and therefore this record still has a considerable portion of missing data, which presented some challenges in its reconstruction. As such, the reconstructions for this station are discussed separately.

A few other patches were made in order to extend the Antarctic records and make them as complete as possible. Both Casey and Rothera stations were extended further back into time than available on the READER archive with nearby temporary records compiled together in the Global Historical Climatological Network [GHCN, *Peterson and Vose*, 1997; *Peterson et al.*, 1998]. The Bellingshausen station record was extended back to 1958 using earlier records from nearby (within 20 km) Arturo Prat and Deception stations. A few missing months after 1996 in the Vostok record were patched using linear regression with the Dome C II AWS (75.1°S, 123.4°E), situated in proximity on the high East Antarctic plateau, available on the READER archive. Two other stations were also combined using linear regression by month to aid in missing data: on the Antarctic Peninsula, the O'Higgins record was patched with the Marsh record situated nearby on King George Island, and on Ross Island the adjacent records of McMurdo and Scott Base were merged to form a more complete pressure record there. A listing of the Antarctic stations reconstructed is provided in Table 1. Mean sea level pressure data were used for all but the three stations on the Antarctic plateau (Amundsen-Scott, Byrd, and Vostok), where surface pressure was used.

The Antarctic pressure reconstructions at each station in Table 1 were based on pressure records from the Southern Hemisphere, mostly in the midlatitude regions. To be considered, all must have records that extend back until 1905, and be more than 75% complete. These monthly Southern Hemisphere midlatitude pressure records were obtained from either the GHCN [*Peterson and Vose*, 1997; *Peterson et al.*, 1998], the University Corporation for Atmospheric Research (UCAR) research data archive dataset ds570.0, or quality-controlled observations from the Climatic Research Unit [*Jones*, 1987; *Jones et al.*, 1999]. Many of these data were used previously in reconstructions of the SAM index [*Jones et al.*, 2009], and further details on these records are given in Table 2. We made primary use of the ds570.0 dataset, and merged records from other datasets for individual stations as needed (including a few that had slight location changes in time) in order to obtain the most complete records for the midlatitude stations as possible. Data for Auckland, New Zealand during 1905-1915 were patched with daily surface data at Auckland from the International Surface Pressure Databank version 3

(ISPDv3), available from UCAR research data archive dataset ds132.1. The patching / merging does not substantially alter the reliability of the long-term midlatitude records since there are a relatively small number of gaps in the ds570.0 archive at most stations (Table 2), and strong similarities exist between the merged records during periods of overlap.

Both the Antarctic and midlatitude data have been updated through 2013. Seasonal means were calculated from the monthly data if there were at least two months present, otherwise the seasonal data were treated as missing. All seasons refer to the standard Southern Hemisphere meteorological seasons: austral summer, December-February (DJF); autumn, March-May (MAM); winter, June-August (JJA); and spring, September-October (SON).

b) Gridded pressure data

We employ four different gridded pressure datasets, namely ERA-Int [*Dee et al.*, 2011] and the ECMWF 20th century reanalysis (ERA-20C), the Hadley Centre gridded mean sea level pressure version 2 [HadSLP2; *Allan and Ansell*, 2006], and the National Oceanic and Atmospheric Administration 20th – Cooperative Institute for Research in Environmental Studies (NOAA-CIRES) century reanalysis, version 2c [20CR, *Compo et al.*, 2011]. In all cases, monthly or seasonal mean data are utilized.

3. Reconstruction methodology

The Antarctic historical pressure values at individual stations are reconstructed using principal component regression (PCR). This method has been successful in many climatological reconstructions such as the SAM index [*Jones et al.*, 2009; *Fogt et al.*, 2009], U.S. drought characteristics [*Cook et al.*, 1999], and well-known temperature reconstructions [*Mann et al.*, 1998, 1999]. For each station and season, we employ two subsets of the midlatitude pressure data (the predictors): those stations that are significantly correlated with the Antarctic station at $p<0.05$ and $p<0.10$, termed the 5% and 10% networks. The PCR methodology then uses the covariance matrix of these established predictors and conducts principal component analysis to partition the covariance matrix into distinct (i.e., orthogonal) modes of variability. Inherent in each of these modes is a spatial pattern, called empirical orthogonal functions (EOFs), and a time series that represents the amplitude of this pattern, termed the PCs. A subset of these PCs is regressed, using ordinary least squares linear regression, onto the Antarctic station being reconstructed (the predictand) in order that each predictor may be precisely calibrated to the predictand. The reconstruction is obtained using these regression coefficients back in time through the length of the midlatitude pressure records; alternatively the reconstruction can be thought of as a weighted sum of the predictors, where the weights are determined through matrix multiplication of the relationship each predictor shares with the retained PCs and the relationship these PCs share with the predictand. The precise number of retained PCs is determined through an independent validation technique (described below). While other approaches could be employed to reconstruct the stations, it is anticipated that the error in the reconstructions dominates any error observed from the various reconstruction models employed. Seasonal mean reconstructions are the primary focus, as these lower-frequency timescales have shown the best skill for the SAM reconstructions [*Jones et al.*, 2009; *Visbeck*, 2009] as well as pan-Antarctic temperature reconstructions [*Chapman and Walsh*, 2007; *Monaghan et al.*, 2008; *Steig et al.*, 2009; *O'Donnell et al.*, 2011; *Nicolas and Bromwich*, 2014]; annual mean reconstructions were not attempted since the high-to-midlatitude atmospheric relationships vary seasonally and are thus underrepresented with the use of annual means [*Jones et al.*, 2009; *Fogt et al.*, 2012b]. Seasons with missing data (in either the predictand or predictor stations) are not included for model calibration. When calculating the reconstruction back in time, missing values in the predictor stations are replaced with the climatological seasonal mean. We performed the PCR method originally using data extending through 2011 and then with updated data through 2013; both are discussed here to assess the sensitivity and consistency of this approach. Additionally, to analyze the sensitivity of the model to trends, we have performed the PCR using detrended and original / trended data for both the predictors and predictand.

The uncertainty in the PCR model is obtained through several validation tests. For each reconstruction, the reduction of error (RE), coefficient of efficiency (CE), and the correlations during both model calibration and validation, are calculated. Both RE and CE values range from negative infinity to positive 1; an RE or CE value greater than zero indicates reconstruction skill better than using the climatological mean and values of 1.0 indicate a perfect reconstruction [*Cook et al.*, 1999]. In our 'full' reconstructions, the full period of 1957-2011/ 2013 is used for model calibration; these reconstructions are validated using the leave-one-out cross validation procedure as in *Jones et al.* [2009]. In this technique, the PCR is performed as many times as there are values during the calibration period (here, 55 or 57 times). Each time PCR is performed, a year and its two neighboring years are left out to account for potential autocorrelation. The center year is then predicted using PCR, and all of the predicted years are subsequently concatenated to produce an independent validation time series. Uncertainty in the final reconstruction is determined by how well the final concatenated reconstruction compares with the calibration series during the period of overlap.

Two other techniques, 'early' and 'late', where data are withheld during model calibration to provide an independent validation series, are also performed to address the model uncertainty and reliability. In these schemes, the PCR model will be calibrated separately to both the first 30 years (1957-1986, the 'early' reconstructions) and the last 30 years of station data (1982-2011 or 1984-2013, the 'late' reconstructions). A similar approach of performing the calibration on the beginning and end half of the data for model validation was employed by *Steig et al.* [2009] in their Antarctic surface temperature reconstructions. In each case, a validation series of at least 25 years is made available, and a reconstruction computed from the PCR model can be compared to these 25 years of direct observations in order to assess the model performance.

4. Results

a) Reconstructions except Byrd Station

We first evaluate the pressure reconstruction performance at all stations except Byrd station, which is evaluated separately due to the larger percentage of missing data. Figure 2 provides box plots of the reconstruction skill metrics, by season, for the highest performing reconstructions across the 8 different approaches (detrended vs. trended; 5% vs. 10% networks, and data ending in 2011 vs. 2013). Furthermore, Fig. 2 compares the reconstruction skill from reconstructions based on the full calibration period, as well as the PCR model calibrated on the first 30 years (‘early period’) and last 30 years (‘late period’) as described previously.

Notably, the reconstruction skill is remarkably high in DJF: except for the early period reconstructions, all skill metrics are above 0.40, and correlations (both calibration and validation) are above 0.50. In comparison, the reconstruction skill is lowest in SON: despite these being the best performing reconstructions, the CE for the late period reconstruction at Amundsen-Scott is weakly negative (-0.035); the median in many other metrics is ~0.40 in this season (not shown). The reconstructions during MAM and JJA are of comparable skill, falling between DJF and SON. Examining across the different calibration / validation periods, there are not significant differences. However, the late period reconstructions often outperform the early period reconstructions outside of SON, especially in DJF. Calibrating over the last 30 years (as is done in the late period reconstructions) produces slightly higher skill since many stations have pressure trends during the last 30 years [*Turner et al.*, 2005; and discussed here later], and the missing data, if present in the Antarctic observations (Table 1), occurs more frequently in the earlier part of the record rather than the later (i.e., a few stations have start dates after 1957). Nonetheless, Fig. 2 demonstrates that skillful Antarctic pressure reconstructions are possible, and even the lower values of RE and CE in some seasons exceed the continent-average RE and CE values observed in the *Steig et al*., [2009] temperature reconstructions.

Given the similarity between the full, early, and late reconstructions, and that if anything the full reconstructions provide a more conservative estimate of the overall reconstruction performance (since the late reconstructions often perform better), the remainder of the paper will focus on a further evaluation of the full reconstructions only. The sensitivity of the reconstruction methodology is further examined in Fig. 3, which shows the full reconstructions for three key stations, for all 8 different approaches tested, as well as the mean cross-correlation between all possible reconstruction pairs from the 8 methods in the upper right above each panel. The three stations were chosen to provide a rough geographic sampling of the Antarctic continent: Bellingshausen represents the Antarctic Peninsula; Amundsen-Scott provides a representation of reconstruction performance on the Antarctic Plateau; Casey highlights the performance along coastal East Antarctica. While there are naturally differences between the reconstructions, these are most marked in the extent of the variability or shifts in the mean values of the reconstruction back in time. At Bellingshausen, the mean correlation between the various approaches is above 0.90 in all seasons. Similarly high mean cross-correlations are also seen at Amundsen-Scott (except in JJA and SON) and all but SON for Casey. Therefore, the reconstruction performance is fairly stable and independent of the precise approach employed; this is true except for where the reconstruction skill is particularly low, as at Amundsen-Scott in SON, for example.

To demonstrate the stability graphically, boxplots comparing the performance of the full reconstructions from the detrended / raw (trended) data as well as the 5% and 10% predictor networks are displayed in Fig. 4. The differences are rather small and statistically insignificant between these trials, although the raw data do have smaller ranges and a slightly higher performance than the detrended data in DJF and MAM. This again relates to the fact that Antarctic pressure trends are strongest in these two seasons. Figure 4 also slightly suggests that the 10% network performs better in general compared to the 5% network, but this is dependent on the station. In most cases, including additional stations strengthens the PCR model only slightly, and the median values and overall range are comparable in most panels in the bottom row of Fig. 4. Additionally, using data that ended in 2011 or 2013 had negligible influence on the overall performance, with mean differences in the various skill metrics of about +0.03 in DJF (2013 performed slightly better) to -0.01 in MAM (2011 performed slightly higher). We therefore deem the PCR model to provide stable seasonal Antarctic pressure reconstructions through the various methods tested.

Not surprisingly, there are differences in the reconstruction performance spatially across Antarctica, as demonstrated by the calibration correlations of the best full period reconstructions in Fig. 5. The skill is highest near the Antarctic Peninsula in all seasons, due to the proximity of the predictor station Orcadas located northeast of the Antarctic Peninsula (Fig. 1b). In DJF, as indicated by the box plot in Fig. 2, the skill is high across the entire continent, including the interior. Notably, the reconstruction skill drops considerably outside of DJF across much of coastal East Antarctica, especially for Novolazarevskaya and Syowa stations in SON. The skill in the Antarctic plateau is also notably lower in SON. The reason for the lower skill in these regions will be discussed later.

