University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Slivinski, LC;Compo, GP;Sardeshmukh, PD;Whitaker, JS;McColl, C;Allan, RJ;Brohan, P;Yin, X;Smith, CA;Spencer, LJ;Vose, RS;Rohrer, M;Conroy, RP;Schuster, DC;Kennedy, JJ;Ashcroft, L;Brönnimann, S;Brunet, M;Camuffo, D;Cornes, R;Cram, TA;Domínguez-Castro, F;Freeman, JE;Gergis, J;Hawkins, E;Jones, PD;Kubota, H;Lee, TC;Lorrey, AM;Luterbacher, J;Mock, CJ;Przybylak, RK;Pudmenzky, C;Slonosky, VC;Tinz, B;Trewin, B;Wang, XL;Wilkinson, C;Wood, K;Wyszynski, P

Title:
An evaluation of the performance of the twentieth century reanalysis version 3

Date:
2021-02-15

Citation:
Slivinski, L. C., Compo, G. P., Sardeshmukh, P. D., Whitaker, J. S., McColl, C., Allan, R. J., Brohan, P., Yin, X., Smith, C. A., Spencer, L. J., Vose, R. S., Rohrer, M., Conroy, R. P., Schuster, D. C., Kennedy, J. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D. ,... Wyszynski, P. (2021). An evaluation of the performance of the twentieth century reanalysis version 3. Journal of Climate, 34 (4), pp.1417-1438. https://doi.org/10.1175/JCLI-D-20-0505.1.

Persistent Link:
https://hdl.handle.net/11343/288710

License:
CC-BY

An Evaluation of the Performance of the Twentieth Century Reanalysis Version 3

L. C. SLIVINSKI,[a,b] G. P. COMPO,[a,b] P. D. SARDESHMUKH,[a,b] J. S. WHITAKER,[b] C. MCCOLL,[a,b] R. J. ALLAN,[c] P. BROHAN,[c] X. YIN,[d] C. A. SMITH,[a,b] L. J. SPENCER,[a,b] R. S. VOSE,[e] M. ROHRER,[f,g] R. P. CONROY,[h] D. C. SCHUSTER,[h] J. J. KENNEDY,[c] L. ASHCROFT,[i] S. BRÖNNIMANN,[f] M. BRUNET,[j,k] D. CAMUFFO,[l] R. CORNES,[m] T. A. CRAM,[h] F. DOMÍNGUEZ-CASTRO,[n,o] J. E. FREEMAN,[d] J. GERGIS,[p,q] E. HAWKINS,[r] P. D. JONES,[k] H. KUBOTA,[s] T. C. LEE,[t] A. M. LORREY,[u] J. LUTERBACHER,[v] C. J. MOCK,[w] R. K. PRZYBYLAK,[x,y] C. PUDMENZKY,[z] V. C. SLONOSKY,[aa] B. TINZ,[bb] B. TREWIN,[cc] X. L. WANG,[dd] C. WILKINSON,[k] K. WOOD,[ee,ff] AND P. WYSZYŃSKI[x,y]

[a] *Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado*
[b] *NOAA/Physical Sciences Laboratory, Boulder, Colorado*
[c] *Hadley Centre, Met Office, Exeter, United Kingdom*
[d] *Riverside Technology, Inc., Asheville, North Carolina*
[e] *National Centers for Environmental Information, Asheville, North Carolina*
[f] *Oeschger Centre for Climate Change Research and Institute of Geography, University of Bern, Bern, Switzerland*
[g] *Axis Capital, Zurich, Switzerland*
[h] *National Center for Atmospheric Research, Boulder, Colorado*
[i] *School of Earth Sciences, University of Melbourne, Parkville, Victoria, Australia*
[j] *Centre for Climate Change, Universitat Rovira i Virgili, Tarragona, Spain*
[k] *Climatic Research Unit, University of East Anglia, Norwich, United Kingdom*
[l] *National Research Council, Institute of Atmospheric Sciences and Climate, Padua, Italy*
[m] *National Oceanography Centre, Southampton, United Kingdom*
[n] *Aragonese Agency for Research and Development (ARAID), Zaragoza, Spain*
[o] *Department of Geography, University of Zaragoza, Zaragoza, Spain*
[p] *Fenner School of Environment and Society, Australian National University, Canberra, Australian Capital Territory, Australia*
[q] *ARC Centre of Excellence for Climate Extremes, Australian National University, Canberra, Australian Capital Territory, Australia*
[r] *National Centre for Atmospheric Science, Department of Meteorology, University of Reading, United Kingdom*
[s] *Hokkaido University, Sapporo, Japan*
[t] *Hong Kong Observatory, Hong Kong, China*
[u] *National Institute of Water and Atmospheric Research, Auckland, New Zealand*
[v] *World Meteorological Organization, Science and Innovation Department, Geneva, Switzerland*
[w] *Department of Geography, University of South Carolina, Columbia, South Carolina*
[x] *Department of Meteorology and Climatology, Nicolaus Copernicus University, Toruń, Poland*
[y] *Centre for Climate Change Research, Nicolaus Copernicus University, Toruń, Poland*
[z] *Centre for Applied Climate Sciences, University of Southern Queensland, Toowoomba, Queensland, Australia*
[aa] *Centre for Interdisciplinary Studies of Montreal, McGill University, Montreal, Quebec, Canada*
[bb] *Deutscher Wetterdienst, Hamburg, Germany*
[cc] *Australian Bureau of Meteorology, Melbourne, Victoria, Australia*
[dd] *Climate Research Division, Environment and Climate Change Canada, Toronto, Ontario, Canada*
[ee] *NOAA/Pacific Marine Environmental Laboratory, Seattle, Washington*
[ff] *Cooperative Institute for Climate, Ocean, and Ecosystem Studies, University of Washington, Seattle, Washington*

(Manuscript received 2 July 2020, in final form 13 November 2020)

ABSTRACT: The performance of a new historical reanalysis, the NOAA–CIRES–DOE Twentieth Century Reanalysis version 3 (20CRv3), is evaluated via comparisons with other reanalyses and independent observations. This dataset provides global, 3-hourly estimates of the atmosphere from 1806 to 2015 by assimilating only surface pressure observations and prescribing sea surface temperature, sea ice concentration, and radiative forcings. Comparisons with independent observations, other reanalyses, and satellite products suggest that 20CRv3 can reliably produce atmospheric estimates on scales ranging from weather events to long-term climatic trends. Not only does 20CRv3 recreate a ''best estimate'' of the weather, including extreme events, it also provides an estimate of its confidence through the use of an ensemble. Surface pressure statistics suggest that these confidence estimates are reliable. Comparisons with independent upper-air observations in the

Denotes content that is immediately available upon publication as open access.

This article is licensed under a Creative Commons Attribution 4.0 license (http://creativecommons.org/licenses/by/4.0/).

Corresponding author: Laura C. Slivinski, laura.slivinski@noaa.gov

DOI: 10.1175/JCLI-D-20-0505.1

© 2021 American Meteorological Society

Unauthenticated | Downloaded 10/14/21 01:07 AM UTC

Northern Hemisphere demonstrate that 20CRv3 has skill throughout the twentieth century. Upper-air fields from 20CRv3 in the late twentieth century and early twenty-first century correlate well with full-input reanalyses, and the correlation is predicted by the confidence fields from 20CRv3. The skill of analyzed 500-hPa geopotential heights from 20CRv3 for 1979–2015 is comparable to that of modern operational 3–4-day forecasts. Finally, 20CRv3 performs well on climate time scales. Long time series and multidecadal averages of mass, circulation, and precipitation fields agree well with modern reanalyses and station- and satellite-based products. 20CRv3 is also able to capture trends in tropospheric-layer temperatures that correlate well with independent products in the twentieth century, placing recent trends in a longer historical context.

KEYWORDS: Synoptic-scale processes; Surface pressure; Surface observations; Data assimilation; Reanalysis data; Decadal variability

1. Introduction

A detailed understanding of past weather and climate, including variability and trends, is essential to better understand and predict ongoing changes in climate and weather statistics. Historical observational datasets intended to accomplish this are spatially and temporally incomplete, and often have inhomogeneity issues (Brönnimann et al. 2013; Cram et al. 2015; Jones et al. 1999; Parker et al. 1997; Rennie et al. 2014; Thorne et al. 2017; Noone et al. 2021). Reanalyses can provide complete and consistent atmospheric fields by objectively combining historical observations with modern numerical weather prediction model forecasts, while accounting for estimated errors in both (Kalnay et al. 1996). Most reanalyses, however, only go back to circa 1950 or 1979 to use the most comprehensive observing network while avoiding inconsistencies arising from major changes in it, such as the introduction of extensive upper-air observations or satellite data (Bengtsson et al. 2004; Bosilovich et al. 2011; Kinter et al. 2004; Kistler et al. 2001; Zhang et al. 2012).

By assimilating only long-term surface observations, historical reanalyses can avoid some of these inconsistencies and extend further back in time. In partnership with the international Atmospheric Circulation Reconstructions over the Earth initiative (ACRE; Allan et al. 2011), the University of Colorado Boulder's Cooperative Institute for Research in Environmental Sciences (CIRES) and the National Oceanic and Atmospheric Administration (NOAA) were the first to generate a dynamically consistent ''sparse input'' global atmospheric reanalysis, the Twentieth Century Reanalysis (20CR), based on only surface pressure observations. The preliminary first version spanned 1908–58 to demonstrate the feasibility of a surface-pressure-only reanalysis (Compo et al. 2006; Whitaker et al. 2004). The second version, 20CRv2, spanned from 1871 to the present and was kept up to date until 2012 (Compo et al. 2011). The follow-up, 20CRv2c, improved upon 20CRv2 and extended the reanalysis period to 1851–2014. The European Centre for Medium-Range Weather Forecasts (ECMWF) subsequently generated a sparse-input reanalysis, the ECMWF Twentieth Century Reanalysis (ERA-20C), assimilating both surface pressure and marine winds and extending back to 1900 (Poli et al. 2016). Their most recent historical reanalysis, the Coupled ECMWF Reanalysis of the Twentieth Century (CERA-20C), spans 1901–2010 and uses a coupled ocean–atmosphere forecast model to also assimilate subsurface ocean profile observations (Laloyaux et al. 2018).

The latest version of the Twentieth Century Reanalysis has been generated by NOAA, CIRES, and the U.S. Department of Energy (DOE). This NOAA–CIRES–DOE 20CR version 3 (20CRv3), uses a newer, higher-resolution model, assimilates a larger set of observations, and includes an improved data assimilation system relative to its predecessor 20CRv2c. The 20CRv3 system further extends the reanalysis period to 1836–2015, with an experimental extension spanning 1806–35. Slivinski et al. (2019a) provide an in-depth description of the system that generated the 20CRv3 reanalysis product, as well as preliminary evaluations of a subset of the 20CRv3 product (mainly via comparisons with 20CRv2c). Here, we provide a first evaluation of the entire 1806–2015 time span of the 20CRv3 dataset through comparisons with independent observations, full- and sparse-input reanalyses, and satellite products. The focus is on the synoptic and climatic behavior of a few key atmospheric variables from the surface to the upper atmosphere; further work will provide more exhaustive evaluations.

By going back to 1806, one can potentially study trends in the longest instrument-based reanalysis generated to date. However, one must first understand the accuracy and reliability of the dataset relative to other products and observations. To that end, this work provides an initial evaluation of the 20CRv3 dataset on weather and climate scales, but in-depth investigations of particular phenomena and variability are left for future research. This work is organized as follows: Section 2 reviews relevant details of 20CRv3 and introduces the other reanalyses and observational datasets used for comparison. Section 3 investigates an extreme weather event in the nineteenth century, illustrating that 20CRv3 has efficacy for evaluating the dynamics of past weather when data density is sufficient. Section 4 considers synoptic-scale performance metrics, including errors in surface pressure and upper-air geopotential heights. Section 5 focuses on climate time scales and discusses long-term structures of mass, circulation, and precipitation in 20CRv3. Section 6 concludes with a discussion of implications.

2. Data and methods

The 20CRv3 system consists of a numerical weather prediction model, an observational dataset, and an assimilation method. We review this system briefly; further details are given by Slivinski et al. (2019a). Using an 80-member ensemble Kalman filter, the 20CRv3 system assimilates only surface

Unauthenticated | Downloaded 10/14/21 01:07 AM UTC

pressure observations from the open, unrestricted, and publicly available International Surface Pressure Databank (ISPD) version 4.7 (Compo et al. 2019; Cram et al. 2015), into the U.S. National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) model, version 14.0.1, with a spectral horizontal resolution of T254 (effectively 60 km at the equator) and a vertical atmospheric resolution of 64 levels up to about 0.3 hPa. Sea surface temperature (SST) fields are prescribed from two eight-member ensembles: version 3 of the Simple Ocean Data Assimilation with sparse input (SODAsi.3), reanalysis for 1836–1980 (Giese et al. 2016); and the Hadley Centre Sea Ice and Sea Surface Temperature dataset, version 2.2 (HadISST2.2), for 1981–2015. These SST fields, originally available as 5-day averages, were interpolated to daily resolution for 20CRv3. Sea ice concentration fields are prescribed from monthly HadISST2.3 (Titchner and Rayner, 2014) interpolated to daily resolution. Solar radiation is determined from the Total Solar Irradiance (TSI) Reconstruction based on the Naval Research Laboratory TSI (NRLTSI2; Coddington et al. 2016), and time-varying atmospheric constituents of volcanic aerosols (Crowley and Unterman, 2013), stratospheric ozone (Cionni et al. 2011), and atmospheric carbon dioxide (CO_2) levels (Saha et al. 2010) are also specified. Output subsurface land fields (such as soil moisture), surface fields, and atmospheric fields are provided at 3-hourly resolution.

The experimental extension spanning 1806–35 uses an identical system, with SST fields for 1815–35 specified from the eight-member ensemble of SODAsi.3 and for 1806–14 specified as the 30-yr average climatological fields from HadISST2.1 for years 1861–90 (to be consistent with previous iterations of 20CR; see Giese et al. 2016). At the time of production, 1804 was the first year with at least one observation globally every six hours (mainly from stations in western Europe and occasionally North America). With a 16-month spinup period, 1806 is the first full year available in the extension.

To evaluate the performance of 20CRv3, its fields are compared with a variety of atmospheric reanalyses and observational datasets. Fujiwara et al. (2017) organize reanalyses into three general categories: full input, conventional input, and surface input. Full-input reanalyses assimilate all available surface and upper-air conventional observations as well as satellite observations as they become available in time. Conventional-input reanalyses assimilate surface and upper-air conventional observations, but not satellite data; we do not consider any conventional-input reanalyses here. Surface-input reanalyses assimilate only surface conventional observations (such as surface pressure and marine winds.)

In addition to reanalyses, 20CRv3 is also compared with independent upper-air, station-based, satellite-based, and satellite-station blended observational datasets of upper-air fields, precipitation, and lower-tropospheric temperature. A summary of all reanalyses and datasets used in this study is shown in Table 1. This list is far from exhaustive; see Fujiwara et al. (2017) for more details.

3. Case study: The Great Blizzard of 1888

A key feature of 20CRv3 is its 3-hourly resolution for the span of 210 years, allowing users to investigate weather extremes from across the nineteenth to twenty-first centuries in a consistent framework. Several studies have shown how well previous versions of 20CR were able to reconstruct individual weather events (e.g., Brönnimann et al. 2013; Gergis et al. 2020; Lorrey and Chappell 2016; Moore and Babij 2017; Stucki et al. 2015). Slivinski et al. (2019a) investigated how well the 20CRv3 system represents the 1915 Galveston Hurricane, and found that 20CRv3 has the strongest intensity (i.e., lowest central pressure) of the four historical reanalyses considered there: 20CRv2c, 20CRv3, ERA-20C, and CERA-20C. Here, we investigate an extratropical winter storm that impacted North America in the nineteenth century and has been the focus of several previous studies (Kocin 1983, 1988; Michaelis and Lackmann 2013), but 20CRv3 could be used to study a variety of weather extremes back to 1806 (e.g., the 1816 year without a summer; Brugnara et al. 2015; Skrynyk et al. 2021).

