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

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.

69   Other tools to examine historical Antarctic climate variability during the last century

70   include gridded datasets that span the 20th century, such as existing historical reanalyses that only

71   assimilate surface pressure, like the National Oceanic and Atmospheric

72   Administration/Cooperative Institute for Research in Environmental Sciences (NOAA-CIRES)

73   Twentieth Century Reanalysis version 2c (hereafter 20CRv2c, Compo *et al.*, 2011) and the new

74   NOAA-CIRES-DOE version 3 (20CRv3, Slivinski *et al.*, 2019). Similar "sparse-input"

75   reanalyses have been completed by the European Centre for Medium Range Weather Forecasts

76   (ECMWF), including their 20th-century reanalysis (ERA-20C, Poli *et al.*, 2016) and a coupled

77   ocean-atmosphere reanalysis of the 20th century (CERA-20C, Laloyaux *et al.*, 2018), each of

78   which assimilates marine wind observations and surface pressure. As a coupled reanalysis,

79   CERA-20C also assimilates subsurface ocean temperature and salinity observations (Laloyaux *et*

80   *al.*, 2018). While these are spatially and temporally complete throughout the 20th century,

81   unsurprisingly they are sensitive to the number of assimilated surface pressure observations, and

82   therefore the skill changes considerably over the high southern latitudes throughout the 20th

83   century (Schneider and Fogt, 2018; Fogt *et al.*, 2019; Slivinski *et al.*, 2019).

84   An alternative approach to address these challenges is to generate seasonal pressure

85   reconstructions, both at individual Antarctic stations (Fogt *et al.*, 2016a, 2016b), as well as

86   spatially complete poleward of 60°S (Fogt *et al.*, 2017a, 2019). When compared to gridded

87   climate datasets, these reconstructions showed far less sensitivity to changes in the number of

88   observations across Antarctica (Schneider and Fogt, 2018), which is perhaps not surprising given

89   that the reconstructions were based on statistical relationships with a spatially and temporally

90   fixed network of stations in the mid and high latitudes of the Southern Hemisphere (Fogt *et al.*,

91   2016a). However, the skill of the reconstructions was only thoroughly evaluated after the IGY,

92   (1957-1958), when most Antarctic observations began; spatially complete comparisons in

93   reconstruction skill were even more limited and only possible after 1979 through evaluating

94   against the ECMWF Interim reanalysis, ERA-Interim (ERA-Int; Dee *et al.*, 2011). Therefore,

95   the skill of these reconstructions in the early 20th century before the IGY remains unknown,

96   although the underlying relationships governing them are likely stationary (Clark and Fogt,

97   2019). This work aims to provide a more complete evaluation of these reconstructions poleward

98   of 60°S by using available pressure observations that are independent from the reconstructions in

99   the early 20th century.

100

101   **2. Data and Methods**

102   We make use of the best performing seasonal spatially-complete Antarctic pressure

103   reconstructions, which are available from 1905-2013 and based on surface pressure anomalies

104   from the 1981-2010 climatology from ERA-Int (Fogt *et al.*, 2019). The reconstructions are

105   based on a kriging interpolation of 18 Antarctic station reconstructions and observations at

106   Orcadas (locations plotted in bottom right panel of Fig. 1) to an 80km x 80 km Cartesian grid

107   centered over the South Pole. For comparison here, this grid has been converted to a 0.75°x0.75°

108   latitude-longitude grid. Importantly, comparisons with ERA-Int after 1979 (Fogt *et al.*, 2017a,

109   2019) demonstrated that the reconstruction skill varies seasonally, with the highest skill (defined

110   as the best agreement with ERA-Int) in austral summer (December – February, DJF), and the

111   lowest skill in the austral autumn and spring (March-May, MAM, and September-November,

112   SON, respectively). In general, higher reconstruction skill is found over the Antarctic continent

113   (especially near the Antarctic Peninsula), and lower skill at the northern edge of the domain near

114   60°S, especially in the South Pacific (Fogt *et al.*, 2019).

115   Historical pressure data prior to 1957 poleward of 60°S were extracted from three

116   primary datasets to further evaluate the reconstruction skill in the early 20th century. The largest

117   source of pressure data comes from the International Surface Pressure Databank version 3.2.9

118   (ISPD; Cram *et al.*, 2015), which contains subdaily measurements of mean sea level and/or

119   surface pressure from both ships and temporary Antarctic bases (during field expeditions) and

120   early Antarctic stations that began prior to the IGY. Additional sea level and/or surface pressure

121   data were obtained from the International Comprehensive Ocean-Atmosphere Data Set release 3

122   (ICOADS; Freeman *et al.*, 2017); we only make use of ship records in ICOADS that were not

123   available in ISPD. As many new historical observations from both ships and early Antarctic

124   expeditions are still being recovered and digitized, we also make use of a few newly digitized

125   additional pressure observations stemming from the Atmospheric Circulation Reconstructions

126   over the Earth (ACRE) initiative (Allan *et al.*, 2011), obtained directly from Rob Allan, the lead

127   of the ACRE project. Only those ACRE pressure observations that were not part of the current

128   releases of ISPD and ICOADS are utilized here, particularly from Antarctic expeditions in the

129   first few decades of the 20th century. Most historical observations reported both SLP and surface

130   pressure; however, a few observations only reported SLP, and therefore the reconstruction is

131   compared to historical SLP observations throughout. We directly use the historical observations

132   as reported in ISPD, ICOADS, or in the ACRE data for SLP and surface pressure, and do not

133   reduce any surface pressure observations to SLP.

134   In making comparisons, it is important to consider that historical observations can have

135   their own error or bias for many reasons, potentially including incorrect reported location

136   (latitude, longitude, or elevation); problems with the barometer; errors associated with

137   adjustments to sea level pressure; or misspecified corrections to pressure measurements from

138   temperature or gravity. Because measurement errors likely vary in time and space, it is difficult

139   to fully quantify how much these errors contribute to the reported pressure observations (Kent

140   and Berry, 2005). Reanalyses generally define instantaneous, random observation errors that

141   depend on platform and/or time (Compo *et al.*, 2011; Poli *et al.*, 2016; Laloyaux *et al.*, 2018;

142   Slivinski *et al.*, 2019), and analyses comparing these expected errors with background statistics

143   suggest estimates of about 1.5-2 hPa in the early 20th century are reasonable (Slivinski *et al.*,

144   2019). However, Slivinski *et al.* (2019) and more recent comparisons (not shown) also suggest

145   that there is a relatively large systematic error in observations in the Southern Hemisphere in the

146   early 20th century. We therefore assume a conservative estimate of 2 hPa total error in the

147   seasonally-average observation estimates (or an observation error variance of 4 hPa^2).

148   In order to compare the historical observations to the reconstructions, additional analysis

149   was required. First, all historical pressure data were converted to hPa, the same units as the

150   pressure reconstructions. Monthly means for all the historical observations were calculated if at

151   least 75% of the days within a given month had at least one observation available, to ensure a

152   reliable monthly pressure estimate. Seasonal means (following the traditional seasonal divisions,

153   DJF, MAM, JJA (June – August), and SON) were calculated to compare to the seasonal pressure

154   reconstructions if at least two monthly means existed using our 75% daily data threshold; data

155   for the third month was added to calculate the seasonal mean if 2 months had met the 75% daily

156   data threshold, and the third month had more than 50% of its daily data. This approach allows

157   for comparison with the maximum amount of data possible. Comparisons of a single monthly

158   mean historical observation (for cases when historical data only had one month meeting our

159   threshold) with the seasonal mean pressure reconstruction produced larger differences (not

160   shown), and so this paper only focuses on comparisons of at least two months of historical data

161   to define a seasonal mean. Altogether, these constraints yielded 271 seasonal means from

162   historical pressure observations prior to 1957. The majority (>75%) of these historical

163   observations occur only in DJF (austral summer), although it is noted that there are several

164   observations after 1940 complete for the entire year. For reference, the year of austral summer

165   refers to the year of December throughout this study.

166   Two more steps were taken prior to making comparisons. First, seasonal mean locations

167   for the historical records were calculated using the coordinates provided in the data archives.