Before examining a few individual reconstructions, we provide two additional evaluation metrics. First, we examine how well the best full period reconstructions produce the observed pressure trends at each Antarctic station in Fig. 6. It should be noted that this figure is comprised of several reconstructions where the PCR model was calibrated using detrended data, so in these cases the reproduction of the observed trends is an additional independent test on the reconstruction reliability. In Fig. 6, the trends and 95% confidence intervals are grouped geographically, with the observed values in black and the reconstructions in red. The trends are calculated during the period 1957-2013, from the observation starting year through the end of the reconstruction (either 2011 or 2013, depending on which reconstruction performed better at each station). The confidence intervals in Fig. 6 reflect the 'goodness of fit' of the regression lines, not the observation uncertainty or skill in the reconstructions; this facilitates comparisons between the imprecise trend estimates in both observations and reconstructions.

In all seasons and stations, there is not a statistically significant difference between the observed and reconstructed trend, even despite some of the lower reconstruction performance in SON. In the majority of the locations, the reconstructed trends are nearly identical to the observed trends, giving further confidence in the ability to reconstruct (in particular) the low-frequency variability. However, in a number of locations in East Antarctica, the reconstructions produce a statistically significant trend (p<0.05) while the observations do not. This is most likely due to the slightly dampened interannual variability in the reconstructions, which reduce the extent of the 95% confidence intervals. Nonetheless, every significant (p<0.05) observed Antarctic pressure trend (mostly in DJF, as discussed earlier) is also reproduced as a significant trend in the reconstructions. As before, the performance overall is higher for the Antarctic Peninsula and Halley stations.

Although the reconstructions were done individually by season and station, to be deemed reliable they should capture the inherent cross-correlations that are present in the observations between the stations. To measure this, a scatter plot of the cross correlations for various station pairs, grouped geographically, is presented in Fig. 7. The station groupings are as follows: Western Antarctic Peninsula (Faraday, Rothera); Northern Antarctic Peninsula (Bellingshausen, Esperanza, O'Higgins / Marsh, Marambio); East Antarctica (Casey, Davis, Mawson, Mirny); Dronning Maud Land (Halley, Novolazarevskaya, Syowa); Ross Sea Region (Dumont d'Urville, McMurdo / Scott Base); Antarctic Plateau (Amundsen-Scott, Vostok). A perfect reconstruction would mirror exactly the correlation between these pairs in the observed records, and would therefore fall on the thickened x=y line in Fig. 7.

In DJF, the reconstructions again perform very well on this metric, reproducing the high correlation between the observed station pairs, with only a very slight (but insignificant) stronger cross-correlation in some of the East Antarctic stations. The reconstruction performance based on this metric is also fairly high in JJA, with the reconstruction correlations being slightly lower in the East Antarctic stations. The ability to capture the spatial structure of the Antarctic pressure correlations is more mixed in MAM and SON. In MAM, the reconstructions have a much stronger correlation between McMurdo and Dumont d'Urville than observed, but weaker correlations again in East Antarctica, particularly for the Mirny-Casey pair, r=0.84 in observations but r=0.63 in the reconstruction, the greatest absolute difference in Fig. 7b. In SON, the weaker reconstruction skill is again captured by this metric, and in particular the lower skill at Syowa, Novolazarevskaya, and Amundsen-Scott (Fig. 5, bottom panel) stand out with large absolute differences in Fig. 7d. The challenges in reconstructing these locations will be investigated in section 4c.

The time series for the full period reconstructions with the highest and lowest skill (overall measures presented in the boxplots in Fig. 2) for each season are presented in Fig. 8; low-frequency version of these time series, smoothed with an 11-year Hamming filter, are provided in Fig. 9. A Hamming filter was used over the more commonly used running mean in order to give higher weight to the middle of the averaging period and therefore better highlight decadal-scale variability. In both Fig. 8 and 9, the gray shading represents the 95% confidence intervals for the reconstructions. These were calculated as 1.96 times the standard deviation of the residuals, defined as the difference between the reconstruction and observations (smoothed differences in Fig. 9). In all but DJF, Bellingshausen is the station with the highest reconstruction skill, due to its proximity and strong relationship with the predictor station Orcadas near the northeastern Antarctic Peninsula (Fig. 1). The best reconstruction consistently has a calibration and validation correlation above r=0.85 and r=0.80, respectively. The interannual variability is well captured, including several years with rapid pressure changes and notable extreme pressure minima or maxima. In DJF, the downward trend at Halley since the mid 1960s is captured remarkably well (Fig. 8a; Fig. 6a). The validation correlations (from the leave-one-out cross validation procedure) are particularly high in DJF and JJA for these highest performing stations. This indicates the very robust nature of these reconstructions, higher than that seen for temperature reconstructions [*Steig et al.*, 2009] or SAM index reconstructions [*Jones et al.*, 2009].

For the stations with the lower skill, the reconstruction performance in DJF is still higher than or comparable to Antarctic temperature reconstructions [*Steig et al.* 2009] or SAM index reconstructions. Outside of this season, the lowest performing reconstructions capture portions of the interannual variability in the observed records, but notably underestimate large portions of the observed variability. The Syowa reconstruction in JJA (Fig. 8f) and Amundsen-Scott reconstruction in SON (Fig. 8h) have much smaller ranges of pressure values than the observed records, and as such the RE and CE values for these stations are only weakly positive (the minimum values in the full period boxplots in Figs. 2c and d). Nonetheless, despite challenges in reproducing the interannual variability for these locations, at most locations and seasons the low-frequency variability is well captured, as indicated in Fig. 9 (note different y-axis scale in Fig. 9 compared to Fig. 8). The median of the correlations is above r=0.82 in all but SON for the smoothed time series, again reflecting the fact that low-frequency variability, including trends over longer time periods, are reasonably well captured (Fig. 5). For some stations, as at Bellingshausen in SON (Figs. 8g and 9g), the interannual variability is better captured than the low-frequency variability; this usually arises due to subtle differences in the timing of mean pressure changes. For example, at Bellingshausen during SON, although the reconstruction has similar interannual variability, the values are slightly lower on average for much of the late 1970s and early 1980s (Fig. 8g). When smoothed, this creates a notable difference in the low-frequency variability (Fig. 9g), and gives rise to a much weaker smoothed correlation than interannual (r=0.853 for interannual, r=0.477 for smoothed). It is noted, however, that such large decreases in the low-frequency performance are rare; the overwhelming majority of stations capture the low-frequency variability as well or even better than the interannual variability. This is an encouraging result, as the PCR model was calibrated on interannual variability, and no specific constraints were made on the ability of the model to directly capture low-frequency variability.

b) Byrd station reconstructions

As mentioned earlier, reconstructions were also performed for Byrd station in West Antarctica, but have not been included in the evaluation presented thus far since the observational gap in the 1970s presented additional challenges in reconstructing pressure at the station. Additionally, ERA-Interim surface pressure data were used to patch the observed record, and this station therefore is unlike the others in being a blend of direct observations and reanalysis data. Despite these challenges, the reconstructions at Byrd station only once have the lowest skill compared to the other stations, and demonstrated modest skill in DJF, as shown in the full period reconstruction time series in Fig. 10. The early reconstructions calibrated during the first 30 years performed lower than both the full period and late period reconstructions, which were comparable in overall skill (not shown).

During DJF and SON, much of the interannual variability is captured by the Byrd reconstruction, although the extent of individual peaks and troughs in the observed record is not fully captured (especially in the early observation record during SON, Fig. 10d). The overall skill in SON, however, is considerably higher than at Amundsen-Scott (Fig. 8h): the validation correlation for Byrd in SON was higher than the calibration correlation at Amundsen-Scott. In contrast, the JJA Byrd reconstruction was altogether the lowest skill full period reconstruction (Fig. 10c), and it is clear that much of the large interannual variability, if present at all, is severely dampened in the reconstruction. Because of the reduced variability in the reconstruction, the smoothed (low-frequency) performance at Byrd station (not shown) is often lower than the interannual, although this metric is harder to compare directly with other reconstructions due to the gap in the Byrd observations. However, in all seasons, the full period reconstructions at Byrd maintained positive RE and CE values (the lowest being the CE of 0.17 in JJA), indicating that the reconstructions are still performing better than compared to the climatological mean. By comparison, Antarctic temperature reconstructions by *Steig et al.* [2009] were not always able to produce positive RE and CE values within the Antarctic interior..

c) Challenges in reconstructing stations

While the reconstruction skill is remarkable in DJF, and fairly high in all other seasons near the Antarctic Peninsula, the skill declines markedly in East Antarctica and to some extent over the Antarctic plateau, particularly in SON. One of the lowest performing stations is Syowa station (Figs. 5 and 8), and we use this station as an example here to address some of the broader issues in maintaining the high level of reconstruction skill outside of DJF. Figure 11 displays the correlation of ERA-Int MSLP from the gridpoint closest to Syowa station (indicated with the large circle) with every other gridpoint south of 15°S, during 1979-2013. The gray shading highlights correlations that are significantly different from zero at $p<0.10$. A similar method was used to determine the predictor stations for each reconstruction by season, except that these correlations were based solely on observed values and during 1957-2013. To highlight the predictors used in the Syowa (full period) reconstruction, individual filled circles are plotted across the Southern Hemisphere, with the color and size representing the weight each predictor location had in the final reconstruction. In general, the sign and magnitude follow the correlation pattern, although some stations close together may be weighted differently depending on the number of principal components retained in the PCR model, and the relationships each of these stations had with these principal components (i.e., some stations in New Zealand are strongly weighted, and not all are of the same sign).

In DJF, as with nearly every station, there is a strong relationship between the pressures across Antarctica and the midlatitudes of the Southern Hemisphere, and the significant correlations extend over many of the continents where the majority of the predictor data are available. As such, there are considerably more stations included in the reconstruction, and the reconstruction is better constrained and more stable. Due to changes in the climatological jet, the correlation pattern changes seasonally [*Fogt et al.*, 2012b], and in particular for stations in East Antarctica and in the Antarctic interior, the relationship between Antarctica and the Southern Hemisphere midlatitudes weakens considerably. Near the Antarctic Peninsula where the skill is higher in all seasons (Fig. 5), stations in nearby South America as well as the Orcadas station are able to more strongly constrain the reconstruction. In SON, the position of the climatological wave-three pattern is such that only portions of South America and the south island of New Zealand show significant correlations with the Syowa station (Fig. 11d, and also at Novolazarevskaya and Amundsen-Scott, not shown). The smaller number of stations therefore used in the reconstruction make it more challenging to capture the full extent of variability at the Antarctic station being reconstructed, and the reconstruction skill drops as a result.

Nonetheless, Fig. 11 highlights an important concept that has been further exploited in order to improve reconstruction skill, namely that there are strong correlations that occur over the midlatitude ocean basins. Unfortunately, no long-term continuous direct observations are available at these locations, even if they fall near an island (as much of meteorological records from the islands in the southern Atlantic and Indian Oceans start in the 1940s or 1950s). We therefore make use of gridded data to estimate the pressure in these regions of highly significant correlations to provide additional sources of data input for our PCR model, and term these 'pseudo-reconstructions' since they are made of a blend of direct observations and gridded pressure data.

5. Pseudo-reconstructions

a) Methodology

We conducted four pseudo-reconstructions for each station / season, one based on HadSLP2 and another based on 20CR for both the 5% and 10% midlatitude station networks; we have evaluated the performance of recently released ERA-20C and found this reanalysis to be an outlier compared to the other two products in the early 20th century, as will be discussed in more detail in our companion paper [*Fogt et al.*, 2016]. We therefore have not conducted pseudo-reconstructions for ERA-20C. For each product, we calculated correlation maps of the model gridpoint closest to the Antarctic station of interest by season, during 1979-2013, as in Fig. 11. From these maps, we selected regions (encompassing several gridpoints) of large, highly significant correlation ($r>0.40$, $p<0.01$) over the ocean basins; these regions were then area-averaged to provide an individual time series.

