RMetS
Royal Meteorological Society

Review Article
The Twentieth Century Reanalysis Project

G. P. Compo,[a,b]*J. S. Whitaker,[b‡] P. D. Sardeshmukh,[a,b] N. Matsui,[a,b] R. J. Allan,[c†] X. Yin,[d] B. E. Gleason, Jr.,[e‡] R. S. Vose,[e‡] G. Rutledge,[e‡] P. Bessemoulin,[f] S. Brönnimann,[g,h] M. Brunet,[i,l] R. I. Crouthamel,[j] A. N. Grant,[g] P. Y. Groisman,[e,k] P. D. Jones,[l] M. C. Kruk,[d] A. C. Kruger,[m] G. J. Marshall,[n] M. Maugeri,[o] H. Y. Mok,[p] Ø. Nordli,[q] T. F. Ross,[r‡] R. M. Trigo,[s] X. L. Wang,[t] S. D. Woodruff,[b‡] and S. J. Worley[u]

[a]*University of Colorado, CIRES, Climate Diagnostics Center, Boulder, CO, USA*
[b]*NOAA Earth System Research Laboratory, Physical Sciences Division, Boulder, CO, USA*
[c]*ACRE Project, Hadley Centre, Met Office, Exeter, UK*
[d]*STG Inc., Asheville, NC, USA*
[e]*NOAA National Climatic Data Center, Asheville, NC, USA*
[f]*MétéoFrance, Toulouse, France*
[g]*ETH Zurich, Switzerland*
[h]*Oeschger Center for Climate Change Research, University of Bern, Switzerland*
[i]*Centre for Climate Change, Universitat Rovira i Virgili, Tarragona, Spain*
[j]*International Environmental Data Rescue Organization, Deale, MD, USA*
[k]*University Corporation for Atmospheric Research, Boulder, CO, USA*
[l]*Climatic Research Unit, University of East Anglia, Norwich, UK*
[m]*South African Weather Service, Pretoria, South Africa*
[n]*British Antarctic Survey, Cambridge, UK*
[o]*Dipartimento di Fisica, Università degli Studi di Milano, Milano, Italy*
[p]*Hong Kong Observatory, Hong Kong, China*
[q]*Norwegian Meteorological Institute, Oslo, Norway*
[r]*NOAA Climate Database Modernization Program, NCDC, Asheville, NC, USA*
[s]*Centro de Geofísica da Universidade de Lisboa, IDL, University of Lisbon, Portugal*
[t]*Environment Canada, Toronto, Ontario, Canada*
[u]*National Center for Atmospheric Research, Boulder, CO, USA*
*Correspondence to: Gilbert P. Compo, 325 Broadway R/PSD1, Boulder, CO USA 80305-3328.
E-mail: compo@colorado.edu, gilbert.p.compo@noaa.gov
†The contribution of R. J. Allan was written in the course of his employment at the Met Office, UK, and is published with the permission of the Controller of HMSO and the Queen's Printer for Scotland.
‡The contributions of these authors were prepared as part of their official duties as US Federal Government employees.

The Twentieth Century Reanalysis (20CR) project is an international effort to produce a comprehensive global atmospheric circulation dataset spanning the twentieth century, assimilating only surface pressure reports and using observed monthly sea-surface temperature and sea-ice distributions as boundary conditions. It is chiefly motivated by a need to provide an observational dataset with quantified uncertainties for validations of climate model simulations of the twentieth century on all time-scales, with emphasis on the statistics of daily weather. It uses an Ensemble Kalman Filter data assimilation method with background 'first guess' fields supplied by an ensemble of forecasts from a global numerical weather prediction model. This directly yields a global analysis every 6 hours as the most likely state of the atmosphere, and also an uncertainty estimate of that analysis.

The 20CR dataset provides the first estimates of global tropospheric variability, and of the dataset's time-varying quality, from 1871 to the present at 6-hourly temporal and 2° spatial resolutions. Intercomparisons with independent radiosonde data indicate that the reanalyses are generally of high quality. The quality in the extratropical Northern Hemisphere throughout the century

is similar to that of current three-day operational NWP forecasts. Intercomparisons over the second half-century of these surface-based reanalyses with other reanalyses that also make use of upper-air and satellite data are equally encouraging.

It is anticipated that the 20CR dataset will be a valuable resource to the climate research community for both model validations and diagnostic studies. Some surprising results are already evident. For instance, the long-term trends of indices representing the North Atlantic Oscillation, the tropical Pacific Walker Circulation, and the Pacific–North American pattern are weak or non-existent over the full period of record. The long-term trends of zonally averaged precipitation minus evaporation also differ in character from those in climate model simulations of the twentieth century.

Key Words: data assimilation; Ensemble Kalman Filter; state estimation; surface pressure; sea-level pressure

Received 23 June 2010; Revised 30 October 2010; Accepted 16 December 2010; Published online in Wiley Online Library

Citation: Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N, Allan RJ, Yin X, Gleason Jr BE, Vose RS, Rutledge G, Bessemoulin P, Brönnimann S, Brunet M, Crouthamel RI, Grant AN, Groisman PY, Jones PD, Kruk MC, Kruger AC, Marshall GJ, Maugeri M, Mok HY, Nordli Ø, Ross TF, Trigo RM, Wang XL, Woodruff SD, Worley SJ. 2011. The Twentieth Century Reanalysis Project. *Q. J. R. Meteorol. Soc.* **137**: 1–28. DOI:10.1002/qj.776

1. Introduction

Long-term climate datasets are critical both for understanding climate variations and evaluating their simulation in climate models. Since the 1990s, major national and international efforts have led to the creation of climate datasets called retrospective analyses or 'reanalyses', which for the dataset jointly created by the National Centers for Environmental Prediction and the National Center for Atmospheric Research (NCEP, NCAR; Kalnay *et al.*, 1996) spans the period of availability of substantial upper-air observations (1948 onward). While such datasets have already proved valuable in climate research and applications, the fact that they extend back to only the mid-twentieth century limits their utility in constraining climate models, in understanding the global impacts of major earlier events such as the 1877/1878 El Niño and Indian famine and the 1883 Krakatoa eruption, and in reliably establishing long-term trends in the statistics of extreme weather associated with severe storms, floods, hurricanes, heat waves, and cold spells. Recently, US efforts such as the Global Change Research Program (USGCRP; Climate Change Science Program, 2003, 2008a,b), international efforts to develop the Global Climate Observing System (GCOS, 2010) and Global Earth Observation System of Systems (Group on Earth Observations, 2005), as well as conferences sponsored by the World Climate Research Program (WCRP) and other programs (e.g. Trenberth *et al.*, 2008; Bengtsson *et al.*, 2007) have called for new reanalysis datasets 'spanning the instrumental record' to compare the patterns and magnitudes of recent and projected changes in both the mean climate and climate variability, including especially the changes in extreme-event statistics. It is hoped that such longer reanalysis datasets will enable researchers to more reliably assess the natural range of variation of extreme-event statistics, and also to understand how variations in the El Niño Southern Oscillation (ENSO) and other climate modes affect those statistics (Easterling *et al.*, 2008). The Twentieth Century Reanalysis (20CR) Project described here represents one of the first comprehensive responses to this call.

The concept of 'reanalysis' may be traced back to a proposal by Brandes in 1816 (Monmonier, 1999) to construct synoptic weather maps. Since that time, the mapping of meteorological observations taken at approximately the same time to infer the current state of the atmosphere, what is now referred to as performing a synoptic 'analysis', has become a mainstay of weather and climate research. Until the mid-twentieth century, such analysis methods were subjective, relying on a meteorologist's experience in determining the contours of pressure and temperature, vectors of wind direction and speed, and other meteorological features. The advent of objective methods for what has come to be called 'data assimilation' has enabled the routine generation of thousands of synoptic maps every day by meteorological centres around the world (Daley, 1991). These methods provide an estimate of the state of the atmosphere at any particular time by forming a weighted average that combines millions of observations taken from weather stations, ships, buoys, balloons, radiosondes, aircraft, satellites, and other measurement platforms, with an *a priori* estimate of that state obtained from a short-term forecast generated using a numerical weather prediction (NWP) model. While crucial for generating accurate real-time forecasts, these modern synoptic analyses have limited utility in climate research owing to the artificial variability associated with frequent changes in observing systems, data assimilation methods, and in the NWP models used to generate the *a priori* 'first guess' analysis estimates (Bengtsson and Shukla, 1988; Trenberth and Olson, 1988). Modern reanalyses have been proposed as a remedy for this problem, with an emphasis on regenerating the synoptic analyses over several decades using a fixed data assimilation system and NWP model

(Bengtsson and Shukla, 1988; Trenberth and Olson, 1988; Bengtsson *et al.*, 2007; Thorne and Vose, 2010).

Strictly speaking, the first retrospective analyses (i.e. reanalyses) were generated in 1819, when Brandes followed up on his original proposal and constructed 365 daily maps of pressure contours using barometric pressure observations taken in 1783 at several European stations (Monmonier, 1999). Many retrospective analysis projects followed upon his pioneering accomplishment. Major efforts to recover historical data and construct retrospective hemispheric sea-level pressure (SLP) analyses were undertaken for the Southern Hemisphere for the years 1901–1904 in association with the expeditions of the *Discovery* and *Gauss* to Antarctica (National Antarctic Expedition, 1913). Similar efforts were also undertaken for the Northern Hemisphere, most notably a retrospective SLP reanalysis for 1899–1939 entitled the *US Historical Weather Maps Daily Synoptic Series, Northern Hemisphere, Sea Level* (Wexler and Tepper, 1947). The US Weather Bureau, Army, and Navy produced these maps to better understand weather events in support of aviation and other needs during World War II. The importance of reanalysis was well recognized by Wexler and Tepper (1947), who enthused that 'by breathing life into a mass of inert data' such projects 'provided an indispensable aid to future research'.

Almost 50 years after the publication of the *US Historical Weather Maps*, the techniques of data assimilation were brought to bear on the problem of retrospective analysis. As part of the Global Atmospheric Research Program, pioneering data assimilation systems were used at the European Centre for Medium-Range Weather Forecasts (ECMWF, Bengtsson *et al.*, 1982) and National Oceanic and Atmospheric Administration (NOAA)'s Geophysical Fluid Dynamics Laboratory (Ploshay *et al.*, 1983) to produce reanalysis fields for the coordinated study year of 1979. This vanguard achievement was followed by the modern three-dimensional reanalysis datasets. These widely utilized datasets extend from no earlier than 1948 to the present, e.g. the 1948–present NCEP–NCAR reanalysis dataset (NNR, Kalnay *et al.*, 1996); the 1957–2002 ECMWF reanalysis (ERA-40; Uppala *et al.*, 2005); the 1979–present NCEP–Department of Energy (DOE) reanalysis dataset (Kanamitsu *et al.*, 2002); the 1979–present Japan Meteorological Agency reanalysis Archive 25 (Onogi *et al.*, 2007); the 1979–present NASA Modern Era Retrospective-analysis for Research and Applications (Bosilovich *et al.*, 2008; Rienecker *et al.*, 2008); the 1979–present NCEP Climate Forecast System Reanalysis dataset (Saha *et al.*, 2010); the 1979–1993 ERA-15 (Gibson *et al.*, 1997); the 1980–1996 NASA/Data Assimilation Office reanalysis dataset (Schubert *et al.*, 1993); and the 1989–present ECMWF Reanalysis Interim (ERA-Interim; Dee and Uppala, 2009; Simmons *et al.*, 2007). These reanalysis systems assimilate all independent usable observations available at the reanalysis time. Many employ variational assimilation techniques, the most frequent being 3D-Var, with a fixed error assigned to the first-guess fields throughout the reanalysis period. Such an invariant specification may be suboptimal for reanalyzing over longer periods, within which significant variations of data density (and accuracy) can occur. Over the latter half of the twentieth century, the number of observations available for each state estimation has varied from several thousand surface observations and a few hundred upper-air observations in the 1950s to additional millions of satellite observations by the 1990s. The application of suboptimal assimilation techniques with fixed parameters over periods of such variations in observational networks has caused problems ranging from understated extratropical storm track variability (Harnik and Chang, 2003; Hodges *et al.*, 2003) to incorrect tropical (Newman *et al.*, 2000) and Antarctic variability (Bromwich and Fogt, 2004), to spurious long-term trends (Trenberth and Smith, 2005; Bengtsson *et al.*, 2004a,b; Kinter *et al.*, 2004; Sturaro, 2003; Kistler *et al.*, 2001). As one attempts to reanalyze the early twentieth century, and especially the nineteenth century, scant upper-air data, upon which the existing 3D-Var systems depend to produce reasonable upper-air fields (Bengtsson *et al.*, 2004b; Kanamitsu and Hwang, 2006), will be available (Brönnimann *et al.*, 2005). The reduced data densities will further compromise the ability of 3D-Var systems with *fixed parameters* to produce reliable upper-air reanalyses of these earlier periods.

In contrast, several studies have demonstrated the feasibility of generating reliable reanalyses of these earlier periods using only surface observations and more advanced data assimilation methods such as the Ensemble Kalman Filter (Whitaker *et al.*, 2004; Anderson *et al.*, 2005; Compo *et al.*, 2006) or 4D-Var assimilation (Thépaut, 2006; Whitaker *et al.*, 2009). These studies have emphasized several advantages of using surface pressure and SLP observations for such historical reanalyses. Measurements of surface pressure have been made consistently since the late 1800s, and standard corrections are known for even earlier observations. The surface pressure information, through geostrophy, yields a reasonable approximation to the barotropic part of the flow, which accounts for a substantial part of the total flow. The surface pressure tendency, being related to the vertically integrated mass flux divergence, provides further information about the tropospheric circulation that can be captured by a multivariate data assimilation system (Bengtsson, 1980).

Motivated by the above considerations, the 20CR utilizes an Ensemble Kalman Filter data assimilation system (Whitaker and Hamill, 2002), a new version of the NCEP atmosphere–land model to generate first-guess fields with interpolated monthly sea-surface temperature and sea-ice concentration fields from the Hadley Centre Sea Ice and SST dataset (HadISST; Rayner *et al.*, 2003) as prescribed boundary conditions, and newly compiled surface pressure and SLP reports and observations, to produce a reanalysis dataset spanning 1871 to the present. The pressure data come from the new International Surface Pressure Databank (ISPD), which incorporates pressure observations extracted from leading international archives of meteorological variables and contributed national and international collections. The databank has been established through extensive international cooperation under the auspices of the Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative, and working groups of GCOS and WCRP. A preliminary first version of the reanalysis (20CRv1; Compo *et al.*, 2008) spanned the period 1908 to 1958. In the second and complete version (20CRv2) described here, global atmospheric fields for 1871 to 2008 have been generated.

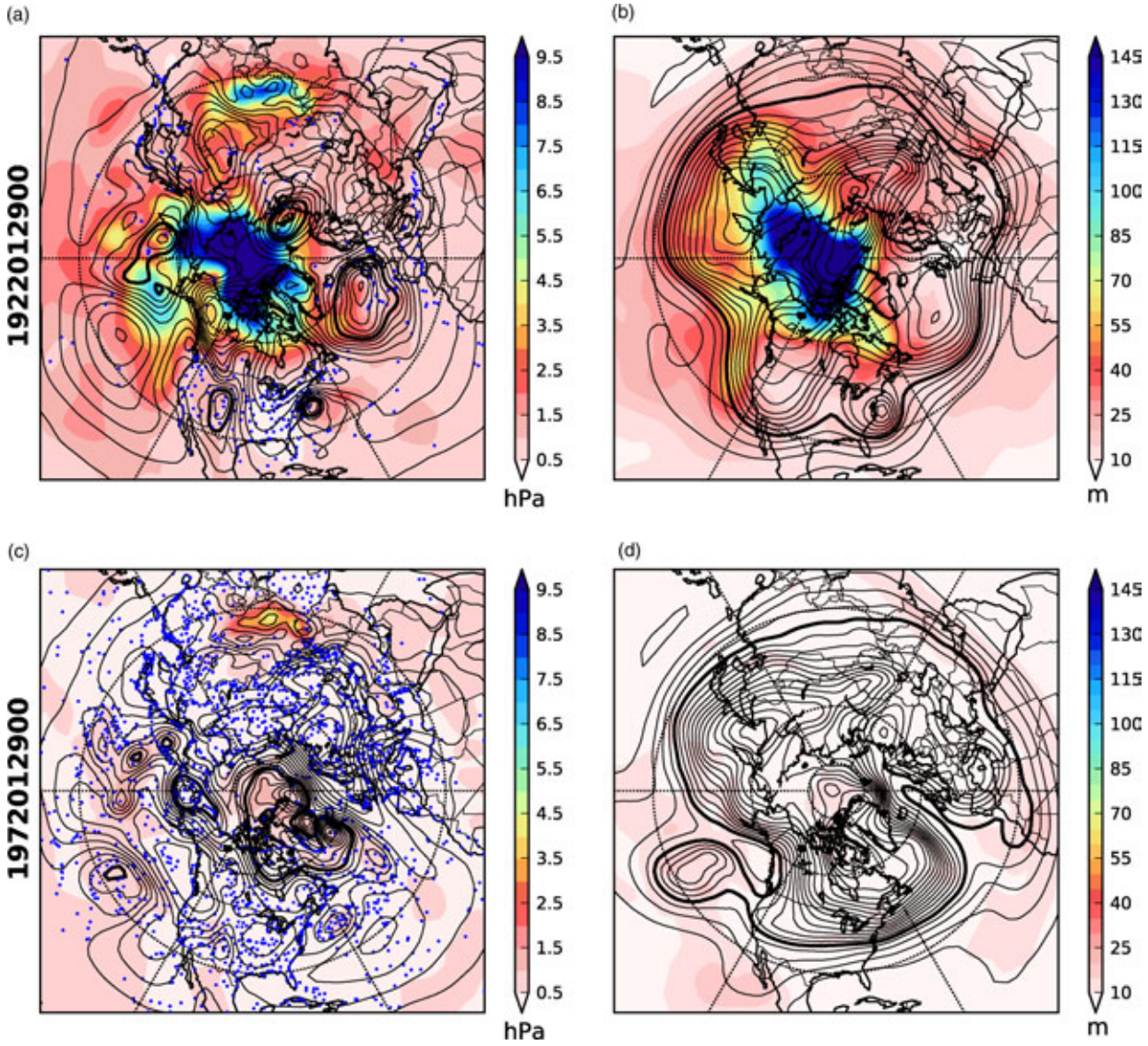

Figure 1. Synoptic charts of the ensemble mean analysis and ensemble uncertainty for (a) sea-level pressure and (b) 500 hPa geopotential mean and spread at 0000 UTC on 29 January 1922. (c, d) are as (a, b), but for 29 January 1972. Line contours indicate the analysis and shading indicates the uncertainty measured as the ensemble standard deviation (or spread) at each location. The line contour interval in (a,c) is 4 hPa, with the 1000 hPa contour thickened, and in (b,d) it is 50 m, with the 5600 m contour thickened. The shading interval in (a,c) is 0.25 hPa and in (b,d) is 5 m.

In our implementation, the Ensemble Kalman Filter provides an estimate of the atmospheric state every six hours and also the uncertainty of that estimation. Figure 1 provides an example of this capability. The maps show the analyzed SLP and 500 hPa geopotential height fields and their estimated uncertainties for 0000 UTC on 29 January 1922 (Figure 1(a, b)) during the time of the so-called 'Knickerbocker' storm along the US East Coast (Kocin and Uccellini, 2004), and also for the same day of 1972 (Figure 1(c, d)), when many more observations were available, to illustrate the impact of observational volumes on state estimation. Similar fields are available for all output variables from the 20CR dataset at all times. As the figure shows, the uncertainty in the estimated state decreases as the network density increases from 1922 to 1972. The uncertainty is also flow-dependent, as evident, e.g. in the relatively high uncertainty in Figure 1(b) along the tightly packed geopotential height contours over the northeast Pacific. In addition to the contribution from atmospheric dynamics, as described below, the uncertainty field also includes a representation of errors arising from imperfect observations and a finite-ensemble first guess generated using an imperfect NWP model.

The remainder of the paper is organized as follows. Section 2 describes the 2008 experimental NCEP Global Forecast System NWP model used for generating the first-guess background fields. Section 3 provides some specifics of our implementation of the Ensemble Kalman Filter for 20CRv2. Section 4 describes the ISPD compilation of surface pressure and SLP observations and the 20CRv2 quality-control (QC) system. Section 5 describes the multi-stream production scheme employed to generate the dataset. Section 6 – the first 'results' section – compares the synoptic variability and uncertainty in the 20CRv2 reanalysis fields with that in the NNR and ERA-40 reanalyses, in the station and marine observations of the ISPD, and in additional upper-air observations. Section 7 presents some basic features of the climatology and climate variability as represented in this dataset, and compares them with other reanalyses and statistical reconstructions. Section 8 provides details of data access and conclusions.