The Great Blizzard of 1888 was a historic snowstorm impacting the northeast United States between 11 and 14 March 1888, delivering up to 125 cm of snow across parts of New England along with strong winds and low temperatures (Kocin and Uccellini 2004; Kocin 1983, 1988). For reference, in the original development of the Northeast Snowfall Impact Scale (NESIS), the Blizzard of 1888 received a score of 8.34, while the Superstorm of March 1993 received a value of 12.52, and the Blizzard of January 1996 received a score of 11.54. Michaelis and Lackmann (2013) downscaled 20CRv2 using the Weather Research and Forecasting (WRF) Model and were able to reconstruct offshore cyclogenesis and heavy snowfall, albeit with a significant position error. Here, we illustrate the ability of the 20CRv3 system to reconstruct this storm at its output resolution. Figure 1 shows synoptic maps over North America for 0000 UTC 13 March 1888. Figure 1a shows ensemble mean sea level pressure (SLP), the locations of observations assimilated within the previous 24 h, and confidence. Confidence of an atmospheric field is defined as in (Slivinski et al. 2019a) but for a single time t:

$$\mathrm{conf}(t) = 1 - \frac{\mathrm{spread}_{\mathrm{ens}}(t)}{\mathrm{spread}_{\mathrm{clim}}(t)}, \tag{1}$$

where $\mathrm{spread}_{\mathrm{ens}}$ is the standard deviation of the ensemble of analyzed fields from 20CRv3 valid for the given date and time, and $\mathrm{spread}_{\mathrm{clim}}$ is the temporal standard deviation of the 20CRv3 ensemble mean field, over the given month, day, and hour, for 1981–2010. This metric is intended to provide information on the ensemble spread (not on mean biases), while normalizing out the effects of intrinsic variability. A value of zero in this metric corresponds to confidence equal to that of a climatological estimate, while negative (positive) values correspond to less (more) confidence than a climatological estimate. In particular, negative values can occur when the instantaneous ensemble spread is larger than the ensemble mean's temporal variability. This does not necessarily imply poor performance of the data assimilation system, unless long-term averages of confidence are regularly below zero (see section 3.1 of Slivinski et al. 2019a). Figure 1b shows ensemble mean 500-hPa geopotential height (Z500) and its confidence field. Figures 1c and 1d show ensemble mean 2-m air temperature and 6-hourly accumulated precipitation, respectively.

Unauthenticated | Downloaded 10/14/21 01:07 AM UTC

TABLE 1. Summary of reanalyses and observational datasets discussed in this study.

Dataset	Reference	Category	Description
NOAA–CIRES 20CRv2, v2c	Compo et al. (2011)	Surface-input reanalysis	Produced by NOAA–CIRES. Available globally from 1871 to 2012 (20CRv2) or from 1851 to 2014 (20CRv2c).
NOAA–CIRES–DOE 20CRv3	Slivinski et al. (2019a)	Surface-input reanalysis	Produced by NOAA–CIRES–DOE. Available globally from 1836 to 2015 (experimental extension from 1806 to 1835).
ERA-20C	Poli et al. (2016)	Surface-input reanalysis	Produced by ECMWF. Available globally from 1900 to 2010.
CERA-20C	Laloyaux et al. (2018)	Surface-input reanalysis	Produced by ECMWF. Available globally from 1901 to 2010.
NNR1	Kalnay et al. (1996) and Kistler et al. (2001)	Full-input reanalysis	Produced by NCEP–NCAR. Available globally from 1948 to the present.
ERA-40	Uppala et al. (2005)	Full-input reanalysis	Produced by ECMWF. Available globally from 1957 to 2002.
JRA-55	Kobayashi et al. (2015)	Full-input reanalysis	Produced by the Japan Meteorological Agency (JMA). Available globally from 1958 to the present.
ERA-Interim	Dee et al. (2011)	Full-input reanalysis	Produced by ECMWF. Available globally from 1979 to 2019.
ERA5	Hersbach et al. (2020) and Hersbach et al. (2016)	Full-input reanalysis	Produced by ECMWF. Available globally from 1979 to the present.
IGRA2	Durre et al. (2016)	Upper-air conventional observations	Full dataset produced by NOAA National Centers for Environmental Information (NCEI). The Lindenberg, Germany, station used in this study is available from 1905 to the present.
GPCP v2.3	Adler et al. (2003, 2018)	Global satellite-station blended precipitation dataset	Produced under the World Climate Research Program (WCRP) and its Global Energy and Water Cycle Experiment (GEWEX). Available globally from 1979 to the present.
CRU TS v4.03	Harris et al. (2020) and Harris and Jones (2020)	Global station-based precipitation dataset	Produced by the University of East Anglia CRU. Available globally over land from 1901 to 2019.
PRISM	http://prism.oregonstate.edu	Regional station-based precipitation dataset	Produced by the PRISM Climate Group of Oregon State University. Available over the continental United States from 1895 to the present.
AWAP	Jones et al. (2009)	Regional station-based precipitation dataset	Produced by the AWAP. Available over Australia from 1900 to the present.
RSS	Mears and Wentz (2009)	Satellite-based temperature lower troposphere	Produced by RSS. Available for 70°S–82°N from 1979 to the present.
UAH	Christy et al. (2017)	Satellite-based mean-layer temperature	Produced by the UAH. Available for 82°S–82°N from 1979 to the present.

Unauthenticated | Downloaded 10/14/21 01:07 AM UTC

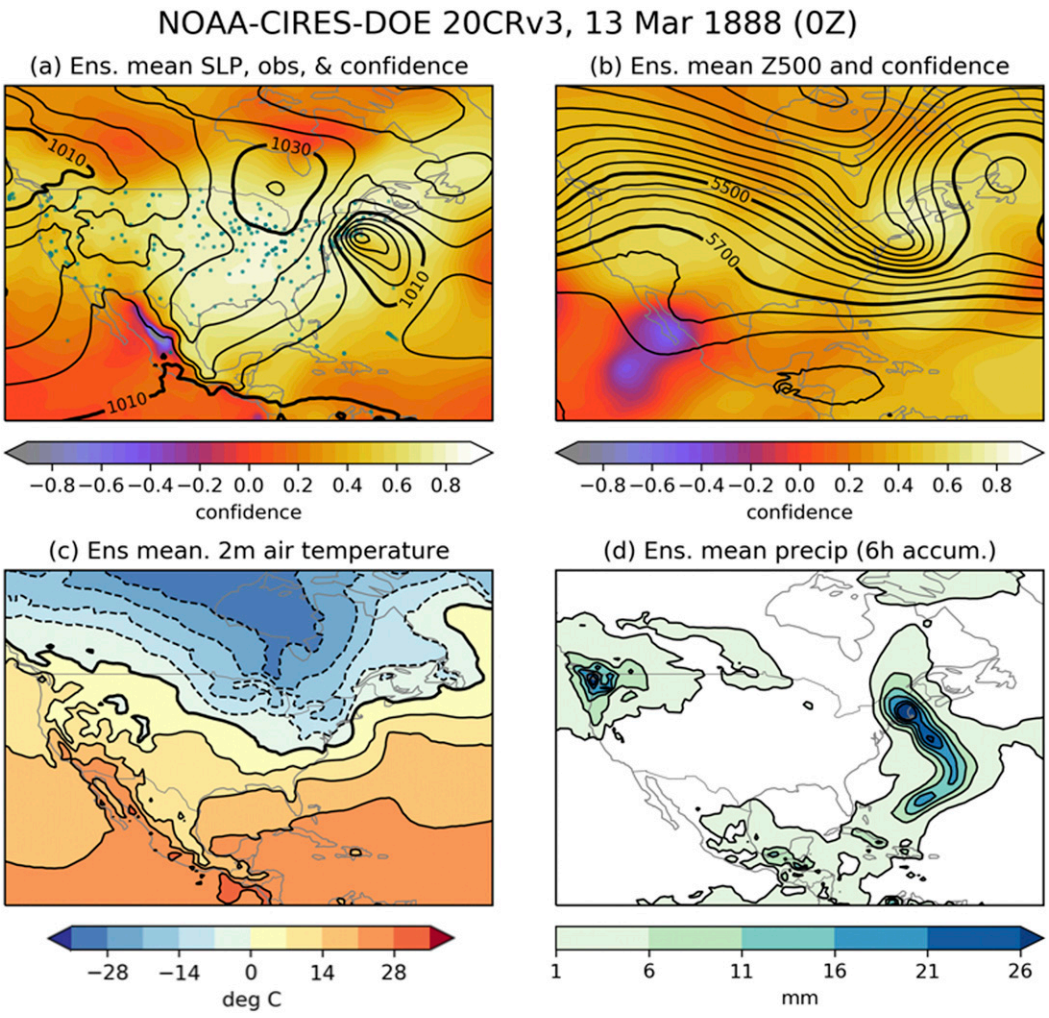


FIG. 1. Synoptic maps from 20CRv3 for 0000 UTC 13 Mar 1888 (the Great Blizzard of 1888). (a) Ensemble mean sea level pressure (contours; interval 5 hPa), observation locations from previous 24 h (teal circles), and confidence (shading). (b) Ensemble mean geopotential height at 500 hPa (contours; interval 50 gpm) and confidence (shading). (c) Ensemble mean 2-m air temperature with the 0°C contour thickened. (d) Ensemble mean precipitation (mm) accumulated over the prior 6 h.

While the confidence shading in Figs. 1a and 1b demonstrates the spread of the ensemble, Fig. 2 provides further evidence of the ensemble variability by illustrating the SLP and precipitation fields from the first 20 ensemble members; this set is representative of the full 80-member ensemble. A storm is evident in all ensemble members shown here, but the variability within the ensemble suggests that there is more uncertainty in precipitation than in SLP.

Comparisons with preexisting reconstructions demonstrate that the overall structure and location of the storm in 20CRv3 are realistic (Fig. 3). The real-time U.S. Daily Weather Map, a reconstruction created by Kocin (1983), and the 20CRv3 fields valid at 1200 UTC 13 March 1888 compare well with each other. Note that this is 12 h later than the fields shown Figs. 1 and 2, after the storm began to move offshore, in order to compare with the U.S. Daily Weather Maps (which are only available once per day at 1200 UTC). Though the 20CRv3 ensemble mean is arguably an imperfect estimate of an extreme weather event since ensemble averaging tends to dampen gradients in position-dependent features, there is still good agreement across all three reconstructions, particularly in the location of the storm.

This example illustrates how well 20CRv3 can represent historic extreme weather events at the surface on subdaily scales. In the remainder of the paper, we investigate how well 20CRv3 performs on synoptic to climatic scales from the surface to high levels of the atmosphere.

4. Synoptic skill evaluation

a. Surface pressure statistics

To investigate the overall performance, we initially consider background errors in surface pressure. The root-mean-squared

Unauthenticated | Downloaded 10/14/21 01:07 AM UTC

errors (RMSEs) between independent surface pressure observations that have not yet been assimilated and the 20CRv3 ensemble mean background equivalents provide a simple metric for how well the system is performing. We refer to this as the ''actual'' error:

$$\mathrm{RMSE}_{\mathrm{actual}} = \left[\frac{1}{N_{\mathrm{obs}}}\sum_{i=1}^{N_{\mathrm{obs}}}(x_{o,i} - x_{b,i})^2\right]^{1/2}, \tag{2}$$

where i indexes all N_{obs} observations considered (here, in a given region for a single year), $x_{o,i}$ is the ith observation, and $x_{b,i}$ is the ensemble mean background field interpolated to the ith observation time and location. By comparing to expected errors, we can determine how well the uncertainty of the system is being estimated as well. Here, ''expected'' errors are calculated as the root mean of the sum of the observation error $\sigma^2_{o,i}$ and the estimated background error (variance of the background ensemble interpolated to the observation's location and time) $\sigma^2_{b,i}$:

$$\mathrm{RMSE}_{\mathrm{exp}} = \left[\frac{1}{N_{\mathrm{obs}}}\sum_{i=1}^{N_{\mathrm{obs}}}\sigma^2_{o,i} + \sigma^2_{b,i}\right]^{1/2}. \tag{3}$$

As shown by Desroziers et al. (2005), if the observation and background errors are uncorrelated and unbiased, then $\mathrm{RMSE}_{\mathrm{actual}}$ should be equivalent to $\mathrm{RMSE}_{\mathrm{exp}}$.

Figure 4 shows observation–forecast difference statistics for surface pressure and the associated expected RMSEs for three zonal regions and for the full experimental and production span of 20CRv3, 1806–2015. Here the Northern Hemisphere (NH) is defined as 20°–90°N, the tropics as 20°S–20°N, and the Southern Hemisphere (SH) as 90°–20°S. The RMSEs generally decrease in time and are consistent with the expected errors even as the available pressure observations vary over four orders of magnitude, demonstrating a high level of performance in this metric and an improvement over the performance of 20CRv2c due to the upgrades in forecast model and data assimilation algorithms (see Slivinski et al. 2019a, their Fig. 12 and related discussion). This also suggests that the prescribed observation errors are realistic, though the disagreement (particularly when the expected errors are larger than the actual errors) suggests that there is still room for improvement; this will be investigated in future tests.

There is a strong negative correlation between the RMSE and the log of the number of observations assimilated (Fig. 4), demonstrating that the errors generally decrease as the observation network density increases. This motivates global and regional data rescue efforts, particularly in regions that are currently data sparse (e.g., Allan et al. 2011; Brönnimann et al. 2019a; Williamson et al. 2018; https://climatehistory.com.au); it is estimated that there could be millions of pressure observations that exist but are currently unavailable for assimilation (i.e., they have not yet been scanned or digitized). If early paper record observations were digitized and made available for assimilation, Fig. 4 suggests that errors could be significantly decreased in the corresponding time periods. The continued decrease in actual error in the tropics and Southern Hemisphere also suggests that even now, the observation networks in these regions are not dense enough for the error to have saturated at its minimum level.

b. *Modern upper-air comparisons*

In Fig. 5 we evaluate the skill of Z500 estimates in 20CRv3, a large-scale variable that governs atmospheric circulation patterns (such as troughs and ridges), with respect to ERA5, the latest full-input reanalysis from ECMWF (Hersbach et al. 2020; Hersbach and Dee 2016). ERA5 assimilates nearly all available observations including extensive satellite-based, radiosonde, aircraft, and other conventional upper-air and surface observations. To compare with the available operational forecast errors, which are unavailable poleward of 80°, here we define ''NH'' as 20°–80°N and ''SH'' as 80°–20°S. In both the NH and SH, 20CRv3 has smaller root-mean-squared differences (RMSDs) with respect to ERA5 than 20CRv2c, and similar to Fig. 4, RMSDs are more consistent with the ensemble spread in 20CRv3. This is in contrast to 20CRv2c, whose small ensemble spread relative to its RMSD suggests overconfidence, likely due to the imperfect covariance inflation algorithm used in the 20CRv2c system that was updated for 20CRv3 (Slivinski et al. 2019a). The seasonal variability in 20CRv3 differences relative to ERA5 has also been diminished from 20CRv2c in the NH. For reference, the 1981–2010 climatological variability (standard deviation over time) in the 20CRv3 ensemble mean field is 79.24 m averaged over the Northern Hemisphere and 91.09 m averaged over the Southern Hemisphere.