168   These locations varied little for the early Antarctic stations; for observations collected on moving

169   ships, we calculated a seasonal mean location by year for comparison since the reconstruction

170   data are only available seasonally, which limits comparisons at higher temporal and spatial

171   frequency. Of the 130 seasonal means from the ship records, more than 75% of the subdaily

172   pressure observations showed less than a 2 degree standard deviation in the latitude, and half of

173   the ships' standard deviations were less than 15 degrees longitude. In terms of pressure, there

174   was no significant difference between ships that had a standard deviation of more and less than

175   20 degrees longitude in the standard deviations of sub-daily pressure observations. We therefore

176   suggest the error associated with the mean location is more strongly related to the number of

177   strong storms a ship encountered, which are more unique to a specific location rather than the

178   mean location of the ship. As a rough estimate of this error, the mean standard deviation of

179   pressure from the subdaily ship observations, 3.45 hPa, can be used, although more than half of

180   the ships have a pressure standard deviation below 3.0 hPa. This standard deviation of subdaily

181   pressure within one season is much smaller than the interannual standard deviation of monthly

182   pressure across the South Pacific, which ranges from 3-7 hPa based on ERA5 data (contours in

183   Fig. 1, bottom right panel).

184   Lastly, to compare the pressure reconstructions to the historical pressure observations,

185 elevation adjustments were needed. The pressure reconstructions, originally constructed as

186 surface pressure anomalies based on the underlying topography of ERA-Int at 0.75°x0.75°

187 latitude / longitude resolution, were adjusted to sea level pressure using a rough linear estimate

188 of 12 hPa pressure decrease per 100 m elevation gain. The 12 hPa approximation is larger than

189 the 10 hPa per 100m assumed in the middle latitudes in the lower troposphere given the colder

190 and drier (and therefore denser) air commonly observed poleward of 60°S. Without simultaneous

191 measurements of temperature and humidity, it is challenging to provide a more accurate

192 adjustment to sea level pressure. Furthermore, reduction to sea level pressure is known to be

193 problematic in nearly all gridded climate datasets (and observations themselves) over Antarctica,

194 including 20CRv3 (Slivinski *et al.*, 2019). This adjustment to sea level was necessary even for

195 historical observations that had both surface and sea level pressure data, as due to the smoothing

196 of the topography of ERA-Int at even the 0.75° resolution, model elevations and observed

197 elevations (where known – this value is not always given in the historical data) are often quite

198 different; these elevation differences similarly make it challenging to adequately compare

199 surface pressure observations with the surface pressure (anomaly) reconstructions. Importantly,

200 elevation differences can also arise due to the smoothing of the steep Antarctic coastline in ERA-

201 Int, and therefore influence comparisons from ships close to the Antarctic continent. This

202 adjustment to sea level introduces the greatest error and uncertainty in our comparison, as will be

203 discussed in detail throughout. Nonetheless, the simple linear adjustment allows us to evaluate

204 interannual (or intra-annual) variability and provide further information on reconstruction

205 performance than otherwise would be possible.

206   In the comparisons, the reconstruction data are extracted from the gridpoint closest to the

207   seasonal mean location in the reconstruction field. To provide further comparison for select

208   observations that span more than one season, from the reanalyses we also extract both surface

209   and sea level pressure data from the closest gridpoint using the ensemble mean in all but ERA-

210   20C (which has only one estimate). This approach also introduces some error, called the 'error

211   of representativeness', reflecting how well a reanalysis gridpoint represents the point location

212   from the observation; this error is assumed to be one of the largest components of observation

213   errors in the reanalyses, at least in 20CRv3 (Slivinski *et al.*, 2019), and is similar to the

214   elevation-based errors for the reconstruction evaluations described previously. Importantly,

215   these reanalyses are not independent of the observations used for comparison here: ERA-20C

216   and CERA-20C assimilate surface pressure and marine surface wind observations from ISPD

217   version 3.2.6 and ICOADS version 2.5.1 (Poli *et al.*, 2016; Laloyaux *et al.*, 2018); 20CRv2c

218   assimilates pressure from ISPD version 3.2.9 and ICOADS version 2.5.P; and 20CRv3

219   assimilates pressure from ISPD version 4.7 and ICOADS version 3+v2 (Slivinski *et al.*, 2019).

220   Since the comparisons are made with ISPD v3.2.9 and ICOADS v3, the observations were at

221   least available to be assimilated in both version of 20CR, and likely for ERA-20C and CERA-

222   20C. However, the assimilation combines the instantaneous observations with a background

223   guess from the forecast model, so fields from different reanalyses are expected to differ from

224   each other (as they each use different forecast models and assimilation algorithms), as well as

225   from the seasonally-averaged observation estimates analyzed here. In addition, each reanalysis

226   system has its own quality control algorithm that will blacklist observations deemed unfit for

227   assimilation; thus, even if a given set of observations are available to the assimilation algorithm,

228   they may not all have been assimilated. Finally, the reanalyses each employ their own

229 observation bias correction scheme, which removes significant, consistent differences between

230 the observations and the background fields. For simplicity, we only consider the uncorrected

231 observations here, but note that systematic differences between the reanalyses and observations

232 in our comparison may have been ameliorated by taking into account the observation bias

233 corrections calculated within each reanalysis.

234   Reconstruction performance compared to the historical observations is evaluated using

235 three primary statistics: overall bias, defined as the mean difference between the seasonal

236 historical observation estimates and reconstructions at sea level (reconstruction minus

237 observations); the mean absolute error (MAE), defined as the mean absolute difference to

238 remove offsetting effects of positive and negative difference; and the root mean square

239 difference (RMSD), which due to the squaring of the differences prior to calculating the mean,

240 gives slightly higher weighting to larger absolute differences than the MAE. For observations

241 that span only one season or have the bias of the same sign across all seasons, the absolute value

242 of the bias and MAE will yield identical results. In all cases, these three statistics were

243 calculated over the full length of each observational record. Further averaging both spatially and

244 temporally is conducted to assess the overall reconstruction performance. Since the

245 reconstruction error determined by Fogt et al. (2019) over 1979-2013 is assumed to be constant

246 in time, it is assumed that there are no (temporally) correlated errors in the reconstruction; the

247 evaluations in this paper help to determine the persistence of reconstruction errors through the

248 early 20th century.

249

250   **3. Results**

251 *3.1. Historical data availability*

252   Prior to estimating the seasonal pressure reconstruction skill in the early 20th century, it is

253   important to recognize the large changes in the number of observations poleward of 60°S,

254   especially before 1957. Figure 1 displays the location and data type for all of the 271 seasonal

255   means that were compared to the reconstruction; open circles represent seasonal mean positions

256   of ship data while filled circles denote base or station records that were stationary or only had

257   minor movements over time. As expected, there are very few observations prior to 1930,

258   although there is a relative increase in the 1910-1919 during the height of the 'Age of Heroes'

259   which includes expeditions to the South Pole, the Australian Antarctic Expeditions, and the

260   British Imperial Trans-Antarctic Expedition led by Ernest Shackleton. The decrease in the

261   number of observations in the 1940s is not only a challenge in the high southern latitudes but

262   also worldwide, associated with the Second World War. Nearly half of the historical

263   observations used for comparison here come from the years just prior to the IGY (1950-1957),

264   with several early Antarctic stations established over the continent (especially along the

265   Antarctic Peninsula), and frequent ships along the East Antarctic coastline. A notable gap in

266   observational coverage is within the Weddell Sea (east of the Antarctic Peninsula) and

267   throughout much of the South Atlantic. In addition to persistent sea ice in the Weddell Sea

268   (Cavalieri and Parkinson, 2008), the South Atlantic area is far from commercial sailing or

269   whaling routes and is likely one of the greatest spatial data voids globally (for example, see Fig.

270   1 of Allan and Ansell, 2006).

271

272

273   *3.2. Overall reconstruction performance*

274   With this temporal evolution of historical data in mind, the reconstruction performance is

275   displayed as a bar chart in Fig. 2; the overall statistics (averaged over all available observations)

276   are given at the bottom: mean bias = -0.79 hPa; mean MAE = 4.05 hPa; mean RMSD = 4.26 hPa.

277   Recalling that the overall comparison is primarily during austral summer, these values are

278   considerably higher than the skill assessed in earlier work in comparison to ERA-Int (Fogt *et al.*,

279   2017a, 2019), which estimated an MAE over the Antarctic continent generally less than 1.5 hPa

280   and over the Southern Ocean around 2-3 hPa. Indeed, the values seem even higher than the

281   MAE in Fogt *et al.* (2019) in the non-summer months in the South Pacific, the region of highest

282   MAE compared to ERA-Int which ranged from 2-4 hPa.

283   This quick comparison initially suggests that the seasonal pressure construction

284   performance is of much lower quality prior to 1957 than it is after 1957. However, when

285   considering that the observation estimates have their own errors (assumed to be approximately

286   ~2 hPa in the early 20th century), the higher MAE prior to 1957 is not far from the assumed

287   observational error. Further, when looking at the average skill as a function of latitude (Fig. 2),

288   the bias is near zero in both of the latitude bands that primarily lie over the Southern Ocean (60°-

289   65°S and 60°-70°S), potentially suggesting that the historical observation estimates tend to

290   fluctuate near the reconstruction seasonal mean overall (both above and below, cancelling to near

291   zero). The MAE in these latitude bands, near 3 hPa, better reflects the overall performance as it

292   removes the cancellation between positive and negative differences. The MAE of 3 hPa is

293   consistent with the skill reflected in Fogt *et al.* (2019), and is also close to the assumed

294   observation error of 2 hPa. In contrast, where Fogt *et al.* (2019) demonstrate lower MAE over

295   the Antarctic continent (due to more station observations constraining the reconstruction skill),

296   comparisons with the historical observation estimates clearly show an increase in MAE poleward

297   of 65°S, which is primarily from stations over the Antarctic continent (poleward of 70°S). As

298   noted earlier, this increase in MAE is an artifact of the elevation sensitivity when comparing the

299   historical observations at (often unknown) elevations different from the seasonal pressure

300   reconstruction based on the underlying topography of the ERA-Int 0.75°x0.75° grid. The simple

301   linear offset of 12 hPa per 100 m creates sea level pressures that are consistently too low

302   (evidenced by the negative bias), which gives rise to large MAE and RMSD over the Antarctic

303   continent.