2. Global numerical weather prediction model and boundary conditions

The coupled atmosphere–land model used here is the April 2008 experimental version of the NCEP Global Forecast System (GFS), courtesy of the NOAA NCEP Environmental Modeling Center (EMC). The model has a complete suite of physical parametrizations, as described in Kanamitsu *et al.* (1991), with recent updates detailed in Moorthi *et al.* (2001). Additional updates to these parametrizations are described in Saha *et al.* (2006). The version used here includes an updated prognostic cloud condensate scheme (Moorthi *et al.*, 2001) and revisions to the solar radiative transfer (Hou *et al.*, 2002), boundary-layer vertical diffusion (Hong and Pan, 1996), Simplified Arakawa–Schubert convection scheme with momentum mixing (Moorthi *et al.*, 2001), gravity wave drag (Kim and Arakawa, 1995), mountain blocking (Lott and Miller, 1997; Alpert, 2004), and long-wave radiative transfer (Mlawer *et al.*, 1997). A prognostic ozone scheme includes parametrizations of ozone production and destruction (Saha *et al.*, 2010). The model contains a complex representation of land processes through coupling with the four-layer Noah land model (Ek *et al.*, 2003). We used an experimental version of the GFS model primarily because of its ability to treat time-varying CO_2 and volcanic aerosols, features not found in the operational version at that time. Annual averages of the time-varying global mean CO_2 concentration, volcanic aerosols, and incoming solar radiation were specified as in Saha *et al.* (2010). For computational efficiency in the 20CRv2 ensemble system, we used the model at a horizontal resolution of T62 and a vertical resolution of 28 vertical hybrid sigma-pressure levels (Juang, 2005).

The lower boundary conditions of sea surface temperature and sea-ice concentration fields needed to run the model in atmosphere–land-only mode were specified using the UK Met Office HadISST1.1 dataset (Rayner *et al.*, 2003). The monthly mean data were interpolated to daily resolution using a monthly-mean preserving algorithm (Taylor *et al.*, 2000). The values used for the first fortnight of the reanalysis, 1–15 January 1871, were determined by interpolating between the average of 1871–1900 December values and the mid-month value of 15 January 1871 from HadISST. As noted by Rayner *et al.* (2003), the quality of the HadISST varies throughout the period from a sparse network of marine SST observations in the 1870s to a network comprising additional numerous marine, buoy, and satellite SST observations in the late twentieth century (with the International Comprehensive Ocean–Atmosphere DataSet (ICOADS; Worley *et al.*, 2005) providing the bulk of the *in situ* SST observations). The sea-ice concentrations used have a similar time-dependence in their quality. As described in Appendix A, the sea-ice concentration ingested into 20CR differs from that provided in HadISST, owing to our accidental mis-specification of the concentration near coastal regions.

3. Ensemble Kalman Filter implementation

The Kalman Filter provides a rigorous method for optimally combining, in a least-squares sense, imperfect observations and an imperfect estimate of the current state (the 'background' or 'first guess') assuming linear error models (Daley, 1991). Owing to computing limitations, for many years it did not have widespread use in meteorological applications. Rapid progress has been made since Evensen (1994) proposed a Monte Carlo approximation, now called the Ensemble Kalman Filter, in which the background-error covariance is approximated using an ensemble of, e.g., short-range NWP forecasts. Recent reviews (e.g. Evensen, 2003; Houtekamer and Mitchell, 2005; Hamill, 2006; Evensen, 2009) discuss many theoretical and practical issues associated with this data assimilation methodology. In particular, when the observations and background errors are Gaussian, then the Kalman Filter is the solution to the Bayesian problem of combining a given observation distribution and a prior background distribution to form a posterior distribution of the state of the atmosphere. There are two distinct classes of algorithms for implementing an Ensemble Kalman Filter: 'stochastic' algorithms that account for observational errors by assimilating perturbed observations as part of the ensemble generation (e.g. Houtekamer and Mitchell, 1998; Burgers *et al.*, 1998; Houtekamer and Mitchell, 2005; Houtekamer *et al.*, 2005), and 'deterministic' algorithms that modify the update equation to reduce potential sampling errors in the stochastic approach, as reviewed, e.g., by Tippett *et al.* (2003). Both classes are areas of active research (e.g. Evensen, 2009).

The Ensemble Kalman Filter algorithm used in 20CRv2 is of the 'deterministic' type (specifically, through the use of $\tilde{\mathbf{K}}$ instead of $\mathbf{K}$ in Eq. (2) below), as described in Whitaker *et al.* (2004) and Compo *et al.* (2006), and is based on the 'Ensemble Square Root Filter' algorithm of Whitaker and Hamill (2002). It has a simple parallelizable implementation when observations are processed one at a time (Whitaker and Hamill, 2002; Anderson and Collins, 2007).

To fix ideas, using the notation of Ide *et al.* (1997), assume that we have an n-member ensemble of first-guess or background fields, with the jth member $\mathbf{x}_j^{\mathrm{b}}$ representing the complete m-dimensional state vector of the NWP forecast model (e.g. wind, geopotential height, humidity, and surface pressure fields on the model domain) in a 6-hour window centred on the analysis time. We use $n = 56$, based on tests which found that ensembles smaller than this degraded the quality of the upper-air analyses (not shown). The sample mean of these fields is denoted $\overline{\mathbf{x}}^{\mathrm{b}} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_j^{\mathrm{b}}$, and the deviations from the mean are

$$\mathbf{x}_j{}'^{\mathrm{b}} = \mathbf{x}_j^{\mathrm{b}} - \overline{\mathbf{x}}^{\mathrm{b}}.$$

The sample background-error covariance matrix is denoted

$$\mathbf{P}^{\mathrm{b}} = \frac{1}{n-1}\sum_{j=1}^{n}\mathbf{x}_j{}'^{\mathrm{b}}\mathbf{x}_j{}'^{\mathrm{bT}}.$$

We denote the first surface pressure observation to be assimilated as y^{o} and the ensemble mean and deviations interpolated to the observation location and time as

$$\overline{y}_j^{\mathrm{b}} = \mathbf{H}\overline{\mathbf{x}}_j^{\mathrm{b}} \text{ and } y_j{}'^{\mathrm{b}} = \mathbf{H}\mathbf{x}_j{}'^{\mathrm{b}},$$

respectively, with $\mathbf{H}$ being the operator that interpolates the first-guess surface pressure field to the observation location and time. We then combine the first-guess ensemble and the

observation to form an n-member analysis ensemble, whose mean $\overline{\mathbf{x}}^{\mathrm{a}}$ and deviations $\mathbf{x}_j'^{\mathrm{a}}$ are calculated via

$$\overline{\mathbf{x}}_j^{\mathrm{a}} = \overline{\mathbf{x}}_j^{\mathrm{b}} + \mathbf{K}\left(y^{\mathrm{o}} - \overline{y}_j^{\mathrm{b}}\right), \tag{1}$$

and

$$\mathbf{x}_j'^{\mathrm{a}} = \mathbf{x}_j'^{\mathrm{b}} - \tilde{\mathbf{K}}\left(y_j'^{\mathrm{b}}\right), \tag{2}$$

where the sample Kalman gain $\mathbf{K}$ is given by

$$\begin{aligned}\mathbf{K} &= \mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} + R)^{-1} \\ &= \frac{1}{n-1}\sum_{j=1}^{n}\mathbf{x}_j'^{\mathrm{b}}y_j'^{\mathrm{b}}\left(\frac{1}{n-1}\sum_{j=1}^{n}y_j'^{\mathrm{b}}y_j'^{\mathrm{b}} + R\right)^{-1},\end{aligned} \tag{3}$$

and the sample modified Kalman gain $\tilde{\mathbf{K}}$ is given by

$$\tilde{\mathbf{K}} = \left(1 + \sqrt{\frac{R}{\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} + R}}\right)^{-1}\mathbf{K}, \tag{4}$$

as in Whitaker and Hamill (2002). The specified observational-error variance R accounts for both the measurement error associated with the observation and the error associated with representing a large area (i.e. an NWP model grid box) by a point measurement. We further assume that the error in y^{o} is uncorrelated with the errors in all other pressure observations to be assimilated. Note that the background-error variance at the observation location,

$$\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} = \frac{1}{n-1}\sum_{j=1}^{n}y_j'^{\mathrm{b}}y_j'^{\mathrm{b}} = {\sigma_{\mathrm{b}}}^2,$$

in (3) and (4) is a scalar. In a significant modification to the system of Whitaker *et al.* (2004) and Compo *et al.* (2006), we incorporated an improved accounting of hourly observations made throughout the 6-hour window centred on the analysis time, but not necessarily at the analysis time itself. This was done, without changing Eqs (1)–(4), by representing time-dependent information between a $y_j'^{\mathrm{b}}$ at a possibly different time than $\mathbf{x}_j'^{\mathrm{b}}$ in the covariance in Eq. (3). Such a modification allows information from, e.g., fast-moving weather systems to be used to improve the analysis, letting the analysis 'know' where the system was and will be within the 6-hour window. Technically, this makes our implementation a light Ensemble Kalman Smoother, rather than an Ensemble Kalman Filter, but the equations are unaltered (Sakov *et al.*, 2010).

The model means, variances, and covariances in Eqs (1)–(4) are all unbiased sample estimates from the $n = 56$ member ensemble. Within the limitations of using an imperfect model and finite ensembles, this formulation represents a minimum-error estimate of the 'true state' (Lorenc, 1986), represented here by the analysis ensemble mean $\overline{\mathbf{x}}^{\mathrm{a}}$. The uncertainty of this estimate is given by the covariance of the analysis ensemble deviations $\mathbf{x}_j'^{\mathrm{a}}$,

$$\mathbf{P}^{\mathrm{a}} = \frac{1}{n-1}\sum_{j=1}^{n}\mathbf{x}_j'^{\mathrm{a}}\mathbf{x}_j'^{\mathrm{aT}}.$$

The uncertainty at the observation location is $\mathbf{H}\mathbf{P}^{\mathrm{a}}\mathbf{H}^{\mathrm{T}} = {\sigma_{\mathrm{a}}}^2$.

To assimilate subsequent observations, the 56 members of the analysis ensemble $\mathbf{x}_j^{\mathrm{a}}$ become the new first-guess ensemble $\mathbf{x}_j^{\mathrm{b}}$ and Eqs (1)–(4) are applied iteratively for each observation. After all observations have been assimilated, the 56-member set of analyses $\mathbf{x}_j^{\mathrm{a}} = \overline{\mathbf{x}}^{\mathrm{a}} + \mathbf{x}_j'^{\mathrm{a}}$ becomes the 56 initial conditions for the subsequent forecast/analysis cycle, with 9-hour forecasts made to generate analyses every 6 hours (Whitaker *et al.*, 2004, give a complete description). As well as archiving the analysis fields, R, ${\sigma_{\mathrm{b}}}^2$, and ${\sigma_{\mathrm{a}}}^2$ are archived for every observation and observation location.

As discussed by, e.g., Anderson and Anderson (1999) and Whitaker and Hamill (2002), sampling and model errors prevent the ensemble-estimated background-error covariances from being optimal in the Kalman update equation (1). Such issues must be addressed to prevent 'filter divergence' wherein the update equations (1)–(4) weight the background too much and the observations too little. Cycling during such a condition reinforces the background and causes the filtered 'analysis' estimate to drift farther and farther from observations. Two methods are used to account for these sources of error: covariance inflation (Anderson and Anderson, 1999) and distance-dependent covariance localization (Houtekamer and Mitchell, 2001; Hamill *et al.*, 2001). Covariance inflation partly corrects for under-specified sampling error and also serves as a crude parametrization of neglected model error, assuming that the model-error covariance is in the state space of the ensemble. Although Whitaker *et al.* (2008) found that so-called 'additive inflation' (adding random perturbations with a specified covariance structure to individual ensemble members) was superior to a simple multiplicative covariance inflation (Whitaker *et al.*, 2004) for a modern complete observing network, our own tests of this method with the limited surface pressure network did not reveal a significant difference in analysis quality (not shown). We therefore used the simpler and less computationally expensive method of a simple multiplicative covariance inflation, as described in Whitaker *et al.* (2004).

Covariance localization (Houtekamer and Mitchell, 2001; Hamill *et al.*, 2001) is a spatial filter that smoothly sets the ensemble covariances to zero beyond a specified distance. This reduces the potential for filter divergence arising from spurious long-distance correlations obtained using finite ensemble sizes. We set the localization distance to 4000 km in the horizontal using the same localization function as Whitaker and Hamill (2002), i.e. Eq. (4.10) of Gaspari and Cohn (1999). In the vertical, localization is set to 4 scale heights (units of $-\ln(p/p_{\mathrm{s}})$), which amounts to smoothly reducing the effect of a 1000 hPa surface pressure observation to zero at about 18 hPa. We note that we localized the variable geopotential height instead of virtual temperature. Localization of temperature gave large geopotential height increments ($\overline{\mathbf{x}}^{\mathrm{a}} - \overline{\mathbf{x}}^{\mathrm{b}}$) that were not in geostrophic balance with the wind increments at locations where the latter had been localized to zero. The relatively large localization distance may also improve the overall balance of the increments compared to smaller localizations of less than 1000 km sometimes used.

The inflation of first-guess error covariance by a factor $r > 1.0$ was applied to the first-guess ensemble deviations $\mathbf{x}_j'^{\mathrm{b}}$ before assimilating the first observation in a 6-hour window. The specified variation of r was designed to account for the enormous range of observational densities from the

Table I. Temporal and spatial variation of the covariance inflation parameter r.

	NH	Tropics	SH
1871–1890	1.05	1.01	1.01
1891–1920	1.09	1.02	1.01
1921–1951	1.12	1.03	1.02
1952–2008	1.12	1.07	1.07

NH = 90°–30°N; Tropics = 30°N–30°S; SH = 30°–90°S.

1870s to 2000s, and also different observational densities in the Tropics and the extratropical Northern and Southern Hemispheres, as indicated in Table I. These values were set based on the observing system experiments of Compo *et al.* (2006) and additional experiments using September 1938. They were not adjusted further or 'tuned' during production. As a hard limit to prevent spiraling inflation in the complete absence of observations in an analysis window, r was set to 1.0 in a region if the first-guess ensemble standard deviation of surface pressure exceeded 17 hPa anywhere in that region. This limit was set to be larger than typical regional maximum values on a global map of the climatological standard deviation of surface pressure, which were estimated using ERA-40 for the 1981–2000 period as ∼16 hPa in the vicinity of the Icelandic low and South Pacific storm track (not shown).

An important issue to address when assimilating surface pressure observations over land is to make appropriate adjustments to these observations to account for the difference between the model orography and the station elevation at the observation location (Ingleby, 1995). Our approach is similar to that of Ingleby (1995). The adjustment was done using the pressure reduction algorithm of Benjamin and Miller (1990), as

$$\hat{p}_{\mathrm{ob}} = p_{\mathrm{ob}} \left(\frac{T_0 + \Gamma(z_{\mathrm{station}} - z_{\mathrm{model}})}{T_0} \right)^{\frac{g}{R_{\mathrm{d}}\Gamma}}, \qquad (5)$$

where z_{model} is the model orography interpolated to the station location, T_0 is a virtual temperature determined from the model first-guess forecast interpolated to the station location, Γ is a model first-guess virtual temperature lapse rate interpolated to the station location, g is gravity, R_{d} is the gas constant for dry air, and z_{station} is the station elevation. The virtual temperature T_0 was determined from

$$T_0 = T_k \left(\frac{p_{\mathrm{ob}}}{p_k} \right)^{\frac{R_{\mathrm{d}}\Gamma}{g}}, \qquad (6)$$

where T_{k} and p_{k} are model first-guess virtual temperature and pressure at model level k, horizontally interpolated to the station location. The virtual temperature lapse rate Γ was estimated as

$$\Gamma = \frac{T_{k-1} - T_k}{z_{k-1} - z_k} \qquad (7)$$

where z_k is the model first-guess geopotential height at model level k interpolated to the station location. The model level k in this calculation was chosen to be the first model level that is usually above the planetary boundary layer, to avoid diurnal effects on the pressure adjustment. Benjamin and Miller (1990) used values at 700 hPa. We used values at level $k = 16$, approximately 250 hPa above the surface, with $k = 1$ being the model surface.

The adjusted observations $\hat{p}_{\mathrm{ob}}$ were computed using ensemble mean first-guess fields. In order to modify the observation-error variance R to reflect the uncertainty introduced by the adjustment, an ensemble of observations and station elevations was constructed. The ensemble of observations was generated by perturbing p_{ob} with Gaussian random noise with zero mean and standard deviation given by the nominal observation error. The ensemble of station elevations was generated by perturbing z_{station} with Gaussian random noise with zero mean and standard deviation given by the uncertainty in the station elevation, specified here as 3 m (J. Comeaux, pers. comm., 2006). This specification may have been conservative, given that the uncertainty in station elevation is often between 10 and 20 m and may be as large as 80 m in some historical observations (Ingleby, 1995). The pressure adjustment was done separately for each ensemble member, using each ensemble member's virtual temperature and pressure interpolated to the station location. The adjusted R was then simply taken to be the variance of the adjusted surface pressure observations. Larger differences between z_{model} and z_{station} led to larger increases of R above the nominal value, reflecting the increased uncertainty introduced by the adjustment. Increases of R were associated with increases of the ensemble variances of T_0 and Γ, since uncertainty in these quantities implies uncertainty in the observation adjustment.

Alternatively, instead of adjusting the surface pressure observations, we could have used the same algorithm to adjust the first-guess surface pressure to the station elevation as part of the computation of the forward observation operator. We tested this alternative approach, and found that it produced similar results.

4. International Surface Pressure Databank collection of observations and quality control

The International Surface Pressure Databank (ISPD) is the world's largest collection of pressure observations. It has been developed by extracting observations from established international archives of meteorological variables and by combining observations made available through additional international cooperation with data recovery facilitated by the ACRE initiative*, and other contributing organizations and projects listed in Table II. ISPD version 2 was used here.

The ISPDv2 comprises three components: station (land) and marine observations and tropical cyclone 'best track' pressure observations and reports. The station component is a new blend of many national and international collections of SLP and surface pressure, with the largest contributor being the International Surface Database (Lott *et al.*, 2008). Yin *et al.* (2008) describe the duplicate elimination procedures for this component. To produce the blend, the ISPDv2 station component employs a two-step duplicate elimination process that first removes duplicates within a collection and then the duplicates among collections. Stations within a 0.1° radius are considered for duplicate elimination to account for possible location errors in the station metadata. For the marine component, observations of SLP are extracted from

*http://www.met-acre.org

Table II. Organizations contributing pressure observations to the International Surface Pressure Databank version 2.

All-Russian Research Institute of Hydrometeorological Information World Data Center
Atmospheric Circulation Reconstructions over the Earth (ACRE)
Australian Bureau of Meteorology
British Antarctic Survey
Danish Meteorological Institute
Deutscher Wetterdienst (DWD; German Weather Service)
European and North Atlantic Daily to Multidecadal Climate Variability (EMULATE)
Environment Canada
ETH Zurich, Switzerland
GCOS Atmospheric Observation and Ocean Observation Panels for Climate WG on Surface Pressure
Hong Kong Observatory
International Comprehensive Ocean–Atmosphere Data Set (ICOADS)
Instituto Geofisico da Universidade do Porto, Portugal
Japanese Meteorological Agency
Jersey Met Department
Koninklijk Nederlands Meteorologisch Instituut (KNMI; Royal Netherlands Meteorological Institute)
MetéoFrance
Meteorological and Hydrological Service, Croatia
Met Office Hadley Centre, UK
National Center for Atmospheric Research (NCAR), USA
NOAA Climate Database Modernization Program (CDMP), USA
NOAA Earth System Research Laboratory (ESRL), USA
NOAA National Climatic Data Center (NCDC), USA
NOAA National Centers for Environmental Prediction (NCEP), USA
NOAA Northeast Regional Climate Center at Cornell Univ., USA
NOAA Midwest Regional Climate Center at UIUC, USA
Norwegian Meteorological Institute
Ohio State Univ. – Byrd Polar Research Center, USA
Proudman Oceanographic Laboratory, UK
Signatures of environmental change in the observations of the Geophysical Institutes (SIGN)
South African Weather Service
Univ. of Colorado – Climate Diagnostics Center (CDC) of the Cooperative Institute for Research in Environmental Sciences (CIRES)
Univ. of East Anglia – Climatic Research Unit, UK
Univ. of Lisbon – Instituto Geofisico do Infante D. Luiz, Portugal
Univ. of Lisbon – Instituto de Meteorologia, Portugal
Univ. of Milan – Department of Physics, Italy
Univ. Rovira i Virgili – Center for Climate Change (C3), Spain
ZentralAnstalt für Meteorologie und Geodynamik (ZAMG; Austrian Weather Service)

ICOADS, which applies extensive duplicate elimination procedures (Worley *et al.*, 2005). The latest available version of ICOADS was used as the 20CRv2 was produced, resulting in ICOADS version 2.4 being used for the period 1952–2008 and version 2.5 for 1871–1951 (Woodruff *et al.*, 2010). Similarly, for the tropical cyclone component, the latest available version of the International Best Track Archive for Climate Stewardship (IBTrACS; Knapp *et al.*, 2010) was used as our reanalysis fields were generated. This resulted in the use of version 'Beta' for 1952–2006, 'v01r 01' for 1871–1883, and 'v02r 01' for 1884–1885 and 2007–2008. In the IBTrACS entries of tropical cyclone central pressure, some data values represent actual measurements while others represent time-interpolated values provided by the tropical cyclone warning centres. We refer to the latter as pressure 'reports'. For IBTrACS entries where neither a central pressure observation nor report was provided, a central pressure report was derived from the wind speed entry using a second-order polynomial approximation to the gradient wind equation with empirically estimated parameters (N. Matsui, pers. comm., 2010). Parameters were estimated separately for each IBTrACS sub-basin using wind and pressure data from the period 1979–2007, to account for the different characteristics of the tropical cyclones in the sub-basins. These pressure values (from observations, interpolated values, and derived from wind speed) included in the ISPD represent less than 0.22% of the total number of data points in 1878 and less than 0.009% in 2008 in the complete ISPD compilation assimilated here. For simplicity, the combination of these scant reports related to tropical cyclone central pressure, along with the actual measurements of pressure, are referred to as 'observations' throughout the remainder of the article, where there is no possibility of confusion.