To put these numbers in context, we also compare with the 2019 annual average errors of operational forecasts of Z500 for 2-, 3-, and 4-day leads. The 20CRv3 analysis errors for 1979–2015 are comparable to modern 3–4-day operational forecast skill in the NH, and 4-day forecast skill in the SH, consistent with the expected skills predicted by Compo et al. (2006). The degraded performance in the SH for the period 1980–85 may be due to a lack of observations in this period (see the small drop in number of assimilated observations in Fig. 4c around 1980), causing the errors and the ensemble spread of 20CRv3 to increase. However, this may also affect the performance of the benchmark ERA5, as there are fewer satellite and conventional upper-air observations available in the Southern Hemisphere in this time period than in more recent periods. This reduction in skill is consistent with Fig. 1 of (Hersbach et al. 2020), which shows a slight decrease in the range of 365-day mean anomaly correlations of Z500 forecasts for ERA5 in the Southern Hemisphere from about 1982–85 and considerable interannual variability in the quality of its Z500 fields before 1987.

c. *Mid-twentieth-century surface and upper-air comparisons*

Next, we compare surface and upper-air fields from 20CRv3 with the Japanese 55-year Reanalysis Project (JRA-55; Kobayashi et al. 2015). Figure 6 shows maps of local anomaly correlation between 20CRv3 and JRA-55 for SLP, Z500, and 300-hPa geopotential heights (Z300). Anomaly fields are computed for each reanalysis relative to its climatology over the stated time period

Unauthenticated | Downloaded 10/14/21 01:07 AM UTC

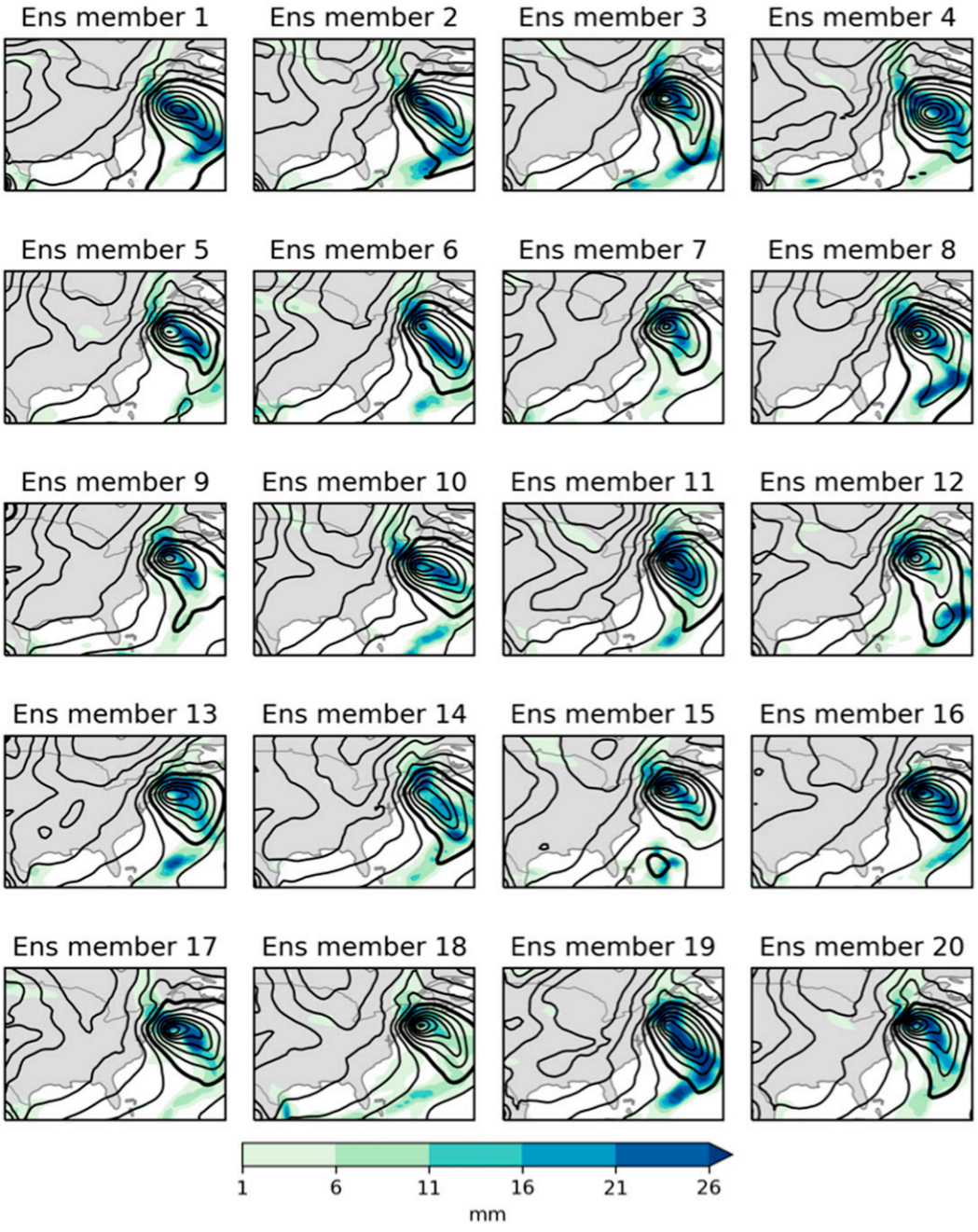


FIG. 2. Sea level pressure (contours; interval 5 hPa with the 1010-hPa contour thickened) and 6-h accumulated precipitation (shading) for the first 20 ensemble members of 20CRv3, 0000 UTC 13 Mar 1888.

and interpolated to a 1° × 1° grid; correlations in time are then computed for each grid point. Correlations are high in the Northern Hemisphere even before considerable satellite data are available for use in JRA-55 (1958–78; left column), though the extent of high correlation regions decreases higher in the atmosphere. While JRA-55 mainly assimilates conventional surface and upper-air data in this period, note that sparse satellite observations from the Vertical Temperature Profile Radiometer (VTPR) are included from 1973 to 1979 (see appendix A of Kobayashi et al. 2015). The correlation increases during the era of significant satellite observations (1979–2015; right column), particularly in the Southern Hemisphere, likely due to the lack of available observations in high southern latitudes prior to 1979 (Bromwich et al. 2007).

Stippled regions illustrate where 20CRv3 has low confidence. The high pattern correlation between the 20CRv3 confidence fields and the local anomaly correlation fields suggest that 20CRv3 uncertainty estimates are a good predictor of

Unauthenticated | Downloaded 10/14/21 01:07 AM UTC

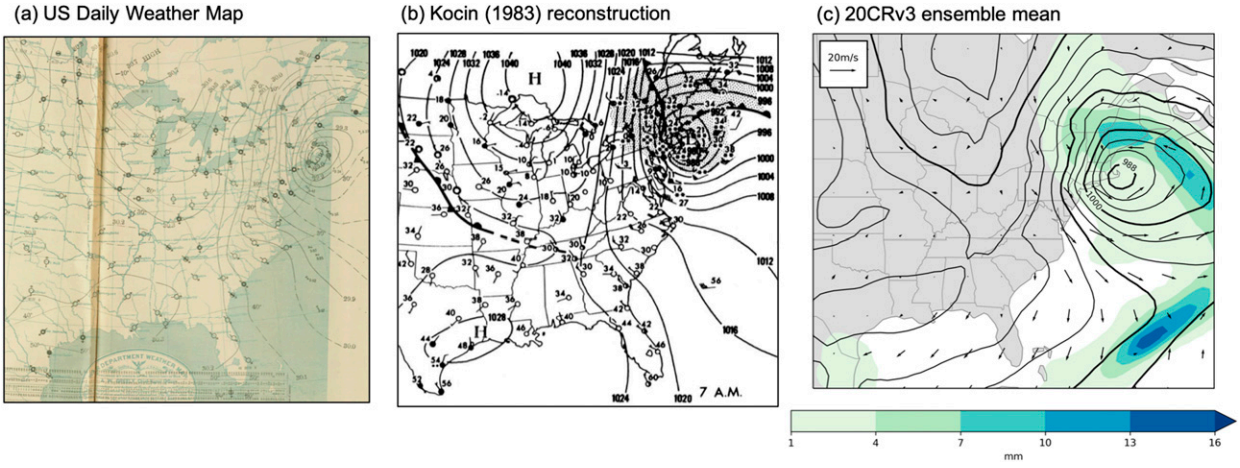


FIG. 3. Comparison of historical reconstructions valid at 1200 UTC 13 Mar 1888. (a) Real-time U.S. Daily Weather Map. (b) Reconstruction by Kocin (1983) showing fronts, surface observations, and area of reconstructed precipitation (shading). (c) 20CRv3 ensemble mean analysis of 6-h accumulated precipitation (shading), sea level pressure [contours; interval 4 hPa as in (b)], and 10-m wind (vectors).

skill relative to JRA-55, at the surface as well as at upper levels. Note that the low correlations in the high southern latitudes may be influenced by poor performance of JRA-55 before 1979 in the Southern Hemisphere (Kobayashi et al. 2015); in addition, several studies have shown that full-input reanalyses often disagree in this region, especially before 1979 (Bromwich et al. 2007; McDonald and Cairns 2020). While this does not provide evidence regarding which reanalysis is more accurate, the 20CRv3 confidence fields can provide guidance as to expected regions of disagreement. In particular, the 20CRv3 confidence fields can provide useful information independent of the verifying dataset: the region of low confidence in SLP for the period 1979–2015 in the eastern tropical Pacific (Fig. 6b) also corresponds to low correlations between 20CRv3 and ERA5 SLP (not shown). Indeed, the full field of SLP anomaly correlation between 20CRv3 and ERA5 (not shown) is quite similar to the correlation between 20CRv3 and JRA-55 (Fig. 6b).

d. Comparisons with independent upper-air observations

Upper-air skill prior to 1958 can be evaluated by comparisons with independent upper-air observations. We begin with observations at Lindenberg, Germany (52.22°N, 14.12°E), the station with the longest observational record available in the Integrated Global Radiosonde Archive, version 2 (IGRA2; Durre et al. 2016), from 1905 to the present. The observations are taken from a mix of kite, airplane, balloon, and radiosonde platforms (Adam and Dier 2005; Stickler et al. 2010). Even in the early twentieth century, the analyzed Z500 anomalies from both 20CRv3 and 20CRv2c correlate highly with observed anomalies (Fig. 7). Errors decrease in time for both systems, but RMSDs with 20CRv3 are consistently smaller than in 20CRv2c, and correlations with these observations are higher.

Figure 8 further demonstrates that 20CRv3 is able to predict its own skill by showing the RMSDs of analyzed Z500 anomalies with respect to the Lindenberg observations as a function of ensemble spread, for 20CRv3 and 20CRv2c. Only the latter two time periods considered in Fig. 7 are shown in Fig. 8, as it is unclear whether large errors in the early time period are due to instrumental errors in the observations or errors in the reanalysis, and the smaller number of observations does not allow for calculation of robust statistics. The gray diagonal line shows the expected RMSD for perfect observations, and the gray shaded area shows where the RMSD would be expected to fall for observations with error between 15 and 25 m (Wartenburger et al. 2013) assuming accurate ensemble spread. Thus, when the scatterplot lies above the gray swath, the reanalysis is overconfident (ensemble spread is too small); when the scatterplot lies below the gray swath, the reanalysis is underconfident (ensemble spread is too large). In the mid- to late twentieth century, 20CRv2c was nearly always overconfident in this field while 20CRv3 more accurately represents uncertainty (1952–80) or is underconfident (particularly in the more recent period from 1981 to 2001).

To investigate whether these results hold in other locations and at other levels in the latter half of the twentieth century, geopotential heights from eleven other stations from IGRA2 are analyzed at 300, 500, and 850 hPa. Station names, locations, and maximum available time periods that include complete metadata are listed in Table 2. The RMSD and correlation with respect to 20CRv3 for each station at each level are also included, as well as the local variability (as measured by the standard deviation of all observed anomalies). The RMSDs are consistently smaller than the local variability, and the correlations are generally high, though in both measures, the performance of 20CRv3 worsens in the tropics and at higher levels, as also seen in (Compo et al. 2011). Figure 9 demonstrates that these errors are often well predicted by the 20CRv3 ensemble spread, even at 300 hPa. Similar to Fig. 8, Fig. 9 shows RMSDs of 20CRv3 analyzed Z300, Z500, and Z850 anomalies with respect to the labeled IGRA station observations as a function of 20CRv3 ensemble spread, for level-dependent observation errors defined in (Wartenburger et al.

Unauthenticated | Downloaded 10/14/21 01:07 AM UTC

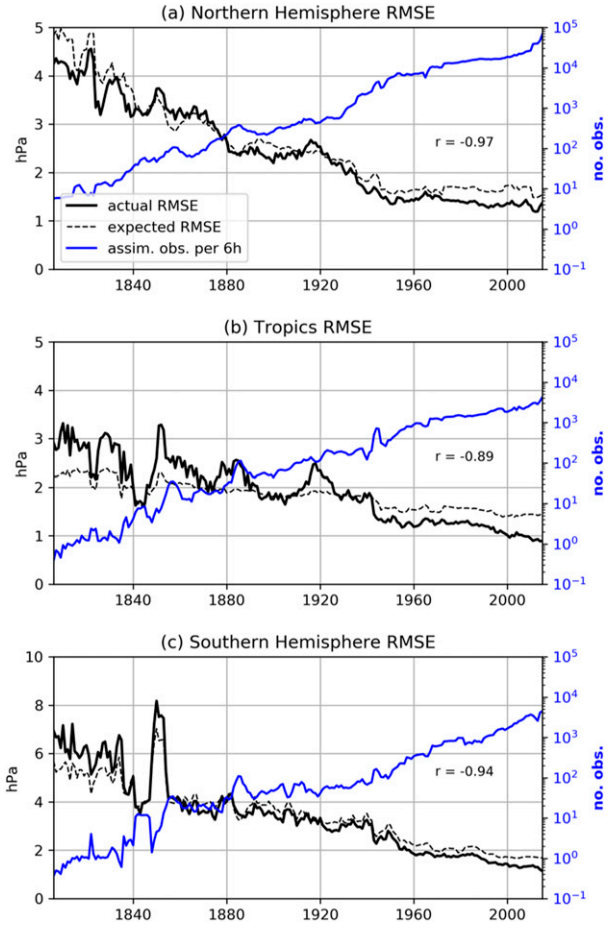


FIG. 4. Actual (solid black) and expected (dashed) annual first-guess RMSE of surface pressure averaged for (a) the Northern Hemisphere (20°–90°N), (b) the tropics (20°S–20°N), and (c) the Southern Hemisphere (90°–20°S). The annual average number of observations assimilated within a 6-h window is shown in blue (right-hand axis). The Pearson correlation (r) between the actual error and log of the number of observations is also given. Note that the left-hand y axis differs in (c).

2013). As in Fig. 8, there are several cases where 20CRv3 is over or underconfident. However, the 20CRv3 analyzed anomalies are interpolated to the instrument release time and location, and thus do not account for balloon drift, which is likely to influence the results (McGrath et al. 2006).

5. Climatic skill evaluation

The previous section illustrates the performance of 20CRv3 for synoptic variability. Here we evaluate the performance of long-term, climatic structures of mass, circulation, and precipitation.

a. Vertical structure of mass and circulation

To investigate possible systematic biases in 20CRv3 throughout the atmosphere, we compare its vertical profiles of long-term means of temperature and zonal wind with those from two full-input reanalyses, JRA-55 and ERA5, during the period 1979–2015 (Fig. 10). On average, 20CRv3 is warmer than both JRA-55 and ERA5 by about 1°–1.5°C from 600 to 300 hPa and below 900 hPa, pointing to a possible model bias in this version of the GFS (though this bias may be due to incorrectly tuned parameters in the model used at this resolution). Also note that midtroposphere (around 800–600 hPa) tropical biases, in both variables, are of opposite sign between the two full-input reanalyses, with magnitudes of about 0.5°C and 0.5 m s^{-1}. In particular, 20CRv3 is cooler than JRA-55 by less than 0.5°C between about 20°S and 20°N and from 800 to 600 hPa, but warmer than ERA5 in the same region, while ERA5 is cooler than JRA-55 by 0.5°–1.0°C (Fig. 10e); similar patterns hold in the same regions for zonal wind. In addition, note that the differences south of 60°S and below 600 hPa are larger between ERA5 and JRA-55 than between 20CRv3 and ERA5. In both zonal wind and temperature, 20CRv3 has much larger differences in the upper atmosphere than in the middle to lower atmosphere, possibly demonstrating the limitations of a system that assimilates only surface observations.