304   The sensitivity to elevation is also apparent in the comparisons as a function of longitude.

305   There is a strong negative bias in the 330°-30°E range, along the mountainous Antarctic

306   Peninsula. In ERA-Int, this terrain was greatly smoothed but still elevated, while the

307   observations are often along the coast near sea level. In other longitude zones, the mean bias is

308   typically in the range of ±1 hPa, with MAE around 3-4 hPa. Given that Fig. 1 indicates that

309   outside of 330°-30°E, the seasonal mean observation estimates primarily are over the ocean and

310   less influenced by elevation corrections, the comparison by longitude again reflects a similar

311   skill of the reconstruction (over the Southern Ocean at least) as observed in Fogt *et al.* (2019)

312   during 1979-2013. Comparisons by decade typically have MAE also in the 3-4 hPa range,

313   consistent with the previous evaluation. One outlier is the 1910s, which from Fig. 1 has a

314   relatively high percentage of historical observations on or near the Antarctic continent,

315   suggesting that elevation corrections are again giving a misleading impression of relatively low

316   reconstruction skill.

317   To further visualize the spatial and elevation dependence of reconstruction performance,

318   Fig. 3 displays the decadal mean RMSD for each station by decade. Several key points emerge

319   from Fig. 3. First, there is an indication from ships in the South Pacific north of the Amundsen

320   and Bellingshausen Seas (roughly 60°-70°S, 150°-90°W) that the reconstruction skill is

321   considerably lower, with RMSD values often above 6 hPa. Since these are seasonal pressure

322   estimates from ships (Fig. 1), elevation correction is not an issue. Rather, the larger errors are

323   consistent with the much lower reconstruction performance in this region of high interannual

324   variability (which often exceed observational errors based on interannual monthly pressure

325   standard deviations as large as 7 hPa in Fig. 1), as discussed in previous evaluations (Fogt *et al.*,

326   2017a, 2019). The larger interannual pressure variability here may also suggest that using the

327   mean ship location could be introducing more error in this sector compared to other regions

328   around Antarctica (Fig. 1). Second, for comparisons in the Southern Ocean for nearly all other

329   regions, the RMSD is generally below 4 hPa, and the majority of the comparisons with the

330   seasonal mean ship observation estimates show RMSD in the 3-4 hPa range. These values are

331   consistent with the slightly lower performance of the reconstruction over the Southern Ocean.

332   However, these values of RMSD still fall within expected error when considering errors in both

333   observation estimates and the reconstruction (taken as the square root of the sum of the

334   observational and reconstruction error variances). Lastly, it is clear that many of the higher

335   errors (exceeding 6 hPa) are found along the coast, the Antarctic Peninsula, or in the interior of

336   the continent. Even on the Ross Ice Shelf, RMSD values are typically larger than 4 hPa. As will

337   be more clearly demonstrated later, these larger differences are due to elevation corrections to

338   the historical observations in or near areas of high or vastly varying terrain. We will show that

339   the reconstruction skill is likely much higher in these areas than suggested by the RMSD values

340   in Fig. 3. Remarkably, in every decade since the 1930s, there are multiple locations where the

341   RMSD values (over the ocean) are less than 1hPa, which is an excellent agreement with the

342   historical observations. Given that we make comparisons with the seasonal mean ship position,

343   that the historical observation estimates are rarely complete for the full season, and that they

344   themselves have errors, this high agreement is noteworthy. It suggests that the reconstruction

345   can be used as a reliable approximation of Antarctic pressure in most regions on and around the

346   continent back to at least the 1930s.

347

348   *3.3 Reconstruction performance evaluated with selected historical observations*

349   The reconstruction performance is perhaps best evaluated by comparing to individual

350   historical observations, which is the focus for the remainder of the study. To facilitate these

351   assessments, a subset of representative stations and ship seasonal locations has been selected;

352   their average location, name, and mean RMSD values over the entire record length are plotted in

353   Fig. 4. We have purposely selected historical pressure estimates with longer records than a

354   single season, and as many complete records from earlier portions of the 20th century as possible,

355   although the sample size is quite limited (Fig. 1). In general, the reconstruction performance

356   indicated by the mean RMSD in Fig. 4 at these locations is comparable to the overall decadal

357   mean performance in Fig. 3, although we do not closely examine any stations with RMSD less

358   than 2 hPa, and only one station where inadequate elevation corrections challenge the assessment

359   of reconstruction skill (Cape Denison). Recall, all of these data were available for assimilation

360   into 20CRv2c and 20CRv3 (and likely available for CERA-20C and ERA-20C), so the

361   comparisons with the reanalyses are not necessarily always independent as discussed in section

362   2.

363   Figure 5 displays three seasonally-averaged historical observation estimates where the

364   mean RMSD values are 2-3 hPa, consistent with the reconstruction skill and uncertainty assessed

365   in Fogt *et al.* (2019), and within the assumed observational error variance of 4 hPa^2. Plotted with

366   the historical observations (in red) are the mean sea level pressure (solid) and surface pressure

367   (dashed, when available) from the nearest gridpoint of 20CRv2c (green), 20CRv3 (orange),

368   ERA-20C (purple), and CERA-20C (blue). To compare with the skill evaluation in Fogt *et al.*

369   (2019), the correlations (if more than 10 values are available) and MAE (in hPa) for each gridded

370   dataset are given at the bottom of each panel: the first number is based on the MSLP, and the

371   second number (where available) is based on surface pressure.

372   Station 889340 (from the ISPD archive) is situated along the Antarctic Peninsula (Fig. 4)

373   and has a continuous pressure record beginning in 1948 for all seasons. As such, it is one of

374   many such stations that provide a useful evaluation of reconstruction skill, and the overall MAE

375   from the reconstruction is less than 2 hPa. Moreover, the temporal variability is well captured (r

376   > 0.80 in all datasets). The various reanalysis products have similar MAE (generally from 1-2

377   hPa), well within the likely observational uncertainty, with CERA-20C having the lowest for

378   MSLP, and 20CRv2c having the lowest for surface pressure. Despite the low MAE / RMSD at

379   this station, adjustment of the reconstruction to sea level pressure may play a small role in its

380   performance, as there are differences in both the historical MSLP and surface pressure, as well as

381   the MSLP and surface pressure in the reanalyses. Nonetheless, this station shows the viability of

382   the reconstruction shortly before the IGY along the Antarctic Peninsula, a region of relatively

383   higher reconstruction skill compared to ERA-Int after 1979 (Fogt *et al.*, 2017a, 2019).

384   Deck 215 (Fig. 5b) from ICOADS has adequate observational coverage only during

385   austral summer. The reconstruction skill is comparable to the skill seen in the southern Indian

386   Ocean during much of the 1930s (Fig. 3). The reconstruction MAE is 2.5 hPa, slightly higher

387   than all reanalyses but ERA-20C, which was found to be one of the lower performing reanalyses

388   in the early 20th century near Antarctica (Schneider and Fogt, 2018). Despite the higher MAE

389   for the reconstruction (which, unlike the reanalysis products, is entirely independent from these

390   historical observations), the historical observation values nearly all fall within the reconstruction

391   uncertainty (95% confidence interval, gray shading), and the reconstruction uncertainty would

392   clearly overlap with the observational uncertainty throughout time. Together, the comparison

393   with Deck 215 data suggests the original assessment of the reconstruction skill provides a good

394   approximation of its performance in the early 20th century. Although the interannual variability

395   is not as strongly captured as in Fig. 5a (correlation is not provided since only six years of data

396   exist), it should be noted that the reanalysis datasets also do not reproduce the interannual

397   variability well (perhaps due to observational errors or the spatial averaging). We also observe

398   that the surface pressure in 20CRv2c is notably different than the surface pressure from the other

399   reanalyses, probably due to the spectral effects in the 20CRv2c elevation field noted by Slivinski

400   et al. (2019).

401   For a longer observational record over the Southern Ocean, Deck 899 (Fig. 4, from

402   ICOADS) also has pressure observations only for austral summer but over multiple decades (Fig.

403   5c). The location varied during these many voyages, but overall the observational estimate of

404   the seasonal mean falls within or very near the reconstruction uncertainty. Indeed, the

405   reconstruction agrees better with the seasonally-averaged ship-based values than the reanalyses

406   (MAE of 2.23 hPa compared to 2.3 – 3.7 hPa from reanalyses, although almost all datasets

407   would fall within the observational uncertainty at this location). The correlations are much lower

408   for these ship data across all datasets, due to differing fluctuations between a few years (i.e.,

409   1933-1934, 1936-1937, 1945-1946), however the interannual variability is better captured after

410   1946. In general, the reconstruction also aligns well with the MSLP from the reanalyses, except

411   for DJF 1934 when the ship traversed a large span of the southern Indian Ocean (from 88°E in

412 December 1934 to 47°E in February 1935). Therefore, the use of the seasonal mean location

413 could be artificially suggesting a lower reconstruction skill for this observational estimate.

414 Nonetheless, as in Figs. 5a and 5b, this comparison further suggests that the uncertainty of the

415 reconstruction is a good estimate of the reconstruction error in the early 20th century. Overall, the

416 reconstruction agrees well with many historical seasonal-mean observation values that were

417 withheld during the reconstruction's development. As in Fig. 5b, we note that there are larger

418 differences in the surface pressure from 20CRv2c which clearly fall outside the uncertainty from

419 using a seasonal mean ship location.

420   Figure 5 presents several comparisons where elevation corrections did not have a

421 noticeable influence on the evaluation of the reconstruction and are perhaps more representative

422 of the overall reconstruction quality. In many other locations, elevation adjustments appear to

423 play a more important role (especially on or near the Antarctic continent) in comparing the

424 reconstruction and observational estimates. At a few of these locations, a simple linear

425 adjustment of 12 hPa per 100m from the ERA-Int elevation at the reconstruction's gridpoint

426 proved sufficient to readily compare it to the historical values. A subset of these stations is

427 presented in Fig. 6.