All other ISPD observations of surface pressure and SLP were QCed during the assimilation cycle in a five-step process that included basic checks for meteorological plausibility, and comparisons with the first-guess ensemble and neighbours (Appendix B; note that the IBTrACS pressure values were not subjected to the QC procedure). The QC indicators contained in the metadata of, e.g. ICOADS and ISD were not used in 20CRv2 so that the 20CR system's own QC examines all marine and station observations. For those observations passing the QC procedure, surface pressure observations were assimilated at the elevation provided in the ISPD's station metadata, whereas observations of SLP were considered as observations at an elevation of zero metres. When both surface pressure and SLP were available with an accompanying station elevation, surface pressure was used.

As described in Appendix B, the complete ISPDv2 includes for every observation the results of the quality control outlined above plus so-called 'feedback' assimilation quantities: the observation-error variance R, the first-guess error variance at the observation location ${\sigma_b}^2$, the analysis-error variance at the observation location ${\sigma_a}^2$, the observation minus first guess, $y_o - \mathbf{H}\overline{\mathbf{x}}^b$, and the observation minus the ensemble-mean analysis, $y_o - \mathbf{H}\overline{\mathbf{x}}^a$.

5. Production of 20CRv2

To take advantage of the massively parallel computers that the US DOE made available to this project, the 20CRv2 dataset was generated in parallel production 'streams'

Table III. Multi-stream production scheme including start date of the spin-up period and start and end dates of the production period.

Stream number	Spin-up start date 1 November 0000 UTC	Production start date 1 January 0000 UTC	Production end date 31 December 1800 UTC
1	1869	1871	1875
2	1874	1876	1880
3	1879	1881	1885
4	1884	1886	1890
5	1889	1891	1895
6	1894	1896	1900
7	1899	1901	1905
8	1904	1906	1910
9	1909	1911	1915
10	1914	1916	1920
11	1919	1921	1925
12	1924	1926	1930
13	1929	1931	1935
14	1934	1936	1940
15	1939	1941	1945
16	1944	1946	1951
17	1949	1952	1955
18	1954	1956	1960
19	1959	1961	1965
20	1964	1966	1970
21	1969	1971	1975
22	1974	1976	1980
23	1979	1981	1985
24	1984	1986	1990
25	1989	1991	1995
26	1994	1996	2000
27	1999	2001	2008

(Table III). Each stream was started from the same 56-member ensemble drawn from a climatological ensemble of states on 1 November 0000 UTC of the stream start year (Table III). For a particular stream, after incrementing the climatological ensemble using Eqs (1)–(4), an ensemble of 56 9-hour forecasts was generated by integrating the NCEP GFS model described in section 2. Each stream was cycled for 14 months to reduce the effect of the initial condition in the lower layers of the land model. Our implementation of Eqs (1)–(4) did not involve any stochastic perturbations or re-arrangements of the ensemble, which had the effect that the *j*th ensemble member within each stream remained distinct and temporally continuous over the course of a stream's production. However, the *j*th ensemble member was not continuous across streams, e.g. member 1 from stream 5 ending 31 December 1895 1800 UTC was not continuous with member 1 from stream 6 on 1 January 1896 0000 UTC.

Production data for almost all streams began from 1 January 0000 UTC fourteen months after the initial climatological ensemble assimilation, and continued for five years (Table III). One exception to this was production stream 16, spanning 1946 to 1951. Additional station pressure observations, courtesy of the Australian Bureau of Meteorology (B. Trewin, pers. comm., 2009), became available after 1951 had been originally generated in stream 17, so stream 16 was extended including the newly available observations. Additionally, stream 27 was extended to 2008 and will be updated to the present on a regular basis.

Unlike the individual ensemble members, the ensemble mean analysis is continuous across the stream boundaries, as expected. This continuity across production streams was tested by extending several streams past their fifth-year boundary into the January of a year 6 and comparing the ensemble mean analyses to the same dates from the year 1 of the overlapping actual production stream. Comparison of root-mean-square (r.m.s.) differences and visual inspection of these analyses with those from the production stream confirmed that the year 1 January analyses were not significantly different from those that would have been obtained by extending the previous stream (not shown).

6. Evaluation of synoptic variability

The 20CRv2 dataset represents the first attempt to generate synoptic analyses of the global troposphere back to 1871. It is also the first reanalysis dataset to estimate the uncertainty in the analysis fields at each analysis time. The ensemble nature of the Filter makes such an estimate particularly convenient compared to variational methods such as 3D-Var or 4D-Var. Additionally, the ensemble of analyses for any particular time can be viewed. Figure 2 illustrates some of these new aspects of this dataset, showing the 500 hPa geopotential height fields for several ensemble members on 29 January 1922 0000 UTC during the 'Knickerbocker' storm (Kocin and Uccellini, 2004), and also the ensemble mean analysis (corresponding to the line contours in Figure 1(b)) and the ensemble standard deviation (corresponding to the shading in Figure 1(b)). The fields illustrate that the reanalysis ensemble provides a sample of a probability distribution that includes probable upper-air fields that are dynamically consistent with concurrent and previous pressure observations. From the Monte Carlo aspect of the Ensemble Kalman Filter theory, each member is equally likely. When errors are distributed as a Gaussian, the ensemble mean analysis (Figure 2) gives the most likely state, while the ensemble standard deviation is one measure of the uncertainty in that state.

To evaluate whether the ensemble uncertainty varies as expected with the time-changing observation network, in Figure 3 we compare the expected error (red curves) to the actual r.m.s. differences of the first-guess ensemble mean and available observations (blue curves). The expected error is calculated as the sum of the observation-error variance and first-guess ensemble variance at the observation location $\left(\sigma_b{}^2 + R\right)$. This is averaged for all observations during each year, and the square root is taken to display the results in units of hPa. Similarly, the squared difference of the observation and the ensemble mean first guess at the observation location $\left(y^o - \overline{y}^b\right)^2$ is averaged over all observations and each year, and the square root is taken. Results are shown for both Northern Hemisphere ($\geq$20°N, Figure 3(a)) and the Southern Hemisphere ($\leq$20°S, Figure 3(b)). Also shown is the number of observations used in the analysis (black curves; note the log scale). The figure provides a demonstration of the good overall consistency of the expected and actual errors, and also their decrease in time over the twentieth century expected from increases in the number of observations.

Figure 2. Top row shows the 56-member ensemble mean analysis and the ensemble spread for 500 hPa geopotential height from Figure 1 for 0000 UTC on 29 January 1922. Subsequent rows show the 500 hPa geopotential height field from every second ensemble. The contour interval is 50 m, with a bold contour at 5550 m. For the ensemble spread, the contour interval is 5 m, with the 20 m line thickened.

Over the Northern Hemisphere, two periods of discrepancies between the actual and expected errors stand out: 1871 to 1890, and 1960 to 1975. The discrepancy in the earlier period is likely a result of setting the inflation factor r to 1.0 in the extratropical Northern Hemisphere when the first-guess spread is larger than 17 hPa anywhere in the region. In the late nineteenth century, the spread in the sparsely observed Arctic and North Pacific often exceeded this threshold, which triggered this setting and resulted in no inflation being applied over the extratropical Northern Hemisphere despite a reduction in spread occurring in the relatively observation-rich areas of the United States and Europe. In these areas, then, the spread became an underestimate of the first-guess uncertainty.

The discrepancy in the 1960 to 1975 period may have a different explanation. The errors in the marine pressure observations in this period may have been larger than specified in our reanalysis, as suggested through independent

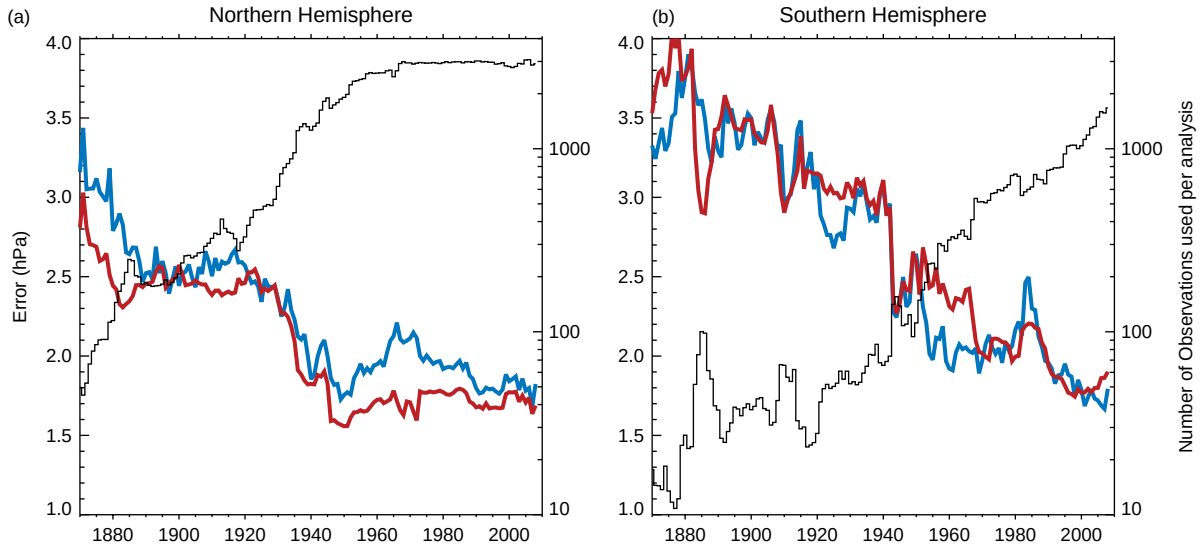

Figure 3. Time series of the 6-hourly first-guess root-mean-square (r.m.s.) difference from pressure observations (blue) and expected r.m.s. difference (red) calculated over individual years from 1870 to 2008 for the extratropical (a) Northern Hemisphere (20°N–90°N) and (b) Southern Hemisphere (20°S–90°S). The square root is calculated on the annual mean square values. The thin black curve shows the average number of pressure observations for each analysis in the indicated year (note the logarithmic scale).

analyses of the marine observations themselves (Kent and Berry, 2005; Chang, 2005, 2007). Assuming that the largest part of the observation error is the so-called error of representativeness, observation errors were kept fixed over the entire reanalysis period. The results of Kent and Berry (2005) and Chang (2005, 2007) combined with Figure 3(a) suggest that some time variation in the specified observation-error variance R may be necessary to improve the consistency of the actual and expected errors in Figure 3.

The reanalysis quality can be additionally assessed by considering the forecast skill for a longer lead time than the 6-hour lead time shown in Figure 3. Every 24 hours, using the 1200 UTC ensemble mean analysis as the initial condition, a 24-hour forecast of surface pressure and SLP was made as the 20CRv2 analyses were generated. Figure 4 shows time series of the forecast skill measured as the r.m.s. differences between the forecasts and surface pressure observations in the Northern and Southern Hemispheres poleward of 20° for the period 1870 to 2008. For comparison, the expected skill of a persistence forecast at each observation location is shown by the black curves. This persistence skill was estimated using the average squared 24-hour pressure tendencies in the period 1971 to 2000 from NNR at the locations of the observations used to verify a given forecast. Because of this, the persistence skill metric varies with the observation network. For the Northern Hemisphere, the 24-hour forecasts are superior to those expected from persistence even in 1870 and continue to improve over the period of record. For the Southern Hemisphere, forecast skill is similar to persistence starting in 1900 but is not better than persistence until about 1950.

It is also important to compare the 20CRv2 analyses with independent measurements. Figure 5 compares the 20CRv2 with a long record of upper-air data: a compiled series of kite, aircraft, balloon, and radiosonde observations of 500 hPa geopotential height and 850 hPa air temperature made from 1905 to 2006 at Lindenberg, Germany (Adam and Dier, 2005) from the Comprehensive Historical Upper Air Network dataset (CHUAN; Stickler *et al.*, 2010). The 20CRv2 analysis is interpolated to the time and location of these observations. The annual cycle has been computed separately for each series and removed to form anomalies. The figure shows that the correlation between the series is quite high, with correlations from the kite, aircraft, pilot balloon, and radiosonde compilation during the period 1905 to 1938 at 0.90 for the geopotential height and 0.86 for temperature, and correlations for the period of radiosonde measurements alone exceeding 0.96 for geopotential and 0.9 for temperature. (We note that the historical upper-air data likely have considerable uncertainties of their own, although quantification of these uncertainties is difficult). The high correlations, even for the early period, suggest that a high-quality analysis has indeed been achieved throughout the twentieth century for this relatively well-observed region. For Figure 5(a), the radiosonde based correlation is similar to the current pattern correlation skill of operation NWP forecasts of 500 hPa geopotential height made by the NCEP GFS three days in advance (NCEP, 2010).

In Figure 6, we examine the r.m.s. error of the 20CRv2 compared to the long series of frequent lower-tropospheric temperature measurements contained in Figure 5(b). Figure 6 shows the expected value of the r.m.s. error (red curves), calculated as the sum of the observation error variance, specified as $2K^2$ (Diamond *et al.*, 1938; Brönnimann, 2003), and the analysis ensemble variance at the observation location. This is averaged for all 850 hPa air temperatures from Lindenberg during each year, and the square root is taken to display the results in units of K. Similarly, the mean-square differences from Figure 5(b) are calculated over all observations and each year, and the square root is taken (blue curves). The actual and expected differences have similar decadal variability, consistent with the increasing analysis quality and improving measurement quality over the twentieth century. In contrast, the interannual agreement is fair before the 1930s, but poorer once the radiosonde era commences and 20CRv2 has a more consistent number of pressure observations (Figure 3(a)). A particularly noteworthy disagreement occurs in 1971 and 1972. Whether this is related to the changing radiosonde observing practices at this time (e.g. Kistler *et al.*, 2001), a

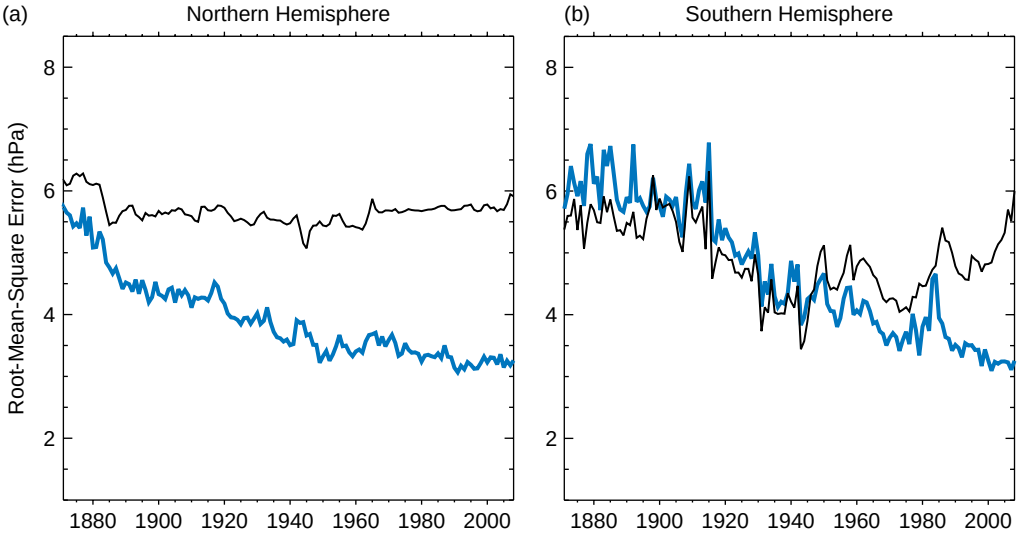

Figure 4. Time series of 24-hour r.m.s. forecast skill verified using all pressure observations (blue) and the expected skill from 24-hour persistence (black) from 1870 to 2008 calculated over the extratropical (a) Northern Hemisphere (20°N–90°N) and (b) Southern Hemisphere (20°S–90°S). The square root is calculated on the annually averaged squared differences between the observation and the forecast. The expected persistence skill is determined from the climatological variance of the 24-hour pressure tendency from NCEP–NCAR reanalyses at the locations of the observations.

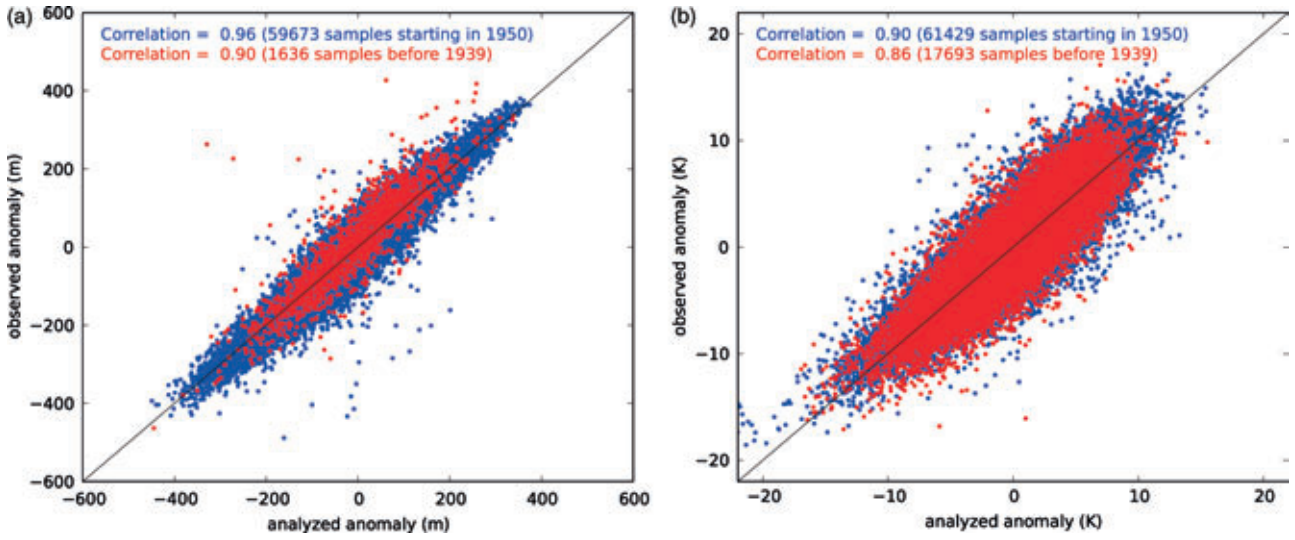

Figure 5. Comparison of (a) 500hPa geopotential height anomalies (m) and (b) 850hPa air temperature anomalies (K) at Lindenberg, Germany (52.22°N, 14.12°E) from upper-air observations in the Comprehensive Upper Air Historical Network (CHUAN; Stickler *et al.*, 2010) and 20CRv2 analyzed anomalies interpolated to the time and location of the observations. Anomalies are with respect to the mean annual cycle of the period shown. Blue dots show anomalies from 1950 to present, and red dots show anomalies from data spanning 1905 to 1938. There are no observations from this station in the CHUAN compilation from 1938 to 1949. The number of red dots prior to the commencement of regular radiosoundings at the Lindenberg Observatory (Dubois *et al.*, 2002) in 1932 is (a) 1166 and (b) 13 715.

change in instrumentation that occurred at the site in 1971 (Adam and Dier, 2005), or to a concurrent discrepancy in the 20CRv2 first-guess spread and error correspondence seen in Figure 3(a), is an issue of current investigation. We suspect that the change in instrumentation is the likely cause, given the good correspondence of the 1971 and 1972 radiosonde data from the nearby Berlin station (52.48°N, 13.4°E) with 20CRv2 (not shown). The generally lower values of actual error compared to expected error in the 1980s to the 2000s probably reflects the increased accuracy of the instrument, as the 20CRv2 use of observations in the extratropical Northern Hemisphere is nearly constant during this time (Figure 3(a)).