Relative to similar comparisons (see Fig. A1 of Compo et al. 2011) between the earlier version 2 of 20CR, ERA-40 (Uppala et al. 2005), and NCEP–NCAR Reanalysis 1 (NNR1; Kalnay et al. 1996; Kistler et al. 2001), the 20CRv3 upper-atmosphere differences in tropical and extratropical temperature appear to be larger, but the high-latitude differences in temperature are much smaller (Fig. 10). The structure of the zonal wind differences are different in Fig. 10 than shown by Compo et al. (2011), but the magnitudes of the differences are smaller between about 400 and 100 hPa. It is possible that the improved representation of the upper atmosphere in high latitudes will also correct systematic upper-level cold biases in vertical temperature stratification for the Arctic detected in previous versions of 20CR in the period 1934–40 (see Klaus et al. 2018, their Fig. 7). The possible causes of these differences between 20CRv3 and full-input reanalyses will be studied in future work.

To provide further context for the numbers in Fig. 10, Fig. 11 shows the global mean absolute biases (MAB) of long-term temperature and zonal wind between 20CRv3, JRA-55, ERA5, and the ensemble mean of a set of 50 nonassimilating atmospheric model simulations, denoted "AMIP." The latter were integrated using an older version of the GFS than 20CRv3 and are provided on a 1° × 1° grid with 64 vertical levels. They have prescribed SST and sea ice concentration fields (Hurrell et al. 2008), time-varying greenhouse gases based on observed data from 1979 to 2005 and RCP6 estimates thereafter (Meinshausen et al. 2011), and time-varying CMIP5 ozone (Cionni et al. 2011). Though the AMIP curves in Fig. 11 use a different version of the GFS and different SSTs than 20CRv3, they provide general magnitudes for biases one would expect from a model simulation relative to a full-input reanalysis. These curves demonstrate that the 20CRv3 biases relative to ERA5 and to JRA-55 are consistently smaller than the respective AMIP biases, with exceptions for surface temperature (where all the biases are between 1° and 1.5°C), and above 150 hPa in both variables. In addition, the magnitudes of the 20CRv3 biases relative to ERA5 in temperature are similar to or smaller than the magnitudes of the absolute differences between

Unauthenticated | Downloaded 10/14/21 01:07 AM UTC

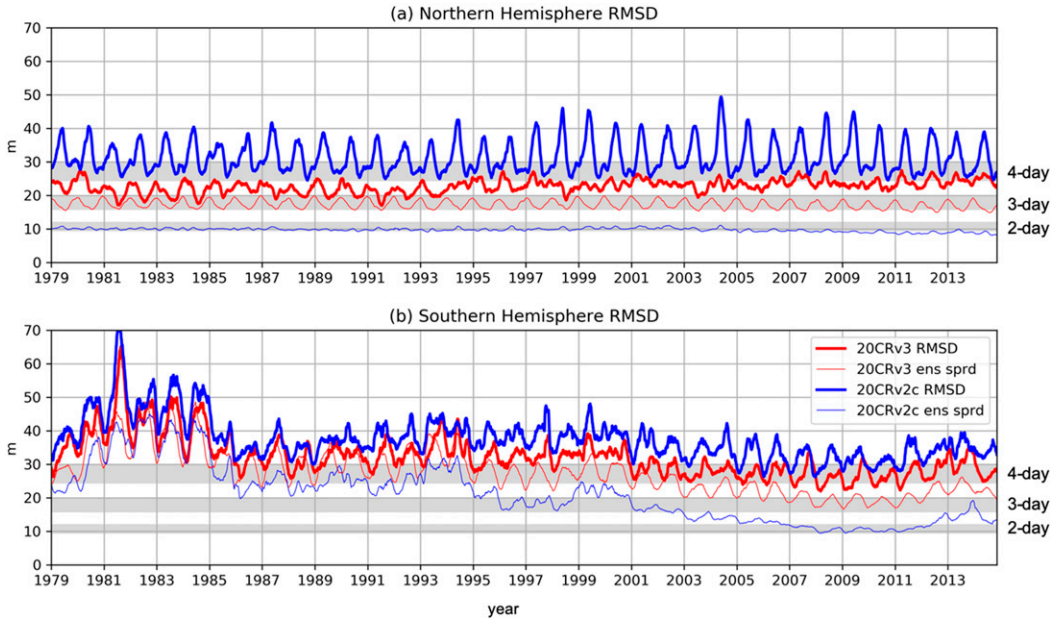


FIG. 5. The 500-hPa geopotential height RMSDs between ERA5 and 20CRv2c (thick blue) and 20CRv3 (thick red) ensemble mean fields for (a) the Northern Hemisphere (20°–80°N) and (b) the Southern Hemisphere (80°–20°S). Root-mean ensemble variance is shown in thin blue and red lines. The 2019 annual-average operational forecast errors for the given lead times are shown in gray swaths: the lower limit of each gray bar is the ECMWF forecast error, and the upper limit is the NCEP forecast error (https://www.emc.ncep.noaa.gov/gmb/STATS_vsdb/longterm/; accessed 29 Jan 2020). The 1981–2010 climatological variability in the 20CRv3 ensemble mean field is 79.24 m averaged over the Northern Hemisphere and 91.09 m averaged over the Southern Hemisphere.

JRA-55 and ERA5 below 700 hPa, and in zonal wind, are similar below 850 hPa.

b. *Precipitation structure*

Figures 12 and 13 illustrate the extent to which the precipitation fields in 20CRv3 could be used for long-term interannual to multidecadal variability and trend studies. Figure 12 shows the structure of the global precipitation field averaged for January 1979–2015 in 20CRv3 as well as differences between 20CRv3 and the satellite-station blend of the Global Precipitation Climatology Project version 2.3 (GPCP; Adler et al. 2003, 2018), and the station blend of the Climatic Research Unit Time Series version 4.03 (CRU TS; Harris et al. 2020; Harris and Jones 2020), over land (Fig. 12d). For reference, the differences between ERA5 and GPCP (Fig. 12c) and CRU TS (Fig. 12e) are also shown. Over land, 20CRv3 differences are largest in high-altitude regions, precipitation in most of North America is overestimated, and precipitation in Australia is underestimated. Relative to GPCP, 20CRv3 also overestimates high-precipitation regions in the tropics and the North Atlantic, and the tropical convection zones differ (see the brown band across the tropical oceans in Fig. 12b). However, over Europe, the Sahel, and much of Russia, the precipitation fields in 20CRv3 lie within the range of magnitudes estimated via GPCP and CRU TS. Relative to CRU TS, 20CRv3 is wetter over all of Europe and Asia and drier over the Sahel; relative to GPCP, though, 20CRv3 is drier over Russia, wetter over the Sahel, and has patchy differences over Europe.

Though the results in Fig. 12 point to likely biases in the precipitation field of 20CRv3, the interannual variability is captured remarkably well in some regions. Figure 13 shows time series of monthly (January and July) precipitation averages over land for 20CRv3, 20CRv2c, GPCP, and CRU TS in two regions: the western United States (WUS; 30°–50°N and 140°–100°W) and Southern Australia (SAus; 50°–26°S and 100°–160°E), shown in black boxes in Fig. 12a. Additionally, two regional station-based datasets are shown: PRISM for the western United States (PRISM Climate Group, Oregon State University, http://prism.oregonstate.edu, created 11 May 2020); and the gridded dataset developed for the Australian Water Availability Project (AWAP; Jones et al. 2009; http://www.bom.gov.au/cgi-bin/climate/change/timeseries.cgi, accessed 11 May 2020). In January in both regions, the correlations are remarkably high. While the correlations are lower in July, especially over the western United States, the correlations between the observational products and 20CRv3 are higher than the respective correlations with 20CRv2c in all cases considered here. The lower correlations over the western United States in July and Southern Australia in January are consistent with (Compo et al. 2006), who predicted lower skill of the reanalysis in summer months. Additionally, Gehne et al. (2016) demonstrate that several global precipitation datasets (including reanalyses and satellite-based products) disagree in their estimates of precipitation over North America, despite the relatively dense network of precipitation observations in this region. In general, estimating large-scale convection in models and full-input reanalyses is difficult (Gehne et al. 2016; Stephens et al. 2010; Trenberth et al. 2003, 2011), and gauge-based precipitation observations often suffer from systematic biases due to wind, evaporation, and snow undercatch, among others (Adam and Lettenmaier 2003; Peterson

Unauthenticated | Downloaded 10/14/21 01:07 AM UTC

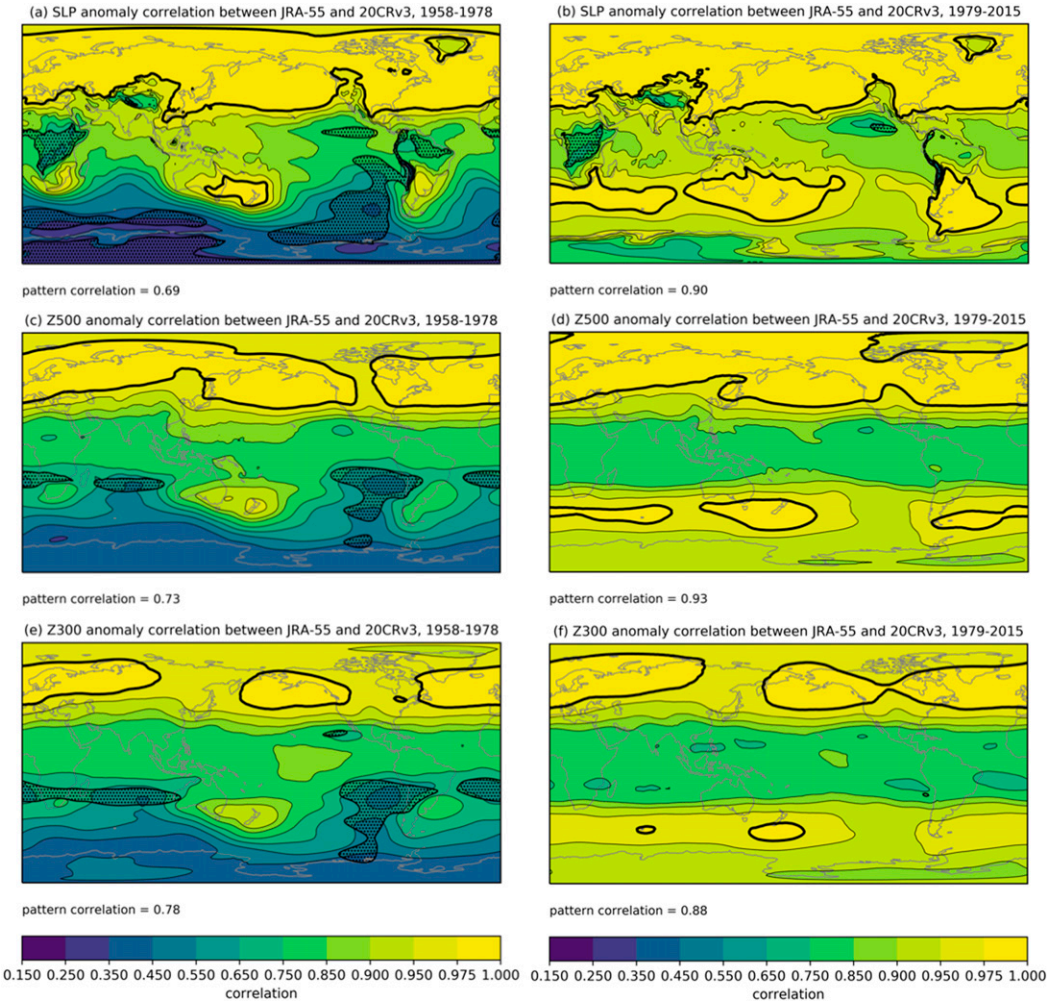


FIG. 6. Maps of local anomaly correlation between JRA-55 and 20CRv3 for (top) SLP, (middle) Z500, and (bottom) Z300 over the years (left) 1958–78 and (right) 1979–2015. Heavy black contours represent correlation values of 0.975; note the nonlinear color scale. Regions where the 20CRv3 confidence field [as calculated by Eq. (1) and averaged in time] for the given variable is less than 0.35 are stippled. The global average pattern correlation between the confidence field and the anomaly correlation field is given for each panel.

et al. 1998; Rasmussen et al. 2011; Sevruk et al. 2009). In particular, precipitation gauges tend to be in valleys, so CRU TS and PRISM may be underestimating in winter WUS due to snow undercatch. While this could explain the consistent difference between 20CRv3 and the observational products in January WUS, note that there is also a consistent difference in January SAus. Despite this, the agreement of interannual variability on a regional, monthly scale between 20CRv3 and three independent precipitation datasets is encouraging.

c. *Long-term variability in mass*

Figure 10 demonstrated the likely biases in air temperature; here we demonstrate how well 20CRv3 can represent multidecadal variability of a comparable mass variable, the mean temperature of the lower troposphere defined as the 500–1000-hPa atmospheric layer, relative to other full- and sparse-input reanalyses and to satellite measurement products. The mean layer temperature from each reanalysis is calculated using the hypsometric equation (Wallace and Hobbs 1977):

$$\overline{T} = (\mathrm{Z500} - \mathrm{Z1000}) \times \frac{\frac{g}{R_d}}{\ln\left(\frac{1000}{500}\right)}, \tag{4}$$

where Z500 and Z1000 are the 500- and 1000-hPa geopotential heights, g is gravitational acceleration, and R_d is the specific gas constant for dry air.

We include two satellite products in this comparison: the Remote Sensing Systems (RSS) Temperature Lower Troposphere calculated via weighted measurements from Microwave Sounding Units (MSU) and Advance Microwave Sounding Units (AMSU) (hereafter denoted RSS; Mears and Wentz 2009); and version 6 of the University of Alabama Huntsville's Mean Layer Temperature record derived from

Unauthenticated | Downloaded 10/14/21 01:07 AM UTC

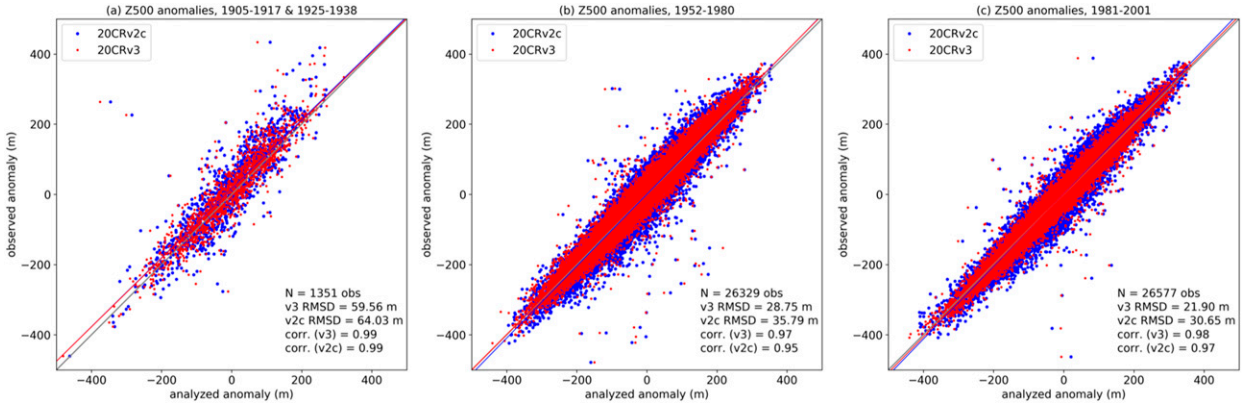


FIG. 7. The 500-hPa geopotential height analyzed anomalies from 20CRv2c (blue) and 20CRv3 (red) vs observed anomalies from upper-air measurements at Lindenberg, Germany, for the (a) early twentieth century (1905–17; 1925–38), (b) mid-twentieth century (1952–80), and (c) late twentieth century (1981–2001). Note that there are no available observations during the period 1918–24. Anomalies are calculated from 1973 to 2001 daily climatologies for the reanalyses and observations, respectively, with the annual cycle smoothed using a Fourier filter with three harmonics. RMSDs between observed and analyzed anomalies, Pearson correlations, and the number of observations for each panel are also given.