428   Perhaps the most famous early American expeditions to Antarctica were those led by

429 Admiral Richard E. Byrd, who set up a station called Little America in several different field

430 campaigns spanning three decades (Byrd, 2003). Although the mean location of Little America

431 was on the Ross Ice Shelf (it varied negligibly during each campaign in comparison to the

432 resolution of the gridded datasets used in this study; Fig. 4), there were notable differences in

433 observed MSLP and surface pressure at the location (Fig. 6a). When adjusting the reconstruction

434 to sea level, a good match is observed both in overall mean pressure (MAE = 3.63 hPa) and

435   interannual variability. Surprisingly, the reconstruction performs much better than the MSLP

436   from 20CRv2c or 20CRv3, most notably due to the much higher MSLP in these two datasets

437   from the 1940s onward compared to the reconstruction and observed values. However, the

438   20CRv2c surface pressure nearly perfectly matches the observed surface pressure (an MAE of

439   only 2.39 hPa), while 20CRv3 has very similar values for surface and sea level pressure

440   observations. We note that although there are significant offsets from the observed values for

441   most of the datasets (MAE > 6 hPa for both), the interannual variability is well-captured, with

442   the reconstruction correlation of 0.85, and surface pressure correlations from the reanalyses all

443   above 0.85. The consistent difference between the observations and 20CRv3 could be due to

444   elevation issues or to a detected bias in the observations; this will be examined in more detail

445   later. CERA-20C, as seen earlier, performs with the highest skill for MSLP (r=0.9). Also note

446   that the offset from the surface pressure to MSLP is not constant, as this difference, through the

447   hypsometric equation, is influenced by both temperature and humidity. Therefore, some of the

448   larger errors in the reconstruction and the reanalyses, particularly in the 1950s, could be due to

449   too strong of an increase in the surface pressure as it was reduced to sea level, since the gap

450   between surface pressure and sea level pressure is much lower in the historical observations

451   during the 1950s. Observational uncertainties, including the various bias corrections employed

452   by the reanalyses as discussed previously, can also create some of these differences between

453   products when compared to the observations.

454   Much earlier in Antarctic history, under the leadership of Jean-Baptiste Charcot, the

455   French conducted their second expedition that wintered over on Peterman Island in 1908-1910.

456   The ship associated with this expedition was Porquoi Pas, with a mean location near the

457   Antarctic Peninsula (Fig. 4). Due to the close proximity of the high terrain of the Antarctic

458 Peninsula, elevation adjustments to the reconstruction were necessary but may not have been

459 sufficient at this location. From historical observations, the difference between surface pressure

460 and sea level pressure ranged from nearly 20 hPa to as little as 5 hPa in DJF 1909 (Fig. 6b).

461 Despite the challenges in correcting for the elevation differences, as seen in Fig. 6b, the

462 adjustment brings the observed MSLP to the far upper-limits of the reconstruction uncertainty,

463 and clearly within the observational uncertainty. Further, the reconstruction agrees well with the

464 observational estimates: the MAE of 3.15 hPa is nearly consistent with time, and the seasonal

465 cycle of the pressure during this portion of the early 20th century is well captured by the

466 reconstruction. It is also encouraging to see high skill in both 20CR datasets, and again in

467 CERA-20C at this location, all within the observational uncertainty; the lower performance of

468 ERA-20C could be related to a cold bias or perhaps resolution issue, as there is a much larger

469 difference between surface pressure (dashed line) and MSLP (solid line) in ERA-20C compared

470 to the observations or CERA-20C. The differences in surface pressure between 20CRv3 and the

471 observations are also noteworthy, with an MAE of 38.60, reflecting differences in model and

472 observed orography.

473   Similarly situated near the Antarctic Peninsula (Fig. 4), Argentine Island provides a long

474 continuous record much like in Fig. 5a, but with a noticeable influence of elevation-based

475 pressure corrections (Fig. 6c). As with Porquoi Pas, the linear SLP adjustment brings the

476 reconstruction close to the historical estimates, but they fall at the upper-bound of the

477 reconstruction uncertainty, and are much closer to the reanalyses overall. All comparisons are

478 within the expected observational uncertainty. The interannual variability is well-captured by

479 the reconstruction (r= 0.82), although it may be slightly dampened in the early 1950s.

480 Interestingly, ERA-20C surface pressure aligns closely with the historical observations (with an

481   MAE of 1.63 hPa), but this reanalysis again demonstrates the highest MAE compared to MSLP

482   (4.88 hPa). In contrast, CERA-20C shows the lowest MAE based on MSLP (0.66 hPa), but

483   considerably lower surface pressure than observed (an MAE of 18.29 hPa), though not as low as

484   surface pressure in 20CRv3 (an MAE of 33.42). All reanalyses show very high correlations with

485   observations for both MSLP and surface pressure ($r > 0.85$).

486   The long, albeit discontinuous record at the Little America station location (Fig. 4)

487   provides an opportunity to further evaluate the reconstruction in comparison to the historical

488   reanalyses, and importantly, to assess the reconstruction skill in light of the reanalyses' internally

489   estimated uncertainties. For the reanalyses studied here, all but ERA-20 are ensemble

490   reanalyses, and therefore the ensemble spread from the seasonal mean ensemble members can be

491   further employed to assess the quality of these products through time and provide a more

492   complete comparison of their skill relative to that of the reconstruction. The MSLP from the

493   closest gridpoint of 20CRv2c, 20CRv3, and CERA-20C are plotted in Fig. 7, along with 95%

494   confidence intervals calculated as 1.96 times the ensemble standard deviation of seasonally

495   averaged MSLP from the 56, 80, and 10 ensemble members of 20CRv2c, 20CRv3, and CERA-

496   20C, respectively. The performance varies slightly at neighboring gridpoints for MSLP, but

497   varies considerably if using surface pressure, suggesting elevation gradients from the nearby

498   Roosevelt Island or the edge of the Ross Ice Shelf can influence the reanalysis estimate at this

499   location.

500   Similar to but slightly larger than the reconstruction, Fig. 7 shows that CERA-20C has a

501   pronounced annual cycle of MSLP at Little America, while both 20CRv2C and 20CRv3 have a

502   dampened annual cycle. Nonetheless, Fig. 7 also demonstrates that there is a marked decrease in

503   the reanalyses' uncertainties at the times when the Little America data were assimilated, which

504   effectively constrained the reanalysis estimates. There are also other reductions in the reanalysis

505   uncertainties prior to the establishment of Little America, including the data assimilated from the

506   South Pole race of 1911-1912 (the Norwegian base Framheim was very close to Little America

507   (Fogt *et al.*, 2017b)) and the data from the first British Antarctic Expedition (1908-1909, which

508   established a base near present day McMurdo station west of Little America, at Cape Royds).

509   At these times, the reanalyses and reconstruction show considerable agreement, and the

510   reanalysis uncertainty (colored shading in Fig. 7) overlaps the reconstruction uncertainty (gray

511   shading in Fig. 7). Exceptions are in 1929 for all reanalyses and in 1911 for 20CRv2c and

512   20CRv3 when the reconstruction and reanalysis uncertainties do not overlap, with the

513   reconstruction seasonal mean MSLP lower than the reanalyses. Nonetheless, the overall

514   agreement at all other times suggests that although there may still be larger MAE in the

515   reanalyses MSLP when compared only during the times of direct observations (as in Figs. 5-6),

516   the observations and reconstruction both fall within the reanalyses uncertainty (estimated by the

517   ensemble spread), and the reanalyses have benefited greatly from assimilating this historical data

518   (moving much closer to, if not within, the observational uncertainty). Clearly, at other periods

519   when data are not available, the reanalysis spread is considerably larger, but the reconstruction

520   nearly always falls within the reanalyses' uncertainties (the high positive pressure values in

521   CERA-20C in 1944 are one exception). We note that the reanalysis uncertainty estimates

522   themselves require further improvement, particularly in the Southern Hemisphere: Slivinski et al

523   (2019) show artificial signals in the uncertainty of 20CRv2c, and demonstrate that the

524   uncertainty in this region still remains too large in 20CRv3. Conversely, Laloyaux *et al.* (2018)

525   acknowledge that the small ensemble of CERA-20C can result in overly-confident estimates of

526   uncertainty. Consistent with the comparison to observations (Fig. 6a), the CERA-20C MSLP and

527   reconstruction (Fig. 7c) show the most similarity throughout the early 20th century at the Little

528   America location (RMSD = 4.32 hPa), and for this location 20CRv3 performs better than

529   20CRv2c (RMSD of 5.95 hPa compared to 8.22 hPa).