A global assessment of the quality of the 20CRv2 upper-air fields can be made by comparing the geopotential height fields with other reanalyses that also assimilated upper-air observations: ERA-40 (Uppala *et al.*, 2005) and NNR (Kalnay *et al.*, 1996). We first compare the 20CRv2 geopotential height fields with the ERA-40 fields for the period 1958–1978, i.e. before substantial amounts of satellite observations were available to the ERA-40 system (Figure 7). Shaded contours show the local anomaly correlations of the four-times-daily 20CRv2 and ERA-40 anomalies of 300 hPa geopotential height (anomalies are computed with respect to each dataset's annual cycle). We obtain similar results for comparisons with NNR, with higher correlations in the Southern Hemisphere (not shown). The correlations with both reanalysis datasets are also higher at lower levels in the troposphere (not shown). To illustrate where the ERA-40 might be considered a highly reliable estimate of the observed variability, the thick black line contour in Figure 7 shows the region poleward of which the correlation of the ERA-40 and NNR 300 hPa geopotential height anomalies is greater

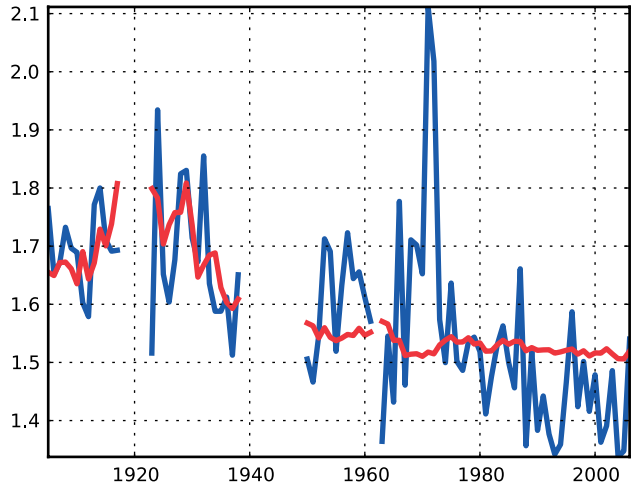

Figure 6. Time series of r.m.s. difference between 850 hPa air temperature anomalies at Lindenberg, Germany (52.22°N, 14.12°E) taken from upper-air observations in the Comprehensive Upper Air Historical Network (CHUAN; Stickler *et al.*, 2010) and 20CRv2 analyzed anomalies interpolated to the time and location of the observations shown in Figure 5 (blue curve). Also shown is the expected r.m.s. difference calculated over individual years from 1905 to 2008 (red curve). Anomalies are computed separately for each dataset with respect to the mean annual cycle of the period shown. The square root is calculated on the annual mean square values. Units are K.

than 0.975 for this period. Over most of this region, the sub-daily anomalies of 20CRv2 and ERA-40 correlate higher than 0.9, with some large areas, mainly over the Pacific and Atlantic storm track regions, correlating at higher than 0.95. Correlations over the Tropics and Oceania region exceed 0.65.

In contrast, away from the Southern Hemisphere midlatitude land regions where ERA-40 would have been expected to have assimilated radiosonde observations (Uppala *et al.*, 2005), the correlations of the 20CRv2 and ERA-40 anomalies are generally below 0.5. Indeed they are lower than expected from our previous feasibility study using observing system experiments (Compo *et al.*, 2006).

To further investigate this issue and expand the comparison with independent observations in general, Figure 8 shows correlations computed over the same 1958–1978 period, but between 20CRv2 and 300 hPa geopotential height observations obtained directly from the CHUAN dataset of radiosonde observations (Stickler *et al.*, 2010), which during this period consists of observations from the Integrated Global Radiosonde Archive (Durre *et al.*, 2006) with the RAOBCOAREv1.4 correction applied (Haimberger, 2007). Correlations are plotted in Figure 8 only if a sounding site contained at least 730 observations over the 21-year period. While the correlations over the Northern Hemisphere are consistent with those obtained with the upper-air-based reanalyses (e.g. Figure 7), correlations for the Tropics and extratropical Southern Hemisphere are not. In particular, correlations with the *in situ* observations from the Southern Hemisphere Extratropics are considerably higher than with ERA-40. For example, the correlation with the radiosonde data exceeds 0.85 for the station on the Antarctic peninsula where the correlations with ERA-40 are less than 0.5. In contrast, correlations with the *in situ* data near the Equator are lower than with ERA-40.

Correlations between 20CRv2 and ERA-40 for the period 1979 to 2001, when the latter used substantial amounts of satellite observations and a considerably larger number of radiosonde observations (Uppala *et al.*, 2005), are shown in Figure 9. The correlations are now high in both extratropical hemispheres, with values in some regions exceeding 0.95, and generally exceeding 0.9 in the Northern Hemisphere middle and high latitudes where ERA-40 and NNR agree well. The overall high correspondence over the globe, with most regions correlating at greater than 0.75, provides further evidence of the overall quality of the 20CRv2 fields when pressure observations are available.

Another common method of examining analysis quality is to examine the forecast skill compared to other forecasting systems. In Figure 10, the r.m.s. errors of 24-hour forecasts of marine SLP observations are shown separately for the Northern and Southern Hemispheres for three different sets of forecasts: those from ERA-40 (red curve), NNR (yellow curve), and 20CRv2 (blue curve). Note that NNR forecasts are actually 21-hour forecasts for the period 1948–1957, as analyses were generated at 1500 UTC instead of 1200 UTC during this period (Kistler *et al.*, 2001). Also note that while daily forecasts are shown for 20CRv2 and ERA-40, 12-hour and longer forecasts were generated only every 5 days in the NNR project. The expected r.m.s. errors of 24-hour persistence forecasts, computed at the observation locations as described previously, are also shown (black curve). It is readily apparent that over the Northern Hemisphere, the ERA-40 and NNR forecasts are superior to those of 20CRv2 for the entire period, with the exception of NNR forecasts during 1972, when the forecast skill was comparable. A similarly degraded forecast skill in the NNR system during the early 1970s is also seen in 5-day forecasts of 500 hPa heights verifying against the NNR analyzed fields (Kistler *et al.*, 2001). Comparing with the ERA-40 forecast skill, we interpret these results as suggestive of additional undetected coding errors (Kistler *et al.*, 2001) in the NNR radiosonde archive of this period that were corrected in the ERA-40 archive, rather than of real variations in predictability.

The Southern Hemisphere forecast skill in Figure 10(b) presents a surprising contrast to that of the Northern Hemisphere. Forecasts made with 20CRv2 initial conditions now have smaller errors than both persistence forecasts and forecasts made with NNR and ERA-40 initial conditions until about 1975 and 1979, respectively, and remain comparable until 1984. The ERA-40 and NNR begin to show improved skill with the assimilation of satellite data, with observations from the Vertical Temperature Profile Radiometer (VTPR) satellite being assimilated into ERA-40 starting in 1973 (Uppala *et al.*, 2005) and into the NNR starting in mid-1975 (R. Kistler, pers. comm., 2010). Note that the time variation of the persistence errors reflects the increasing density and expanded spatial coverage of marine observations for this region (e.g. Woodruff *et al.*, 2010). These lower 24-hour forecast errors for 20CRv2 are consistent with the high correlations of 20CRv2 to radiosonde observations seen in Figure 8. We note a peculiar degradation of the forecasts in 1983 for all three systems. This variation suggests either an issue with the underlying pressure observations or the prescribed SSTs, as 20CRv2 would have no dependency on any other observing system. An alternative explanation, that the lower forecast skill reflects a substantial interannual variation in predictability coincident with the outstanding El Niño episode of 1982–1983, seems unlikely given the absence of another such unanimous drop in forecast skill in any other year in Figure 8.

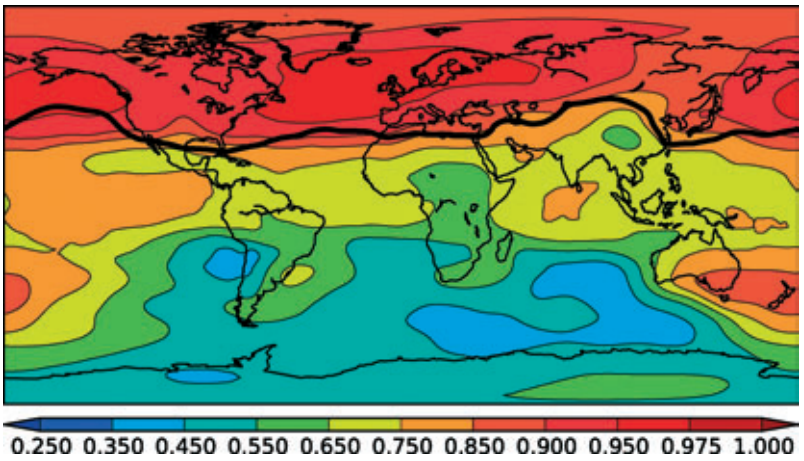

Figure 7. Map of the local anomaly correlation between four-times-daily anomalies of 300 hPa geopotential from ERA-40 and 20CRv2 over the period 1958 to 1978. Anomalies are computed separately for each dataset with respect to the mean annual cycle of the period shown. The area north of the thick black line is where ERA-40 and NCEP–NCAR reanalyses correlate highly ($\geq$0.975).

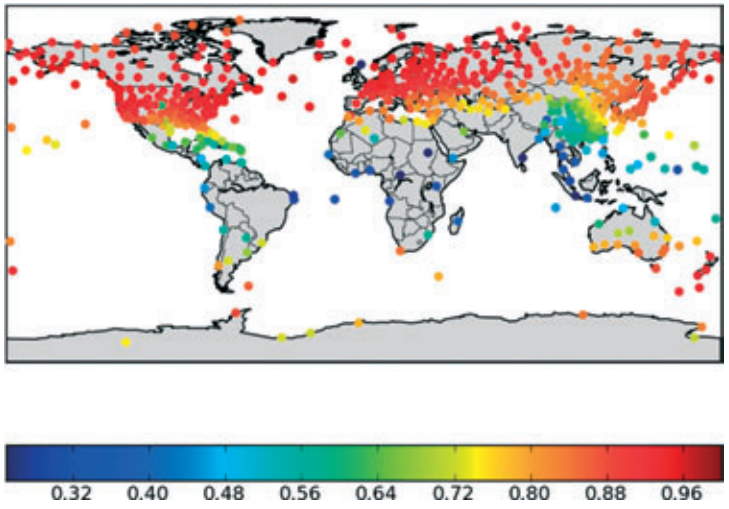

Figure 8. Correlation between 300 hPa geopotential height subdaily anomalies from 20CRv2 and radiosondes for 1958–1978. Anomalies are computed separately for each dataset with respect to the mean annual cycle of the period shown. Values are only shown for radiosonde stations having at least 730 observations during the 21-year period.

The results in Figures 7, 9, and 10 are indicative of improvement in the quality of ERA-40 in the Southern Hemisphere as satellite data become available (Bromwich and Fogt, 2004; Bengtsson *et al.*, 2004b; Uppala *et al.*, 2005). A reasonable hypothesis for this effect is that the use of constant background-error covariances that were consistent with higher-quality forecasts in the satellite era was inappropriate for, and reduced the quality of, the ERA-40 analyses in the pre-satellite era. The lower skill of 24-hour forecasts prior to the 1970s using ERA-40 and NNR initial conditions, compared to 20CRv2 initial conditions, in the Southern Hemisphere lends support to this hypothesis. To the extent that better upper-level fields lead to better SLP forecasts beyond a few hours, these results also suggest that the upper-tropospheric 20CRv2 fields before about 1975 may be more accurate in the Southern Extratropics.

Overall, the analysis quality shown in Figures 7–10 is largely consistent with previous observing system experiments using only surface pressure observations (Whitaker *et al.*, 2004; Compo *et al.*, 2006; Whitaker *et al.*, 2009). These results are expected to be representative of the 20CRv2 quality during early periods for regions where similar pressure data densities are available.

7. Representation of mean climate and climate variability

Assimilating only surface pressure and SLP observations and prescribing monthly-mean SST and sea-ice concentration, we have generated full three-dimensional estimates of the state of the troposphere every six hours from 1871 to 2008 in the 20CRv2. Figure 11 shows latitude–height sections of the zonally averaged 138-year mean zonal wind and temperature in this reanalysis dataset. The principal features of these zonally averaged quantities are generally as expected, and also generally consistent with those in the ERA-40 and the NNR datasets. The mean differences of the 20CRv2 from these datasets, over a common 1979 to 2001 post-satellite era, are detailed in Appendix A. One common difference worth highlighting here is a warm lower-tropospheric polar temperatures bias of the 20CRv2 with respect to both the ERA-40 and the NNR. We have identified this as an error arising from a mis-specification of the sea-ice concentration near coastal areas discussed in section 2, which will be corrected in future versions of our reanalysis.

We recognize that the reproduction of well-known climatological features in Figure 11, while reassuring, does not provide a hard test for this reanalysis: as already

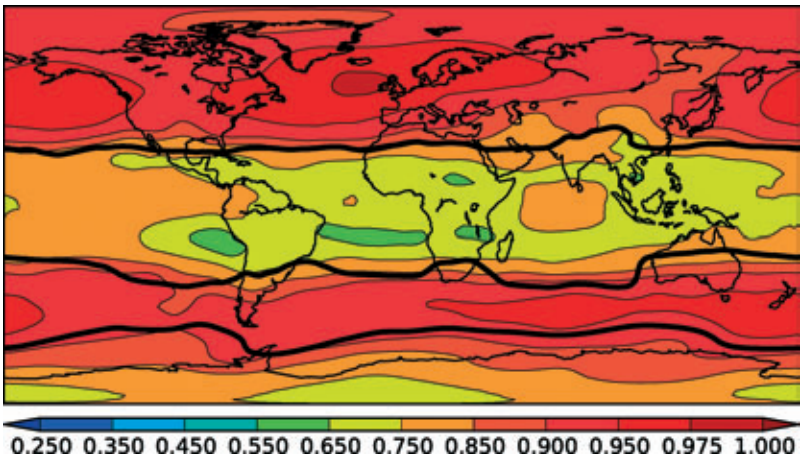

Figure 9. As Figure 7, but calculated for the period 1979–2001. In the Southern Hemisphere, the area between the two thick black lines is where ERA-40 and NCEP–NCAR reanalyses correlate highly (≥0.975).

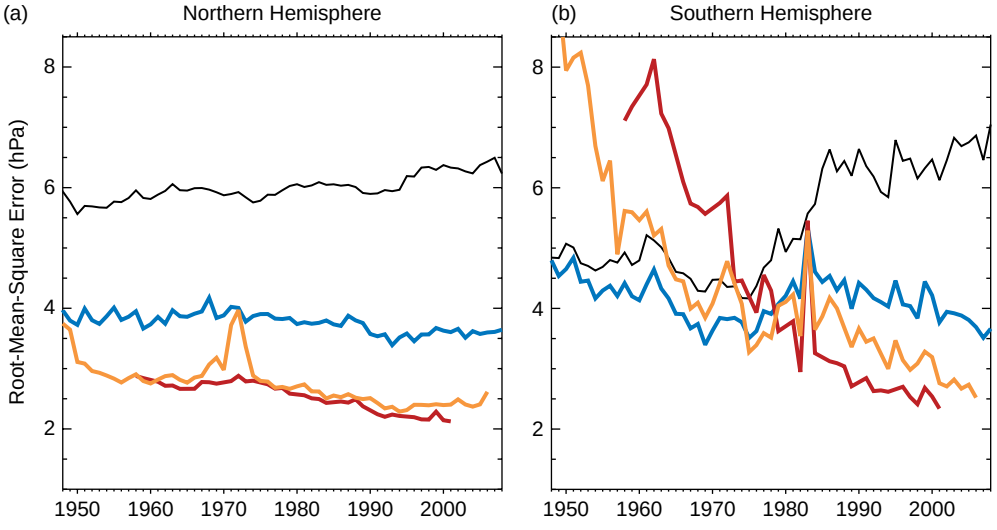

Figure 10. Time series showing 24-hour r.m.s. difference between 24-hour forecasts of SLP and observations of SLP from marine platforms over the extratropical (a) Northern Hemisphere (20°N–90°N) and (b) Southern Hemisphere (20°S–90°S). Forecast verification is shown for forecasts from20CRv2 (blue), ERA-40 (red), and NCEP–NCAR (orange) reanalyses. The thin black curve shows the expected skill of 24-hour persistence forecasts at the observation locations. Mean square differences are computed for each calendar year and then the square root is taken.

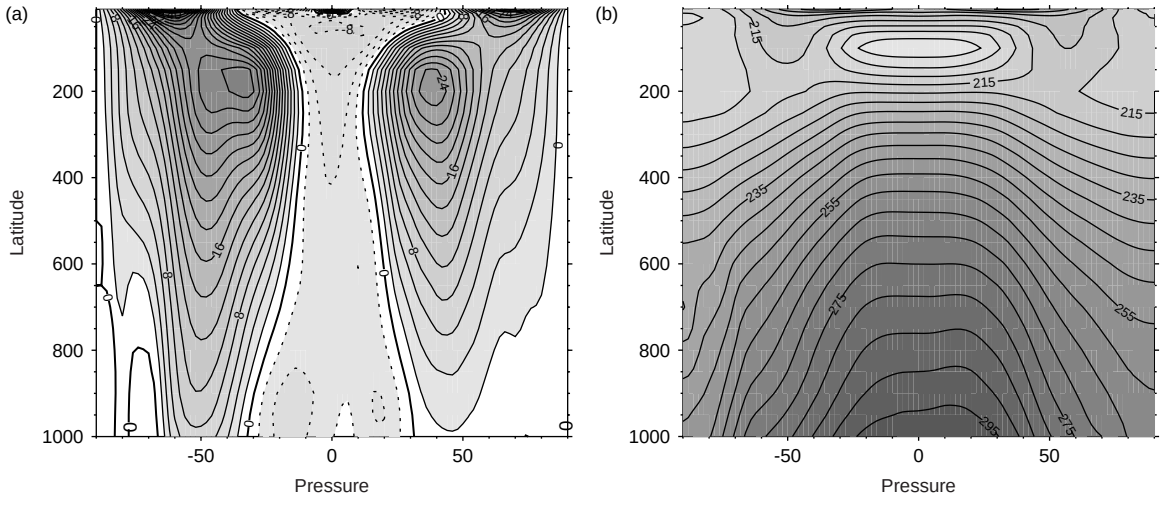

Figure 11. Zonal mean of (a) zonal wind speed and (b) air temperature from 20CRv2 averaged over the period 1871 to 2008. The contour interval is (a) 2 m s^{-1} and (b) 5 K. In (a), the zero contour is thickened and negative contours are dotted.

demonstrated in numerous studies, most of those features can also be captured in 'AMIP'-style atmospheric general circulation model (AGCM) integrations with prescribed observed SSTs and sea-ice concentrations alone (e.g. Gates *et al.*, 1999; Anderson *et al.*, 2004; Hurrell *et al.*, 2006), i.e. without any other observational input. A harder test would be the reproduction of the climatological features of synoptic variability. If the surface pressure observations were having no impact on our analyses, then the synoptic variability in our 56-member ensemble-mean analysis would be muted by the averaging out of random phase errors in each ensemble member, as is the case with the ensemble mean of 'AMIP' integrations. Indeed, for an infinite-member ensemble, the resulting variability would consist of low-frequency variations associated only with the imposed SST and sea-ice variations. It is therefore of considerable interest to examine the statistics of synoptic variability present in the *ensemble-mean* 20CRv2 analyses. Figure 12 provides important reassurance in this regard. It shows that the statistics of extratropical 'storm tracks' in the 20CRv2, as represented by the variance of 2–7 day band-pass filtered anomalies (e.g. Blackmon, 1976; Compo and Sardeshmukh, 2004) of 500 hPa geopotential height, are very similar in both the Northern and Southern Hemispheres to those obtained from NNR for the entire 1948–2008 period. Similar results are obtained (Figure 13) for the 'storm tracks' represented in the variance of the band-pass filtered 500 hPa vertical velocity. This is a pleasant surprise, in that the 500 hPa vertical velocity was not considered a well-analyzed variable in NNR (a so-called class 'B' quantity, in the terminology of Kalnay *et al.*, 1996), but the storm-track features are remarkably robust in the two datasets. Analysis of this 'vertical velocity storm track' measure is particularly useful; its interannual variability is closely related to that of precipitation (Compo and Sardeshmukh, 2004) and, as noted by Chang (2009), it does not appear to suffer from a Doppler effect observed in 'geopotential height storm track' measures (Burkhardt and James, 2006).

The storm tracks estimated from the ensemble-mean 20CRv2 analyses for the previous 61 years (1887–1947) are notably weaker than for the later (1948–2008) period, in terms of both 500 hPa 'geopotential height' (Figures 12(c, f)) and for 500 hPa 'vertical velocity' (Figure 13(c)) storm tracks. Such a result should not be taken as indicative of an actual climate change. Rather, as the observational density gets lower, less synoptic variability is present in the ensemble mean analyses as fewer observations are available to increment the ensemble-mean first guess in Eq. (1). In the limit of no observations, the 20CRv2 ensemble mean analysis fields become the ensemble mean of a 56-member AGCM integration forced by SST and sea ice. In this limit, the synoptic variability in Figures 12(c), 12(f), and 13(c) should be reduced to about $1/(56)^{0.5} = 0.13$ times that observed from a more complete observing system, or an individual ensemble member. The storm track estimates for 1887–1947 have considerably greater amplitude than this limit, particularly in the Northern Hemisphere, suggesting that even the sparse synoptic information in this period has made an impact on the analyzed synoptic variability. Nonetheless, it is clear that caution is indicated in investigations of the interannual variability and trends of this and other storm-track-related quantities using the ensemble-mean 20CRv2 analyses rather than each ensemble member. A Monte Carlo method to calculate derived quantities, such as tracking storm features, in which the statistic is calculated for each ensemble member and then averaged, may prove more fruitful in this regard.