To address the uncertainty and reliability in these time series, particularly in the earlier portions of the 20th century, we make use of the number of observations available from HadSLP2 [totaling the total number of direct observations in the each of the gridboxes; *Allan and Ansell*, 2006] and the ensemble spread from 20CR [*Compo et al.*, 2011]. For HadSLP2, seasonal data were only included in these time series if at least one direct observation was included in the region, otherwise the data were treated as missing. For the 20CR, the data were only included if the area-averaged ensemble spread was less than 4 standard deviations from the variability in the area-averaged pressure time series. If the ensemble spread exceeded four standard deviations in any season, the data were similarly treated as missing. As with observations, if more than 75% of the pseudo data were missing, the time series was not used in the reconstruction. For the pseudo-reconstructions, areas were chosen over a single grid point since it is more likely that observations were included in HadSLP2, and generally the ensemble spread decreased with increasing area (most likely tied to the inclusion also of more ship data *in situ*). Finally, these time series, usually from 2-4 regions, were appended to the direct observation matrix, and the PCR model was re-run and the pseudo-reconstructions were produced. While we continued to construct full, early, and late reconstructions based on both 5% and 10% midlatitude networks to more fully evaluate the stability of the reconstructions, we only worked with raw / trended data through 2011, as these changes had much smaller effect on the overall reconstruction skill (Fig. 4).

b) Pseudo-reconstruction performance

Figure 12 displays the reconstruction statistics for the best full period reconstruction (as in Fig. 2), as well as the pseudo-reconstructions based on both 20CR and HadSLP2. The numbers on each panel provide the mean difference between the best original full period reconstruction and the best overall pseudo-reconstruction; a positive value indicates that the pseudo-reconstructions performed higher on average. In DJF, there are small differences between the original method and the pseudo-reconstructions, and the pseudo-reconstructions actually produce slightly lower RE and CE values. This is perhaps not surprising, given the high skill obtained in the original reconstructions. In all other seasons, the pseudo-reconstructions improve the original reconstructions. The improvement is most marked in JJA, where the pseudo-reconstructions are of high quality and comparable to the original reconstructions in DJF. The improvement is less in other seasons as many of the significant correlations in MAM and SON occur in more southern portions of the ocean basins, where the HadSLP2 and 20CR are more uncertain (and therefore less pseudo data are used). Despite the coarse resolution of HadSLP2 (5°x5° latitude-longitude), there is no clear preference between HadSLP2 and 20CR in terms of their reconstruction skill; both perform similarly. This is likely a reflection that even with the addition of the pseudo data from these products, the majority of the data used in the reconstructions is the same, coming from direct observations.

The pseudo-reconstruction performance is examined spatially in Fig. 13 through the calibration correlations of the best full period pseudo-reconstruction, as in Fig. 5. The numbers by each station indicate the difference between the best original full period reconstruction and the best pseudo-reconstruction, with positive values indicating an improvement in the pseudo-reconstruction. As noted earlier, the improvements are small in DJF, and similarly because of the relatively high skill in the original reconstructions along the Antarctic Peninsula, the improvements here are also modest in all seasons. In

MAM, the pseudo-reconstructions lead to widespread improvements in the reconstruction skill, most notably at Halley, Amundsen-Scott, and McMurdo / Scott Base, where the calibration correlations all increase by more than 0.10. In JJA, as indicated in Fig. 12, the pseudo-reconstructions dramatically improve the overall reconstruction skill, and many stations see improvements in the calibration correlation of more than 0.10 outside of the Antarctic Peninsula; the improvement at Syowa station is remarkable with an increase in the calibration correlation of 0.23 (this was previously the worst performing station in this season, Fig. 8f). In SON, the pseudo-reconstructions have a mixed performance, increasing the skill across the Antarctic Peninsula, but only slight changes (including a few decreases in the calibration correlation) along much of the Antarctic coast. Notably, the pseudo-reconstruction skill at Amundsen-Scott in SON, the worst performing original reconstruction in this season (Fig. 8h), decreases slightly (the calibration correlation drops by 0.07, Fig. 13). As with Syowa station (Fig. 11), there are few predictor stations, and the uncertainties in both HadSLP2 and 20CR (and therefore missing pseudo data) make this interior station challenging to reconstruct by these approaches in SON.

6. Conclusions

This paper has examined the reliability of several seasonal Antarctic station-based pressure reconstructions since 1905, using midlatitude pressure data as predictors in the model calibration. The reconstructions were based on principal component regression, a method successfully used in previous temperature [*Mann et al.*, 2008, 2009] and SAM index reconstructions [*Jones et al.*, 2009, *Fogt et al.*, 2009]. To further test this approach, we performed multiple reconstructions at each station and season: separate reconstructions based on two networks of midlatitude pressure data, reconstructions based on all data detrended before model calibration and another based on the original data with any trends, as well as one reconstruction with data ending in 2011, and another updated with all data ending in 2013. The reconstructions were validated in three ways: one in which the model was calibrated over the full period of observations, and validated using the leave-one-out cross validation procedure, as in *Jones et al.* [2009] for SAM index reconstructions, and two other approaches where only 30 years were used for model calibration and the remaining 25 or 27 years used independently for model validation. In all seasons, reconstructions that outperformed the climatological mean were possible at all stations across Antarctica, indicated by positive values in both the reduction of error and coefficient of efficiency. However, there are important differences in reconstruction skill, both seasonally and regionally.

In general, reconstruction skill was considerably higher in austral summer, where calibration correlations were frequently high ($r>0.80$), and other skill metrics were consistently above 0.50, due to stronger relationships with midlatitude predictor data. Spatially, reconstruction skill was highest in and near the Antarctic Peninsula, due to the strong weight of the nearby station Orcadas (situated east and slightly north of the Antarctic Peninsula) in the final reconstructions. The reconstruction skill tended to be the lowest during austral spring, especially along coastal East Antarctica. Comparing across the various reconstruction techniques, all produced very similar reconstructions, although there was a tendency for reconstructions based on the original / trended data to perform better in austral summer and autumn, when recent pressure trends are the strongest in the Antarctic station data. In addition to the main metrics of reconstruction skill, we also demonstrated that the majority of the pressure trends in the Antarctic records were reproduced by the reconstructions during 1957-2013, as well as the inherent spatial correlation between subgroups of Antarctic stations, especially outside of austral spring.

The reconstruction performance is naturally, and strongly, impacted by the location and relationship of midlatitude pressure data with the individual Antarctic station being reconstructed. For cases along coastal East Antarctica and the Antarctic plateau during austral spring (and to some extent winter), these relationships are moderately weaker, and stronger relationships are observed over the ocean basins rather than on islands / continents with long-term meteorological measurements. In an attempt to improve the reconstruction skill, pseudo-reconstructions were therefore performed, employing area-averaged gridded pressure data from HadSLP2 [*Allan and Ansell*, 2006] or the NOAA 20th century reanalysis [*Compo et al.*, 2011]. Both of these pseudo-reconstructions performed similarly, and outside of austral summer where the original reconstruction skill was already high, the pseudo-reconstructions improved the original reconstructions. The improvement was largest in austral winter, with the pseudo-reconstructions having similar skill across all of Antarctica as the original reconstructions in austral summer.

These reconstructions, especially in austral summer and along the Antarctic Peninsula, as well as the pseudo-reconstructions in austral winter, now afford a much longer investigation of the Antarctic atmospheric circulation variability during the 20th century. This analysis is the heart of our companion paper [*Fogt et al.,* 2016], where comparisons are made to other long-term climatological datasets to better evaluate their performance prior to Antarctic observations. Recent trends and variability are also placed in a longer historical context from the near doubling of the length of the Antarctic pressure reconstructions. Future work includes creating an Antarctic continental wide pressure reconstruction, as well as comparing to several unique new climate model simulations currently being processed in order to better understand the relative roles of various external and internal forcing mechanisms in causing pressure changes over Antarctica during the 20^{th} century.

Acknowledgments

All authors but JMJ acknowledge support from NSF grant #1341621. Thanks are extended to ECMWF for their reanalysis data (http://apps.ecmwf.int/datasets/), the British Antarctic Survey for hosting the Antarctic READER data (https://legacy.bas.ac.uk/met/READER/), NOAA Earth System Research Laboratory for the 20^{th} Century Reanalysis data (http://www.esrl.noaa.gov/psd/data/gridded/data.20thC_ReanV2c.html), the UK Hadley Centre for HadSLP2 data (http://www.metoffice.gov.uk/hadobs/hadslp2/), and the NCAR/UCAR research data archive for the majority of mid-latitude pressure data (http://rda.ucar.edu/datasets/ds570.0/#!description). The pressure reconstructions generated and evaluated here can be made available by request through email to the corresponding author RLF (fogtr@ohio.edu).

References

Allan, R. J., and T. J. Ansell (2006), A new globally complete monthly historical mean sea level pressure data set (HadSLP2): 1850-2004, *J. Clim.*, *19*, 5816-6842.

Bromwich, D. H., and R. L. Fogt (2004), Strong trends in the skill of the ERA-40 and NCEP/NCAR Reanalyses in the high and middle latitudes of the Southern Hemisphere, 1958-2001, *J. Clim.*, *17*, 4603-4619.

Bromwich, D. H., R. L. Fogt, K. I. Hodges, and J. E. Walsh (2007), A tropospheric assessment of the ERA-40, NCEP, and JRA-25 global reanalyses in the polar regions, *J. Geophys. Res.*, *112*, doi:10.1029/2006JD007859.

Bromwich, D. H., J. P. Nicolas, A. J. Monaghan, M. A. Lazzara, L. M. Keller, G. A. Weidner, and A. B. Wilson (2013), Central West Antarctica among the most rapidly warming regions on Earth, *Nat. Geosci.*, *6*, 139-145, doi:10.1038/ngeo1671.

Bromwich, D. H., J. P. Nicolas, A. J. Monaghan, M. A. Lazzara, L. M. Keller, G. A. Weidner, and A. B. Wilson (2014), Corrigendum: Central West Antarctica among the most rapidly warming regions on Earth, *Nat. Geosci.*, *7*, doi:10.1038/ngeo2016.

Chapman, W.L. and J. E. Walsh (2007), A synthesis of Antarctic temperatures, *J. Clim.*, *20*, 4096-4117.

Compo, G. P. et al. (2011), The Twentieth Century Reanalysis Project, *Q. J. R. Meteorol. Soc.*, *137*, 1-28, doi:10.1002/qj.776.

Cook, E. R., D. M. Meko, D. W. Stahle, and M. K. Cleveland (1999), Drought reconstructions for the continental United States, *J. Clim.*, *12*, 1145-1163.

Dee, D. P., et al. (2011), The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, *137*, 553-597.

Fogt, R. L., J. Perlwitz, A. J. Monaghan, D. H. Bromwich, J. M. Jones, G. J. Marshall (2009), Historical SAM Variability. Part II: 20th century variability and trends from reconstructions, observations, and the IPCC AR4 Models, *J. Clim.*, *22*, 5346-5365.

Fogt, R. L., A. J. Wovrosh, R. A. Langen, and I. Simmonds (2012a), The Characteristic Variability and Connection to the Underlying Synoptic Activity of the Amundsen-Bellingshausen Seas Low, *J. Geophys. Res.*, *117*, doi:10.1029/2011JD017337.

Fogt, R.L., J. M. Jones, and J. Renwick (2012b), Seasonal zonal asymmetries in the Southern Annular Mode and their impact on regional climate anomalies, *J. Clim.*, *25*, 6253-6270.

Fogt, R. L., J. M. Jones, C. A. Goergens, M. E. Jones, G. Witte, and M. Y. Lee (2016), Antarctic Station Based Seasonal Pressure Reconstructions Since 1905, Part 2: Variability and trends during the 20th century, *J. Geophys. Res.,* in review.

Holland, P. R., and R. Kwok (2012), Wind-driven trends in Antarctic sea-ice drift, *Nat. Geosci.*, *5*, 872-875, doi:10.1038/NGEO1627.

Hosking, J. S., A. Orr, G. J. Marshall, J. Turner, and T. Phillips (2013), The influence of the Amundsen-Bellingshausen Seas Low on the climate of West Antarctica and its representation in coupled climate model simulations, *J. Clim.*, *26*, 6633-6648.

Jones, J. M., and M. W. Widmann (2004), Atmospheric Science- Early peak in Antarctic oscillation index, *Nature*, *432*, 290-291.

Jones, J. M., R. L. Fogt, M. Widmann, G. Marshall, P. D. Jones, and M., Visbeck (2009), Historical SAM Variability. Part I: Century Length Seasonal Reconstructions, *J. Clim.*, *22*, 5319-5345.