MSU and AMSU radiances (UAH; Christy et al. 2017). To calculate their temperature of the lower-troposphere (TLT) product, RSS calculates a weighted difference between single-channel measurements from MSU2 and AMSU5, while UAH calculates TLT as a linear combination of their calculated temperatures in the midtroposphere, tropopause, and lower stratosphere (Christy et al. 2017).

To consistently compare RSS and UAH to each other as well as to the reanalyses, both RSS and UAH are regressed onto the ERA5 $\overline{T}$ anomalies for 1979–2018. While the most accurate method of comparing a satellite-based product to a reanalysis would be to use a weighted vertical profile of the reanalysis temperatures, the vertical structures of the weights used for RSS and UAH differ; it is therefore impossible to consistently compare all datasets simultaneously. For this reason, instead of projecting the reanalysis data onto the two different satellite-based observation spaces, we instead project the satellite data into the "reanalysis space" of $\overline{T}$.

Figure 14 shows annual anomalies in $\overline{T}$ relative to 1981–2010 for the maximum available region over the globe, 70°S–82°N. By taking anomalies, the biases illustrated in Figs. 10 and 11 are effectively removed. Uncertainty swaths are included as twice the ensemble standard deviation for the ensemble products 20CRv2c, 20CRv3, CERA-20C, and RSS. Note that, for RSS, the central estimate is provided by version 4 of the data. The RSS uncertainty swath was calculated by applying the above regression to the anomaly of each member of the v3.3 ensemble, since there is currently no version 4 ensemble available. This ensemble is only available until 2015, so the RSS spread for 2016 onward is calculated as the average spread for 1998–2012. Finally, we find similar results using the levels 300–850 hPa (not shown), which are arguably closer to the "lower troposphere" as defined in the UAH and RSS weighting functions.

In the modern period, the decadal variability of atmospheric-layer temperature from 20CRv3 correlates well with that of other reanalyses and with satellite products. The

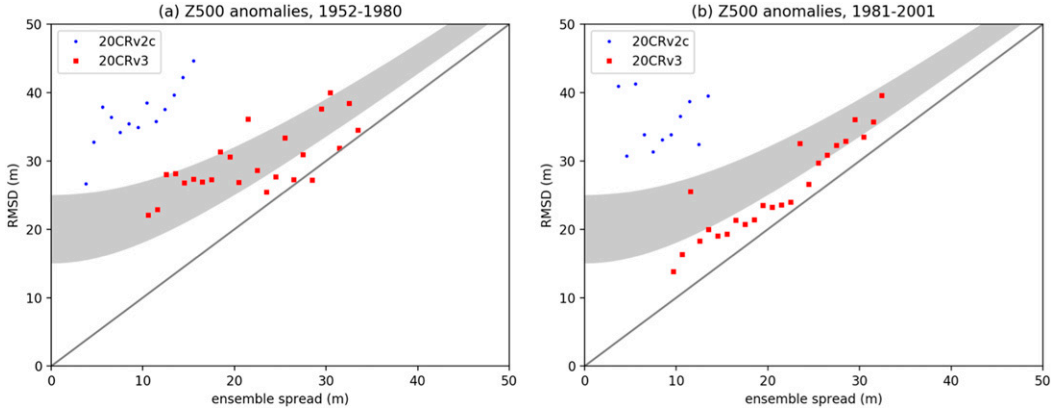


FIG. 8. RMSDs between analyzed Z500 anomalies and observed Z500 anomalies at Lindenberg, Germany, from 20CRv2c (blue) and 20CRv3 (red) as a function of 20CR ensemble spread (where bin width is 1 m). The gray line shows expected RMSD for perfect observations; gray shading shows region of expected RMSD for observation errors between 15 and 25 m. Only points with 30 observations or more are plotted.

Unauthenticated | Downloaded 10/14/21 01:07 AM UTC

TABLE 2. Station details and statistics for IGRA stations plotted in Fig. 9. RMSDs and correlations are calculated between the observed anomalies and the 20CRv3 anomalies. Local variability is calculated as the temporal standard deviation across all observations in the statistic.

Station	Lat (°)	Lon (°)	Years	No. of observations	RMSD (m)	Correlation	Local variability (m)
Z300							
Moscow	55.93	37.52	1956–2015	51 312	42.25	0.97	170.47
Vienna	48.25	16.36	1949–2015	47 035	39.09	0.96	146.64
Payerne	46.81	6.94	1943–2015	44 581	40.65	0.96	152.57
Buffalo	42.94	−78.72	1948–2015	50 625	43.52	0.97	167.03
Salt Lake	40.77	−111.96	1956–2015	43 658	38.2	0.96	136.59
Osan	37.08	127.03	1957–2015	71 787	48.48	0.92	120.56
Kowloon	22.33	114.17	1949–2015	42 032	30.01	0.67	39.83
Dakar	14.73	−17.5	1949–2015	34 647	32.69	0.51	36.77
Niamey	13.48	2.17	1950–2015	29 211	30.09	0.39	30.15
Darwin	−12.42	130.89	1943–2015	30 100	23.25	0.65	29.58
Sydney	−33.95	151.17	1976–2014	10 247	48.44	0.9	109.58
Z500							
Moscow	55.93	37.52	1956–2015	51 888	25.23	0.98	128.35
Vienna	48.25	16.36	1949–2015	47 955	22.73	0.98	103.73
Payerne	46.81	6.94	1943–2015	45 347	23.5	0.98	108.82
Buffalo	42.94	−78.72	1948–2015	51 556	24.47	0.98	116.03
Salt Lake	40.77	−111.96	1956–2015	40 353	23.05	0.97	90.83
Osan	37.08	127.03	1957–2015	73 371	29.12	0.93	80.89
Kowloon	22.33	114.17	1949–2015	42 380	18.88	0.76	28.81
Dakar	14.73	−17.5	1949–2015	35 248	18.89	0.64	24.17
Niamey	13.48	2.17	1950–2015	29 582	18.43	0.53	21.08
Darwin	−12.42	130.89	1943–2015	31 597	15.6	0.69	21.47
Sydney	−33.95	151.17	1976–2014	10 680	31.35	0.92	78.11
Z850							
Moscow	55.93	37.52	1956–2015	52 013	14.69	0.99	84.8
Vienna	48.25	16.36	1949–2015	47 973	13.37	0.98	64.13
Payerne	46.81	6.94	1943–2015	44 971	11.59	0.99	65.26
Buffalo	42.94	−78.72	1948–2015	51 880	10.74	0.98	59.75
Salt Lake	40.77	−111.96	1956–2015	43 970	13.83	0.96	51.17
Osan	37.08	127.03	1957–2015	73 024	15.56	0.93	40.86
Kowloon	22.33	114.17	1949–2015	41 725	9.51	0.92	24.53
Dakar	14.73	−17.5	1949–2015	36 126	10.56	0.74	15.68
Niamey	13.48	2.17	1950–2015	30 052	10.12	0.72	14.63
Darwin	−12.42	130.89	1945–2015	33 398	8.36	0.87	17.21
Sydney	−33.95	151.17	1976–2014	11 332	18.79	0.93	52.57

high correlation is not solely a result of trends: detrended correlations (Fig. 15 and Table 3) between 20CRv3 and the other datasets remain higher than 0.8, with many higher than 0.9. The relatively low correlation between 20CRv3 and ERA5 over land suggests that ERA5 may not represent interannual fluctuations of this variable over land as well as 20CRv3, relative to the RSS satellite estimate. While intriguing, though, these low correlations are not statistically significantly different from the respective (higher) correlations over the oceans. Zonal averages show similar correlations, though correlations are generally lower in the Southern Hemisphere (Table 3).

Not only does 20CRv3 correlate highly with upper-air reanalyses and satellite data products in the mid- to late twentieth century and early twenty-first century, it also has reasonable correlations (above 0.8) with CERA-20C in the early twentieth century (Fig. 15), though CERA-20C has a stronger warming trend throughout the twentieth and early twenty-first centuries. The variability in 20CRv3 and in 20CRv2c in the last quarter of the nineteenth century behave similarly, though they differ significantly from 1851 to 1870. This difference in behavior will be explored further in future work, but the confidence intervals for 20CRv2c are less reliable than those for 20CRv3; note the artificial jumps in the 20CRv2c ensemble spread in 1871 and 1951 (Fig. 14) corresponding to suboptimal parameter changes (see Slivinski et al. 2019a).

The warming trend in 20CRv3 during 1835–65 may be an effect of the slow recovery of the Earth system from multiple volcanic eruptions between 1808 and 1835 (Brönnimann et al. 2019b), and the subsequent cooling from 1875 to 1905 is intriguing, though several merged SST/land temperature datasets commonly used to describe temperature variability exhibit a similar trend (Hegerl et al. 2018; Vose et al. 2012; Wen et al. 2011). These trends also exist in the 20CRv3 time series restricted both to land only and ocean only (not shown). Additional work will be needed to determine the sensitivity of these trends to the specified SODAsi.3 SSTs, errors in the

Unauthenticated | Downloaded 10/14/21 01:07 AM UTC

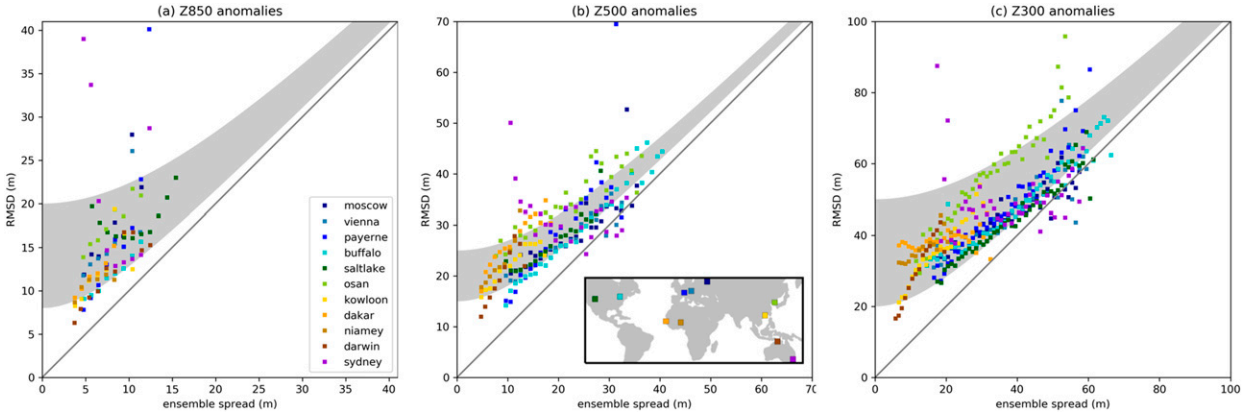


FIG. 9. As in Fig. 8, but for (a) Z850, (b) Z500, and (c) Z300 anomalies from 20CRv3 and the labeled IGRA stations. Observation errors are assumed to be between 8 and 20 m for Z850, between 15 and 25 m for Z500, and between 20 and 50 m for Z300. Locations of the stations are shown in map inset. Exact latitude, longitude, and time period for each station plotted, as well as other statistics, are shown in Table 2.

assimilated observations, and other factors. However, the overall high correlations between 20CRv3 and other products throughout the twentieth century suggest that the multidecadal variability seen in the full 210-yr span of 20CRv3 is representative of the true variability on these time scales.

6. Conclusions

The latest Twentieth Century Reanalysis version 3, has been evaluated on synoptic to climatic time scales via comparison with independent observations, satellite products, and full- and sparse-input reanalyses. Although the availability of comparable datasets decreases further back in time, the results shown here demonstrate that 20CRv3 can produce useful state estimates for its full time span and that its internal estimates of uncertainty are informative and reliable in the illustrated situations. In particular, 20CRv3 has smaller errors that are more consistent with its expected errors than 20CRv2c.

In the late twentieth and early twenty-first centuries, 20CRv3 represents several surface and upper-air fields and their variability well, relative to satellite products, reanalyses, and independent observations. Comparisons with a modern, full-input reanalysis suggest that the upper-air skill of 20CRv3 in the late twentieth and early twenty-first centuries is comparable to modern 3–4-day forecasts in both the Northern and Southern Hemispheres.

In the early to mid-twentieth century, 20CRv3 is evaluated via comparisons with independent surface and upper-air observations, and with the full-input reanalysis JRA-55 in the mid-twentieth century, which assimilates conventional surface, upper-air, and early satellite observations. The errors in 20CRv3 generally increase further back in time, but these errors are mainly predicted by the uncertainty in the 20CRv3 ensemble.

In the nineteenth century, 20CRv3 can reconstruct individual storms in the Northern Hemisphere and as in later time periods, internal background statistics demonstrate good agreement between the synoptic-scale errors in surface pressure and the prediction of those errors from the background ensemble spread. These results, along with the evaluations in the twentieth and early twenty-first centuries, suggest that 20CRv3 can provide reliable estimates of the illustrated atmospheric fields on synoptic scales, as well as their uncertainties, even in the nineteenth century.

Further, there is evidence that the interannual variability of mass and precipitation fields from 20CRv3 on climatic scales is also reliable. Comparisons with station-based precipitation datasets over regional and monthly scales indicate that 20CRv3 captures variability remarkably well for 1901–2015, while comparisons with a satellite-station blended product over 1979–2015 further support this assessment. Similarly, the large-scale mass variable of 500–1000-hPa layer temperature shows high correlations globally in the modern time period with three full-input reanalyses and two satellite data products. It is interesting to note the large differences in this variable between 20CRv2c and 20CRv3 prior to about 1875 in light of the analyses above, though previous work suggests that the confidence estimates of 20CRv2c in this time period are less reliable than those of 20CRv3. These results suggest that the interannual variability of many fields from 20CRv3 and their uncertainty estimates on global and regional scales may be reliable throughout the twentieth and twenty-first centuries, even for difficult-to-estimate variables like precipitation, but caution is still needed to interpret results in the nineteenth century. Further comparisons with early observations are required to more carefully evaluate 20CRv3 in this time period, when there are as yet no other comparable instrument-based reanalyses. Additional evaluation of 20CRv3 in the nineteenth century would allow us to place modern trends of these variables in a consistent, long-term context. Future work, currently in progress, will allow further quantification of these results, particularly on smaller regional scales.