Beyond the synoptic scale, the statistics of lower frequency variations, including blocking events, are also often represented in terms of the variance of band-pass-filtered anomalies (e.g. Blackmon, 1976; Compo *et al.*, 2001). Figure 14 shows the standard deviation of 500 hPa geopotential height anomalies in the 7.5–45 day period band. The 20CRv2 fields for 1948–2008 compare well with NNR fields albeit with slightly weaker maxima, similar to what was seen in Figure 12. Also as in Figure 12, the standard deviation for 1887 to 1947 is generally smaller than for 1948 to 2008 over the North Pacific and western North America, again reflecting the weaker variability of the ensemble-mean analyses in the case of sparse observations. Still, over the relatively observation-rich North Atlantic region, Figures 14(b,c) are quite similar in the two periods.

One might suspect that while synoptic and submonthly variability throughout the troposphere could be reasonably well captured using only synoptic surface observations, lower-frequency variability might be more poorly represented (Kanamitsu and Hwang, 2006). However, Figure 15 shows that the spatial patterns of the Northern Hemisphere monthly anomalies of 300 hPa geopotential height for 20CRv2 correspond well with those of ERA-40 and NNR, and continue to show the impact of the surface pressure observations in the reanalysis even at this relatively long time-scale beyond that associated with the prescribed observed SST and sea-ice boundary conditions. Time series of the pattern correlation between monthly anomalies from 20CRv2 and NNR for the months of December and June are shown. The results for all other months lie between these two extremes. The correlations with the ERA-40 fields for December and June are similar. The increase of June correlations from an average of 0.84 in 1958–1978 to an average of 0.89 in 1979–2001 most likely reflects the increasing use of satellite observations in ERA-40 and the NNR.

Figure 15 further demonstrates the impact of synoptic surface pressure observations even on upper-tropospheric monthly mean anomalies beyond that associated with prescribed observed SST and sea-ice boundary conditions. To assess whether the pattern correlations of the 20CRv2 fields with the ERA-40 and NNR fields merely reflect the correlations associated with response patterns to boundary forcing, the same pattern correlations were computed, but in a perfect model context using a 24-member ensemble of ECHAM4.5 AGCM (Roeckner *et al.*, 1996) integrations generated by the International Research Institute using prescribed observed SSTs. To calculate the perfect model curves, the monthly-mean 300 hPa geopotential height anomaly field from a single ensemble member was treated as the 'observed' field and the mean of the remaining 23 anomaly fields was correlated with it. This procedure was repeated for all 24 members and all months. The average pattern correlation is plotted for the months of December and June. They are clearly much lower than the pattern correlations of the 20CRv2 with the ERA-40 and NNR fields. The slight apparent increase in those correlations from $\sim$0.18 before 1979 to $\sim$0.23 afterwards is not statistically significant at the 5% level.

Climate variability is often represented by climate indices of seasonally averaged data (e.g. as recently reviewed by Brönnimann *et al.*, 2009). In Figure 16, we examine

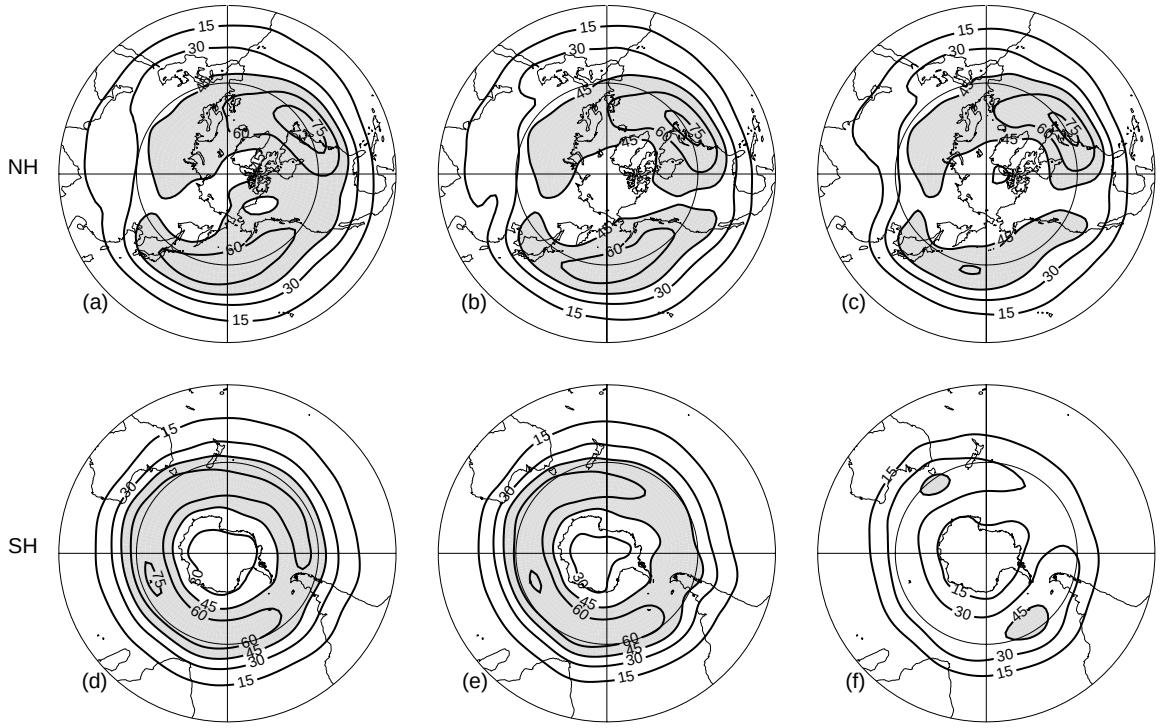

Figure 12. Climatological winter mean (December–February) variance of 2–7-day bandpass filtered Northern Hemisphere 500 hPa geopotential height. Note that the square root of each field is plotted. In (a) and (b), the climatology is calculated for the period 1948–2008 using (a) NCEP–NCAR and (b) 20CRv2 reanalysis data. (c) is as (b) but for the period 1887–1947. (d)–(f) are as (a)–(c), but for the Southern Hemisphere. Contour interval in all panels is 15 m.

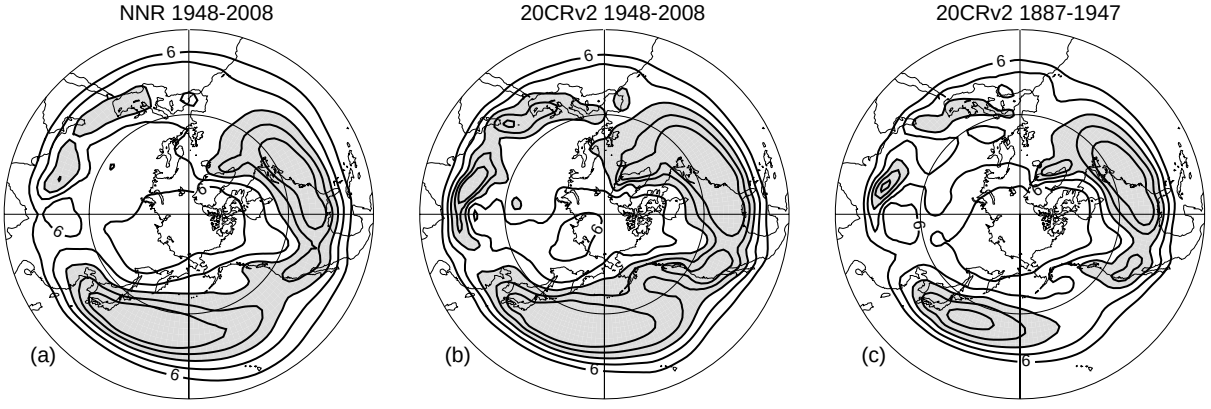

Figure 13. As Figure 12, but for the 500 hPa vertical velocity (omega) for the Northern Hemisphere only. The contour interval in all panels is 2 cPa s^{-1}.

three climate indices calculated from the 20CRv2 and compare to five other estimates. Time series representing the Pacific Walker Circulation index (PWC; Figure 16(a)), the North Atlantic Oscillation (NAO; Figure 16(b)), and the Pacific–North America Pattern (PNA; Figure 16(c)) are shown from the reanalysis datasets of 20CRv2, ERA-40, NNR, and ERA-Interim, from statistical reconstructions, and from a nine-member ensemble of boundary- and chemistry-forced integrations using the Solar Climate Ozone Links (SOCOL) atmospheric-chemical GCM (Schraner *et al.*, 2008). The SOCOL chemistry-climate GCM was integrated in the 'all forcings' configuration over the period 1901–1999 (Fischer *et al.*, 2008), with a T30 horizontal truncation and 39 vertical levels. Each integration had prescribed boundary conditions of monthly mean SST and sea ice from HadISST (Rayner *et al.*, 2003) and prescribed land surface conditions, stratospheric aerosols, solar variability, surface concentrations of greenhouse gases and ozone-depleting substances, emissions of short-lived species, and the Quasi-Biennial Oscillation in the stratosphere. Their climate variability was previously examined by (e.g.) Brönnimann *et al.* (2009) and Scaife *et al.* (2009).

The definitions of the indices are the same as used in Brönnimann *et al.* (2009). The PWC is defined following Oort and Yienger (1996) as the difference in the area-averaged 500 hPa vertical pressure velocity between the regions of (10°S–10°N, 180–100°W) and (10°S–10°N, 100–150°E). The NAO is defined as the difference in the standardized monthly SLP anomalies at Ponta Delgada (Azores) and Reykjavik (Iceland). As the datasets are gridded, the nearest grid points are used. The PNA is defined following Wallace and Gutzler (1981) as

$$PNA = \frac{(Z_P - Z_Q + Z_R - Z_S)}{4}, \qquad (8)$$

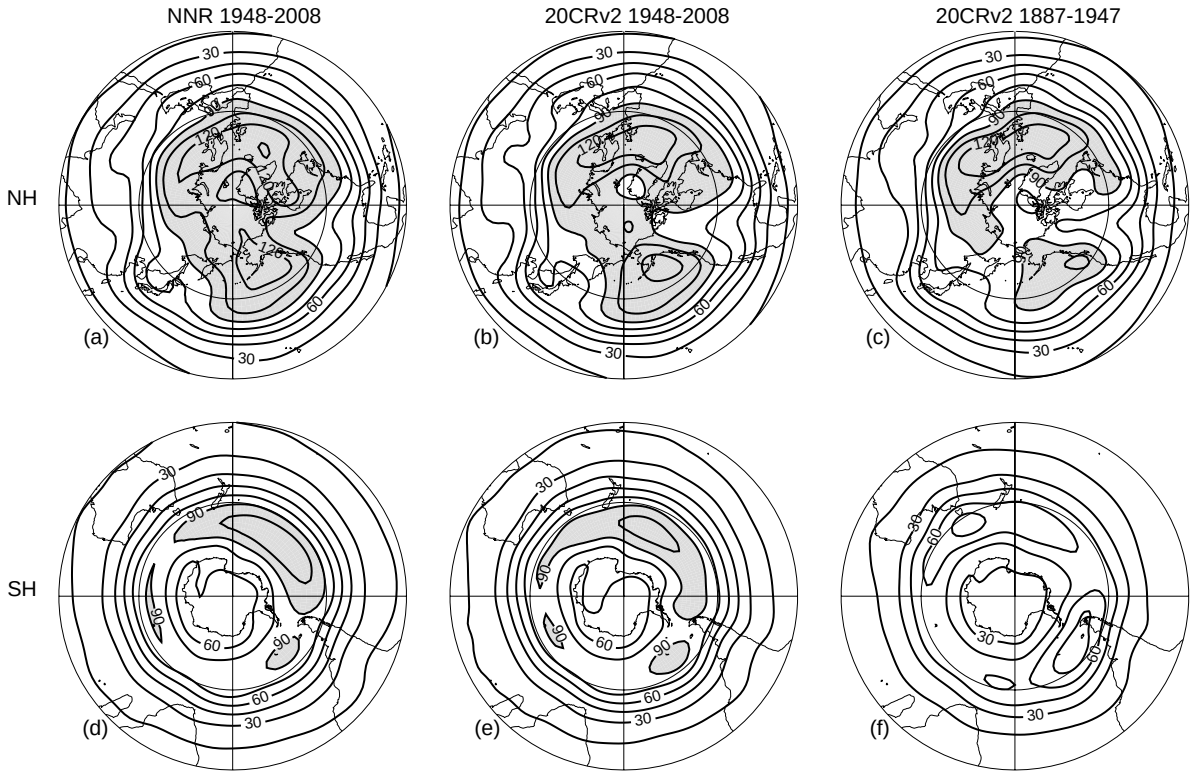

Figure 14. As Figure 12, but showing the square root of the variance of 7.5–45-day bandpass filtered 500 hPa geopotential height.

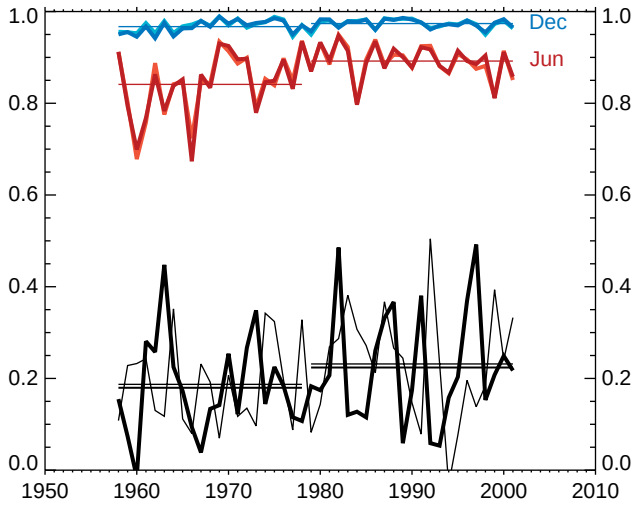

Figure 15. Time series showing anomaly pattern correlations between monthly mean anomaly fields of Northern Hemisphere extratropical 300 hPa geopotential height from two upper-air-based reanalyses and 20CRv2 for the months of December (cool colours) and June (warm colours). Correlations with ERA-40 (NCEP–NCAR) reanalysis fields are shown by the blue (cyan) and red (orange) curves. For each ERA-40 series, the horizontal coloured lines show the mean value of the correlations when averaged before the period of substantial satellite observations (1958–1978) and then afterwards (1979–2001). The black curves show the expected anomaly pattern correlations when only the observed sea-surface temperature fields are available for the months of (thick) December and (thin) June. The horizontal black lines show the mean value of these expected correlations, with the thicker line for December and the thinner for June. Anomalies are with respect to the mean annual cycle of the period shown.

where Z is the standardized 500 hPa monthly geopotential height anomaly at points P (20°N, 160°W), Q (45°N, 165°W), R (55°N, 115°W), and S (30°N, 85°W). All anomalies were determined with respect to the maximal overlap period for all datasets (1989–1999). Note that the statistical reconstructions (RECs) of the PWC (Brönnimann *et al.*, 2009) and PNA (Griesser *et al.*, 2010) span the period of 1901 to 1947 and 1880 to 1957, respectively. Unlike the reanalyses and SOCOL integrations, neither REC uses SST information, employing only the available monthly-mean SLP, land air temperature, and upper-air observations (Brönnimann *et al.*, 2009; Griesser *et al.*, 2010). The statistics for reconstructing PWC are referenced to the NNR (1948–2004) and those for the PNA are referenced to the ERA-40 dataset (1958–2001). The Hadley Centre sea-level pressure dataset (HadSLP2; Allan and Ansell, 2006), used as the statistical reconstruction for the NAO index, spans the full period. HadSLP2 also does not use SST information, incorporating only monthly mean SLP observations from marine and land platforms.

Correlations for the period of overlap between the estimates of the indices from the different datasets are shown in Table IV. It is no surprise that the SLP-based NAO agrees extremely well with the observational estimates, all of which use pressure observations. For example, the correlation between ERA-40 and 20CRv2 is 0.998 for the 1958–2001 period. In contrast, the forced SOCOL integration agrees poorly with all of the observational estimates, consistent with relatively weak SST forcing of the NAO (e.g. Scaife *et al.*, 2009). Interestingly, there is also good agreement among the observational estimates for the time series representing the upper-air pattern of the PNA. In all cases, the observation-based estimates are in better agreement with each other than they are with the anomalies from the SOCOL integrations. While the relatively high correlation of 0.701 between the SOCOL PNA and ERA-40 PNA provides evidence of the well-known SST influence on this pattern of variability (e.g. Alexander *et al.*, 2002; Barsugli and Sardeshmukh, 2002), the substantially higher

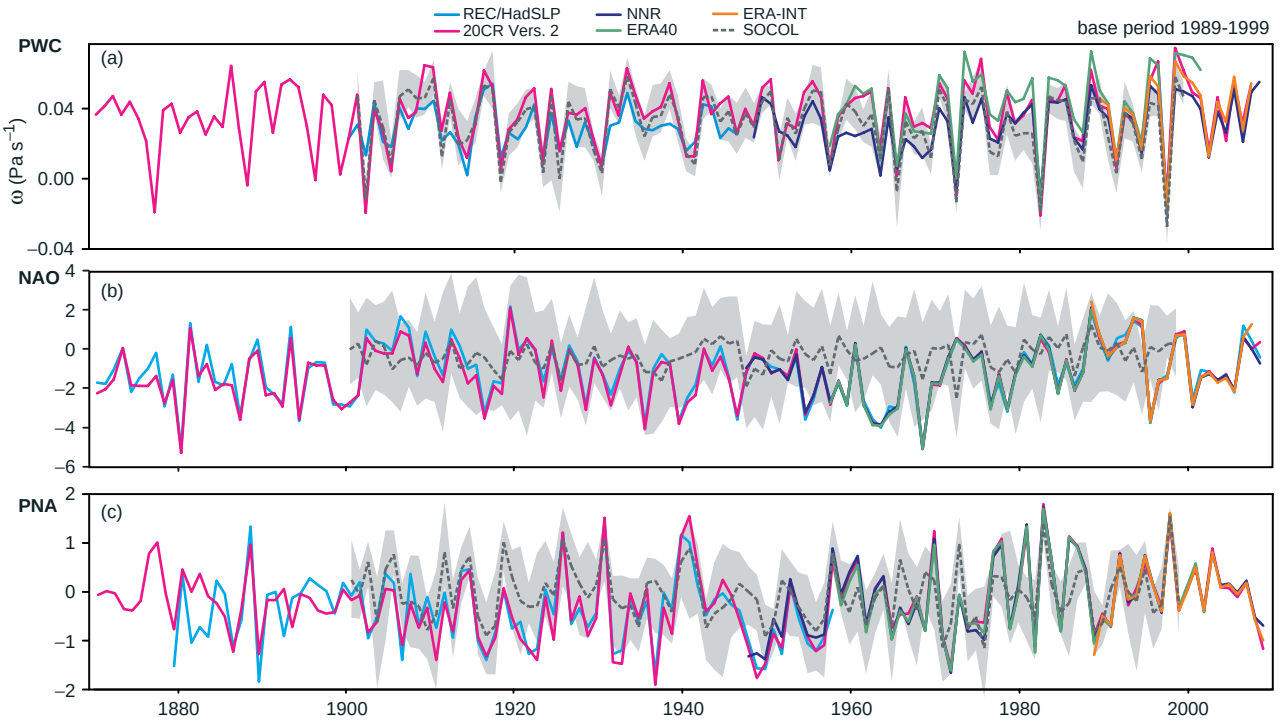

Figure 16. Time series of seasonally averaged climate indices representing (a) the tropical September to January Pacific Walker Circulation (PWC), (b) the December to March North Atlantic Oscillation (NAO), and (c) the December to March Pacific North America (PNA) pattern. Indices are calculated from various sources: 20CRv2 (pink); statistical reconstructions using Bronnimann *et al.* (2009) for the PWC, Griesser *et al.* (2010) for the PNA, and HadSLP2 (Allan and Ansell, 2006) for the NAO (all cyan); NCEP–NCAR reanalyses (NNR; dark blue); ERA-40 (green); ERA-Interim (orange); and SOCOL ensemble mean (dark grey). The light grey shading indicates the minimum and maximum range of the SOCOL ensemble. All indices are computed with respect to the overlapping 1989–1999 period. Indices are defined as in Brönnimann *et al.* (2009).

correlation of 0.992 of the 20CRv2 with ERA-40 attests to the considerable observational information in the 20CRv2.

In contrast to the other two indices, the PWC index agrees well with all the estimates, including those from SOCOL. This is not surprising, as there is a considerable SST influence on the variations of the Walker Circulation (e.g. Oort and Yienger, 1996). Note that the higher REC correlation with SOCOL than with 20CRv2 is not statistically significant at even the 20% level, assuming 45 degrees of freedom (the number of years of overlap between the PWC REC and SOCOL minus two).

Considering longer-term variations in Figure 16, it is readily apparent that none of the indices have demonstrable trends over the 1871 to 2008 period of 20CRv2, and a more rigorous statistical trends significance analysis supports this (not shown). Additional studies incorporating the uncertainty estimates in the 20CRv2 are under way to quantify further any significant decadal variability in Figure 16.