Jones, P. D. (1987), The early twentieth century Arctic high- Fact or fiction? *Clim. Dyn.*, *1*, 63-75.

Jones, P. D., M. J. Salinger, and A. B. Mullan (1999), Extratropical circulation indices in the Southern Hemisphere based on station data, *Int. J. Climatol.*, *19*, 1301-1317.

Mann, M. E., R. E. Bradley, and M. K. Hughes (1998), Global-scale temperature patterns and climate forcing over the past six centuries, *Nature*, *392*, 779-787.

Mann, M. E., R. E. Bradley, and M. K. Hughes (1999), Northern Hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitation, *Geophys. Res. Lett.*, *26*, 759-762.

Monaghan, A. J., D. H. Bromwich, W. Chapman, and J. C. Comiso (2008), Recent variability and trends of Antarctic near-surface temperature, *J. Geophys. Res.*, *113*, doi:10.1029/2007JD009094.

Nicolas, J. P., and D. H. Bromwich (2014), New reconstruction of Antarctic near-surface temperatures: multidecadal trends and reliability of global reanalyses, *J. Clim.*, *27*, 8070-8093, doi:10.1175/JCLI-13-00733.1

O'Donnell, R., N. Lewis, S. McIntyre, J. Condon (2011), Improved Methods for PCA-Based Reconstructions: Case Study Using the Steig et al. (2009) Antarctic Temperature Reconstruction, *J. Clim.*, *24*, 2099-2115.

Peterson, T. C., and R. S. Vose (1997), An overview of the global historical climatology network temperature database, *Bull. Amer. Meteorol. Soc.*, *78*, 2837-2849.

Peterson, T. C., R. Vose, R. Schmoyer, V. Razuvaev (1998), Global historical climatology network (GHCN) quality control of monthly temperature data, *Int. J. Climatol.*, *18*, 1169-1179.

Raphael, M. N., G. J. Marshall, J. Turner, R. L. Fogt, D. Schneider, D. A. Dixon, J. S. Hosking, J. M. Jones, W. R. Hobbs (2015), The Amundsen Sea Low: Variability, change, and impact on Antarctic climate, *Bull. Amer. Meteorol. Soc.*, in press.

Steig, E. J., D. P. Schneider, S. D. Rutherford, M. E. Mann, J. C. Comiso, D. T. Shindell (2009), Warming of the Antarctic ice-sheet surface since the 1957 International Geophysical Year, *Nature*, *457*, 459-462.

Turner, J., S.R. Colwell, G.J. Marshall, T.A. Lachlan-Cope, A.M. Carleton, P.D. Jones, V. Lagun, P.A. Reid, S. Iagovkina (2004), The SCAR READER project: Toward a high-quality database of mean Antarctic meteorological observations, *J. Clim.*, *17*, 2890-2898.

Turner, J., S. R. Colwell, G. J. Marshall, T. A. Lachlan-Cope, A. M. Carleton, P. D. Jones, V. Lagun, P. A. Reid, and S. Iagovkina (2005), Antarctic Climate Change during the Last 50 years, *Int. J. Climatol.*, *25*, 279-294.

Turner, J., T. Phillips, J. S. Hosking, G. J. Marshall, and A. Orr (2013), The Amundsen Sea low, *Int. J. Climatol.*, *33*, doi:10.1002/joc.3558.

Visbeck, M. (2009), A station-based Southern Annular Mode index from 1884 to 2005, *J. Clim.*, *22*, 940-950.

Table Captions

Table 1. Details of Antarctic stations reconstructed. Station ID = World Meteorological Organization (WMO) station identifier, and start yr = starting year of pressure observations. Patched with IDs lists any WMO IDs used in patching the record, if applicable. All stations extend through 2013. The percent complete is based on the total number of monthly records available after any patching from the start of the station through 2013.

Table 2. As in Table 1, but for Southern Hemisphere midlatitude pressure records that begin before 1905. Percent complete for Auckland data is calculated during 1900-2013.

Figure Captions

Figure 1. Map of a) Antarctic pressure stations reconstructed and b) midlatitude predictor stations used in the reconstructions, with records extending back until 1905.

Figure 2. Box plots from the 17 main stations of seasonal reconstruction skill metrics for the highest performing reconstructions for full, early, and late period reconstruction techniques. Cal r = calibration correlation, Val r = validation correlation. RE and CE reflect the performance of the reconstruction and validation series, respectively, against climatology during the verification period.

Figure 3. Time series plots of all 8 reconstruction trials (as indicated in legend) for three select stations: Bellingshausen (left column, representing the Antarctic Peninsula); Amundsen-Scott (middle column, representing the Antarctic Plateau); Casey (right column, representing coastal East Antarctica). The number in the upper right of each panel is the mean cross-correlation among the 8 different reconstructions.

Figure 4. As in Fig. 2, but for boxplots highlighting skill metric comparisons between reconstructions based on detrended / original data (top row), and from the 5% and 10% networks (bottom row).

Figure 5. Spatial plot of calibration correlations for the best full period reconstructions.

Figure 6. Trends and 95% confidence intervals for the observations (black) and best full period reconstructions (red) during 1957-2013, from the start of the observations to the end of the reconstruction (either 2011 or 2013). Some station names have been shortened.

Figure 7. Scatterplot of cross-correlations between various station pairs during 1957-2013, grouped geographically by colors (see text for details). The x-axis denotes the correlations between the station pairs in the observed data, while the y-axis shows the correlations in the reconstructions.

Figure 8. Time series of the reconstructions with the highest (left column) and lowest (right column) skill by season. Also given is the calibration and validation correlation for each reconstruction. The gray shading represents the 95% confidence intervals, taken as 1.96 times the standard deviation of the reconstruction – observation residuals.

Figure 9. As in Fig. 8, but for the observations and reconstructions smoothed with an 11-yr Hamming filter. The correlation between the smoothed observations and reconstructions is given for each panel.

Figure 10. Time series for the Byrd station reconstructions in West Antarctica.

Figure 11. Map of MSLP correlations (contoured, significant correlations at $p<0.10$ shaded in gray) of the ERA-Int gridpoint closest to Syowa station (indicated with large black circle) and every other gridpoint from ERA-Int south of 15°S during 1979-2013. Also shown are the predictor stations used in the full period Syowa reconstruction, with the weight each midlatitude predictor station had in the final reconstruction indicated by the color / size of the circle, as indicated below the figure.

Figure 12. Boxplots of reconstruction statistics across the main 17 Antarctic stations for the best full period reconstruction ('Original') and the full period, 10% trended reconstructions from 20CR (red) and HadSLP2 (green). The numbers in each panel represent the mean difference between the best original and best pseudo-reconstruction (between 20CR and HadSLP2), with positive values indicating improvements in the pseudo-reconstructions.

Figure 13. Maps of the best full period pseudo-reconstruction calibration correlation by season, as in Fig. 5. The number by each station is the difference between the original and pseudo-reconstruction, with positive numbers indicating improvement in the pseudo-reconstructions

Table 1. Details of Antarctic stations reconstructed. Station ID = World Meteorological Organization (WMO) station identifier, and start yr = starting year of pressure observations. Patched with IDs lists any WMO IDs used in patching the record, if applicable. All stations extend through 2013. The percent complete is based on the total number of monthly records available after any patching from the start of the station through 2013.

Station Name	**Lat**	**Lon**	**Station ID**	**Start yr**	**Patched with IDs**	**% complete**
Amundsen-Scott	-90.0	0.0	890090	1957		100
Bellingshausen	-62.2	-58.9	890500	1959	889380 890570	99.85
Byrd	-80.0	-119.4	893240	1957	893240 and ERA-Int	84.30
Casey[+]	-66.3	110.6	896110	1957		99.71
Davis	-68.6	78.0	895710	1957		92.25
Dumont d'Urville	-66.7	140.0	896420	1956		99.57
Esperanza	-63.4	-57.0	889630	1945		94.69
Faraday / Vernadsky	-65.3	-64.3	890630	1947		95.15
Halley	-75.5	-26.7	890220	1957		100
Marambio	-64.2	-56.7	890550	1970		98.48
Mawson	-67.6	62.9	895640	1954		99.86
McMurdo / Scott Base	-77.9	166.8	896640/ 896650	1956		99.71
Mirny	-66.6	93.0	895920	1956		100
Novolazarevskaya	-70.8	11.8	895120	1961		99.84
O'Higgins / Marsh*	-63.3	-57.9	890590/ 890560	1969		97.55
Rothera[+]	-67.6	-68.1	890620	1946		90.69
Syowa	-69.0	39.6	895320	1957		91.67
Vostok	-78.5	106.9	896060	1958	898280	97.02

*For O’Higgins / Marsh, the latitude, longitude, ID, and Start yr given are that of O’Higgins station. Marsh is located very close to the Bellingshausen station on King George Island.

[+]Both Casey and Rothera records start earlier in the GHCN archive than in READER. These earlier records in GHCN were compiled from nearby temporary stations before the bases were established in 1959 and 1976, respectively.

Table 2. As in Table 1, but for Southern Hemisphere midlatitude pressure records that begin before 1905. Percent complete for Auckland data is calculated during 1900-2013.

Station Name	Lat	Lon	Station ID	Start yr	Patched with IDs	% complete
Adelaide	-35.0	138.5	946720	1857		99.68
Alice Springs	-23.8	133.9	943260	1885		99.55
Auckland	-37.0	174.8	931190	1863	ISPDv3 933090	99.63
Bahia Blanca	-38.7	-62.2	877500	1896		97.88
Brisbane	-27.4	153.1	945780	1887	845760	99.93
Buenos Aires	-34.6	-58.5	875850	1858	875760	99.15
Cape Town	-34.0	18.6	688160	1841		99.86
Catamarca	-28.6	-65.8	872220	1901	873450	98.75
Chatham Island	-44.0	-176.6	939870	1878		98.9
Christchurch	-43.5	172.5	937800	1864		98.61
Cordoba	-31.3	-64.2	873440	1873		99.76
Dunedin	-45.9	170.5	938940	1864	938440	99.44
Durban	-30.0	31.0	685880	1884		99.17
Hobart	-42.8	147.5	949750	1866		96.34
Hokitika	-42.7	171.0	936150	1866		91.16
Melbourne	-37.8	145.0	948680	1903	948650	99.92
Orcadas	-60.7	-44.7	889680	1903		99.55
Perth	-31.9	116	946100	1876		99.82
Port Elizabeth	-34.0	25.6	688420	1887		99.87
Punta Arenas	-53.0	-70.8	859340	1889		99.80
Rio de Janeiro	-22.9	-43.2	837430	1851	837810	99.49
Salta	-24.9	-65.5	870470	1901		97.42
Santiago	-33.4	-70.8	855740	1861		99.84
Sarmiento	-45.6	-69.1	878490	1903	878600	98.72
St. Helena Island	-16.0	-5.7	619010	1892		99.25
Sydney	-33.9	151.2	947680	1859	947670	99.3
Tahiti	-17.6	-149.6	919380	1876		100
Valdivia	-39.6	-73.1	857660	1899	857430	100
Wellington	-41.3	174.8	934340	1864	934170 936780	99.33