530   Evaluation of the reconstruction skill in Figs. 2 and 3 often indicates substantial

531   differences, suggesting that the seasonal mean values from the historical observations do not

532   always fall within the reconstruction uncertainty when elevation corrections are applied,

533   particularly near high terrain (even when their uncertainty is accounted for; not shown). To

534   examine this issue further, Figure 8 depicts two cases – a set of Antarctic expedition ship records

535   (Deck 246, Fig. 8a) and a temporary base (Cape Denison, Fig. 8b), where elevation corrections

536   have mixed results; MAE values in Fig. 8 are only calculated for MSLP. Ships included in Deck

537   246 (from ICOADS) operated discontinuously near present day Dumont d'Urville station in the

538   Ross Sea sector of Antarctica (Fig. 4); the data from ICOADS have two disparate locations in

539   January – February 1912 (the mean of these would be over the continent), so instead of

540   averaging data from both locations, each were treated as separate data points in Fig. 8a and the

541   comparison statistics. The two different values for MAE in Fig. 8a are therefore based on the

542   MSLP at these two different locations, with the data closer to the Ross Ice Shelf (second value)

543   agreeing better with all gridded datasets than the data closer to the East Antarctic plateau, near

544   the location plotted in Fig. 4. Even with this potentially conflicting information, the

545   reconstruction performs very well after adjustment to sea level pressure, with an MAE of 2.32 /

546   1.85. A closer look shows that the majority of this error is from DJF 1910, when the

547   reconstruction was about 6 hPa lower than the observational estimate, otherwise it is generally

548   within 1.5 hPa. This performance exceeds the reanalyses' performance for Deck 246, as the

549   reanalyses typically have much higher MSLP values.

24

550   In contrast, the famous base for the Australian Antarctic Expedition, Cape Denison

551   (Mawson, 1998), situated along the East Antarctic coast south of Australia (Fig. 4) is marked

552   with much lower performance across all datasets (Figure 8b). Elevation corrections and the

553   models' orography have a strong effect here, clearly demonstrated by the large differences in

554   surface and sea level pressure in observations and reanalyses (note, surface pressure is plotted on

555   the right axis, but because of the large differences in surface pressure the MAE is not given).

556   Whereas the seasonal means of the historical observations demonstrate surface pressures around

557   920 hPa on average, sea level pressure estimates from the station are around 70 hPa higher, at

558   around 990 hPa. Such high differences in the two over the cold ice sheet plateau compromise the

559   quality of the reduced sea level pressure (not only in observations, but across all datasets). The

560   MAE for the reconstruction (13.50 hPa) is one of the highest of all locations examined.

561   Although the reanalyses' errors are nearly half of this (around 5-7 hPa, likely improved by

562   making further use of temperature and humidity calculated within the reanalyses to reduce

563   surface pressure to sea level), the errors are still large, and the reanalyses and reconstructions all

564   likely fall outside the observational uncertainty. Furthermore, it is difficult to assess how well

565   the interannual variability of the MSLP is reproduced, since there are fewer than ten observations

566   and the reduction to sea level varies considerably by season (affected by temperature and

567   humidity), but the reconstruction was adjusted to sea level uniformly (and linearly) across all

568   seasons. Nonetheless, the limited comparison demonstrates that overall there is still good

569   agreement in the surface pressure interannual variability (despite large differences in magnitude)

570   between the reanalyses and the reconstruction, which potentially suggests that the reconstruction

571   is still capturing aspects of the pressure variability at this location in the early 20th century.

572   Importantly, due to the influence of crude elevation adjustments, it is highly likely that the

573 reconstruction is performing better than indicated by the MAE (Fig. 8b) or RMSD (Fig. 4).

574 Unfortunately, the reconstruction skill is difficult to precisely determine at this or any other

575 location (as suggested in reviewing Fig. 2) where large elevation adjustments make the

576 comparison to historical estimates challenging.

577

578 **4. Discussion and Conclusions**

579 The analysis presented here has compared seasonal Antarctic pressure reconstructions to

580 numerous historical observations and reanalyses throughout the early 20th century. As none of

581 these historical data were included in the reconstruction calibration, they serve as an independent

582 evaluation of the reconstruction skill during a period when relatively little is known about

583 Antarctic climate variability.

584 The results overall confirm that the reconstruction error and uncertainty assessed in

585 earlier work (Fogt *et al.*, 2017b, 2019), a mean absolute error of around 2-4 hPa across the

586 Southern Ocean, is supported when comparing with ship observations. A few ship records

587 suggest even higher reconstruction skill (MAE less than 2 hPa), while others situated north of the

588 Amundsen and Bellingshausen Seas demonstrate lower skill (MAE above 4 hPa), consistent with

589 earlier work. Furthermore, most comparisons with ship observations that span multiple seasons

590 indicate the reconstruction also captures the interannual variability well (correlations often

591 greater than 0.80). Comparison with historical reanalyses provide further evaluation of the

592 reconstruction's performance, as they assimilate most of the historical observations used here but

593 are independent of the reconstruction.

594 While the comparisons with observation estimates taken at sea level are relatively

595 straightforward and further validate the reliability of the early portions of the reconstruction, it is

596   far more challenging to make assessments of the reconstruction skill over areas of higher

597   elevation or near the coastline. In many of these locations, reduction of the reconstruction

598   pressure (which was constructed as surface pressure anomalies relative to the ERA-Int model

599   topography) to sea level pressure using a simple linear adjustment does not satisfactorily agree

600   with the historical data, even after considering the potential observational error. In many of

601   these locations, there are also large differences between 20th century reanalysis MSLP and the

602   historical observational estimates, highlighting the reduced reliability of MSLP from all sources

603   over the cold, high Antarctic continent. While the general comparisons conducted here suggest a

604   larger reconstruction MAE of nearly 6 hPa or more over high elevation, a closer examination at

605   select locations reveals that the reconstruction uncertainty is likely much lower than this and

606   perhaps as low as 1-3 hPa, as indicated in Fogt *et al.* (2019).

607   This work has demonstrated the value of digitizing historical observations from ships and

608   temporary bases for both understanding long term change across the high southern latitudes and

609   evaluating gridded datasets. While their temporary nature may make them difficult to use for

610   assessing long-term variability and change, when coupled with gridded climate datasets like the

611   seasonal Antarctic pressure reconstructions evaluated here, they serve as an independent and

612   valuable tool of documenting historical climate. As one recent example, the use of newly

613   digitized historical observations and pressure reconstructions shed new light on exceptional

614   conditions during the South Pole race of 1911-1912 (Fogt *et al.*, 2017c, 2018; Sienicki, 2018).

615   Future work will hopefully continue to unlock the power of these and other historical

616   observations, so that the ongoing change across the high southern latitudes can be placed in a

617   much-needed longer historical context (Jones *et al.*, 2016).

618

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

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

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

675   Edwards TL, Brandon MA, Durand G, Edwards NR, Golledge NR, Holden PB, Nias IJ, Payne AJ,
676   Ritz C, Wernecke A. 2019. Revisiting Antarctic ice loss due to marine ice-cliff instability. *Nature*,
677   566(7742): 58–64. https://doi.org/10.1038/s41586-019-0901-4.

678   Fogt RL, Goergens CA, Jones JM, Schneider DP, Nicolas JP, Bromwich DH, Dusselier HE. 2017a. A
679   twentieth century perspective on summer Antarctic pressure change and variability and
680   contributions from tropical SSTs and ozone depletion. *Geophysical Research Letters*, 44(19):
681   9918–9927. https://doi.org/10.1002/2017GL075079.

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

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

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

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

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

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

701 reconstructions and CAM5 simulations. *Climate Dynamics*. https://doi.org/10.1007/s00382-
702 019-04674-8.

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

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

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

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

719 Laloyaux P, de Boisseson E, Balmaseda M, Bidlot J-R, Broennimann S, Buizza R, Dalhgren P, Dee
720 D, Haimberger L, Hersbach H, Kosaka Y, Martin M, Poli P, Rayner N, Rustemeier E, Schepers D.
721 2018. CERA-20C: A Coupled Reanalysis of the Twentieth Century. *Journal of Advances in
722 Modeling Earth Systems*, 10(5): 1172–1195. https://doi.org/10.1029/2018MS001273.

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

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

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

733 Purich A, England MH. 2019. Tropical teleconnections to Antarctic sea ice during austral spring
734 2016 in coupled pacemaker experiments. *Geophysical Research Letters*.
735 https://doi.org/10.1029/2019GL082671.

736   Rignot E, Jacobs S, Mouginot J, Scheuchl B. 2013. Ice-Shelf Melting Around Antarctica. *Science*,
737   341(6143): 266–270. https://doi.org/10.1126/science.1235798.

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

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

744   Sienicki K. 2018. Comments on "An Exceptional Summer during the South Pole Race of
745   1911/12." *Bulletin of the American Meteorological Society*, 99(10): 2139–2143.
746   https://doi.org/10.1175/BAMS-D-17-0282.1.

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

755   Steig EJ, Schneider DP, Rutherford SD, Mann ME, Comiso JC, Shindell DT. 2009. Warming of the
756   Antarctic ice-sheet surface since the 1957 International Geophysical Year. *Nature*, 457(7228):
757   459–462. https://doi.org/10.1038/nature07669.

758   Stenni B, Curran MAJ, Abram NJ, Orsi A, Goursaud S, Masson-Delmotte V, Neukom R, Goosse H,
759   Divine D, van Ommen T, Steig EJ, Dixon DA, Thomas ER, Bertler NAN, Isaksson E, Ekaykin A,
760   Werner M, Frezzotti M. 2017. Antarctic climate variability on regional and continental scales
761   over the last 2000 years. *Climate of the Past*, 13(11): 1609–1634. https://doi.org/10.5194/cp-
762   13-1609-2017.

763   Stuecker MF, Bitz CM, Armour KC. 2017. Conditions leading to the unprecedented low Antarctic
764   sea ice extent during the 2016 austral spring season. *Geophysical Research Letters*, 44(17):
765   9008–9019. https://doi.org/10.1002/2017GL074691.