The results in Figures 12–16 demonstrate that the 20CRv2 reanalysis has successfully incorporated the information in synoptic surface pressure observations and its beneficial impact on estimates of the global tropospheric circulation, not only on the synoptic but also much longer time-scales. We end this section with a tantalizing look at perhaps the hardest test for such a surface-pressure-based reanalysis system: its ability to represent the mean hydroclimate and its variability. Figure 17(a) compares the 1980 to 2000 mean of zonally averaged precipitation P in the 20CRv2 and the Global Precipitation Climatology Project (GPCP, Adler *et al.*, 2003) v.2 datasets. The comparison is generally favourable, and within the uncertainties estimated from intercomparisons among other observational precipitation datasets (e.g. Adler *et al.*, 2003). Figure 17(b) shows the 1980–2000 mean of zonally averaged precipitation minus evaporation, $P-E$, in the 20CRv2, and also its change $\Delta(P-E)$ from that during the first 20-year period (1871–1891) of the reanalysis. The surprise here is that the meridional structure of $\Delta(P-E)$ does not resemble that of $P-E$ itself. Such a resemblance might have been anticipated from simple arguments and climate model simulations (Held and Soden, 2006) as a 'robust' feature of the global hydrological response to global warming. Indeed at 10°N the sign of $\Delta(P-E)$ is opposite to that of $P-E$. Assessing the realism of such aspects of the 20CRv2 and other historical reanalysis datasets will clearly continue to be of interest.

8. Summary and concluding remarks

To begin to address the needs of climate science for reanalysis products spanning the instrumental record, NOAA's Earth System Research Laboratory (ESRL) Physical Sciences Division and the University of Colorado CIRES Climate Diagnostics Center have led an international Twentieth Century Reanalysis Project to produce version 2 of the reanalysis (20CRv2) using US DOE supercomputers. The 20CRv2 dataset contains the first synoptic-observation-based estimate of global tropospheric variability spanning 1871 to 2008, and is derived using only observations of synoptic surface pressure and prescribing monthly SST and sea-ice distributions as boundary conditions for the atmosphere. The beneficial impact of these synoptic observations on the analysis of the tropospheric circulation has been demonstrated not only on the synoptic but on even much longer time-scales.

Table IV. Correlations between various estimates of seasonal indices[a,b,c] of climate variability for the full period of overlap from each source[d].

PWC[a]	20CRv2	ERA-40	ERA-Interim	REC	NNR	SOCOL
20CRv2	1	0.951	0.968	0.788	0.904	0.927
ERA-40		1	0.989	NA	0.962	0.935
ERA-Interim			1	NA	0.991	0.975
REC				1	NA	0.828
NNR					1	0.880
SOCOL						1
NAO[b]	20CRv2	ERA-40	ERA-Interim	HadSLP2	NNR	SOCOL
20CRv2	1	0.998	0.983	0.976	0.993	0.252
ERA-40		1	0.997	0.992	0.999	0.293
ERA-Interim			1	0.980	0.983	0.090
HadSLP2				1	0.989	0.213
NNR					1	0.250
SOCOL						1
PNA[c]	20CRv2	ERA-40	ERA-Interim	REC	NNR	SOCOL
20CRv2	1	0.992	0.982	0.876	0.981	0.618
ERA-40		1	0.995	NA	0.993	0.701
ERA-Interim			1	NA	0.992	0.895
REC				1	0.932	0.622
NNR					1	0.691
SOCOL						1

[a] Pacific Walker Circulation; [b] North Atlantic Oscillation; [c] Pacific–North America pattern. [d] Time periods used are 20CRv2 (1871–2007), ERA-40 (1958–2001), ERA-Interim (1989–2007), REC (PWC:1901–1947, PNA:1880–1957), HadSLP2 (1871–2007), NNR (1948–2007), SOCOL (1901–1999).

The first version of the dataset (20CRv1) has already been used to investigate issues as diverse as the US Dust Bowl (Cook *et al.*, 2010), the early twentieth century Arctic warming (Wood and Overland, 2010), historical ENSO events (Giese *et al.*, 2010), decadal Atlantic hurricane variability (Emanuel, 2010), and ocean ecology (Baird *et al.*, 2010). The 138-year span of the 20CRv2 dataset should make it even more useful for a variety of climate applications ranging from assessments of storm track and extreme event variations to studies of drought and decadal variability to investigations into meteorological history. The dataset may also be useful for detecting inhomogeneities in independent observed time series, such as from radiosondes, utilizing, e.g., the method of Haimberger (2007).

The analysis fields are available at six-hourly temporal and 2° horizontal resolution and 24 pressure levels from NOAA ESRL (http://www.esrl.noaa.gov/psd/data/20thC_Rean/) and NCAR (http://dss.ucar.edu/datasets/ds131.1) and will also be distributed via the NOAA National Model Archive and Distribution System (NOMADS; http://nomads.ncdc.noaa.gov). Selected fields from individual ensemble members will be made available via the NERSC science gateway (http://portal.nersc.gov), NCAR, and NOMADS. Additionally, the complete spectral files for every ensemble member were archived, so any additional variable can be obtained for every member. These will be made available by courtesy of NCAR and NOMADS. Finally, the results of the assimilation of each observation in the ISPD, including the first guess and analysis uncertainty at the observation location (σ_b^2 and σ_a^2, respectively), are also available by courtesy of NCAR (http://dss.ucar.edu/datasets/ds132.0).

With the production of the 20CRv2 version of the dataset complete, it is reasonable to consider how, in addition to using an improved NWP model in future versions, the dataset could be further improved. It is readily apparent that the current quality depends on the availability of pressure observations. We have demonstrated that, as more pressure observations become available, the reanalysis fields improve and become more certain (e.g. Figures 1, 3, 4, and 6). Building on a history of national data rescue projects (e.g. Manabe, 1999) and international exchange of marine and terrestrial data, recent 'data archaeology' by university researchers, national meteorological services, and ACRE and its partners suggests that millions of additional surface pressure observations from the early twentieth century and nineteenth century remain to be digitized. As an example, only 62 land stations were used to generate the 1871 fields of 20CRv2. Our recent research has uncovered 363 additional stations that could have been used if the data had been digitized. Considering even earlier periods, García-Herrera *et al.* (2005), for example, document an effort that digitized eighteenth and nineteenth century European national marine meteorological observations, and Wilkinson *et al.* (2010) describe marine SLP pressure observations becoming available from the British East India Company for 1790–1834 and a range of other sources (also Woodruff *et al.*, 2005). Przybylak (2009) describes the network of Polish meteorological stations back to the eighteenth century, part of the network of European stations measuring pressure extending back even to the seventeenth century (Jones, 2001). Organizations such as the World Meteorological Organization and its Joint Technical Commission (with the Intergovernmental Oceanographic Commission) for Oceanography and Marine Meteorology

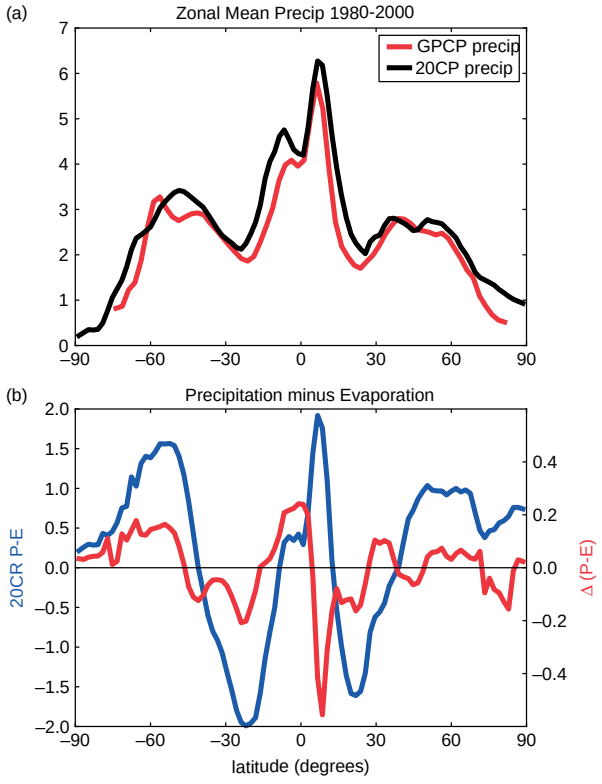

Figure 17. (a) Zonal mean precipitation rate P (mm day^{-1}) averaged over 1980–2000 from 20CRv2 (black curve) and from GPCPv2 (red curve) datasets. (b) Zonal mean P minus evaporation rate E averaged over 1980–2000 from 20CRv2 (blue curve), and the 1980–2000 $P-E$ average minus the 1871–1891 $P-E$ average (red curve).

(JCOMM), the NOAA Climate Database Modernization Program (CDMP, Dupigny-Giroux *et al.*, 2007), the International Environmental Data Rescue Organization, universities, national meteorological services, and ACRE are working to recover these observations and uncover additional observations over land and ocean (e.g. Brunet and Kuglitsch, 2008). Taking advantage of other observation types, such as near-surface wind or temperature may provide another fruitful avenue for improvement (Anderson *et al.*, 2005).

Several components of the current algorithm concerned with accounting for uncertainties arising from the use of a finite ensemble, an imperfect model, and imperfect observations are rather simplistic. In all three areas, there are opportunities for improvement. Instead of the current covariance localization and inflation, a more sophisticated accounting of model and sampling errors, such as using perturbed model parametrizations (e.g. Houtekamer *et al.*, 2009), or prescribing spatially coherent additive stochastic noise (e.g. Whitaker *et al.*, 2008), may reduce the discrepancy between the expected and actual errors in Figures 3 and 6. Allowing the observation error to vary adaptively (Li *et al.*, 2009) may account for the apparent time variations in that quantity, particularly as shown for marine observations (Kent and Berry, 2005; Chang 2005, 2007). Known issues with balance in the Ensemble Kalman Filter (e.g. Mitchell *et al.*, 2002) are indirectly addressed with our rather large localization distance of 4000 km, but additional steps to maintain balance could be implemented, such as the so-called 'Chi correction' (Sardeshmukh, 1993). The uncertainties in the prescribed SST, sea-ice concentration, CO_2, volcanic aerosols, and solar variations are also only mildly accounted for with the present covariance inflation. A Monte Carlo approach that accounts for all sources of uncertainty in these components of the first guess may also improve the representation of uncertainty and the correspondence of the actual and expected errors.

Additionally, using a broader time window may improve the analyses. The implementation of the Ensemble Kalman Filter used here is actually a smoother (Sakov *et al.*, 2010), since it incorporates observations up to 3 hours after the analysis time. Khare *et al.* (2008) have proposed an efficient lagged ensemble smoother algorithm, potentially extending the time window of observations used out to many hours or even days after the analysis time. While errors in the representation of the covariances associated with such long times will be an obstacle, the use of a wider observation window may provide a way to extract more information from sparse observations, particularly from mobile marine platforms. The ERA-CLIM project, which will develop the next ECMWF reanalysis back to 1900, will investigate the benefit of using longer time windows in the analysis step (D. Dee, pers. comm., 2010).

The overall quality of the 20CRv2 dataset may surprise some readers. While the relevance for weather studies appears to be consistent with that anticipated from advanced data assimilation algorithms applied in observing system experiments using only surface observations (Whitaker *et al.*, 2004, 2009; Anderson *et al.*, 2005; Thépaut, 2006; Compo *et al.*, 2006), the relevance for climate studies, e.g. as suggested by the high correlations of monthly-mean anomalies in Figure 15 and climate indices in Figure 16, could not have been anticipated from those short feasibility experiments. The ability to generate skilful 24-hour forecasts of surface pressure (relative to persistence forecasts) even in years as data-poor as 1871 was another pleasant surprise. Still, many additional evaluations could be performed. These are planned as future work, e.g. comparisons with the EMSLP daily SLP dataset which extends from 1850 to the present (Ansell *et al.*, 2006), and with the monthly mean land temperature datasets produced by NCDC (Smith *et al.*, 2008), NASA Goddard Institute for Space Studies (Hansen *et al.*, 2001), and the Climatic Research Unit of the University of East Anglia (Brohan *et al.*, 2006).

With further improvements to the analysis algorithm and the potential for utilizing additional recoverable observations, a backward extension of the dataset to the start of the nineteenth century or even earlier seems plausible. For such earlier periods, the quality of the dataset in the Northern Hemisphere could be expected to be comparable to that shown here for the Southern Hemisphere in the late nineteenth century. We note that in his original 'reanalysis', Brandes (1820) constructed synoptic maps of 1783 using the Meteorological Society of the Palatinate observations, a collection of extensive weather observations, including pressure, spanning 1781 to 1792 taken by weather observers in 18 countries on both sides of the North Atlantic (Monmonier, 1999). Additional observing-system and observing-system-simulation experiments, with careful assessment of the quality of the recovered historical observations, will be needed to provide scientific support for such an undertaking.

Acknowledgements

The Twentieth Century Reanalysis Project received support and observational data from many people, organizations, and projects. The NOAA NCEP/EMC staff's years of work improving the GFS, and particularly the help of S. Moorthi, H.L. Pan, S. Saha, R. Kistler, and S. Lord is gratefully acknowledged. The NOAA ESRL/Physical Sciences Division (PSD) and CIRES/Climate Diagnostics Center (CDC) IT staff provided invaluable computer support, especially K. Healy, R. Jesse, A. McColl, C. McColl, B. McInnes, and N. Wilde. Assistance with ISPD calculations by C. McColl is also gratefully acknowledged. The work of D. Hooper and C. Smith to make the dataset available at NOAA ESRL/PSD and CIRES/CDC, and the work of the NCAR Data Support Section, especially J. Comeaux, D. Schuster and C.-F. Shi, to make the dataset available at NCAR is gratefully acknowledged. Consulting and support provided by the staff of the US DOE National Energy Research Scientific Computing Center, especially H. He and F. Verdier, and by the staff of the US DOE National Center for Computational Sciences, especially I. Carpenter and D. Kothe, is gratefully acknowledged. Useful discussions are acknowledged with colleagues in ESRL/PSD and CIRES/CDC (especially T. Hamill, R. Webb, R. Dole, C. McColl, M. Newman, and W. Neff), colleagues at NCAR (especially J. Anderson, J. Comeaux, R. Jenne, and K. Trenberth), colleagues at ECMWF (especially D. Dee, A. Simmons, and J-N. Thépaut), colleagues at the UK Met Office Hadley Centre (especially P. Brohan and T. Ansell), colleagues at the University of Maryland (especially P. Arkin and E. Kalnay), and colleagues at the US DOE (especially A. Bamzai and K. Yelick). The US Department of Commerce-Boulder and NCAR librarians kindly assisted with manuscript research. The authors would like to thank the organizations and projects listed in Table II and the following individuals for invaluable assistance in exchanging observations for the ISPD: W. Adam of the Lindenberg Observatory, DWD; K. Andsager of NOAA's Midwestern Regional Climate Center; T. Brandsma of KNMI; J. Burroughs, S. Doty, J. Elms, K. R. Knapp, D. H. Levinson, N. Lott, T. C. Peterson, and R. Truesdell of NOAA's NCDC; J. Comeaux and C.-F. Shi of NCAR; F. Le Blancq of the Jersey Met Service; S. J. Lubker of NOAA ESRL; G. Lentini of Universita degli Studi Milano; N. Nichols of Monash University; L. Srnec of the Meteorological and Hydrological Service of Croatia; A. Stickler of ETH Zurich for CHUAN; V. Swail of Environment Canada; B. Trewin and D. Jones of the Australian Bureau of Meteorology; D. Tse of the Hong Kong Observatory; M. A. Valente of the University of Lisbon and M. Barros of the University of Porto for SIGN contributions; J. S. Woollen of NOAA/NCEP; and V. Wagner and R. Zoellner of DWD. The IBTrACS dataset benefitted from the considerable work of C. Landsea and the HURDAT Reanalysis Project. The assembly of the ISPD under the auspices of the GCOS AOPC/OOPC Working Group on Surface Pressure and the WCRP/GCOS Working Group on Observational Data Sets for Reanalysis by NOAA/ESRL, NOAA's National Climatic Data Center (NCDC), and the Climate Diagnostics Center (CDC) of the University of Colorado's Cooperative Institute for Research in Environmental Sciences (CIRES) is gratefully acknowledged. The HadISST fields are courtesy of N. Rayner, BADC, and the UK Met Office Hadley Centre. The assistance of M. Fiorino with daily interpolation of HadISST fields is acknowledged. Access to the ECHAM4.5 AGCM integrations was courtesy of the IRI data library. The ERA-40 and ERA-Interim datasets are courtesy of ECMWF. Access to the NCEP–NCAR Reanalysis dataset is courtesy of NOAA ESRL/PSD. The authors would like to thank referees D. Dee, H. Mitchell, and an anonymous reviewer for constructive comments on an earlier version of the manuscript. R. J. Allan receives support under ACRE from the Queensland Climate Change Centre of Excellence, and from the UK Joint Department of Energy and Climate Change and Department for Environment, Food and Rural Affairs Integrated Climate Programme (DECC/Defra GA01101). NCAR is sponsored by the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation. S. Brönnimann and A. N. Grant were supported by the Swiss National Science Foundation. Support for the Twentieth Century Reanalysis Project dataset is provided by the US Department of Energy, Office of Science Innovative and Novel Computational Impact on Theory and Experiment (DOE INCITE) program, and Office of Biological and Environmental Research (BER), and by the NOAA Climate Goal. The project used resources of the National Energy Research Scientific Computing Center and of the National Center for Computational Sciences at Oak Ridge National Laboratory, which are supported by the Office of Science of the US Department of Energy under contract No. DE-AC02-05CH11231 and contract No. DE-AC05-00OR22725, respectively.

Appendix A: Systematic Differences with Previous Reanalyses (NNR and ERA-40)

The overall structure of the time-mean flow shown in Figure 11 suggests that the basic features of the general circulation are captured in the 20CRv2 dataset. Such a result would be expected even for three-dimensional fields generated by an AGCM forced solely with SSTs (e.g. Gates *et al.*, 1999; Anderson *et al.*, 2004; Hurrell *et al.*, 2006). Figure A1 shows that the features are in general agreement with those estimated from ERA-40 and NNR. Figure A1(a) shows the difference between the 20CRv2 and ERA-40 zonal mean zonal wind speed averaged over 1979–2001. Away from the southern polar troposphere, the overall biases in the troposphere are relatively small. Similar results are seen when comparing 20CRv2 to NNR fields (Figure A1(b)). The south polar tropospheric wind biases throughout the column are smaller relative to NNR than to ERA-40. In this same vein, tropical biases are of opposite sign up to 300 hPa, though the differences are slightly larger against ERA-40. In the stratosphere, large biases are seen in both Figure A1(a) and (b), with a magnitude and structure similar to the AGCM stratospheric biases noted in the Atmospheric Model Intercomparison Project (AMIP; Gates *et al.*, 1999).

Temperature biases (Figures A1(c,d)) are also relatively small throughout the troposphere, except in the lower troposphere in the vicinity of both poles. The tropospheric differences are generally much smaller than when compared to AMIP and other AGCM simulations forced by SSTs (e.g. Gates *et al.*, 1999; Anderson *et al.*, 2004). In the region of the tropical tropopause, some biases are observed, with positive bias near the Equator and negative biases approaching the Subtropics. These are still relatively small and smaller against

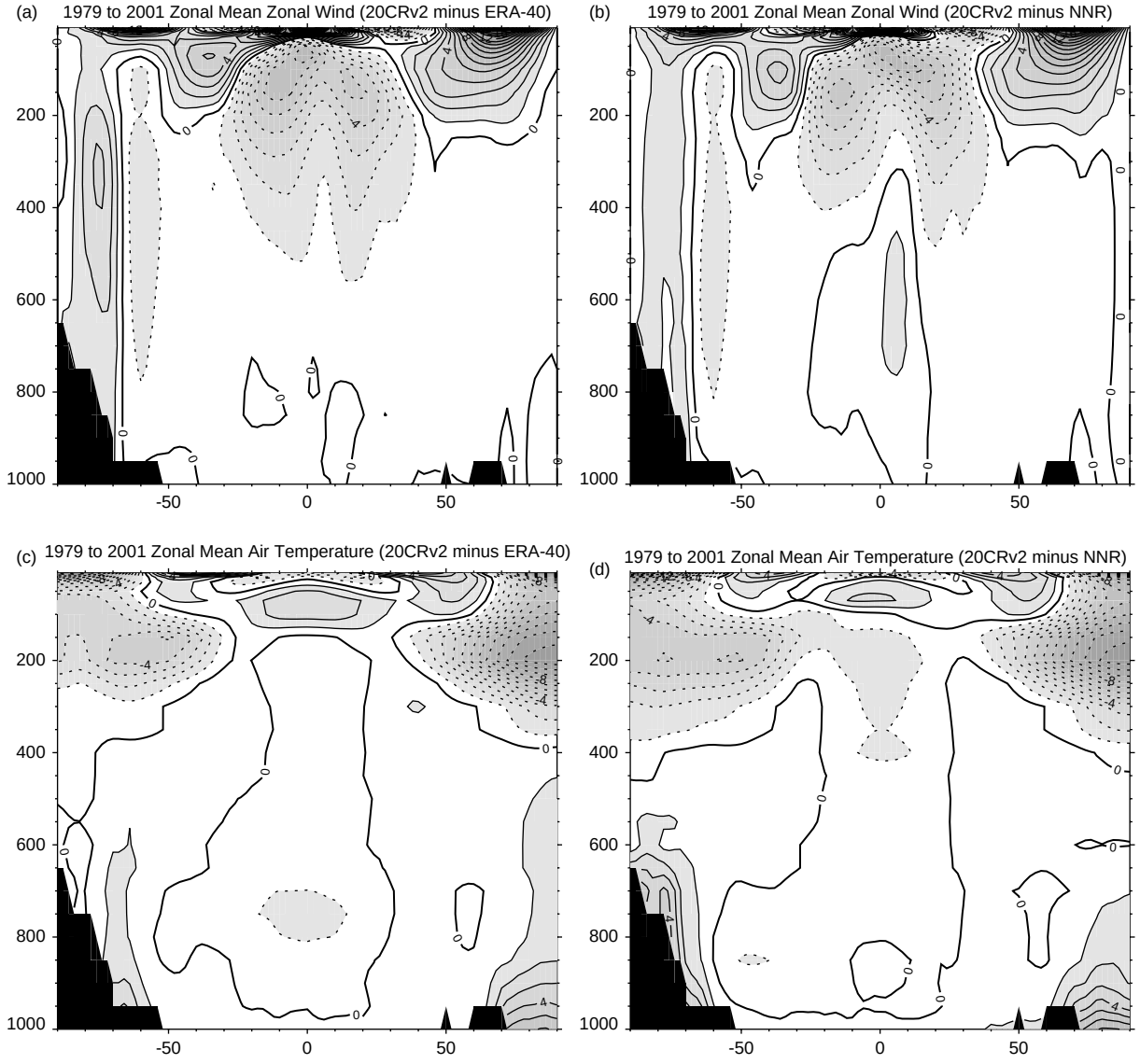

Figure A1. Average difference of 20CRv2 zonal means of (a, b) zonal wind speed and (c, d) air temperature from (a, c) ERA-40 and (b, d) NCEP–NCAR reanalyses (NNR) over the period 1979–2001. Contour interval is (a, b) 1 K and (c, d) 1 m s^{-1}. The zero contour is thickened in all panels, and dotted contours indicate negative differences. Black shaded regions indicate where more than 50% of the pressure-level grid points at that latitude are below ground.