While the overall performance of this dataset has improved over previous versions of 20CR, there are several remaining issues for potential users of the new 20CRv3 to keep in mind. First, although variability of 20CRv3 fields appears to be well

Unauthenticated | Downloaded 10/14/21 01:07 AM UTC

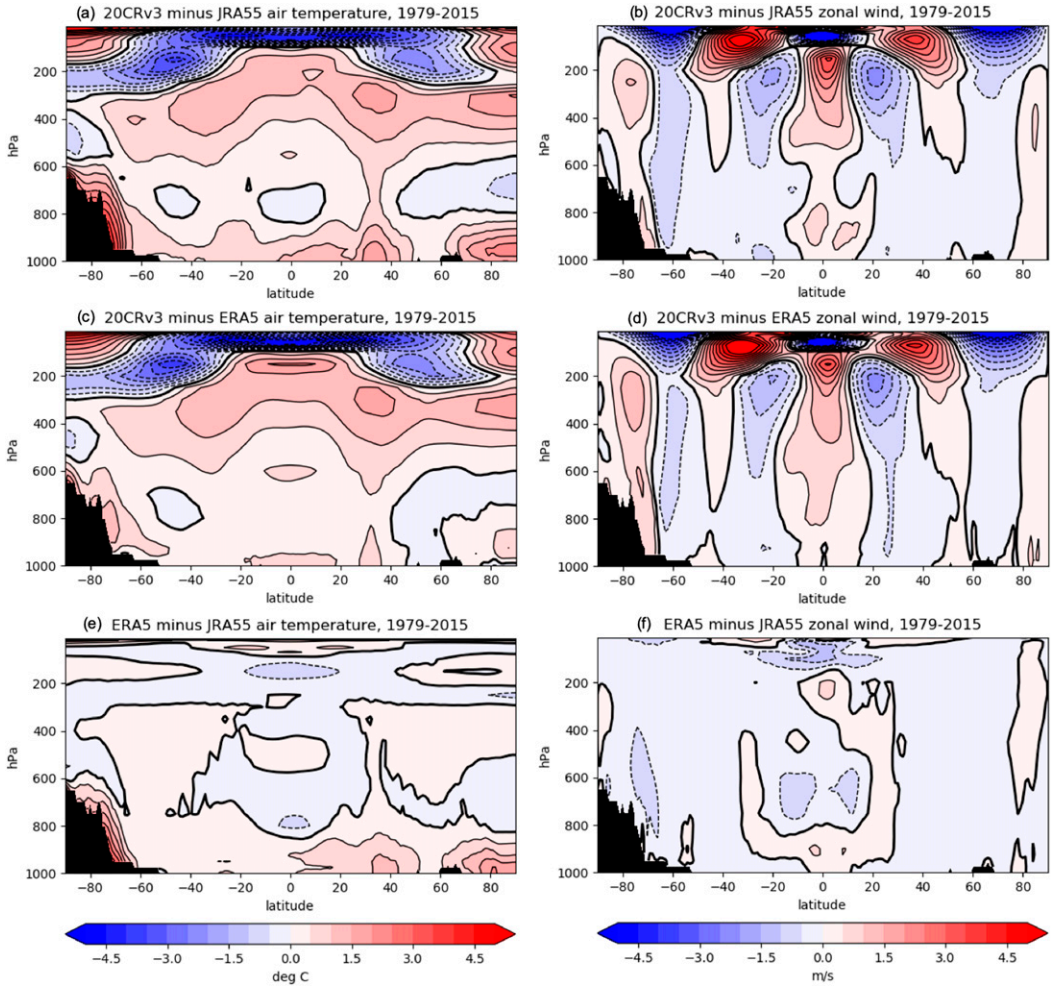


FIG. 10. (Left) Zonal mean temperature and (right) zonal mean zonal wind profile differences between 20CRv3 and (top) JRA-55 and (middle) ERA5 averaged over 1979–2015. (bottom) Differences between ERA5 and JRA-55. Locations where more than half of the grid points for that latitude and pressure level are subsurface are not shown. Contour intervals are 0.5°C and 0.5 m s^{-1}, respectively; the zero contours are bolded.

represented, there are still mean biases in temperature, wind, and precipitation, as well as in the location and orientation of the tropical convergence zones. Second, there are substantial biases in temperature and wind above 300 hPa, suggesting that a reanalysis that assimilates only surface pressure may not be adequate for upper-atmosphere studies. Finally, the Southern Hemisphere fields are generally less accurate than the Northern Hemisphere, and have less confidence. However, there is evidence that acquiring more observations, particularly by digitizing paper records in the eighteenth and nineteenth centuries, could have strong effects on the performance of future versions of 20CR.

In addition to further data rescue efforts, continued improvements to the 20CR system could provide much-needed information to support model development. For example, use of the 20CR system in sparsely populated regions, such as the Pacific Islands, could extend understanding of rainfall dynamics and past weather and climate patterns arising from convergence zones (Harvey et al. 2019; Lorrey et al. 2012). These regions are poorly represented in models but produce extreme conditions with strong impacts on society (Kruk et al. 2015; Lorrey and Fauchereau 2018) and would therefore benefit strongly from improved model representation.

Despite remaining issues, this evaluation of the first global, 3-hourly, instrument-based reanalysis spanning over 200 years has demonstrated that the 20CRv3 system can skillfully represent mass, circulation, and precipitation fields, on synoptic to climatic scales. Importantly, the ensemble-based approach also adds reliable estimates of uncertainty that can often predict this skill. Results demonstrate that 20CRv3 exhibits multidecadal variability in several fields, providing the opportunity to place recent trends in a longer historical context.

Acknowledgments. The NOAA–CIRES–DOE Twentieth Century Reanalysis Project version 3 used resources of the National Energy Research Scientific Computing Center (NERSC)

Unauthenticated | Downloaded 10/14/21 01:07 AM UTC

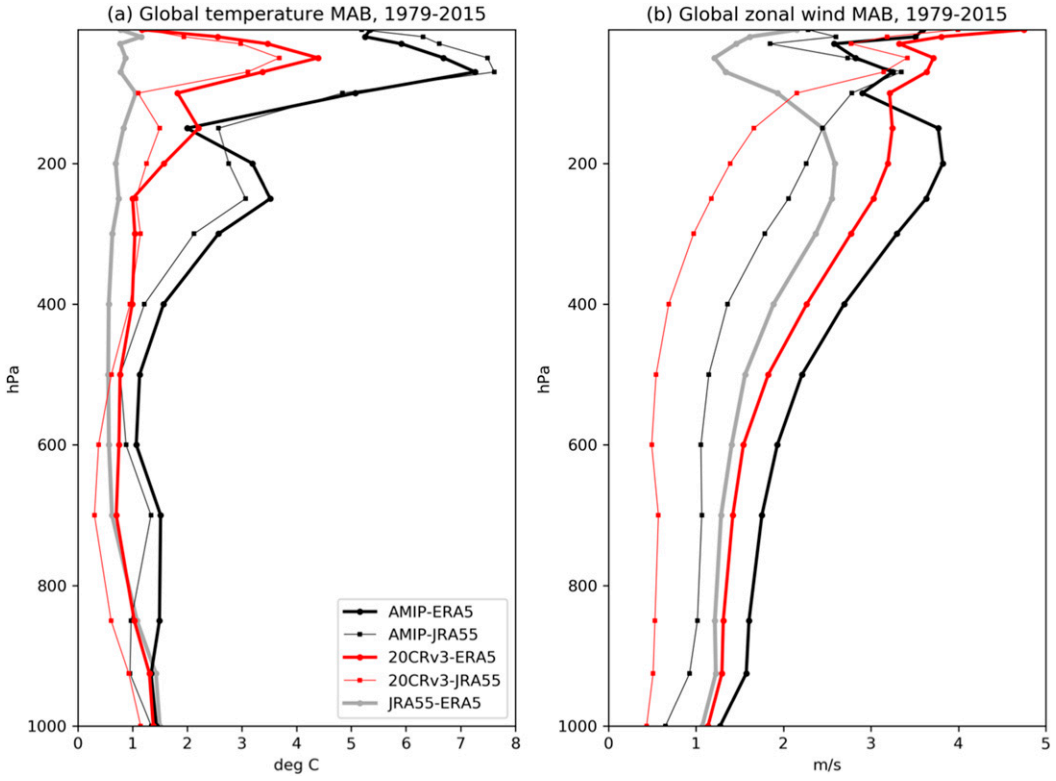


FIG. 11. Vertical profiles of global mean absolute biases for (a) temperature and (b) zonal wind of 20CRv3 (red) and the ensemble mean of a set of 50 nonassimilating atmospheric model simulations denoted "AMIP" (black). Biases relative to ERA5 are shown by thick curves, and biases relative to JRA-55 are shown by thin curves. The mean absolute difference between JRA-55 and ERA5 is shown in thick gray.

managed by Lawrence Berkeley National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract DE-AC02-05CH11231 and used resources of NOAA's Remotely Deployed High-Performance Computing Systems. Support for the Twentieth Century Reanalysis Project version 3 dataset is provided by the U.S. Department of Energy, Office of Science Biological and Environmental Research (BER), by the National Oceanic and Atmospheric Administration Climate Program Office, and by the NOAA Physical Sciences Laboratory.

The efforts of the NERSC consultants are acknowledged. The technical support of the IT group of the NOAA Physical Sciences Laboratory is acknowledged. Don Hooper of NOAA PSL and CIRES is especially acknowledged for his extensive work to make 20CRv3 freely available from PSL. Comments from Steve Penny (CIRES and NOAA) on an earlier version improved this manuscript. John Christy (UAH) is thanked for access to the UAH TLT fields and further discussions about comparing reanalyses to the UAH data. Carl Mears (RSS) is thanked for access to the RSS TLT fields and discussions about their uncertainty.

Collaborations with N. Rayner and H. Titchner of the Met Office in the development, production, and use of HadISST boundary conditions are gratefully acknowledged. Collaborations with B. Giese of Texas A&M University in the iterative development, production, and use of the SODAsi boundary conditions are gratefully acknowledged. The efforts of the National Center for Atmospheric Research (NCAR) Data Engineering and Curation section, especially C.-F. Shih are acknowledged. M. Benoy for the Citizen Science Unit of the Australian Meteorological Association, working with the Australian Bureau of Meteorology, is gratefully acknowledged for ongoing support. H. Mächel's (DWD) contribution of German climate observations is gratefully acknowledged. The authors would also like to thank the following individuals for their invaluable contributions of observations to the ISPD: L. Alexander (University of New South Wales), M. Barriendos (University of Barcelona), T. Brandsma (KNMI), Y. Brugnara (University of Bern), O. Bulygina (All-Russia Research Institute of Hydrometeorological Information), A. Dawson (University of Aberdeen), J. Filipiak (University of Gdansk), P. Groisman (NC State University Research Scholar at NOAA NCEI), J. Holopainen (University of Helsinki), D. Jones (Australian Bureau of Meteorology), T. Jonsson (Icelandic Met Office), S. Jourdain (Météo-France), A. Kaplan (Columbia University), F. Le Blancq (States of Jersey Meteorological Department), J. A. López (Instituto Nacional de Meteorología, Madrid), M. Maugeri (Università degli Studi di Milano), O. Mestre (Météo-France), A. Moberg (Stockholm University), GWK Moore (University of Toronto), Ø. Nordli (Norwegian

Unauthenticated | Downloaded 10/14/21 01:07 AM UTC

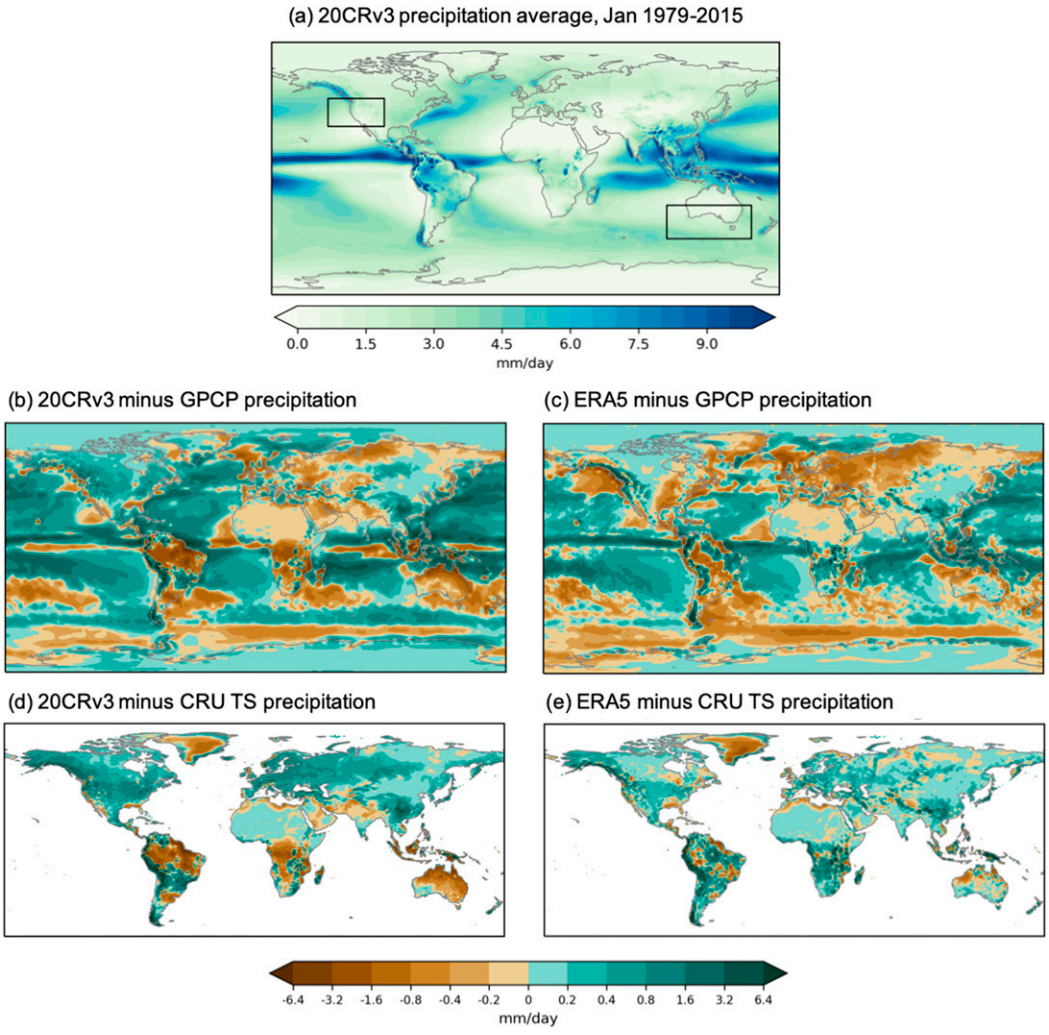


FIG. 12. Comparisons of daily precipitation rate averaged over January, 1979–2015. (a) The 20CRv3 ensemble mean precipitation rate. Black boxes define the WUS (30°–50°N and 140°–100°W) and SAus (50°–26°S and 100°–160°E) domains used for Fig. 13. (b) Difference between 20CRv3 and GPCP. (c) Difference between ERA5 and GPCP. (d) Difference between 20CRv3 and CRU TS. (e) Difference between ERA5 and CRU TS. Note the nonlinear color scale in (b)–(e).

Meteorological Institute), C. Reason (University of Cape Town), M. Rodwell (ECMWF), T. Schmith (Danish Meteorological Institute), L. Srnec (Croatian Meteorological and Hydrological Service), M. Tolstykh (Hydrometcentre of Russia), M. A. Valente (Universidade de Lisboa), N. Westcott (Midwestern Regional Climate Center), and P. Woodworth (National Oceanography Centre, United Kingdom). IBTrACS data are courtesy of K. Knapp (NOAA/NCEI). The authors thank R. Crouthamel (IEDRO) for his continuing support of the Twentieth Century Reanalysis.