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

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

772   Turner J, Phillips T, Marshall GJ, Hosking JS, Pope JO, Bracegirdle TJ, Deb P. 2017.

773   Unprecedented springtime retreat of Antarctic sea ice in 2016: The 2016 Antarctic Sea Ice

774   Retreat. *Geophysical Research Letters*, 44(13): 6868–6875.

775   https://doi.org/10.1002/2017GL073656.

776

777   **Figure Captions**

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

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

791   **Figure 3**. Decadal mean RMSD plotted by decade.

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

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

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

811   **Figure 7.** Time series of seasonal mean MSLP for all four seasons at Little America (from
812   ISPD) on the northern edge of the Ross Ice Shelf for the reconstruction along with values from a)
813   20CRv2c; b) 20CRv3; c) CERA-20C. The gray shading in each panel represents the 95%
814   confidence interval for the reconstruction, while the colored shading represents 95% confidence
815   intervals for each of the reanalyses (calculated as the 1.96 times the standard deviation across the
816   seasonal mean ensemble members). The overall RMSD compared to the reconstruction is given
817   in the upper right for each dataset.

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

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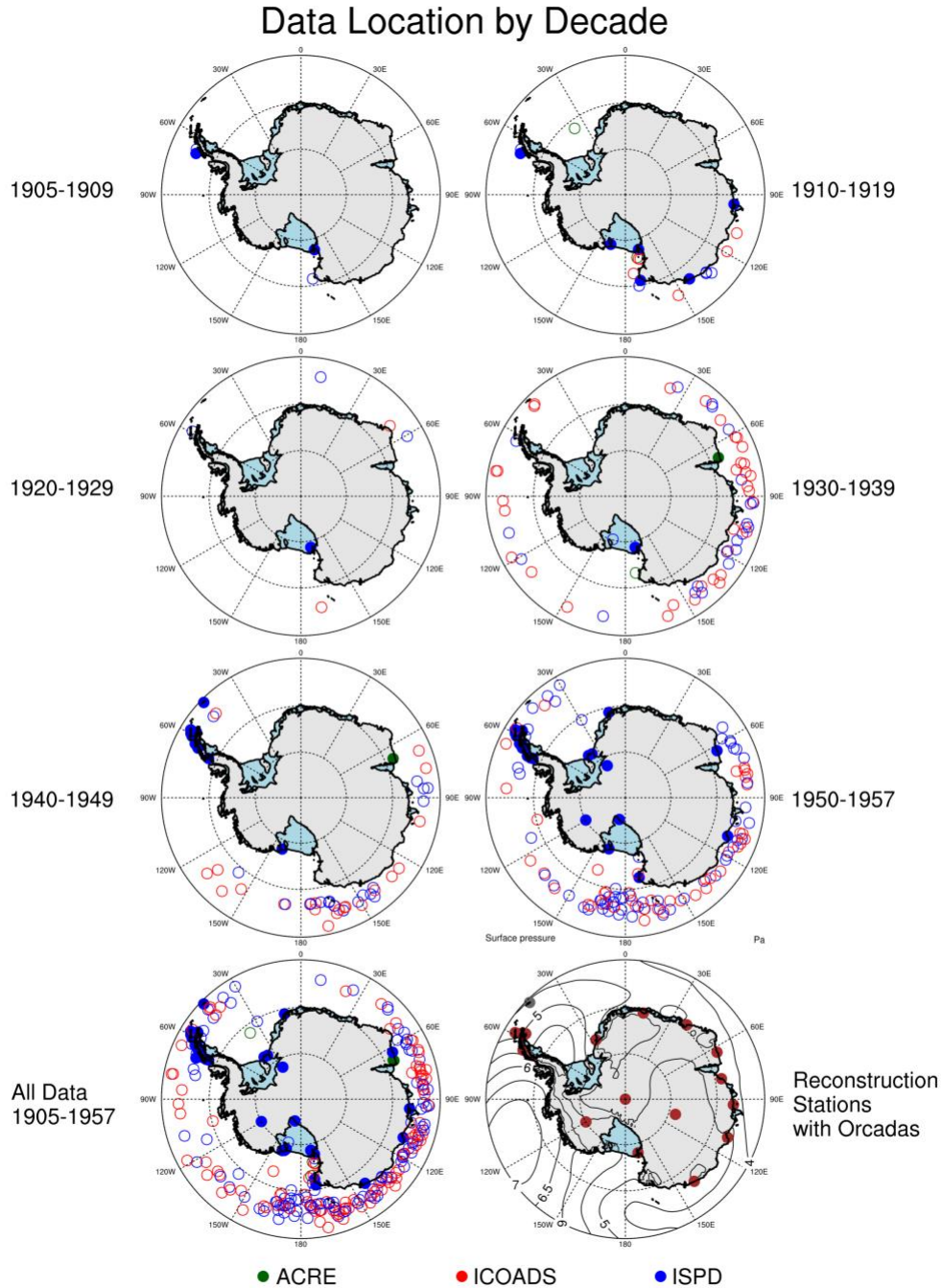826   version 3.

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.
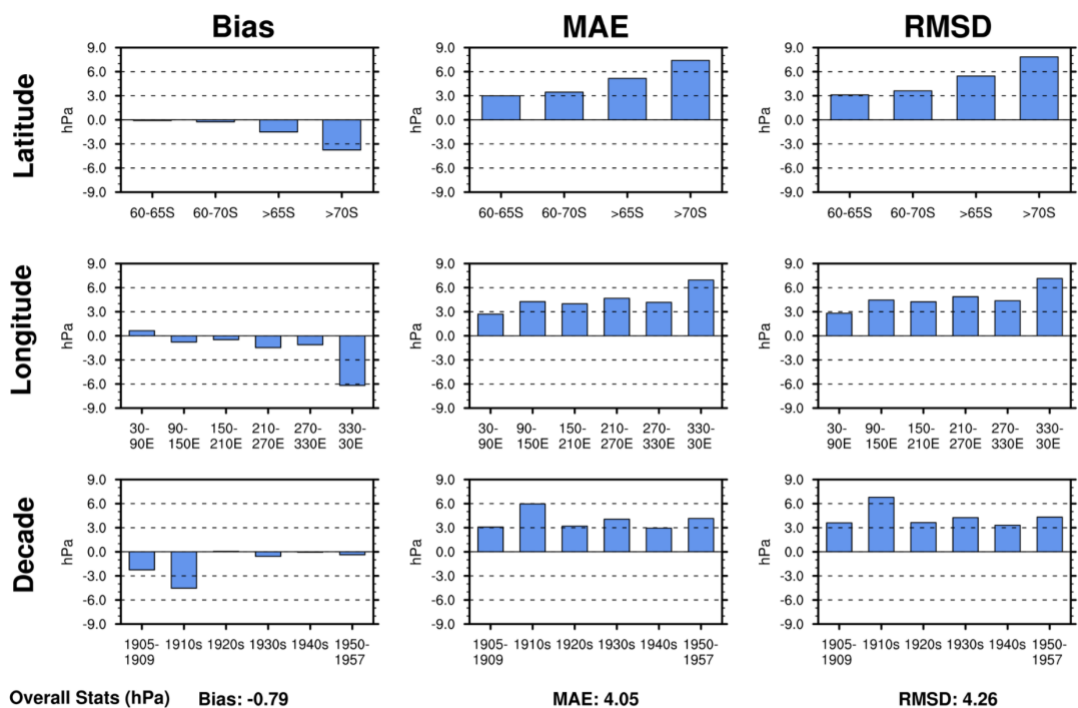
837
Figure 2. Mean reconstruction skill statistics (columns; bias, MAE, and RMSD) compared to
historical observations, and averaged over various latitude bands (top row), longitude bands
(middle row), and decades (bottom row). The statistics calculated over all observations are listed
at the bottom of the figure.
842
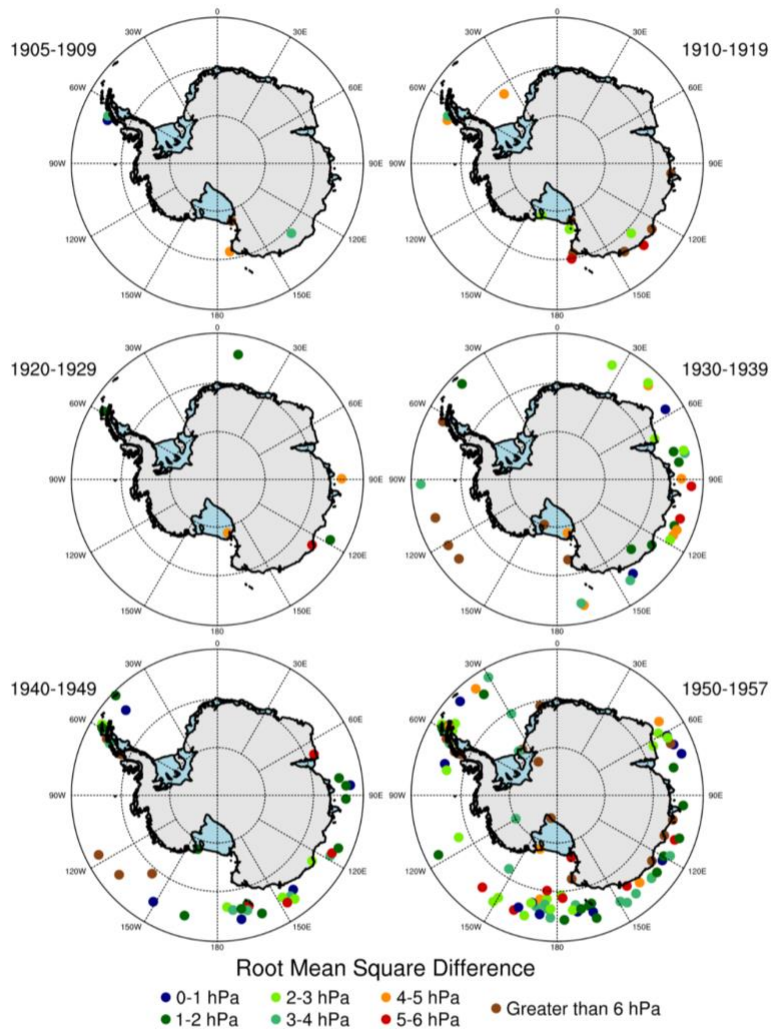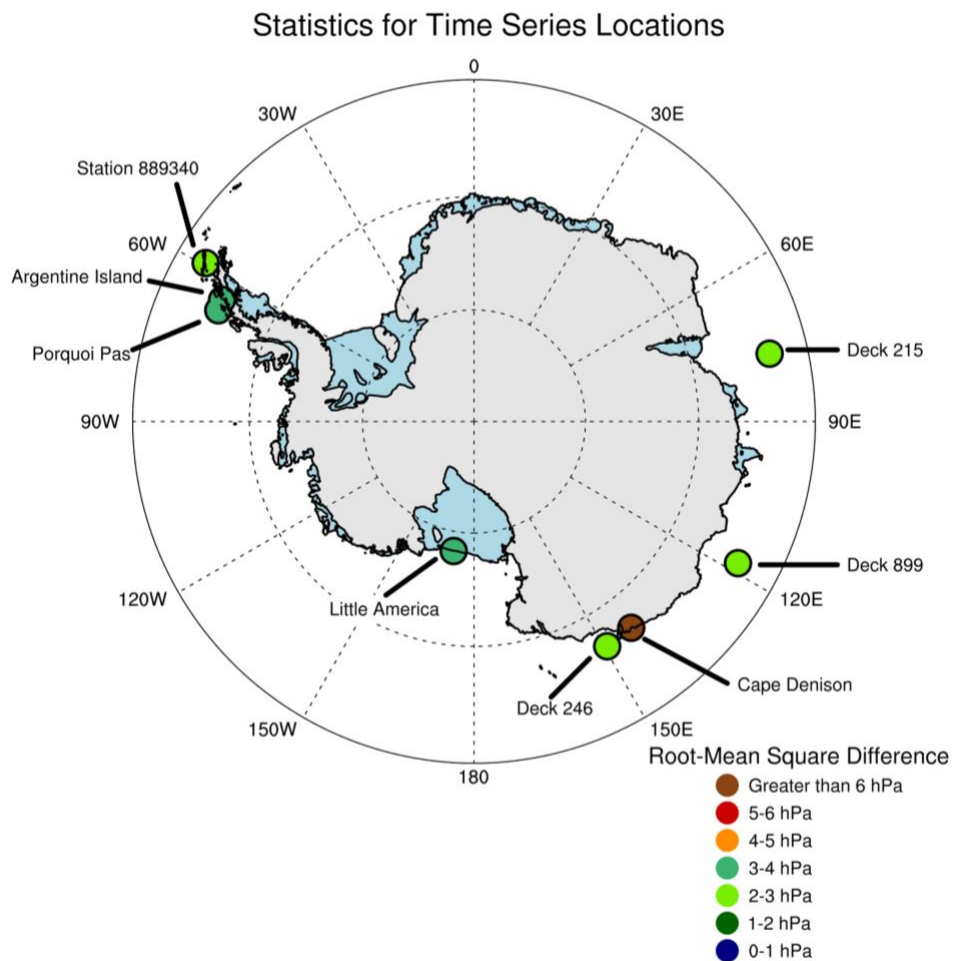
843
844   **Figure 3**. Decadal mean RMSD plotted by decade.
845