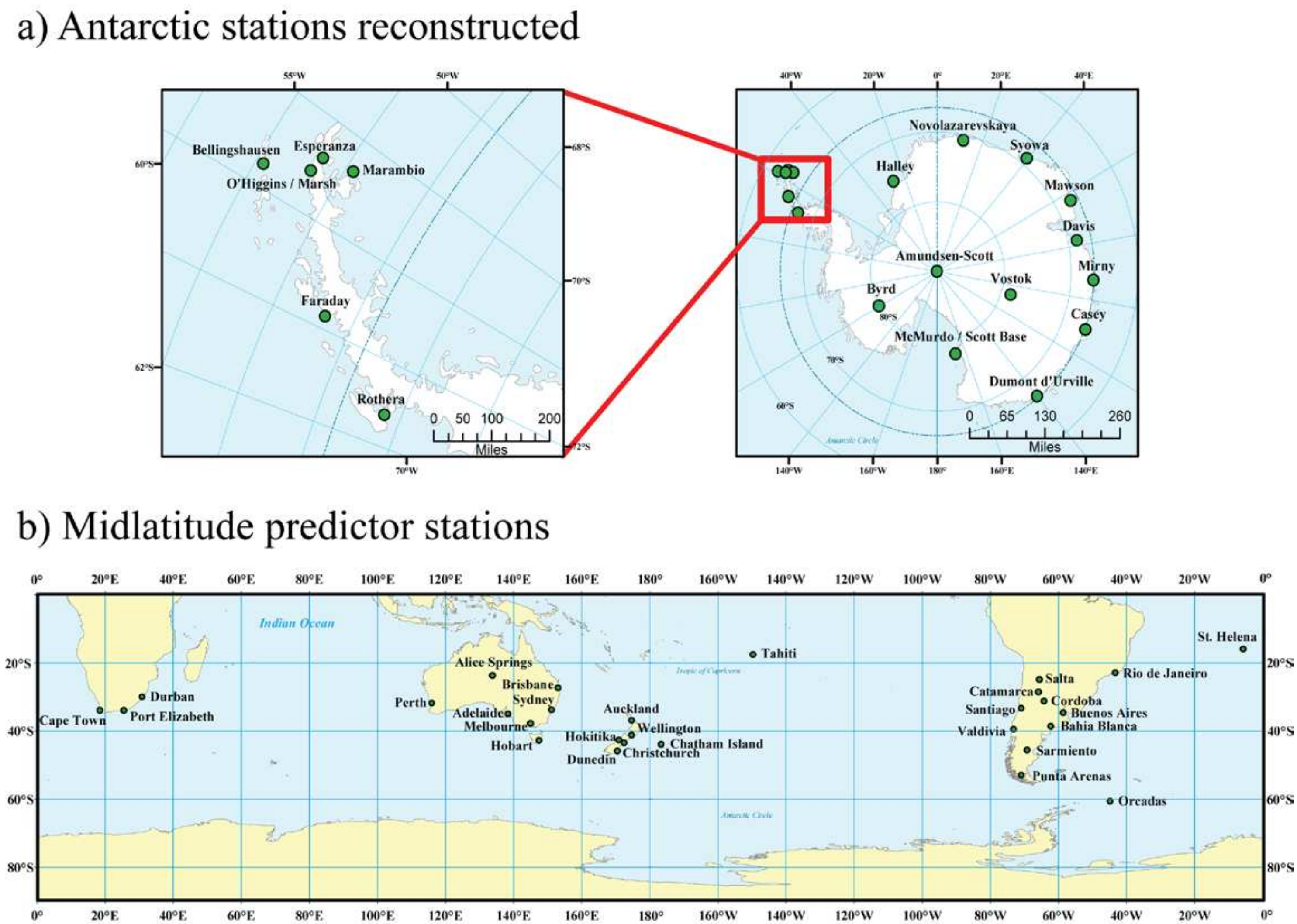

Figure 1. Map of a) Antarctic pressure stations reconstructed and b) midlatitude predictor stations used in the reconstructions, with records extending back until 1905.

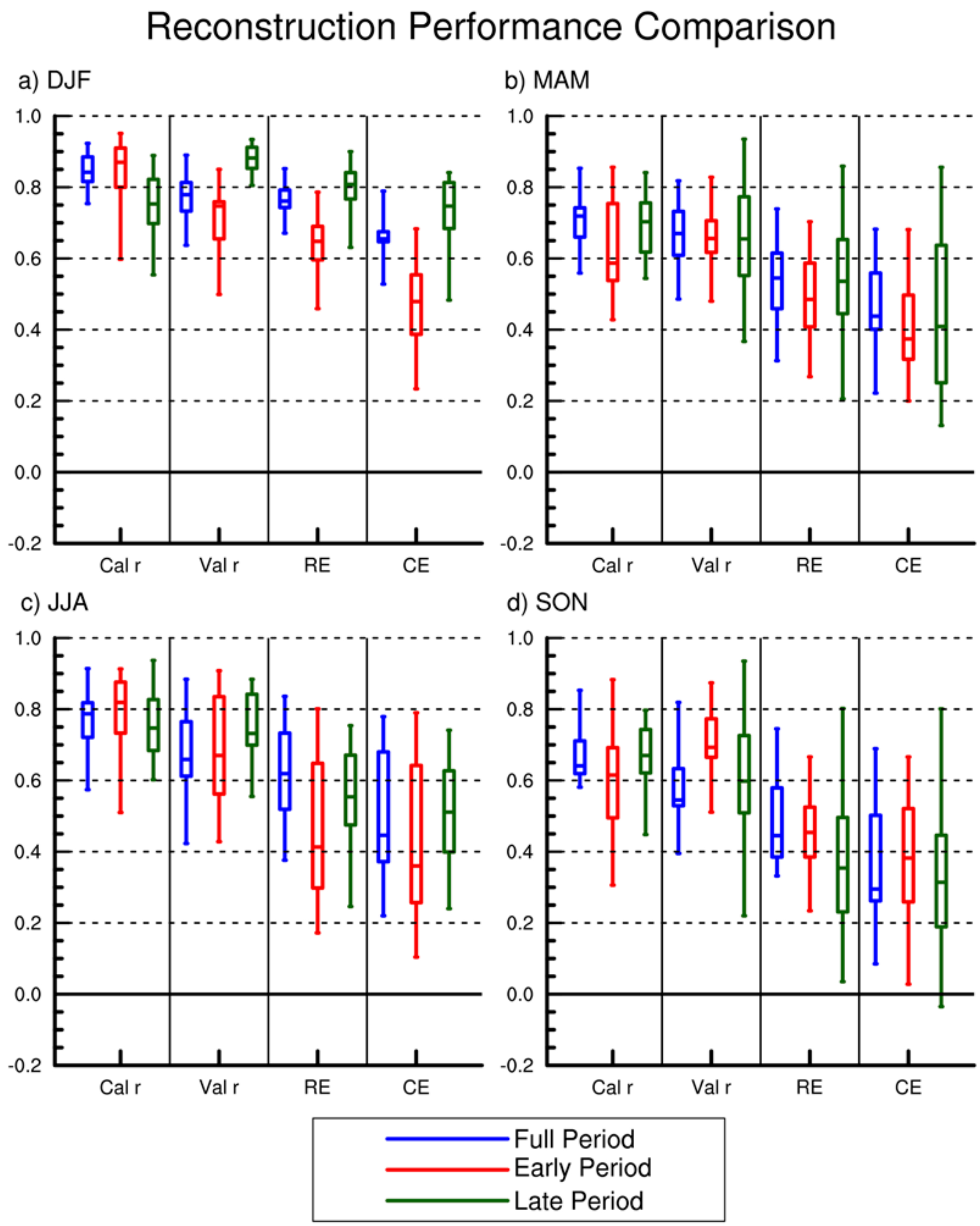

Figure 2. Box plots from the 17 main stations of seasonal reconstruction skill metrics for the highest performing reconstructions for full, early, and late period reconstruction techniques. Cal r = calibration correlation, Val r = validation correlation. RE and CE reflect the performance of the reconstruction and validation series, respectively, against climatology during the verification period.

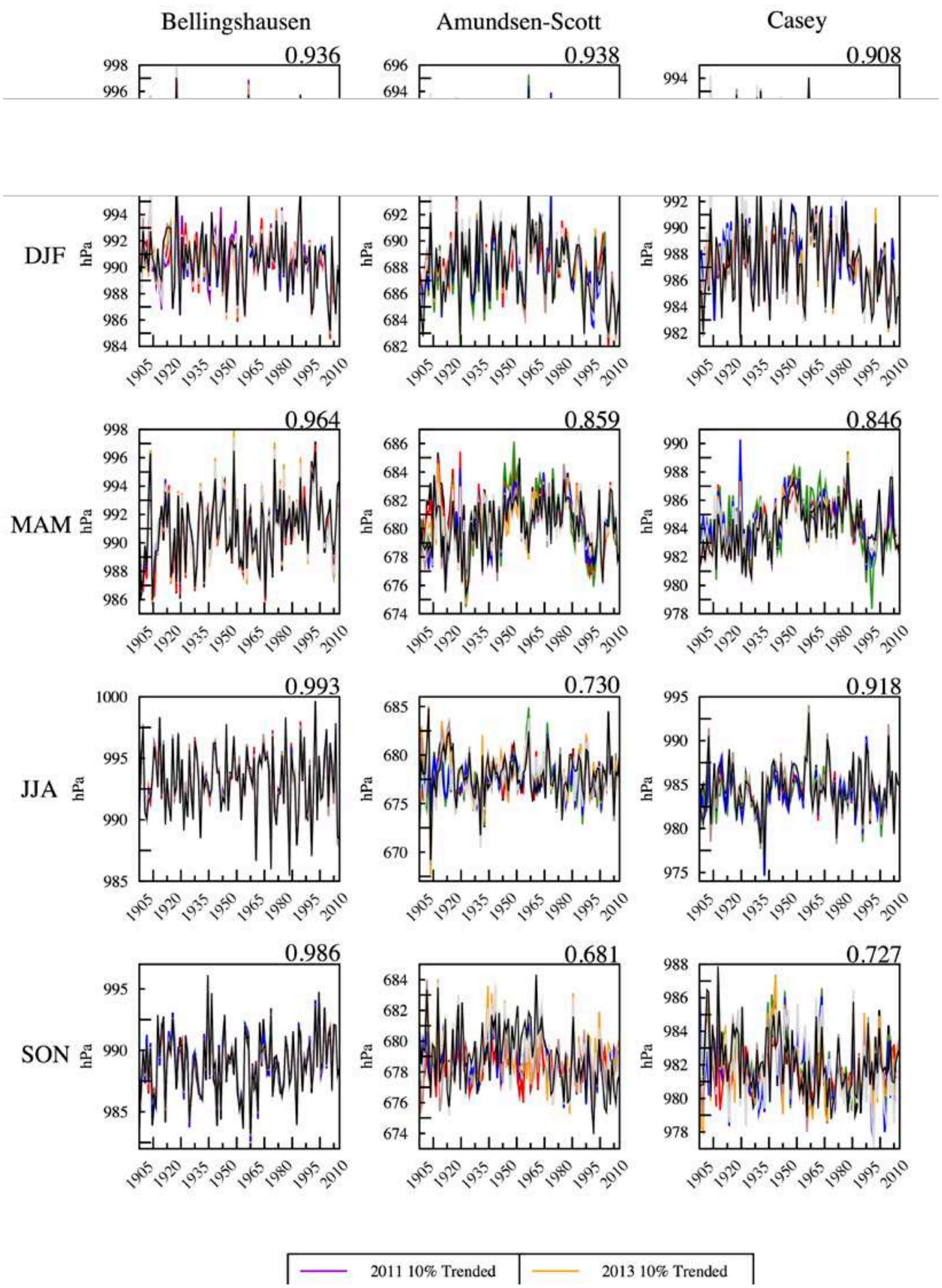

Figure 3. Time series plots of all 8 reconstruction trials (as indicated in legend) for three select stations: Bellingshausen (left column, representing the Antarctic Peninsula); Amundsen-Scott (middle column, representing the Antarctic Plateau); Casey (right column, representing coastal East Antarctica). The number in the upper right of each panel is the mean cross-correlation among the 8 different reconstructions.