R. Allan is supported by funding from the U.K. Newton Fund [which is managed by the U.K. Department for Business, Energy and Industrial Strategy (BEIS)], under its CSSP China and WCSSP South Africa projects, plus the EU Copernicus C3S Data Rescue Service. He also acknowledges the University of Southern Queensland, Toowoomba, Australia, and the Centre for Maritime Historical Studies, University of Exeter, Exeter, United Kingdom, where he is an Adjunct and Honorary Professor, respectively. S. Brönnimann was supported by the Swiss National Science Foundation (project 188701). M. Brunet was funded by the European Commission via the Copernicus Climate Change Service Contract 311a Lot1 for Collection and Processing of In Situ Observations–Data Rescue. J. Gergis was funded by Australian Research Council Project DE130100668 and the Australian National University's Futures Scheme. E. Hawkins was supported by the U.K. National Centre for Atmosphere Science. J. Kennedy was supported by the Met Office Hadley Centre Climate Programme funded by BEIS and DEFRA. H. Kubota was supported by Grant-in-Aid for Scientific Research (20K20328, 19H00562) funded by the Japan Society for the Promotion of Science (JSPS). A. M. Lorrey was supported by the NIWA Strategic Science Investment Fund project ''Climate Present

Unauthenticated | Downloaded 10/14/21 01:07 AM UTC

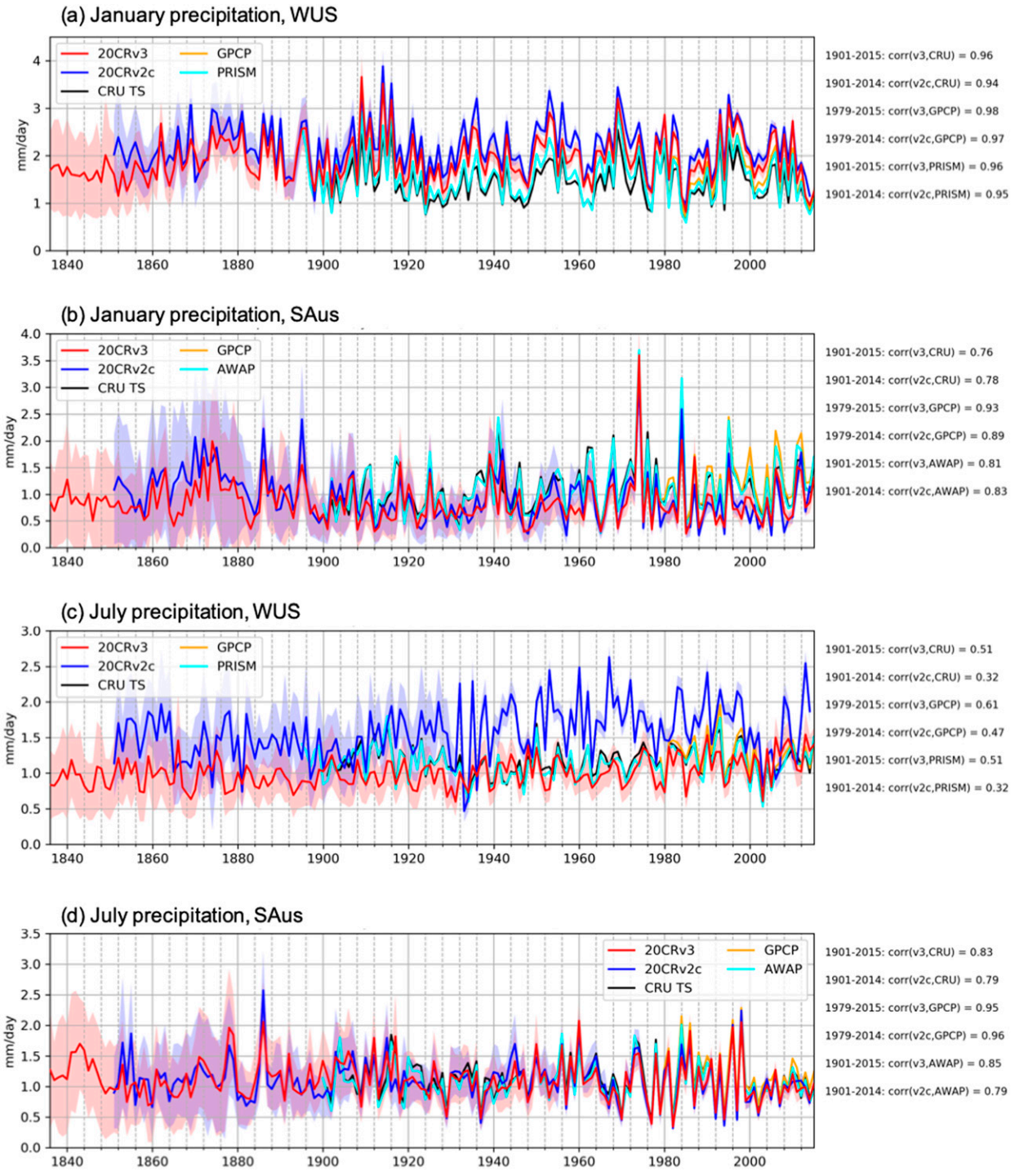


FIG. 13. Regionally averaged, land-only precipitation rate for (a) January in WUS (30°–50°N and 140°–100°W), (b) January in SAus (50°–26°S and 100°–160°E), (c) July in WUS, and (d) July in SAus. The 20CRv3 ensemble mean is shown in red, 20CRv2c ensemble mean in blue, CRU TS in black, GPCP in gold, and PRISM/AWAP in cyan (depending on region). The 2σ confidence intervals (calculated from the ensemble standard deviation) are shaded for 20CRv2c and 20CRv3. Pearson correlations for comparable time periods included.

and Past'' Contract CAOA2001. The research work of R. Przybylak and P. Wyszyński was supported by the National Science Centre, Poland (Grants DEC-2012/07/B/ST10/04002 and 2015/19/B/ST10/02933). The Justus Liebig University of Giessen, Germany is thanked for financial support to digitise, quality control, and analyse early instrumental meteorological data across the world. The following people are financially supported by the University

Unauthenticated | Downloaded 10/14/21 01:07 AM UTC

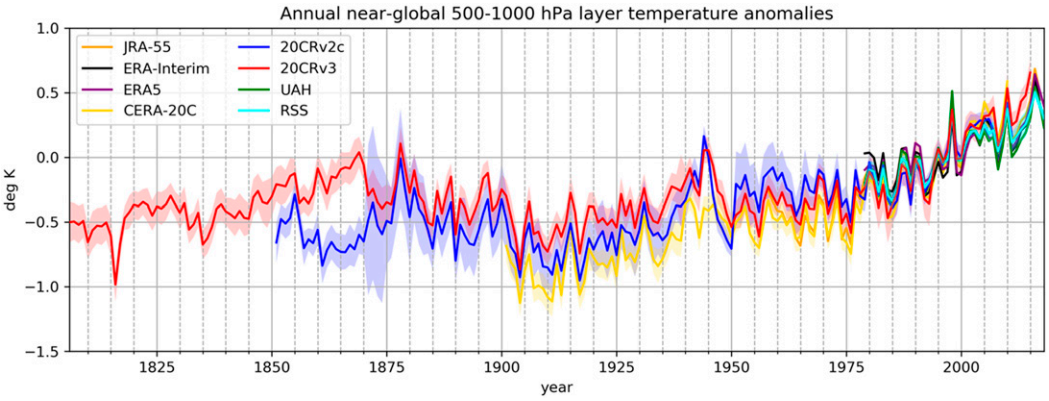


FIG. 14. Near-global (70°S to 82°N) average 500–1000-hPa layer temperature $\overline{T}$ anomalies with respect to each dataset's 1981–2010 climatology, estimated from geopotential heights in all reanalyses. The confidence intervals are calculated as twice the ensemble standard deviation and shown for 20CRv2c, 20CRv3, CERA-20C, and RSS; see text for description of issues in 20CRv2c uncertainty. UAH and RSS temperature in the lower-troposphere products were each regressed onto ERA5 for the available 1979–2018 period.

of Giessen and digitised subdaily pressure data: L. Dergianli, G. Kelly, D. Xoplaki, V. Iakovoglou, E. Kaimasidou, E. Tsalkitzidou, M. Athanasiou, L. Behr, A. Megalou, C. Chandolia, E. Fleitmann, P. Zafeiropoulou, N. Kong, M. Ostheimer, S. Dafka, K. Pometti, N. Gouta, P. Katsaouni, Z. Theorchari, A. Theocharis, F. Williams, J. Schermuly, C. Mett, L. Theile, J. Zuckermann, P. Strehlau, C. Samaras, A. Tsikerdekis, C. Athanasiou, J. Braun, T. Sperzel, J. Damster, N. Luther, M. Miltscheff-Petroff, M. Kelbling, D. Griechbaum, M. Hänsgen, and J. Viezens. L. Dergianli and E. Xoplaki are acknowledged for QC, coordination, management, and preparation of all University of Giessen data. M. Brunet, J. Luterbacher, R. J. Allan, G. P. Compo, P. Jones, S. Brönnimann, and A. Lorrey acknowledge Package 3 of the Copernicus Climate Change Service 311a Lot1 for Collection and Processing of In Situ Observations Data Rescue. J. Luterbacher acknowledges Climate Science for Service Partnership China Project (CSSP).

Use of the U.S. Daily Weather Map is courtesy of NOAA Central Library Data Imaging Project. The 20CR homepage is located at https://go.usa.gov/XTd. The scientific results and conclusions, as well as any views of opinions expressed herein, are those of the authors and do not necessarily reflect the views of the University of Colorado, NOAA, the Department of Commerce, or any other organization associated with this work.

Data availability statement. Data from the Twentieth Century Reanalysis Project are freely accessible. Ensemble statistics for many variables are available in netCDF4 format from the NOAA Physical Sciences Laboratory at https://psl.noaa.gov/data/gridded/data.20thC_ReanV3.html and NCAR's Research Data Archive (RDA) at https://doi.org/10.5065/H93G-WS83 (Slivinski et al. 2019b). Individual ensemble members are available from NERSC at https://portal.nersc.gov/project/20C_Reanalysis/. All observations contained in ISPDv4.7 are open and unrestricted, and the ISPDv4.7 dataset is available from NCAR's RDA: https://doi.org/10.5065/9EYR-TY90 (Compo et al. 2019). RSS upper-air temperature data is available from Remote Sensing Systems at http://www.remss.com/support/data-shortcut/. UAH upper-air temperature data are available from NCEI at https://doi.org/10.7289/V5MC8X31. GPCP and CRU TS data have been provided by the NOAA/OAR/PSL, Boulder, Colorado, from their website at https://www.psl.noaa.gov/. The GFS-AMIP data are available from the PSL Facility for Weather and Climate Assessments (FACTS) website at https://psl.noaa.gov/repository/facts/ (Murray et al. 2020). The ERA-Interim, ERA5, and CERA-20C datasets are courtesy of ECMWF.

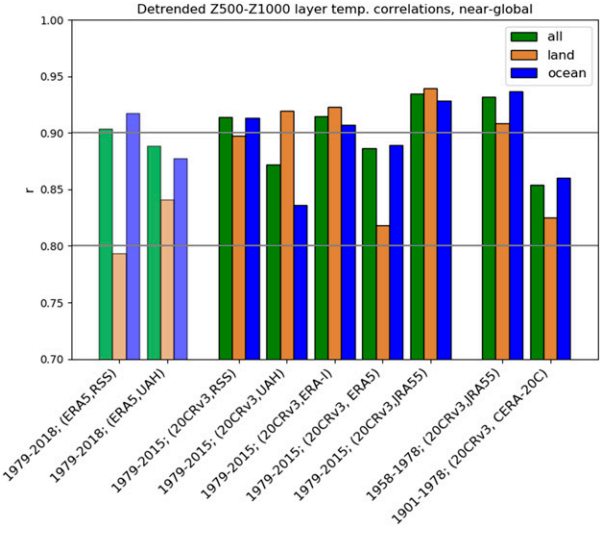


FIG. 15. Correlations between near-global (70°S–82°N) detrended 500–1000-hPa layer temperatures $\overline{T}$ for the given time periods and land/sea masks. Correlations between ERA5, RSS, and UAH are shown on the left in lighter colors for clarity.

Unauthenticated | Downloaded 10/14/21 01:07 AM UTC

TABLE 3. Correlations between zonally averaged detrended 500–1000-hPa layer temperatures $\overline{T}$, calculated over the given time window and land/sea masks. "NH" denotes 30°–82°N, "tropics" denotes 30°S–30°N, and "SH" denotes 70°–30°S. Correlation values above 0.9 are bolded.

Area	1979–2018: (ERA5, RSS)	1979–2018: (ERA5, UAH)	1979–2015: (20CRv3, RSS)	1979–2015: (20CRv3, UAH)	1979–2015: (20CRv3, ERA-I)	1979–2015: (20CRv3, ERA5)	1979–2015: (20CRv3, JRA-55)	1958–1978: (20CRv3, JRA-55)	1901–1978: (20CRv3, CERA-20C)
NH									
All	0.88	**0.92**	0.84	0.89	0.86	0.85	**0.90**	0.89	0.67
Land	0.80	**0.90**	0.87	**0.92**	0.89	0.80	**0.92**	**0.91**	0.71
Ocean	0.79	0.82	0.84	0.86	0.85	0.74	0.88	0.86	0.66
Tropics									
All	**0.94**	**0.90**	**0.93**	0.89	**0.90**	0.88	**0.92**	**0.96**	0.84
Land	0.89	0.84	**0.91**	**0.90**	**0.90**	0.83	**0.92**	**0.92**	0.82
Ocean	**0.94**	**0.90**	**0.92**	0.89	**0.90**	0.88	**0.92**	**0.97**	0.85
SH									
All	0.88	0.81	0.87	0.64	0.85	0.83	0.84	0.64	0.65
Land	0.48	0.53	0.83	0.81	0.81	0.64	0.86	0.61	0.50
Ocean	0.88	0.80	0.88	0.63	0.85	0.83	0.84	0.66	0.66

REFERENCES

Adam, J. C., and D. P. Lettenmaier, 2003: Adjustment of global gridded precipitation for systematic bias. *J. Geophys. Res.*, **108**, 4257, https://doi.org/10.1029/2002JD002499.

Adam, W. K., and H. Dier, 2005: Lange Messreihen zur Wetter- und Klimaforschung am Meteorologischen Observatorium Lindenberg. *Promet*, **31**, 159–170.

Adler, R. F., and Coauthors, 2003: The Version-2 Global Precipitation Climatology project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeor.*, **4**, 1147–1167, https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2.

——, and Coauthors, 2018: The Global Precipitation Climatology Project (GPCP) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere*, **9**, 138, https://doi.org/10.3390/atmos9040138.

Allan, R., P. Brohan, G. P. Compo, R. Stone, J. Luterbacher, and S. Brönnimann, 2011: The international Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative. *Bull. Amer. Meteor. Soc.*, **92**, 1421–1425, https://doi.org/10.1175/2011BAMS3218.1.

Bengtsson, L., S. Hagemann, and K. I. Hodges, 2004: Can climate trends be calculated from reanalysis data? *J. Geophys. Res.*, **109**, D11111, https://doi.org/10.1029/2004JD004536.

Bosilovich, M. G., F. R. Robertson, and J. Chen, 2011: Global energy and water budgets in MERRA. *J. Climate*, **24**, 5721–5739, https://doi.org/10.1175/2011JCLI4175.1.

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**, D10111, https://doi.org/10.1029/2006JD007859.

Brönnimann, S., and Coauthors, 2013: Historical weather extremes in the "Twentieth Century Reanalysis." *Weather Extremes During the Past 140 Years*, S. Brönnimann and O. Martius, Eds., Geographica Bernensia, 7–17, https://doi.org/10.4480/GB2013.G89.01.

——, and Coauthors, 2019a: Unlocking pre-1850 instrumental meteorological records: A global inventory. *Bull. Amer. Meteor. Soc.*, **100** (12), ES389–ES413, https://doi.org/10.1175/BAMS-D-19-0040.1.

——, and Coauthors, 2019b: Last phase of the Little Ice Age forced by volcanic eruptions. *Nat. Geosci.*, **12**, 650–656, https://doi.org/10.1038/s41561-019-0402-y.

Brugnara, Y., and Coauthors, 2015: A collection of sub-daily pressure and temperature observations for the early instrumental period with a focus on the "year without a summer" 1816. *Climate Past*, **11**, 1027–1041, https://doi.org/10.5194/CP-11-1027-2015.

Christy, J. R., and Coauthors, 2017: NOAA Climate Data Record (CDR) of MSU and AMSU-A mean layer temperatures, UAH version 6.0 TLT. NOAA National Centers for Environmental Information, accessed 1 November 2020, https://doi.org/10.7289/V5MC8X31.