NNR than against ERA-40. These biases have opposite signs in some parts of the tropical upper troposphere. Near the surface in the Arctic and Antarctic, however, the differences are relatively large compared even to AGCM biases (e.g. Gates *et al.*, 1999; Anderson *et al.*, 2004). In the polar lower troposphere, the differences are considerably smaller against ERA-40 than against NNR. These differences reflect an issue with the handling of the specified sea-ice concentration. In 20CRv2, the HadISST sea-ice concentration was specified in each gridbox. However, an error in the transformation of this concentration near coastal areas resulted in specified concentrations lower than in the HadISST daily-interpolated dataset. This resulted in a warm bias in the lower troposphere of both poles compared to the upper-air-based reanalyses. In the lowest layers, the difference from ERA-40 is smaller than that from NNR, but, in the mid-troposphere over the Arctic, the difference is smaller from NNR. The sea-ice concentration issue will be corrected in a future version of the dataset.

Appendix B: Quality-Control Procedure

A five-step quality-control procedure was utilized on the SLP and surface pressure observations from marine and station platforms. First, as a basic test, every pressure observation was reduced to sea level using the US Standard Atmospheric lapse rate in the troposphere (6.5 K km^{-1}). If this reduced observation fell outside the plausible range of 880 to 1060 hPa, the observation was flagged and rejected without further consideration. This basic check was intended to prevent grossly erroneous observations from influencing subsequent adaptive quality-control decisions. In the following discussion, our references to observations will refer to only those pressure observations that passed this plausibility test.

In the second step, after an observation was reduced to the model orography using (5) to form y^{o}, its absolute difference from the first guess $\left|y^{\mathrm{o}} - \overline{y}^{\mathrm{b}}\right|$ was compared to the combined error variance $\sqrt{\sigma_{\mathrm{b}}^2 + R}$. If $\left|y^{\mathrm{o}} - \overline{y}^{\mathrm{b}}\right| > 3\sqrt{\sigma_{\mathrm{b}}^2 + R}$, the

observation was flagged as having failed the 'background check'.

Observations that passed the background check were then used in the third step, a 'buddy check' evaluation of all observations (Dee *et al.*, 2001). We considered each observation in turn, constructing a single observation analysis $\overline{\mathbf{x}}_i^{\mathrm{a}}$ for the ith observation y_i^{o} using Eqs (1)–(4). We then determined the ratio B_i of the mean-square observation departure from $\overline{\mathbf{x}}_i^{\mathrm{a}}$ to the mean-square observation departure from the first guess $\overline{\mathbf{x}}^{\mathrm{b}}$ for the K nearby observations y_k^{o} located within a radius D from y_i^{o},

$$B_i = \frac{\frac{1}{K}\sum_{\substack{k=1\\k\neq i}}^{K}(H\overline{\mathbf{x}}_i^{\mathrm{a}} - y_k^{\mathrm{o}})^2}{\frac{1}{K}\sum_{\substack{k=1\\k\neq i}}^{K}(H\overline{\mathbf{x}}^{\mathrm{b}} - y_k^{\mathrm{o}})^2}. \tag{B1}$$

If $B_i < 1$ at the 50% level, according to the F-test, even if the observation had been rejected by the background, it was now accepted, as it would improve the fit of the analysis to the neighbouring observations despite its significant difference from the first-guess ensemble mean. By the same token, if B_i was significantly greater than 1 at the 10% level, then it was rejected on the grounds that it would degrade the analysis fit to the neighbouring observations. In instances of $B_i = 1$, the observation was accepted only if it passed the background check. We iterated this buddy test twice through all observations, so that observations that were accepted became part of the pool of nearby observations and could then influence the automated decision-making on other nearby observations. After trial and error in developing this testing procedure, we set $D = 1000$ km.

The fourth step in this quality-control procedure was data thinning using the F-test on the ratio $F = \sigma_{\mathrm{a}}^2/\sigma_{\mathrm{b}}^2$ of the analysis uncertainty to the first-guess uncertainty at the observation location. Observations were assimilated in the order of increasing F. Since smaller values of F imply a larger impact on the analysis ensemble, this means that observations expected to have a relatively large impact on the analysis were assimilated first. F was recalculated for all remaining observations as each observation was assimilated sequentially, i.e. Eqs (1)–(4) were iterated for all observations that passed the background and buddy checks or were restored by the buddy check. This recalculation accounted for the effect of already assimilated observations on the value of F. If F was not less than 1 at the 55% level (slightly higher than even odds) based on the F-test with 56 degrees of freedom, then the observation was not assimilated. This thinning allowed us to reserve a set of observations that are independent of the assimilation for use in later validation studies. As an illustration of the effect of these parameters, about 98% of the observations were retained in 1891 and 32% in 2005. We performed several tests with different levels of thinning and determined that the thinning used here did not adversely affect the quality of the analysis (not shown). Further, the thinning procedure removed previously undetected duplicate observations, and limited the inhomogeneity in the observing system by effectively capping the number of observations assimilated in the Northern Hemisphere at mid-twentieth century levels (as evidenced by the plateau in the thin black curve starting at roughly 1960 in Figure 3(a)).

The fifth quality-control step was a bias correction of the station observations. Surface pressure observations can have systematic differences from first-guess forecasts. These differences arise from actual measurement error, metadata errors in reported station elevation or position, coding errors, and model errors (e.g. Ingleby, 1995). Such differences will make the assimilation sub-optimal, as Eqs (1)–(4) assume that errors are random. We accounted for these biases through a systematic, time-varying bias correction algorithm. After 60 days of assimilation, at the start of every assimilation cycle, station observations were investigated for statistically significant biases in their differences from the first guess. All available pairs of first-guess forecast and raw observations were aligned in time. The observation bias was estimated as the mean difference between forecast and observations over the previous 60 days. If the difference was significant at the 5% level using a paired sample t-test, the estimated bias was removed from the raw observation. The number of degrees of freedom used in the t-test was the number of days N for which observation/forecast pairs were available. The maximum possible number was 60, if pairs were available for all 240 analysis times in the 60-day window. The bias was estimated only if N exceeded 30.

We expect that many of these biases are related to errors in the station elevation metadata. Occasionally, mean SLP observations appear to be coded as station pressure (or vice versa) or to have some other error, e.g. station elevation misreported by 40 m or more (Ingleby, 1995) or the 50 ft rule whereby Canadian stations reported surface pressure as SLP if the station elevation was 50 ft or less (Slonosky and Graham, 2005); our algorithm was able to account for the systematic effects of situations such as these. While some of the biases are also likely related to systematic model errors, one might expect calculated biases to be larger in regions of large orography, which we did not find (not shown). Additionally, biases can arise from systematic errors in the model's simulation of planetary waves (D. Dee, pers. comm., 2010), or through systematic errors in the temperature field which is used for the reduction to the model orography. Both of these may vary with season and contribute to the time-varying biases removed. We made no attempt to discriminate model and observation bias in this procedure, and as a result we likely included some model systematic errors into our observation bias correction. Distinguishing between these two sources of bias, perhaps using the thinned independent observations, is an ongoing area of research.

There was also no attempt to bias-correct marine observations, as not all marine observations have consistent identification or other observational metadata (e.g. Kent *et al.*, 2007). Future enhancements to ICOADS may permit such corrections in subsequent versions of this reanalysis dataset.

References

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

Adler RF, Huffman GJ, Chang A, Ferraro R, Xie PP, Janowiak J, Rudolf B, Schneider U, Curtis S, Bolvin D, Gruber A, Susskind J, Arkin P, Nelkin E. 2003. The Version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeorol.* **4**: 1147–1167.

Alpert JC. 2004. 'Sub-grid scale mountain blocking at NCEP'. In *Proceedings of 20th Conference on WAF, 16th conference on NWP*, Seattle WA, 10–15 January 2004, P2.4. Amer. Meteorol. Soc: Boston.

Alexander MA, Blade I, Newman M, Lanzante JR, Lau NC, Scott JD. 2002. The atmospheric bridge: The influence of ENSO teleconnections on air–sea interaction over the global oceans. *J. Climate* **15**: 2205–2231.

Allan R, Ansell T. 2006. A new globally complete monthly historical gridded mean sea level pressure dataset (HadSLP2): 1850–2004. *J. Climate* **19**: 5816–6842.

Anderson JL, Anderson SL. 1999. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Weather Rev.* **127**: 2741–2758.

Anderson JL, Collins N. 2007. Scalable implementations of Ensemble Filter algorithms for data assimilation. *J. Atmos. Oceanic Technol.* **24**: 1452–1463.

Anderson JL, Balaji V, Broccoli AJ, Cooke WF, Delworth TL, Dixon KW, Donner LJ, Dunne KA, Freidenreich SM, Garner ST, Gudgel RG, Gordon CT, Held IM, Hemler RS, Horowits LW, Klein SA, Knutson TR, Kushner PJ, Langenhost AR, Lau N-C, Lian Z, Malyshev SL, Milly PCD, Nath MJ, Ploshay JJ, Ramaswamy V, Schwarzkopf MD, Shevliakova E, Sirutis JJ, Soden BJ, Stern WF, Thompson LA, Wilson RJ, Wittenberg AT, Wyman BL. 2004. The new GFDL global atmosphere and land model AM2-LM2: Evaluation with prescribed SST simulations. *J. Climate* **17**: 4641–4673.

Anderson JL, Wyman B, Zhang S, Hoar T. 2005. Assimilation of surface pressure observations using an ensemble filter in an idealized global atmospheric prediction system. *J. Atmos. Sci.* **62**: 2925–2938.

Ansell TJ, Jones PD, Allan RJ, Lister D, Parker DE, Brunet M, Moberg A, Jacobeit J, Brohan P, Rayner NA, Aguilar E, Alexandersson H, Barriendos M, Brandsma T, Cox NJ, Della-Marta PM, Drebs A, Founda D, Gerstengarbe F, Hickey K, Jónsson T, Luterbacher J, Nordli Ø, Oesterle H, Petrakis M, Philipp A, Rodwell MJ, Saladie O, Sigro J, Slonosky V, Srnec L, Swail V, García-Suárez AM, Tuomenvirta H, Wang X, Wanner H, Werner P, Wheeler D, Xoplaki E. 2006. Daily mean sea level pressure reconstructions for the European–North Atlantic region for the period 1850–2003. *J. Climate* **19**: 2717–2742.

Baird ME, Everett JD, Suthers IM. 2010. Analysis of southeast Australian zooplankton observations of 1938–42 using synoptic oceanographic conditions. *Deep Sea Res. II*, DOI:10.1016/j.dsr2.2010.06.002 (in press).

Barsugli JJ, Sardeshmukh PD. 2002. Global atmospheric sensitivity to tropical SST anomalies throughout the Indo–Pacific basin. *J. Climate* **15**: 3427–3442.

Bengtsson L. 1980. On the use of a time sequence of surface pressures in four-dimensional data assimilation. *Tellus* **32**: 189–197.

Bengtsson L, Shukla J. 1988. Integration of space and in situ observations to study global climate change. *Bull. Amer. Meteorol. Soc.* **69**: 1130–1143.

Bengtsson L, Kanamitsu M, Kallberg P, Uppala S. 1982. FGGE four-dimensional data assimilation at ECMWF. *Bull. Amer. Meteorol. Soc.* **63**: 29–43.

Bengtsson L, Hagemann S, Hodges KI. 2004a. Can climate trends be calculated from reanalysis data? *J. Geophys. Res.* **109**: D11111. DOI: 10.1029/2004JD004536.

Bengtsson L, Hodges KI, Hagemann S. 2004b. Sensitivity of the ERA-40 reanalysis to the observing system: Determination of the global atmospheric circulation from reduced observations. *Tellus* **56A**: 456–471.

Bengtsson L, Arkin P, Berrisford P, Bougeault P, Folland CK, Gordon C, Haines K, Hodges KI, Jones P, Kallberg P, Rayner N, Simmons AJ, Stammer D, Thorne PW, Uppala S, Vose RS. 2007. The need for a dynamical climate reanalysis. *Bull. Am. Meteorol. Soc.* **88**: 495–501.

Benjamin SG, Miller PA. 1990. An alternative sea level pressure reduction and a statistical comparison of geostrophic wind estimates with observed surface winds. *Mon. Weather Rev.* **118**: 2099–2116.

Blackmon ML. 1976. A climatological spectral study of the 500 mb geopotential height of the northern hemisphere. *J. Atmos. Sci.* **33**: 1607–1623.

Bosilovich MG, Chen J, Robertson FR, Adler RF. 2008. Evaluation of global precipitation in reanalyses. *J. Appl. Meteorol. Climatol.* **47**: 2279–2299.

Brandes, *Beiträge zur Witterungskunde* (1820, Leipzig) referenced in Frängsmyr T, Heilbron JL, Rider RE (eds.) 1990. *The Quantifying Spirit in the Eighteenth Century.* University of California Press: Berkeley. http://ark.cdlib.org/ark:/13030/ft6d5nb455/.

Bromwich DH, Fogt RL. 2004. Strong trends in the skill of the ERA-40 and NCEP-NCAR reanalyses in the high and midlatitudes of the southern hemisphere 1958–2001. *J. Climate* **23**: 4603–4619.

Brönnimann S. 2003. A historical upper-air data set for the 1939–44 period. *Int. J. Climatol.* **23**: 769–791.

Brönnimann S, Compo GP, Sardeshmukh PD, Jenne R, Sterin A. 2005. New approaches for extending the twentieth century climate record. *Eos* **86**: 2–6, DOI: 10.1029/2005EO010002.

Brönnimann S, Stickler A, Griesser T, Fischer AM, Grant A, Ewen T, Zhou T, Schraner M, Rozanov E, Peter T. 2009. Variability of large-scale atmospheric circulation indices for the northern hemisphere during the past 100 years. *Meteorol. Zeit.* **18**: 379–396.

Brohan P, Kennedy JJ, Harris I, Tett SFB, Jones PD. 2006. Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J. Geophys. Res.* **111**: D12106, DOI: 10.1029/2005JD00654.

Brunet M, Kuglitsch FG. (eds.) 2008. *MEDARE: Proceedings of the international workshop on rescue and digitization of climate records in the Mediterranean basin*. WMO-TD No. 1432. World Meteorological Organization: Geneva.

Burgers G, van Leeuwen PJ, Evensen G. 1998. Analysis scheme in the Ensemble Kalman Filter. *Mon. Weather Rev.* **126**: 1719–1724.

Burkhardt U, James IN. 2006. The effect of Doppler correction on measures of storm track intensity. *Clim. Dyn.* **27**: 515–530. DOI: 10.1007/s00382-006-0146-4.

Chang EKM. 2005. Effects of secular changes in frequency of observations and observational errors on monthly mean MSLP summary statistics derived from ICOADS. *J. Climate* **18**: 3623–3633.

Chang EKM. 2007. Assessing the increasing trend in northern hemisphere winter storm track activity using surface ship observations and a statistical storm track model. *J. Climate* **20**: 5607–5628.

Chang EKM. 2009. Are band-pass variance statistics useful measures of storm track activity? Re-examing storm track variability associated with the NAO using multiple storm track measures. *Clim. Dyn.* **33**: 277–296, DOI: 10.1007/s00382-009-0532-9.

Climate Change Science Program. 2003. *Strategic Plan for the US Climate Change Science Program.* US Climate Change Science Program: Washington DC. Available from http://www.climatescience.gov/Library/stratplan2003/final/ccspstratplan2003-all.pdf.

Climate Change Science Program. 2008a. *Revised Research Plan for the US Climate Change Science Program*. US Climate Change Science Program: Washington DC. Available from http://www.climatescience.gov/Library/stratplan2008/.

Climate Change Science Program. 2008b. *Reanalysis of Historical Climate Data for Key Atmospheric Features: Implications for Attribution of Causes of Observed Change.* A Report by the US Climate Change Science Program and the Subcommittee on Global Change Research. Dole R, Hoerling M, Schubert S (eds.). NOAA, National Climatic Data Center: Asheville, NC, USA, 156pp.

Compo GP, Sardeshmukh PD. 2004. Storm track predictability on seasonal and decadal scales. *J. Climate* **17**: 3701–3720.

Compo GP, Sardeshmukh PD, Penland C. 2001. Changes of subseasonal variability associated with El Niño. *J. Climate.* **14**: 3356–3374.

Compo GP, Whitaker JS, Sardeshmukh PD. 2006. Feasibility of a 100-year reanalysis using only surface pressure data. *Bull. Amer. Meteorol. Soc.* **87**: 175–190.

Compo GP, Whitaker JS, Sardeshmukh PD. 2008. 'The 20th Century Reanalysis Project.' In *Proceedings of the Third WCRP International Conference on Reanalysis, The University of Tokyo, Japan. 28 Jan–1 Feb 2008.* Available at http://wcrp.ipsl.jussieu.fr/Workshops/Reanalysis2008/Documents/V5-511_ea.pdf.

Cook BI, Seager R, Miller RL. 2010. Atmospheric circulation anomalies during two persistent North American droughts: 1932–1939 and 1948–1957. *Clim. Dyn.* DOI: 10.1007/s00382-010-0807-1 (in press).

Daley R. 1991. *Atmospheric Data Analysis*. Cambridge University Press: Cambridge, UK.

Dee DP, Uppala S. 2009. Variational bias correction of satellite radiance data in the ERA-Interim reanalysis. *Q. J. R. Meteorol. Soc.* **135**: 1830–1841.

Dee DP, Rukhovets L, Todling R, da Silva AM, Larson JW. 2001. An adaptive buddy check for observational quality control. *Q. J. R. Meteorol. Soc.* **127**: 2451–2471.

Diamond H, Hinman WS, Lapham EG. 1938. Comparison of soundings with radio-meteorographs, aerographs and meteorographs. *Bull. Amer. Meteorol. Soc.* **19**: 129–141.

Dubois JL, Multhauf RP, Ziegler CA. 2002. *The invention and development of the radiosonde, with a catalog of upper-atmospheric telemetering probes in the National Museum of American History, Smithsonian Institution*. Smithsonian Institution Press: Washington DC. http://hdl.handle.net/10088/2453.

Dupigny-Giroux LA, Ross TF, Elms JD, Truesdell R, Doty SR. 2007. NOAA's climate database modernization program: Rescuing, archiving, and digitizing history. *Bull. Amer. Meteorol. Soc.* **88**: 1015–1017.

Durre I, Vose RS, Wuertz DB. 2006. Overview of the Integrated Global Radiosonde Archive. *J. Climate* **19**: 53–68.

Easterling DR, Anderson DM, Cohen SJ, Gutowski WJ, Holland GJ, Kunkel KE, Peterson TC, Pulwarty RS, Stouffer RJ, Wehner MF. 2008. Measures to improve our understanding of weather and climate extremes. In *Weather and Climate Extremes in a Changing Climate. Regions of Focus: North America, Hawaii, Caribbean, and US Pacific Islands*. Karl TR, Meehl GA, Miller CD, Hassol SJ, Waple AM, Murray WL. (eds.) A report by the US Climate Change Science Program and the Subcommittee on Global Change Research, Washington DC. Available from http://downloads.climatescience.gov/sap/sap3-3/sap3-3-final-Chapter4.pdf

Ek MB, Mitchell KE, Lin Y, Rogers E, Grunmann P, Koren V, Gayno G, Tarpley JD. 2003. Implementation of the Noah land-use model advances in the NCEP operational mesoscale Eta model. *J. Geophys. Res.* **108**: 8851, DOI: 10.1029/2002JD003296

Emanuel K. 2010. Tropical cyclone activity downscaled from NOAA-CIRES reanalysis, 1908–1958. *J. Adv. Model. Earth Syst.* **2**: Art. 1, DOI: 10.3894/JAMES.2010.2.1

Evensen G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**: 10143–10162.