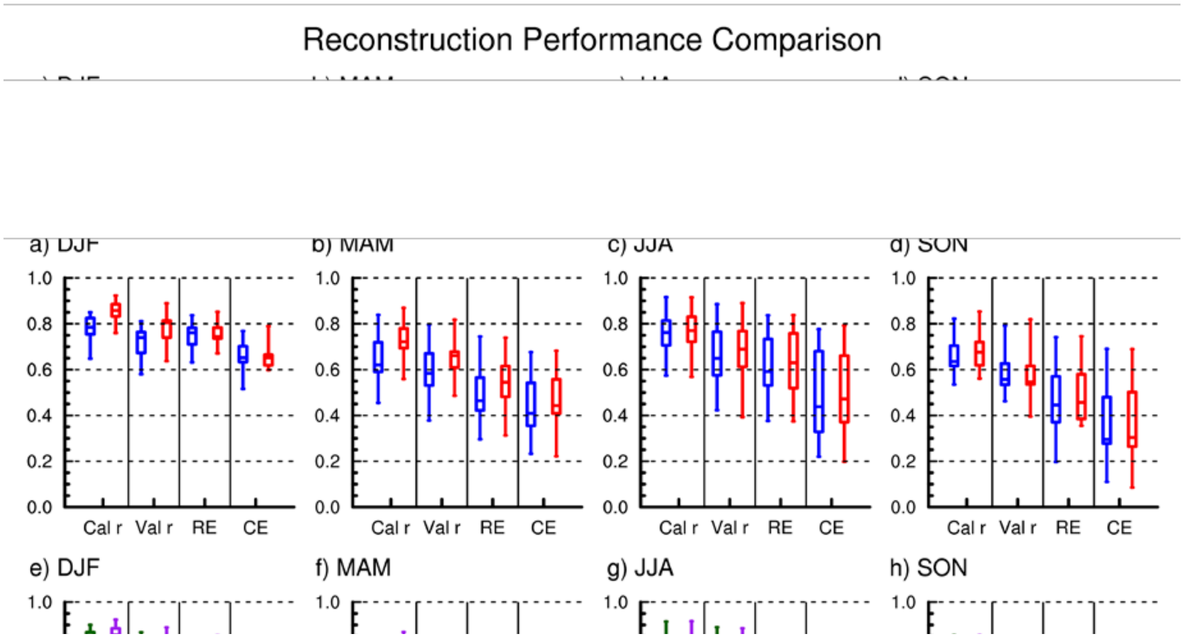

Figure 4. As in Fig. 2, but for boxplots highlighting skill metric comparisons between reconstructions based on detrended / original data (top row), and from the 5% and 10% networks (bottom row).

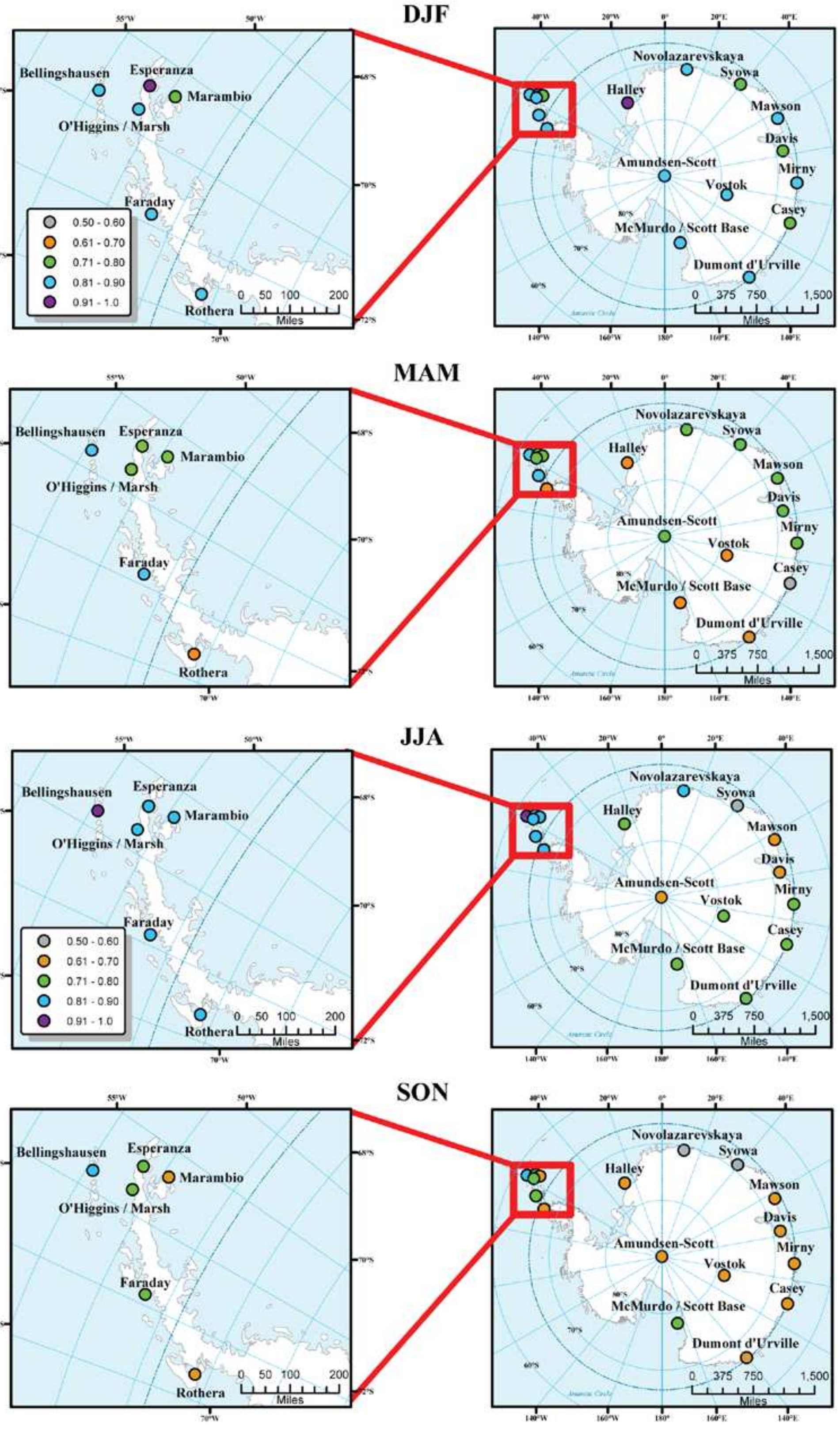

Figure 5. Spatial plot of calibration correlations for the best full period reconstructions.

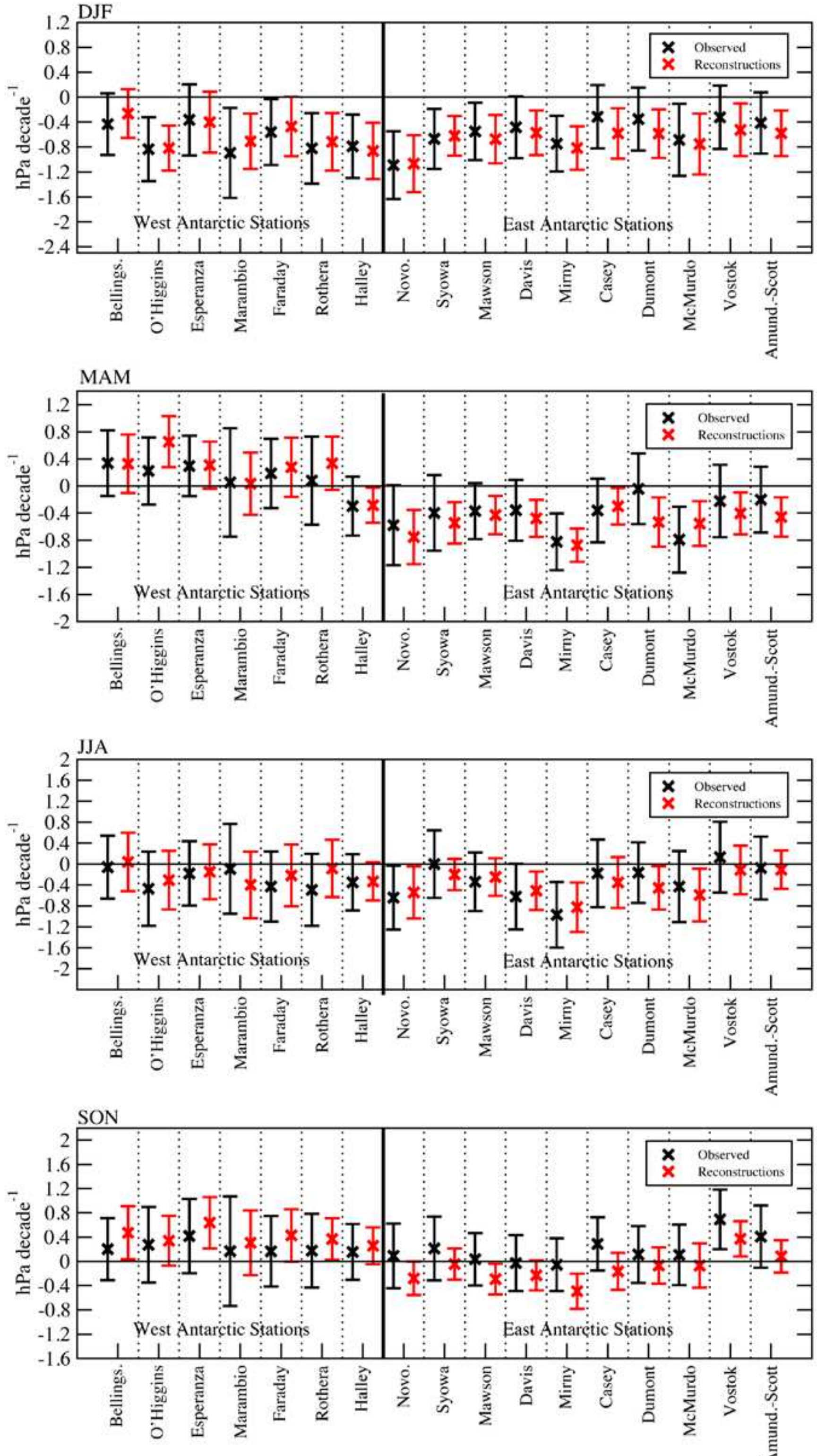

Figure 6. Trends and 95% confidence intervals for the observations (black) and best full period reconstructions (red) during 1957-2013, from the start of the observations to the end of the reconstruction (either 2011 or 2013). Some station names have been shortened.

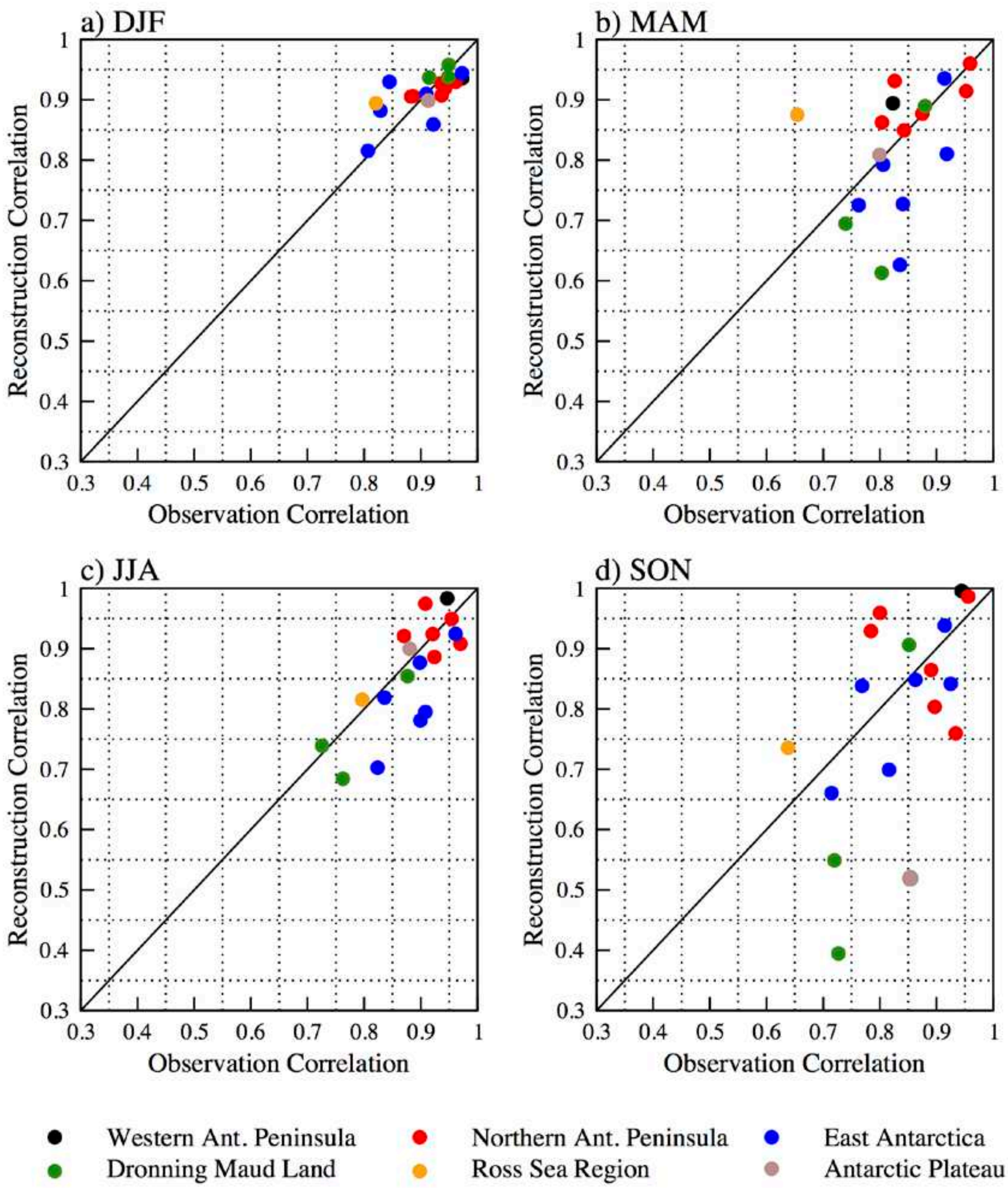

Figure 7. Scatterplot of cross-correlations between various station pairs during 1957-2013, grouped geographically by colors (see text for details). The x-axis denotes the correlations between the station pairs in the observed data, while the y-axis shows the correlations in the reconstructions.