Figure 4. Map showing observational mean (over full length of record) RMSD for select representative locations examined in more detail.
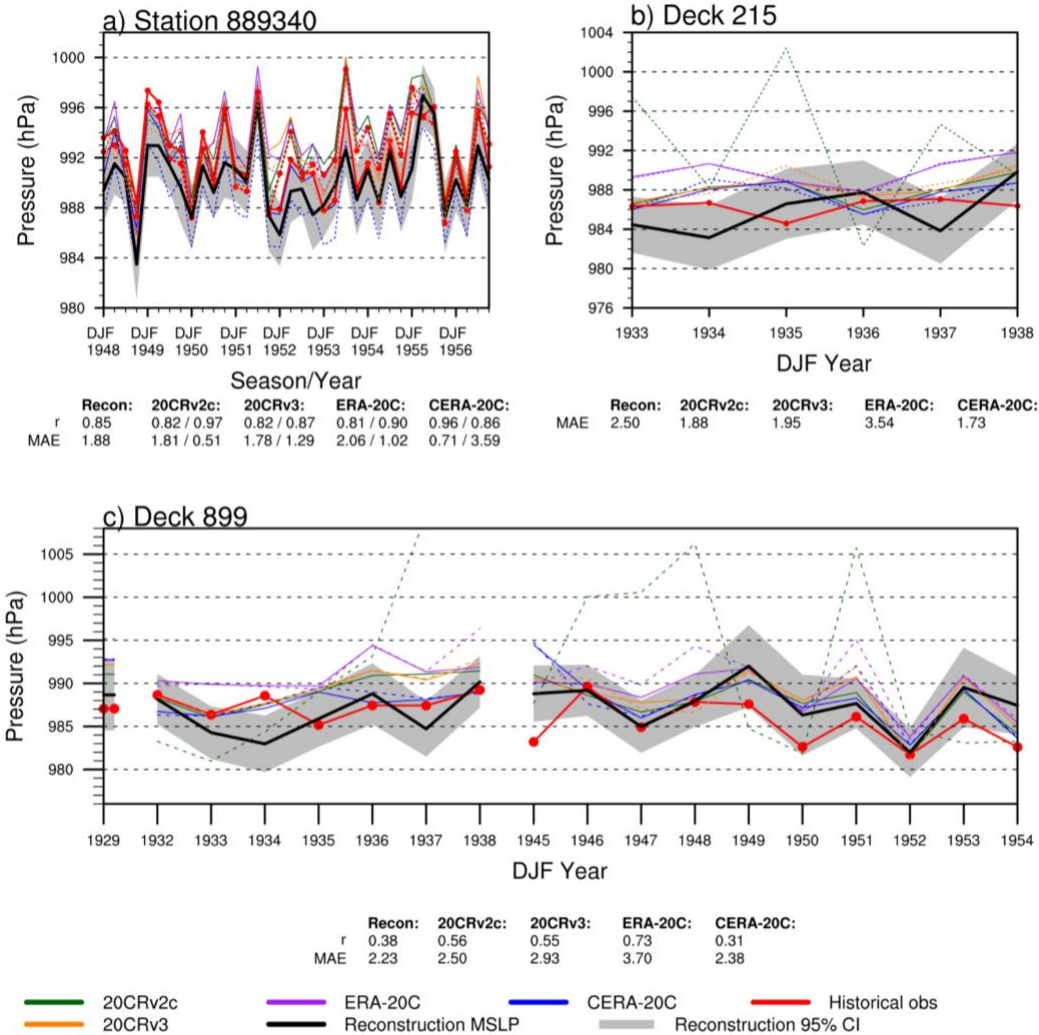
Figure 5. Time series of historical observations, reconstruction (with 95% confidence interval in grey shading), and gridded reanalysis products for observations representative of the lowest RMSD (solid lines for MSLP, dashed lines for surface pressure). The name is the record identifier provided in 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.
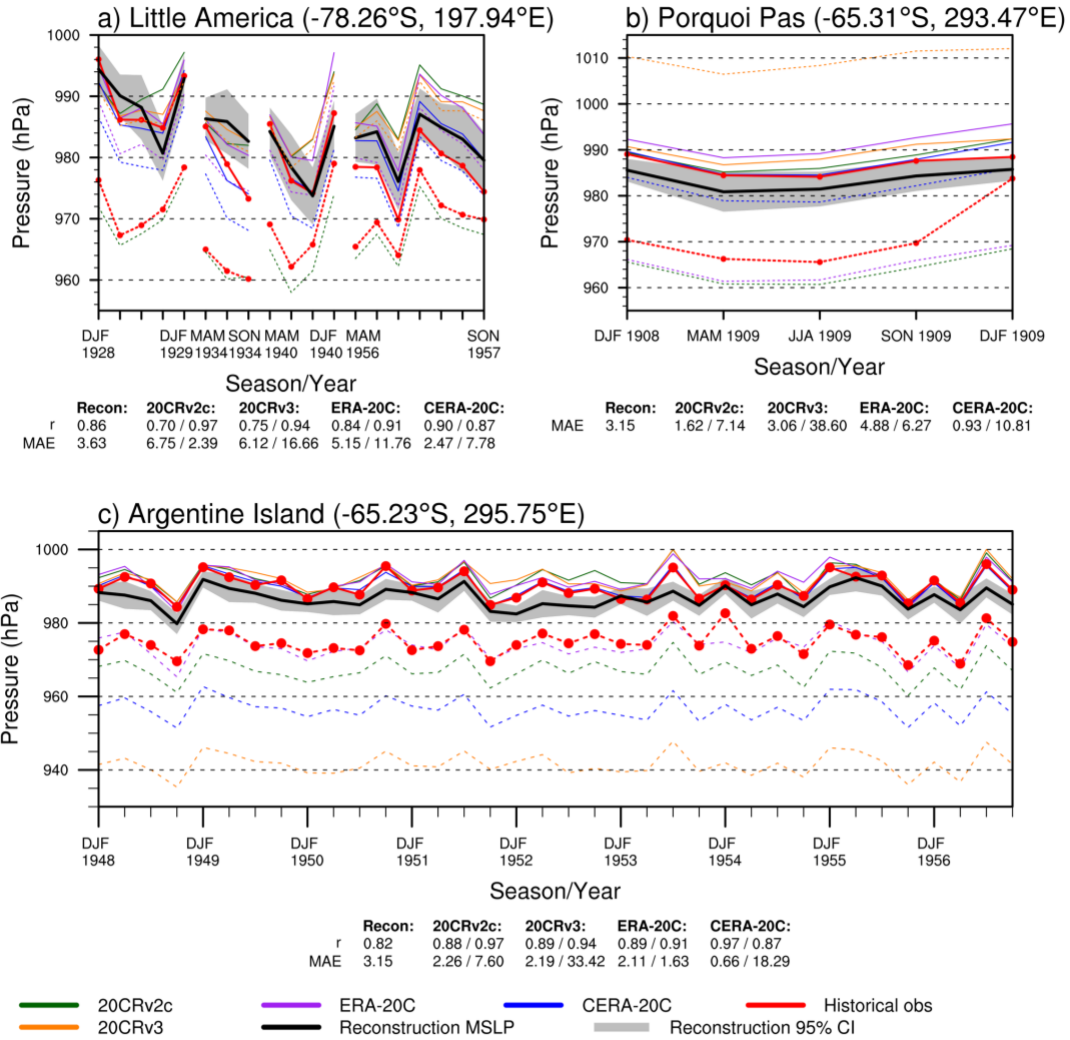
Figure 6. As in Fig. 5, but representative of stations where elevation corrections (reduction to sea level pressure) play an important aspect of the reconstruction performance evaluation. In all panels, the x-axis varies by season. All data in this figure were obtained from ISPD version 3.