Cionni, I., and Coauthors, 2011: Ozone database in support of CMIP5 simulations: Results and corresponding radiative forcing. *Atmos. Chem. Phys.*, **11**, 11 267–11 292, https://doi.org/10.5194/acp-11-11267-2011.

Coddington, O., J. L. Lean, P. Pilewskie, M. Snow, and D. Lindholm, 2016: A solar irradiance climate data record. *Bull. Amer. Meteor. Soc.*, **97**, 1265–1282, https://doi.org/10.1175/BAMS-D-14-00265.1.

Compo, G. P., J. S. Whitaker, and P. D. Sardeshmukh, 2006: Feasibility of a 100-year reanalysis using only surface pressure data. *Bull. Amer. Meteor. Soc.*, **87**, 175–190, https://doi.org/10.1175/BAMS-87-2-175.

——, and Coauthors, 2011: The Twentieth Century Reanalysis Project. *Quart. J. Roy. Meteor. Soc.*, **137**, 1–28, https://doi.org/10.1002/qj.776.

——, and Coauthors, 2019: The International Surface Pressure Databank version 4. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, accessed 1 November 2020, https://doi.org/10.5065/9EYR-TY90.

Cram, T. A., and Coauthors, 2015: The International Surface Pressure Databank version 2. *Geosci. Data J.*, **2**, 31–46, https://doi.org/10.1002/gdj3.25.

Crowley, T. J., and M. B. Unterman, 2013: Technical details concerning development of a 1200 yr proxy index for global volcanism. *Earth Syst. Sci. Data*, **5**, 187–197, https://doi.org/10.5194/essd-5-187-2013.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, https://doi.org/10.1002/qj.828.

Unauthenticated | Downloaded 10/14/21 01:07 AM UTC

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131**, 3385–3396, https://doi.org/10.1256/qj.05.108.

Durre, I., X. Yin, R. S. Vose, S. Applequist, and J. Arnfield, 2016: Integrated Global Radiosonde Archive (IGRA), version 2. NOAA National Centers for Environmental Information, accessed 16 October 2020, https://doi.org/10.7289/V5X63K0Q.

Fujiwara, M., and Coauthors, 2017: Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems. *Atmos. Chem. Phys.*, **17**, 1417–1452, https://doi.org/10.5194/acp-17-1417-2017.

Gehne, M., T. M. Hamill, G. N. Kiladis, and K. E. Trenberth, 2016: Comparison of global precipitation estimates across a range of temporal and spatial scales. *J. Climate*, **29**, 7773–7795, https://doi.org/10.1175/JCLI-D-15-0618.1.

Gergis, J., L. Ashcroft, and P. Whetton, 2020: A historical perspective on Australian temperature extremes. *Climate Dyn.*, **55**, 843–868, https://doi.org/10.1007/s00382-020-05298-z.

Giese, B. S., H. F. Seidel, G. P. Compo, and P. D. Sardeshmukh, 2016: An ensemble of ocean reanalyses for 1815–2013 with sparse observational input. *J. Geophys. Res. Oceans*, **121**, 6891–6910, https://doi.org/10.1002/2016JC012079.

Harris, I. C., and P. D. Jones, 2020: CRU TS4.03: Climatic Research Unit (CRU) Time-Series (TS) version 4.03 of high-resolution gridded data of month-by-month variation in climate (Jan. 1901-Dec. 2018). Centre for Environmental Data Analysis, accessed 1 May 2020, https://doi.org/10.5285/10d3e3640f004c578403419aac167d82.

——, T. J. Osborn, P. Jones, and D. Lister, 2020: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci. Data*, **7**, 109, https://doi.org/10.1038/s41597-020-0453-3.

Harvey, T., J. A. Renwick, A. M. Lorrey, and A. Ngari, 2019: The representation of the South Pacific convergence zone in the Twentieth Century Reanalysis. *Mon. Wea. Rev.*, **147**, 841–851, https://doi.org/10.1175/MWR-D-18-0237.1.

Hegerl, G. C., S. Brönnimann, A. Schurer, and T. Cowan, 2018: The early 20th century warming: Anomalies, causes, and consequences. *Wiley Interdiscip. Rev.: Climate Change*, **9**, e522, https://doi.org/10.1002/wcc.522.

Hersbach, H., and D. J. E. N. Dee, 2016: ERA5 reanalysis is in production. *ECMWF Newsletter*, No. 147, ECMWF, Reading, United Kingdom, 7, https://www.ecmwf.int/sites/default/files/elibrary/2016/16299-newsletter-no147-spring-2016.pdf.

——, and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hurrell, J. W., J. J. Hack, D. Shea, J. M. Caron, and J. Rosinski, 2008: A new sea surface temperature and sea ice boundary dataset for the Community Atmosphere Model. *J. Climate*, **21**, 5145–5153, https://doi.org/10.1175/2008JCLI2292.1.

Jones, D. A., W. Wang, and R. Fawcett, 2009: High-quality spatial climate data-sets for Australia. *Aust. Meteor. Oceanogr. J.*, **58**, 233–248, https://doi.org/10.22499/2.5804.003.

Jones, P. D., M. New, D. E. Parker, S. Martin, and I. G. Rigor, 1999: Surface air temperature and its changes over the past 150 years. *Rev. Geophys.*, **37**, 173–199, https://doi.org/10.1029/1999RG900002.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kinter, J. L., III, M. J. Fennessy, V. Krishnamurthy, and L. Marx, 2004: An evaluation of the apparent interdecadal shift in the tropical divergent circulation in the NCEP–NCAR reanalysis. *J. Climate*, **17**, 349–361, https://doi.org/10.1175/1520-0442(2004)017<0349:AEOTAI>2.0.CO;2.

Kistler, R., and Coauthors, 2001: The NCEP–NCAR 50-Year Reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.*, **82**, 247–267, https://doi.org/10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2.

Klaus, D., P. Wyszyński, K. Dethloff, R. Przybylak, and A. Rinke, 2018: Evaluation of 20CR reanalysis data and model results based on historical (1930–1940) observations from Franz Josef Land. *Pol. Polar Res.*, **39**, 225–254, https://doi.org/10.24425/118747.

Kobayashi, S., and Coauthors, 2015: The JRA-55 reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48, https://doi.org/10.2151/jmsj.2015-001.

Kocin, P. J., 1983: An analysis of the "Blizzard of '88." *Bull. Amer. Meteor. Soc.*, **64**, 1258–1272, https://doi.org/10.1175/1520-0477(1983)064<1258:AAOTO>2.0.CO;2.

——, 1988: Meteorological analyses of the March 1888 "Blizzard of '88." *Eos, Trans. Amer. Geophys. Union*, **69**, 137–147, https://doi.org/10.1029/88EO00088.

——, and L. W. Uccellini, 2004: *The Cases*. Vol. II, *Northeast Snowstorms*, *Meteor. Monogr.*, No. 54, Amer. Meteor. Soc., 297–811.

Kruk, M. C., A. M. Lorrey, G. M. Griffiths, M. Lander, E. J. Gibney, H. J. Diamond, and J. J. Marra, 2015: On the state of the knowledge of rainfall extremes in the western and northern Pacific basin. *Int. J. Climatol.*, **35**, 321–336, https://doi.org/10.1002/joc.3990.

Laloyaux, P., and Coauthors, 2018: CERA-20C: A coupled reanalysis of the twentieth century. *J. Adv. Model. Earth Syst.*, **10**, 1172–1195, https://doi.org/10.1029/2018MS001273.

Lorrey, A. M., and P. R. Chappell, 2016: The "dirty weather" diaries of Reverend Richard Davis: Insights about early colonial-era meteorology and climate variability for northern New Zealand, 1839–1851. *Climate Past*, **12**, 553–573, https://doi.org/10.5194/cp-12-553-2016.

——, and N. C. Fauchereau, 2018: Southwest Pacific atmospheric weather regimes: Linkages to ENSO and extra-tropical teleconnections. *Int. J. Climatol.*, **38**, 1893–1909, https://doi.org/10.1002/joc.5304.

——, G. Dalu, J. Renwick, H. Diamond, and M. Gaetani, 2012: Reconstructing the South Pacific convergence zone position during the presatellite era: A La Niña case study. *Mon. Wea. Rev.*, **140**, 3653–3668, https://doi.org/10.1175/MWR-D-11-00228.1.

McDonald, A. J., and L. H. Cairns, 2020: A new method to evaluate reanalyses using synoptic patterns: An example application in the Ross Sea/Ross Ice Shelf Region. *Earth Space Sci.*, **7**, e2019EA000794, https://doi.org/10.1029/2019EA000794.

McGrath, R., T. Semmler, C. Sweeney, and S. Wang, 2006: Impact of balloon drift errors in radiosonde data on climate statistics. *J. Climate*, **19**, 3430–3442, https://doi.org/10.1175/JCLI3804.1.

Mears, C. A., and F. J. Wentz, 2009: Construction of the remote sensing systems V3.2 atmospheric temperature records from the MSU and AMSU microwave sounders. *J. Atmos. Oceanic Technol.*, **26**, 1040–1056, https://doi.org/10.1175/2008JTECHA1176.1.

Meinshausen, M., and Coauthors, 2011: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, **109**, 213–241, https://doi.org/10.1007/s10584-011-0156-z.

Unauthenticated | Downloaded 10/14/21 01:07 AM UTC

Michaelis, A. C., and G. M. Lackmann, 2013: Numerical modeling of a historic storm: Simulating the Blizzard of 1888. *Geophys. Res. Lett.*, **40**, 4092–4097, https://doi.org/10.1002/grl.50750.

Moore, G. W. K., and M. Babij, 2017: Iceland's Great Frost Winter of 1917/1918 and its representation in reanalyses of the twentieth century. *Quart. J. Roy. Meteor. Soc.*, **143**, 508–520, https://doi.org/10.1002/qj.2939.

Murray, D., and Coauthors, 2020: Facility for Weather and Climate Assessments (FACTS): A community resource for assessing weather and climate variability. *Bull. Amer. Meteor. Soc.*, **101**, E1214–E1224, https://doi.org/10.1175/BAMS-D-19-0224.1.

Noone, S., and Coauthors, 2021: Progress towards a holistic land and marine surface meteorological database and a call for additional contributions. *Geosci Data J.*, https://doi.org/10.1002/gdj3.109, in press.

Parker, D. E., M. Gordon, D. P. N. Cullum, D. M. H. Sexton, C. K. Folland, and N. Rayner, 1997: A new global gridded radiosonde temperature data base and recent temperature trends. *Geophys. Res. Lett.*, **24**, 1499–1502, https://doi.org/10.1029/97GL01186.

Peterson, T. C., and Coauthors, 1998: Homogeneity adjustments of in situ atmospheric climate data: A review. *Int. J. Climatol.*, **18**, 1493–1517, https://doi.org/10.1002/(SICI)1097-0088(19981115)18:13<1493::AID-JOC329>3.0.CO;2-T.

Poli, P., and Coauthors, 2016: ERA-20C: An atmospheric reanalysis of the twentieth century. *J. Climate*, **29**, 4083–4097, https://doi.org/10.1175/JCLI-D-15-0556.1.

Rasmussen, R., and Coauthors, 2011: How well are we measuring snow: The NOAA/FAA/NCAR winter precipitation test bed. *Bull. Amer. Meteor. Soc.*, **93**, 811–829, https://doi.org/10.1175/BAMS-D-11-00052.1.

Rennie, J. J., and Coauthors, 2014: The international surface temperature initiative global land surface databank: Monthly temperature data release description and methods. *Geosci. Data J.*, **1**, 75–102, https://doi.org/10.1002/gdj3.8.

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1058, https://doi.org/10.1175/2010BAMS3001.1.

Sevruk, B., M. Ondrás, and B. Chvíla, 2009: The WMO precipitation measurement intercomparisons. *Atmos. Res.*, **92**, 376–380, https://doi.org/10.1016/j.atmosres.2009.01.016.

Skrynyk, O., and Coauthors, 2021: Ukrainian early (pre-1850) historical weather observations. *Geosci. Data J.*, https://doi.org/10.1002/gdj3.108, in press.

Slivinski, L. C., and Coauthors, 2019a: Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. *Quart. J. Roy. Meteor. Soc.*, **145**, 2876–2908, https://doi.org/10.1002/qj.3598.

——, and Coauthors, 2019b: NOAA–CIRES–DOE Twentieth Century Reanalysis version 3. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, accessed 1 November 2020, https://doi.org/https://doi.org/10.5065/H93G-WS83.

Stephens, G. L., and Coauthors, 2010: Dreary state of precipitation in global models. *J. Geophys. Res.*, **115**, D24211, https://doi.org/10.1029/2010JD014532.

Stickler, A., and Coauthors, 2010: The Comprehensive Historical Upper-Air Network. *Bull. Amer. Meteor. Soc.*, **91**, 741–752, https://doi.org/10.1175/2009BAMS2852.1.

Stucki, P., and Coauthors, 2015: Dynamical downscaling and loss modeling for the reconstruction of historical weather extremes and their impacts: A severe foehn storm in 1925. *Bull. Amer. Meteor. Soc.*, **96**, 1233–1241, https://doi.org/10.1175/BAMS-D-14-00041.1.

Thorne, P. W., and Coauthors, 2017: Toward an integrated set of surface meteorological observations for climate science and applications. *Bull. Amer. Meteor. Soc.*, **98**, 2689–2702, https://doi.org/10.1175/BAMS-D-16-0165.1.

Titchner, H. A., and N. A. Rayner, 2014: The Met Office Hadley Centre Sea Ice and Sea Surface Temperature data set, version 2: 1. Sea ice concentrations. *J. Geophys. Res. Atmos.*, **119**, 2864–2889, https://doi.org/10.1002/2013JD020316.

Trenberth, K. E., A. Dai, R. M. Rasmussen, and D. B. Parsons, 2003: The changing character of precipitation. *Bull. Amer. Meteor. Soc.*, **84**, 1205–1218, https://doi.org/10.1175/BAMS-84-9-1205.

——, J. T. Fasullo, and J. Mackaro, 2011: Atmospheric moisture transports from ocean to land and global energy flows in reanalyses. *J. Climate*, **24**, 4907–4924, https://doi.org/10.1175/2011JCLI4171.1.

Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012, https://doi.org/10.1256/qj.04.176

Vose, R. S., and Coauthors, 2012: NOAA's Merged Land–Ocean Surface Temperature analysis. *Bull. Amer. Meteor. Soc.*, **93**, 1677–1685, https://doi.org/10.1175/BAMS-D-11-00241.1.

Wallace, J. M., and P. V. Hobbs, 1977: *Atmosphere Science: An Introductory Survey*. Elsevier, 467 pp.

Wartenburger, R., S. Brönnimann, and A. Stickler, 2013: Observation errors in early historical upper-air observations. *J. Geophys. Res. Atmos.*, **118**, 12 012–12 028, https://doi.org/10.1002/2013JD020156.

Wen, X., G. Tang, S. Wang, and J. Huang, 2011: Comparison of global mean temperature series. *Adv. Climate Change Res.*, **2**, 187–192, https://doi.org/10.3724/SP.J.1248.2011.00187.

Whitaker, J. S., G. P. Compo, X. Wei, and T. M. Hamill, 2004: Reanalysis without radiosondes using ensemble data assimilation. *Mon. Wea. Rev.*, **132**, 1190–1200, https://doi.org/10.1175/1520-0493(2004)132<1190:RWRUED>2.0.CO;2.

Williamson, F., and Coauthors, 2018: Collating historic weather observations for the East Asian region: Challenges, solutions, and reanalyses. *Adv. Atmos. Sci.*, **35**, 899–904, https://doi.org/10.1007/s00376-017-7259-z.

Zhang, L., A. Kumar, and W. Wang, 2012: Influence of changes in observations on precipitation: A case study for the Climate Forecast System Reanalysis (CFSR). *J. Geophys. Res.*, **117**, D08105, https://doi.org/10.1029/2011JD017347.

Unauthenticated | Downloaded 10/14/21 01:07 AM UTC