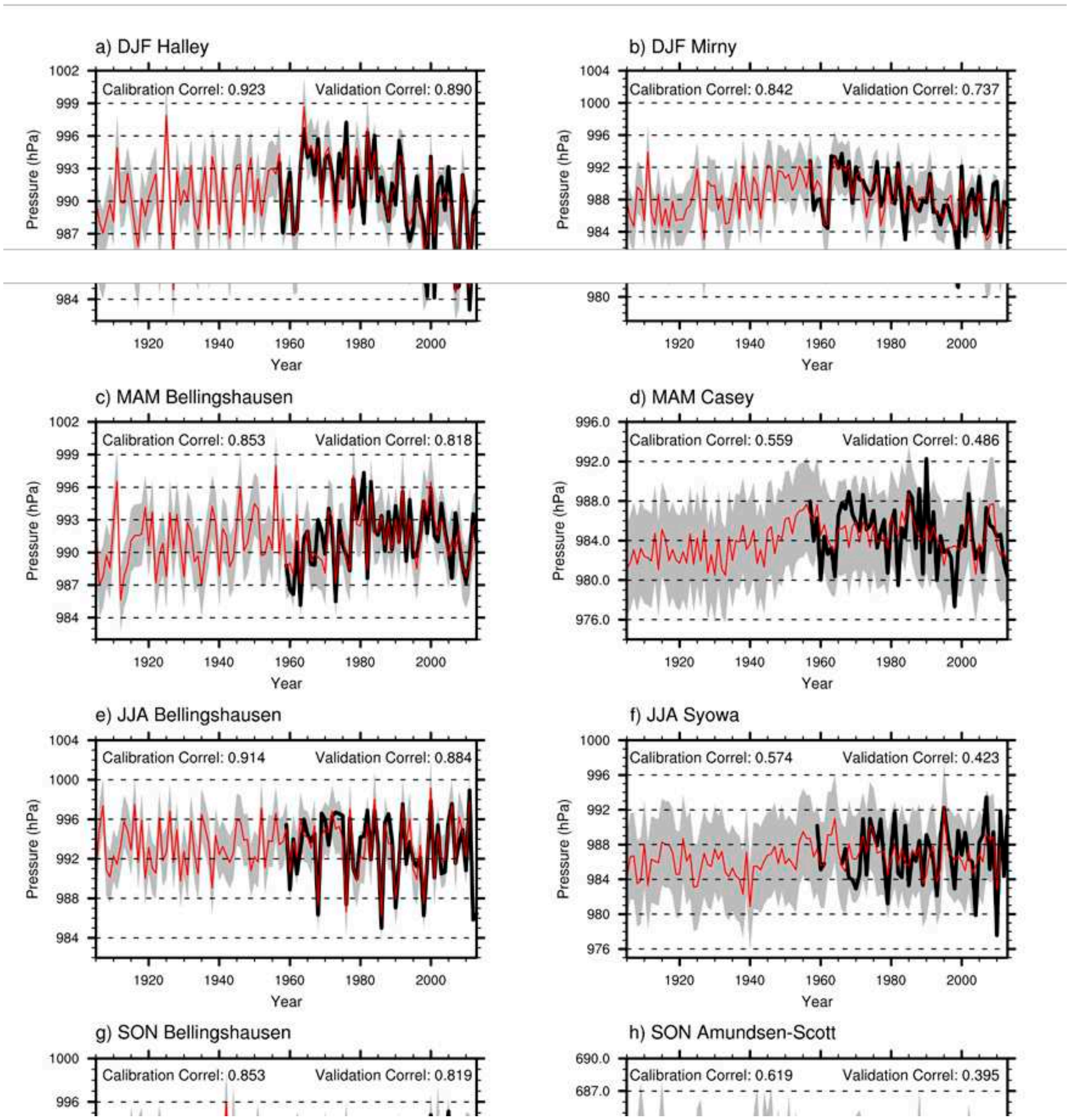

Figure 8. Time series of the reconstructions with the highest (left column) and lowest (right column) skill by season. Also given is the calibration and validation correlation for each reconstruction. The gray shading represents the 95% confidence intervals, taken as 1.96 times the standard deviation of the reconstruction – observation residuals.

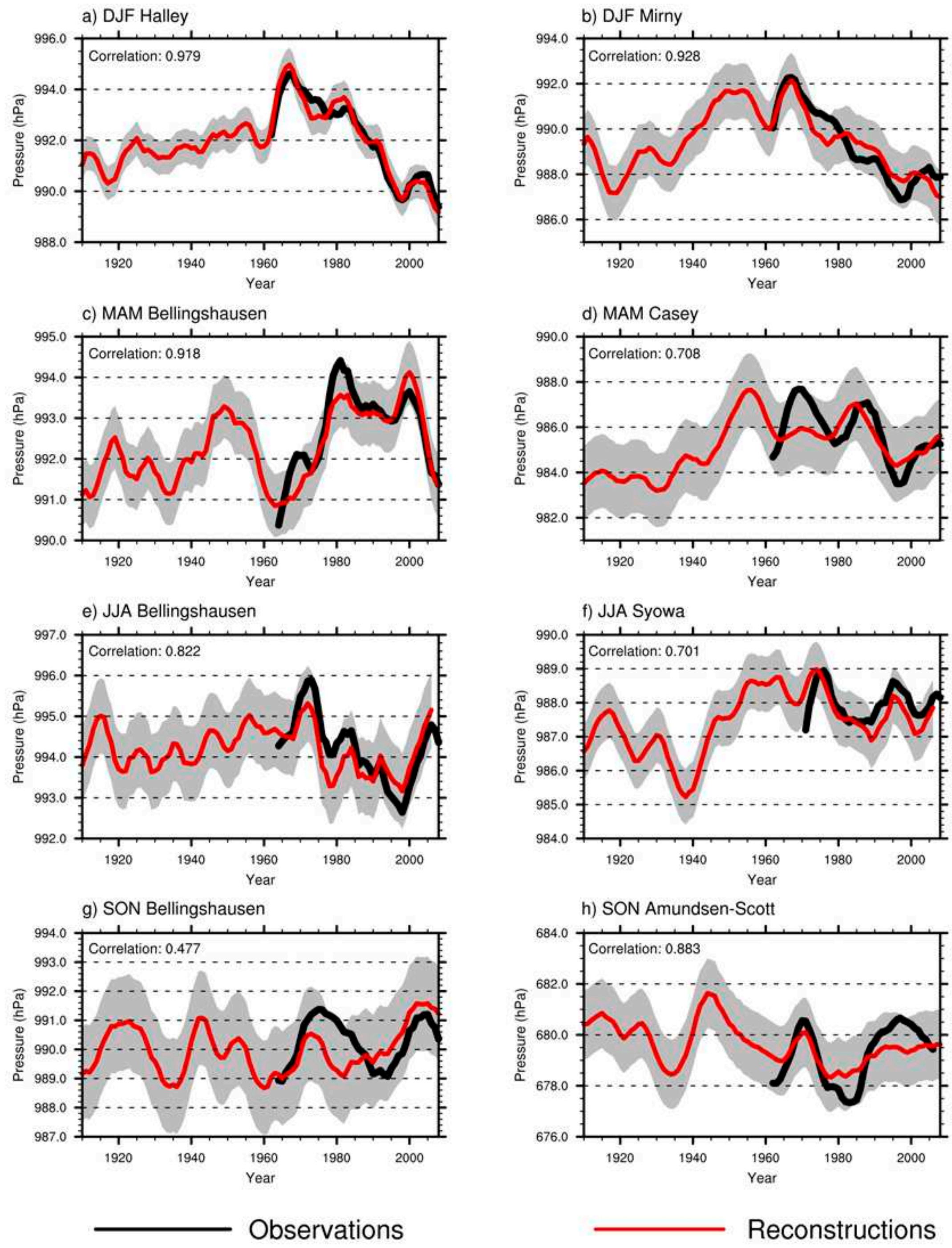

Figure 9. As in Fig. 8, but for the observations and reconstructions smoothed with an 11-yr Hamming filter. The correlation between the smoothed observations and reconstructions is given for each panel.

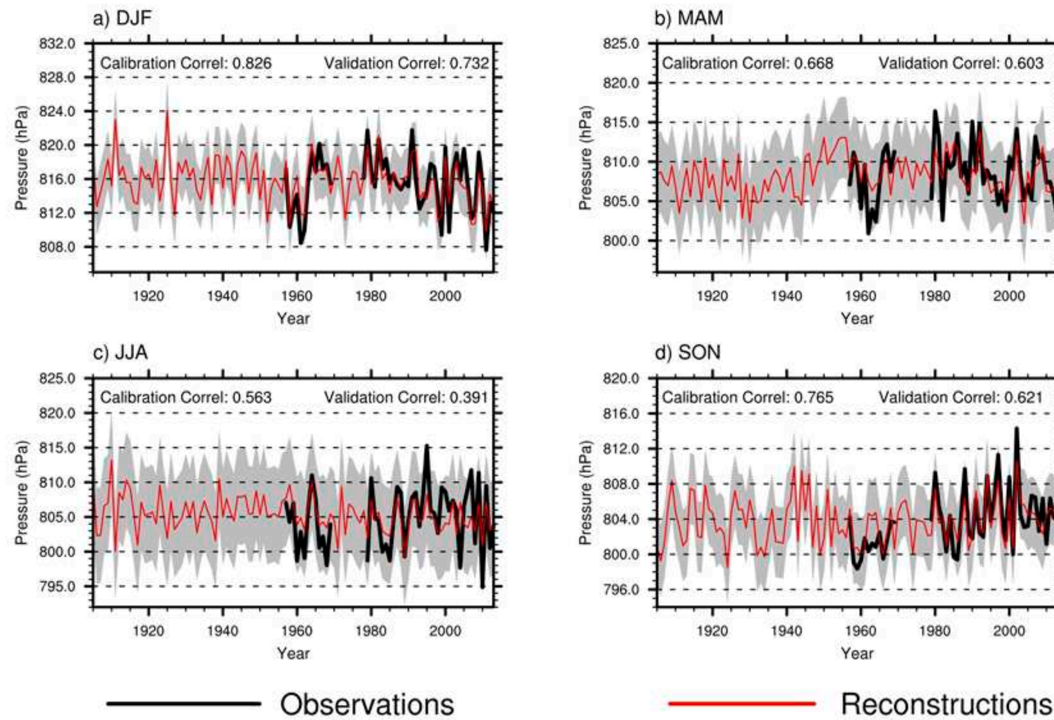

Figure 10. Time series for the Byrd station reconstructions in West Antarctica.