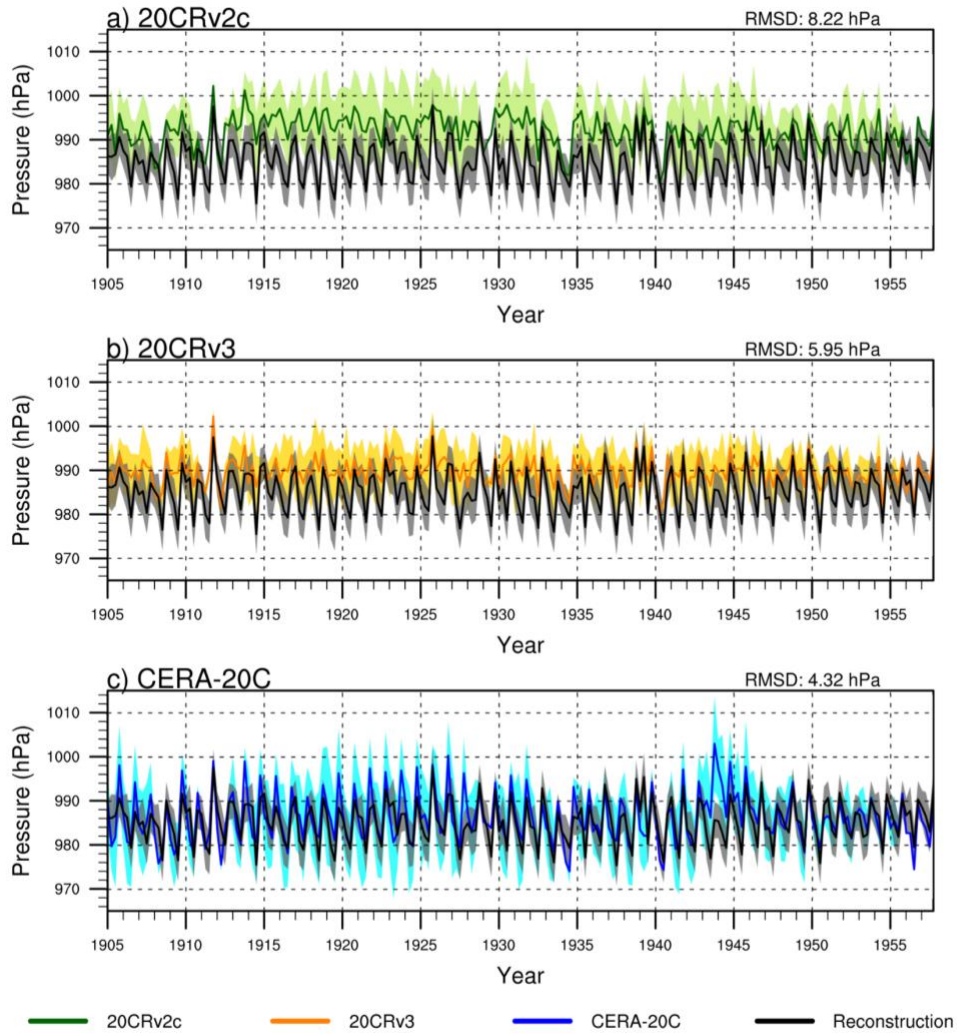
Figure 7. Time series of seasonal mean MSLP for all four seasons at Little America (from ISPD) on the northern edge of the Ross Ice Shelf for the reconstruction along with values from a) 20CRv2c; b) 20CRv3; c) CERA-20C. The gray shading in each panel represents the 95% confidence interval for the reconstruction, 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.
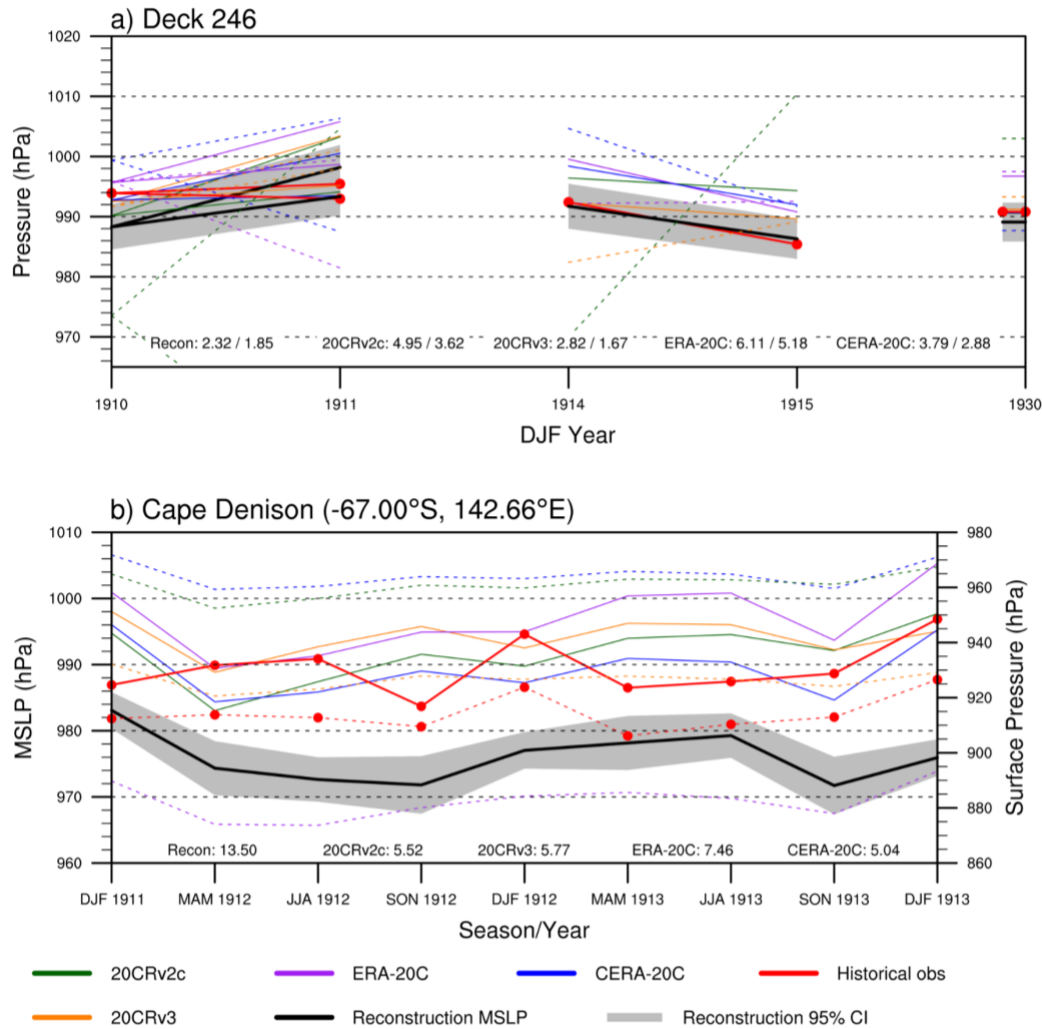
Figure 8. As in Fig. 5, but for a) Deck 246, which operated near the East Antarctic coast discontinuously between 1910-1930, and b) observations at Cape Denison during the Australian Antarctic Expedition of 1911-1914. Note in b) that surface pressure is plotted on the right axis. MAE values are based 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.