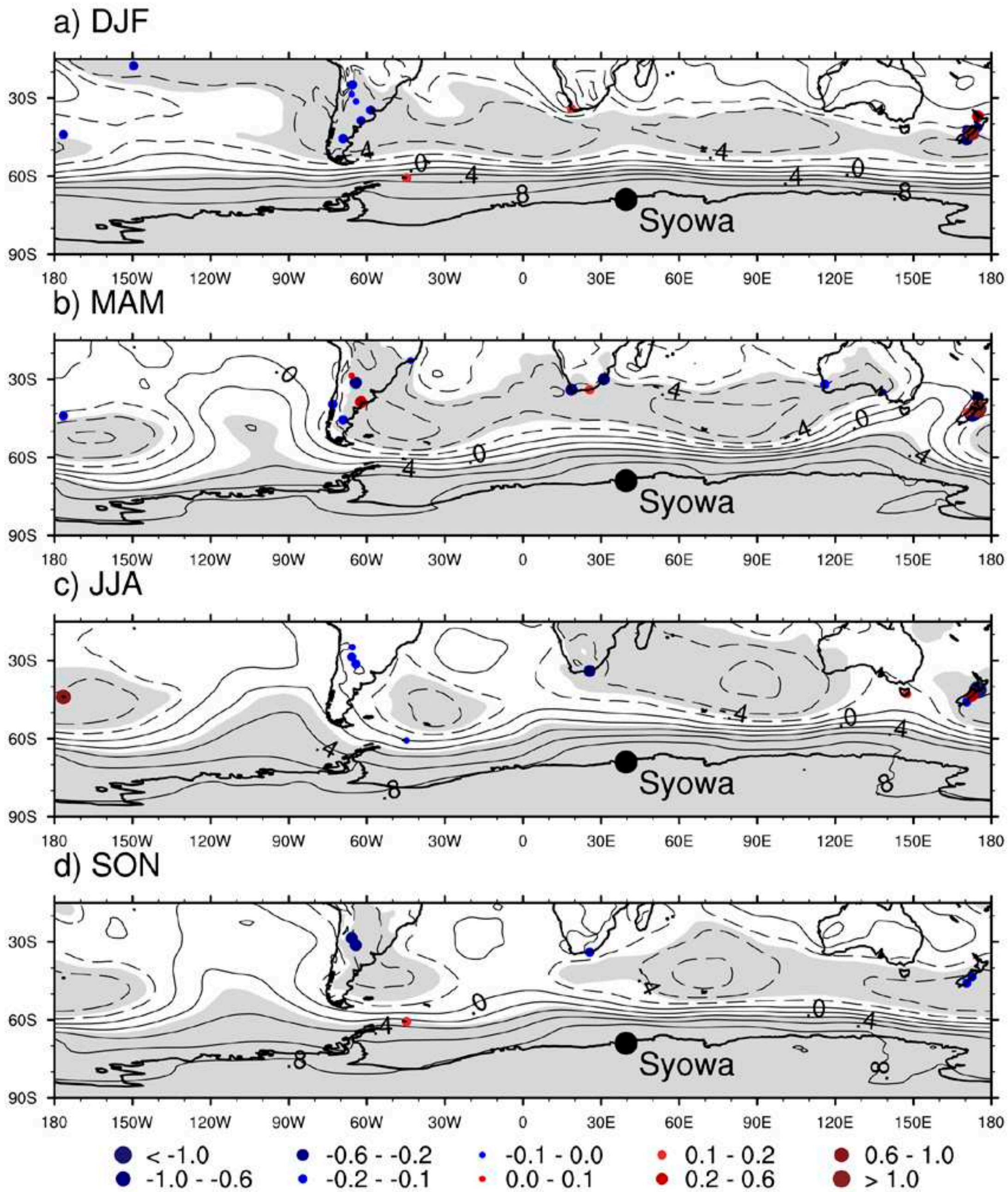

Figure 11. Map of MSLP correlations (contoured, significant correlations at $p<0.10$ shaded in gray) of the ERA-Int gridpoint closest to Syowa station (indicated with large black circle) and every other gridpoint from ERA-Int south of 15°S during 1979-2013. Also shown are the predictor stations used in the full period Syowa reconstruction, with the weight each midlatitude predictor station had in the final reconstruction indicated by the color / size of the circle, as indicated below the figure.

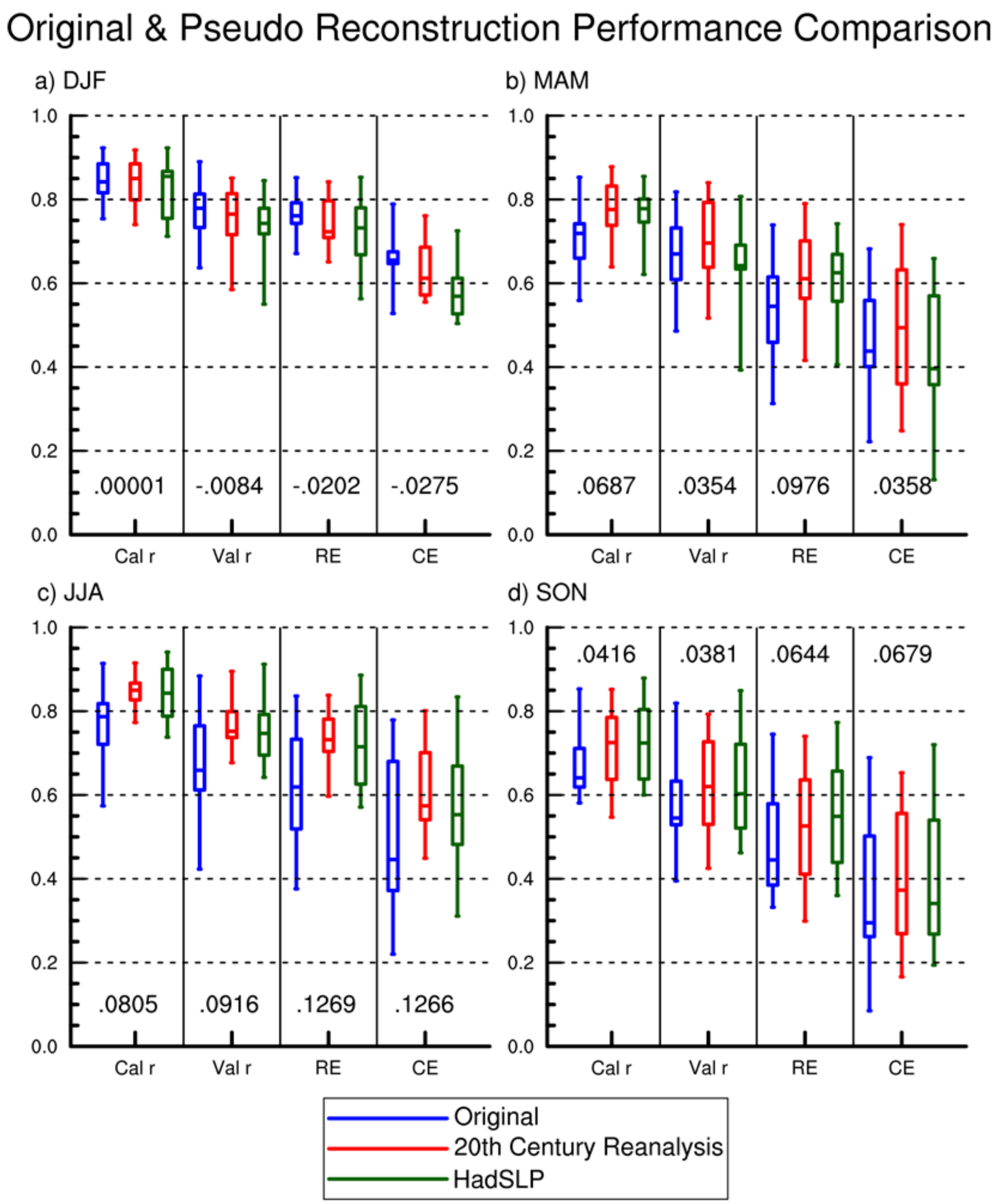

Figure 12. Boxplots of reconstruction statistics across the main 17 Antarctic stations for the best full period reconstruction ('Original') and the full period, 10% trended reconstructions from 20CR (red) and HadSLP2 (green). The numbers in each panel represent the mean difference between the best original and best pseudo-reconstruction (between 20CR and HadSLP2), with positive values indicating improvements in the pseudo-reconstructions.

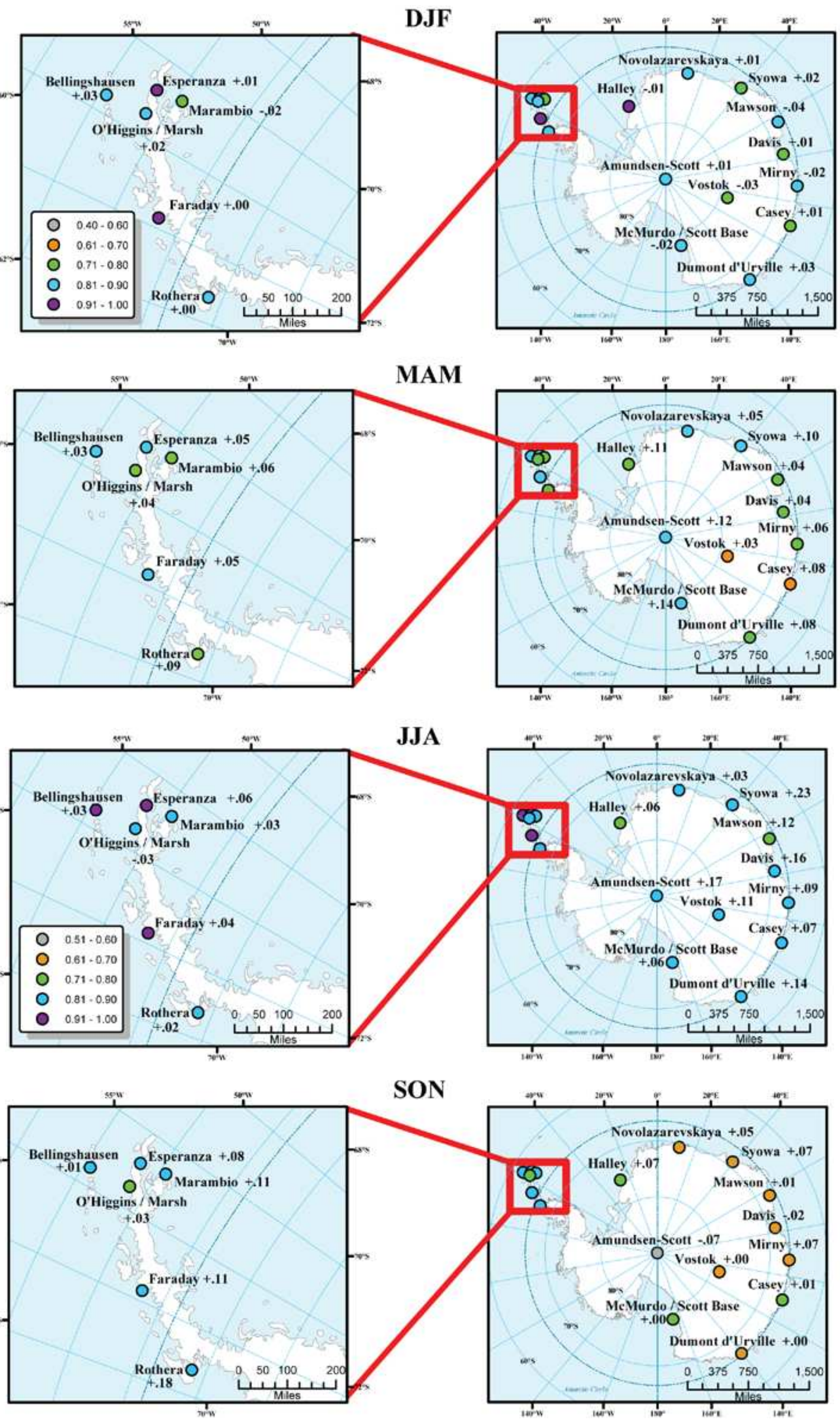

Figure 13. Maps of the best full period pseudo-reconstruction calibration correlation by season, as in Fig. 5. The number by each station is the difference between the original and pseudo-reconstruction, with positive numbers indicating improvement in the pseudo-reconstructions.