Evensen G. 2003. The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dyn.* **53**: 343–367.

Evensen G. 2009. *Data Assimilation: The Ensemble Kalman Filter*. Springer: Berlin.

Fischer AM, Schraner M, Rozanov E, Kenzelmann P, Schnadt Poberaj C, Brunner D, Lustenberger A, Luo BP, Bodeker GE, Egorova T, Schmutz W, Peter T, Brönnimann S. 2008. Interannual-to-decadal variability of the stratosphere during the 20th century: Ensemble simulations with a chemistry-climate model. *Atmos. Chem. Phys.* **8**: 7755–7777.

García-Herrera R, Können GP, Wheeler DA, Prieto MR, Jones PD, Koek FB. 2005. CLIWOC: A Climatological Database for the World's Oceans 1750–1854. *Climatic Change* **73**: 1–12.

Gaspari G, Cohn SE. 1999. Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.* **125**: 723–757. DOI: 10.1002/qj. 49712555417

Gates WL, Boyle JS, Covey C, Dease CG, Doutriaux CM, Drach RS, Fiorino M, Gleckler PJ, Hnilo JJ, Marlais SM, Phillips TJ, Potter GL, Santer BD, Sperber KR, Taylor KE, Williams DN. 1999. An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bull. Amer. Meteorol. Soc.* **80**: 29–55.

GCOS. 2010. *Implementation plan for the global observing system for climate in support of the UNFCCC* (2010 Update). WMO-IOC-UNEP-ICS, GOOS-184, GTOS-76, GCOS-138, WMO TD-1523. World Meteorological Organization: Geneva, Switzerland. Available at http://www.wmo.int/pages/prog/gcos/Publications/gcos-138.pdf

Gibson JK, Kallberg P, Uppala S, Hernandey A, Nomura A, Serano E. 1997. *'ERA description'. Reanalysis Project Report Series* 1, ECMWF: Reading, UK.

Giese BS, Compo GP, Slowey NC, Sardeshmukh PD, Carton JA, Ray S, Whitaker JS. 2010. The 1918/1919 El Niño. *Bull. Amer. Meteorol. Soc.* **91**: 177–183, DOI: 10.1175/2009BAMS2903.1

Griesser T, Brönnimann S, Grant A, Ewen T, Stickler A, Comeaux J. 2010. Reconstruction of global monthly upper-level temperature and geopotential height fields back to 1880. *J. Climate* **23**: 5590–5609. DOI: 10.1175/2010JCLI3056.1

Group on Earth Observations. 2005. *The Global Earth Observation System of Systems 10-year Implementation Plan*. Available at http://earthobservations.org/documents/10-Year Implementation Plan.pdf

Haimberger L. 2007. Homogenization of radiosonde temperature time series using innovation statistics. *J. Climate* **20**: 1377–1403.

Hamill TM. 2006. Ensemble-based data assimilation. In *Predictability of Weather and Climate*. Palmer TN, Hagedorn R (eds.) Cambridge University Press: Cambridge, UK. 124–156.

Hamill TM, Whitaker JS, Snyder C. 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Weather Rev.* **129**: 2776–2790.

Hansen JE, Ruedy R, Sato M, Imhoff M, Lawrence W, Easterling D, Peterson T, Karl T. 2001. A closer look at United States and global surface temperature change. *J. Geophys. Res.* **106**: 23947–23963.

Harnik N, Chang EKM. 2003. Storm track variations as seen in radiosonde observations and reanalysis data. *J. Climate* **16**: 480–495.

Held IM, Soden BJ. 2006. Robust response of the hydrological cycle to global warming. *J. Climate* **19**: 5686–5699.

Hodges KI, Hoskins BJ, Boyle J, Thorncroft C. 2003. A comparison of recent reanalysis datasets using objective feature tracking: Storm tracks and tropical easterly waves. *Mon. Weather Rev.* **131**: 2012–2037.

Hong S-Y, Pan H-L. 1996. Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Mon. Weather Rev.* **124**: 2322–2339.

Hou Y, Moorthi S, Compana K. 2002. 'Parameterization of solar radiation transfer in NCEP models'. Office Note 441. NCEP: Washington DC. Available from http://www.emc.ncep.noaa.gov/officenotes/FullTOC.html#2000

Houtekamer PL, Mitchell HL. 1998. Data assimilation using an Ensemble Kalman Filter technique. *Mon. Weather Rev.* **126**: 796–811.

Houtekamer PL, Mitchell HL. 2001. A sequential Ensemble Kalman Filter for atmospheric data assimilation. *Mon. Weather Rev.* **129**: 123–137.

Houtekamer PL, Mitchell HL. 2005. Ensemble Kalman filtering. *Q. J. R. Meteorol. Soc.* **131**: 3269–3289. DOI: 10.1256/qj.05.135

Houtekamer PL, Mitchell HL, Pellerin G, Buehner M, Charron M, Spacek L, Hansen B. 2005. Atmospheric data assimilation with an Ensemble Kalman Filter: Results with real observations. *Mon. Weather Rev.* **133**: 604–620.

Houtekamer PL, Mitchell HL, Deng X. 2009. Model error representation in an operational Ensemble Kalman Filter. *Mon. Weather Rev.* **137**: 2126–2143.

Hurrell JW, Hack JJ, Phillips AS, Caron J, Yin J. 2006. The dynamical simulation of the Community Atmosphere Model version 3 (CAM3). *J. Climate* **19**: 2162–2183.

Ide K, Courtier P, Ghil M, Lorenc AC. 1997. Unified notation for data assimilation: Operational, sequential, and variational. *J. Meteorol. Soc. Japan* **75**: 181–189.

Ingleby NB. 1995. Assimilation of station level pressure and errors in station height. *Weather Forecasting* **10**: 172–182.

Jones P. 2001. Early European Instrumental Records. In *History and Climate: Memories of the Future.* Jones P, Ogilvie AEJ, Davies TD, Briffa KR. (eds.) Kluwer Academic: New York, NY.

Juang H-M. 2005. 'Discrete generalized hybrid vertical coordinates by a mass, energy and angular momentum conserving vertical finite-differencing scheme'. Office Note 445. NCEP: Washington DC. Available from http://www.emc.ncep.noaa.gov/officenotes/FullTOC.html#2000

Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetmaa A, Reynolds R, Jenne R, Joseph D. 1996. The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77**: 437–471.

Kanamitsu M, Alpert JC, Campana KA, Caplan PM, Deaven DG, Iredell M, Katz B, Pan H-L, Sela J, White GH. 1991. Recent changes implemented into the Global Forecast System at NMC. *Weather Forecasting* **6**: 425–435.

Kanamitsu M, Hwang S-O. 2006. The role of sea surface temperature in reanalysis. *Mon. Weather Rev.* **134**: 532–552.

Kanamitsu M, Ebisuzaki W, Woollen J, Yang SK, Hnilo JJ, Fiorino M, Potter GL. 2002. The NCEP-DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteorol. Soc.* **83**: 1631–1643.

Kent EC, Berry DI. 2005. Quantifying random measurement errors in Voluntary Observing Ships' meteorological observations. *Int. J. Climatol.* **25**: 843–856.

Kent EC, Woodruff SD, Berry DI. 2007. Metadata from WMO Publication No. 47 and an assessment of Voluntary Observing Ship observation heights in ICOADS. *J. Atmos. Oceanic Technol.* **24**: 214–234.

Khare SP, Anderson JL, Hoar TJ, Nychka D. 2008. An investigation into the application of an ensemble Kalman smoother to high-dimensional geophysical systems. *Tellus* **60A**: 97–112.

Kim Y-J, Arakawa A. 1995. Improvement of orographic gravity wave parameterization using a mesoscale gravity wave model. *J. Atmos. Sci.* **52**: 1875–1902.

Kinter JL, Fennessy MJ, Krishnamurthy V, Marx L. 2004. An evaluation of the apparent interdecadal shift in the tropical divergent circulation in the NCEP-NCAR reanalysis. *J. Climate* **17**: 349–361.

Kistler R, Kalnay E, Collins W, Saha S, White G, Woollen J, Chelliah M, Ebisuzaki W, Kanamitsu M, Kousky V, van den Dool H, Jenne R, Fiorino M. 2001. The NCEP–NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bull. Am. Meteorol. Soc.* **82**: 247–267.

Knapp KR, Kruk MC, Levinson DH, Diamond HJ, Neumann CJ. 2010. The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone best track data. *Bull. Amer. Meteorol. Soc.* **91**: 363–376. DOI: 10.1175/2009BAMS2755.1

Kocin P, Uccellini L. 2004. *Northeast Snowstorms*. American Meteorological Society. *Meteorol. Monogr.* **32**: 2 volumes.

Li H, Kalnay E, Miyoshi T. 2009. Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Q. J. R. Meteorol. Soc.* **135**: 523–533.

Lorenc AC. 1986. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.* **112**: 1177–1194.

Lott F, Miller MJ. 1997. A new subgrid-scale orographic drag parameterization: Its formulation and testing. *Q. J. R. Meteorol. Soc.* **123**: 101–127.

Lott N, Vose R, Del Greco SA, Ross TF, Worley S, Comeaux J. 2008. 'The Integrated Surface Database: Partnerships and progress.' In *Proceedings of 88th AMS Annual Meeting*, 20–25 January 2008, New Orleans, Louisiana, combined preprints (CD-ROM), Amer. Meteorol. Soc: Boston, MA. Available from http://www1.ncdc.noaa.gov/pub/data/ish/ams-isd-jan08.pdf

Manabe T. 1999. The digitized Kobe collection, Phase I: Historical surface marine meteorological observations in the archive of the Japan Meteorological Agency. *Bull. Amer. Meteorol. Soc.* **80**: 2703–2715.

Mitchell HL, Houtekamer PL, Pellerin G. 2002. Ensemble size, balance, and model-error representation in an Ensemble Kalman Filter. *Mon. Weather Rev.* **130**: 2791–2808.

Mlawer EJ, Taubman SJ, Brown PD, Iacono MJ, Clough SA. 1997. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.* **102**(D14): 16663–16682.

Monmonier M. 1999. *Air Apparent: How Meteorologists Learned to Map, Predict, and Dramatize Weather*. University of Chicago Press: Chicago.

Moorthi S, Pan H-L, Caplan P. 2001. 'Changes to the 2001 NCEP operational MRF/AVN global analysis/forecast system'. Technical Procedures Bulletin 484, NOAA, NWS: Silver Spring, MD. Available from http://www.nws.noaa.gov/om/tpb/484.htm

National Antarctic Expedition. 1913. *Meteorology Part II: Comprising daily synchronous charts 1 October 1901 to 31 March 1904*. Hepworth MWC (ed.) R. Soc. London.

National Centers for Environmental Prediction. 2010. 'Model performance statistics'. NOAA, NCEP: Washington DC. http://www.emc.ncep.noaa.gov/gmb/STATS/STATS.html

Newman M, Sardeshmukh PD, Bergman JW. 2000. An assessment of the NCEP, NASA and ECMWF reanalyses over the tropical West Pacific warm pool. *Bull. Amer. Meteorol. Soc.* **81**: 41–48.

Onogi K, Tsutsui J, Koide H, Sakamoto M, Kobayashi S, Hatsushika H, Matsumoto T, Yamazaki N, Kamahori H, Takahashi K, Kadokura S, Wada K, Kato K, Oyama R, Ose T, Mannoji N, Taira R. 2007. The JRA-25 Reanalysis. *J. Meteorol. Soc. Japan* **85**: 369–432.

Oort AH, Yienger JJ. 1996. Observed interannual variability in the Hadley Circulation and its connection to ENSO. *J. Climate* **9**: 2751–2767.

Ploshay JJ, White R, Miyakoda K. 1983. 'FGGE level III-b daily global analyses, Part I'. NOAA Data Report ERL GFDL-1.

Przybylak R. 2009. The Climate of Poland in recent centuries: A synthesis of current knowledge: Instrumental observations. In *The Polish Climate in the European Context: An Historical Overview.* Przybylak R, Majorowicz J, Brázdil R, Kejna M (eds.) Springer: Berlin, Heidelberg, New York, NY.

Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A. 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* **108**(D14): 4407, DOI: 10.1029/2002JD002670

Rienecker MM, Suarez MJ, Todling R, Bacmeister J, Takacs L, Liu H-C, Gu W, Sienkiewicz M, Koster RK, Gelaro R, Stajner I, Nielsen JE. 2008. '*The GEOS-5 data assimilation system – Documentation of versions 5.0.1 and 5.1.0, and 5.2.0*'. NASA Technical Report Series on Global Modeling and Assimilation, NASA/T-2008-104606, **27**. NASA Goddard Space Flight Center: Greenbelt, MA.

Roeckner E, Arpe K, Bengtsson L, Christoph M, Claussen M, Dümenil L, Esch M, Giorgetta M, Schlese U, Schulzweida U. 1996. 'The atmospheric general circulation model ECHAM4: Model description and simulation of present-day climate'. Report 218, Max-Planck Institut für Meteorologie: Hamburg, Germany.

Saha S, Nadiga S, Thiaw C, Wang J, Wang W, Zhang Q, van den Dool HM, Pan H-L, Moorthi S, Behringer D, Stokes D, Peña M, Lord S, White G, Ebisuzaki W, Peng P, Xie P. 2006. The NCEP Climate Forecast System. *J. Climate* **19**: 3483–3517.

Saha S, Moorthi S, Pan H-L, Wu X, Wang J, Nadiga S, Tripp P, Kistler R, Woollen J, Behringer D, Liu H, Stokes D, Grumbine R, Gayno G, Wang J, Hou Y-T, Chuang H-Y, Juang H-MH, Sela J, Iredell M, Treadon R, Kleist D, Van Delst P, Keyser D, Derber J, Ek M, Meng J, Wei H, Yang R, Lord S, van den Dool H, Kumar A, Wang W, Long C, Chelliah M, Xue Y, Huang B, Schemm J-K, Ebisuzaki W, Lin R, Xie P, Chen M, Zhou S, Higgins W, Zou C-Z, Liu Q, Chen Y, Han Y, Cucurull L, Reynolds RW, Rutledge G, Goldberg M. 2010. The NCEP Climate Forecast System reanalysis. *Bull. Amer. Meteorol. Soc.* **91**: 1015–1057. DOI: 10.1175/2010BAMS3001.1

Sakov P, Evensen G, Bertino L. 2010. Asynchronous data assimilation with the EnKF. *Tellus* **62A**: 24–29, DOI: 10.1111/j.1600-0870.2009.00417.x.

Sardeshmukh PD. 1993. The baroclinic χ problem and its application to the diagnosis of atmospheric heating rates. *J. Atmos. Sci.* **50**: 1099–1112.

Scaife AA, Kucharski F, Folland CK, Kinter J, Brönnimann S, Fereday D, Fischer AM, Grainger S, Jin EK, Kang IS, Knight JR, Kusunoki S, Lau NC, Nath MJ, Nakaegawa T, Pegion P, Schubert S, Sporyshev P, Syktus J, Yoon JH, Zeng N, Zhou T. 2009. The CLIVAR C20C project: Selected twentieth century climate events. *Clim. Dyn.* **33**: 603–614. DOI: 10.1007/s00382-008-0451-1.

Schraner M, Rozanov E, Schnadt Poberaj C, Kenzelmann P, Fischer AM, Zubov V, Luo BP, Hoyle CR, Egorova T, Fueglistaler S, Brönnimann S, Schmutz W, Peter T. 2008. Technical Note: Chemistry-climate model SOCOL: version 2.0 with improved transport and chemistry/microphysics schemes. *Atmos. Chem. Phys.* **8**: 5957–5974.

Schubert SD, Rood RB, Pfaendtner J. 1993. An assimilated dataset for Earth-science applications. *Bull. Amer. Meteorol. Soc.* **74**: 2331–2342.

Simmons A, Uppala S, Dee D, Kobayashi S. 2007. 'ERA-Interim: New ECMWF reanalysis products from 1989 onwards'. Newsletter 110, ECMWF: Reading, UK.

Slonosky VC, Graham E. 2005. Canadian pressure observations and circulation variability: Links to air temperature. *Int. J. Climatol.* **25**: 1473–1492. DOI: 10.1002/joc.1191.

Smith TM, Reynolds RW, Peterson TC, Lawrimore J. 2008. Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006). *J. Climate* **21**: 2283–2296.

Stickler A, Grant AN, Ewen T, Ross TF, Vose RS, Comeaux J, Bessemoulin P, Jylhä K, Adam WK, Jeannet P, Nagurny A, Sterin AM, Allan R, Compo GP, Griesser T, Brönniman S. 2010. The Comprehensive Historical Upper Air Network (CHUAN). *Bull. Amer. Meteorol. Soc.* **91**: 741–751. DOI: 10.1175/2009BAMS2852.1.

Sturaro G. 2003. A closer look at the climatological discontinuities present in the NCEP/NCAR reanalysis temperature due to the introduction of satellite data. *Clim. Dyn.* **21**: 309–316. DOI: 10.1007/s00382-003-0334-4.

Taylor KE, Williamson D, Zwiers F. 2000. 'The sea surface temperature and sea-ice concentration boundary conditions of AMIP II simulations'. Report 60, Program for Climate Model Diagnosis and Intercomparison. Lawrence Livermore National Laboratory: Livermore, California. Available from http://www-pcmdi.llnl.gov/projects/amip/AMIP2EXPDSN/BCS/amip2bcs.php

Thépaut J-N. 2006. 'Assimilating only surface-pressure observations in 3- and 4D-Var'. In *Proceedings of ECMWF/GEO Workshop on Atmospheric Re-analysis*, Reading, UK. 19–22 June 2006. Available from http://www.ecmwf.int/publications/library/ecpublications/_pdf/workshop/2006/Reanalysis/Thepaut.pdf

Thorne PW, Vose RS. 2010. Reanalyses suitable for characterizing long-term trends. *Bull. Amer. Meteorol. Soc.* **91**: 353–361. DOI: 10.1175/2009BAMS2858.1

Tippett MK, Anderson JL, Bishop CH, Hamill TM, Whitaker JS. 2003. Ensemble square root filters. *Mon. Weather Rev.* **131**: 1485–1490.

Trenberth KE, Olson JG. 1988. An evaluation and intercomparison of global analyses from the National Meteorological Center and the European Centre for Medium-range Weather Forecasts. *Bull. Amer. Meteorol. Soc.* **69**: 1047–1057.

Trenberth KE, Smith L. 2005. The mass of the atmosphere: A constraint on global analyses. *J. Climate* **18**: 864–875.

Trenberth KE, Koike T, Onogi K. 2008. Progress and prospects for reanalysis for weather and climate. *Eos* **89**: 234–235.

Uppala SM, Kållberg PW, Simmons AJ, Andrae U, da Costa Bechtold V, Fiorino M, Gibson JK, Haseler J, Hernandez A, Kelly GA, Li X, Onogi K, Saarinen S, Sokka N, Allan RP, Andersson E, Arpe K, Balmaseda MA, Beljaars ACM, Van De Berg L, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosavac M, Fisher M, Fuentes M, Hagemann S, Hólm E, Hoskins BJ, Isaksen L, Janssen PAEM, Jenne R, McNally AP, Mahfouf J-F, Morcrette J-J, Rayner NA, Saunders RW, Simon P, Sterl A, Trenberth KE, Untch A, Vasiljevic D, Viterbo P, Woollen J. 2005. The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.* **131**: 2961–3012 DOI: 10.1256/qj.04.176.

Wallace JM, Gutzler DS. 1981. Teleconnections in the geopotential height field during the northern hemisphere winter. *Mon. Weather Rev.* **109**: 784–812.

Wexler H, Tepper M. 1947. Results of the Wartime Historical and Normal Map Program. *Bull. Amer. Meteorol. Soc.* **28**: 175–178.

Whitaker JS, Hamill TM. 2002. Ensemble data assimilation without perturbed observations. *Mon. Weather Rev.* **130**: 1913–1924.

Whitaker JS, Compo GP, Wei X, Hamill TM. 2004. Reanalysis without radiosondes using ensemble data assimilation. *Mon. Weather Rev.* **132**: 1190–1200.

Whitaker JS, Hamill TM, Wei X, Song Y, Toth Z. 2008. Ensemble data assimilation with the NCEP Global Forecast System. *Mon. Weather Rev.* **136**: 463–482.

Whitaker JS, Compo GP, Thépaut J-N. 2009. A comparison of variational and ensemble-based data assimilation systems for reanalysis of sparse observations. *Mon. Weather Rev.* **137**: 1991–1999.

Wilkinson C, Woodruff SD, Brohan P, Claesson S, Freeman E, Koek F, Lubker SJ, Marzin C, Wheeler D. 2010. Recovery of Logbooks And International Marine Data: The RECLAIM project. *Int. J. Climatol.* DOI: 10.1002/joc.2102 (in press).

Wood KR, Overland JE. 2010. Early twentieth century Arctic warming in retrospect. *Int. J. Climatol.* **30**: 1269–1279. DOI: 10.1002/joc.1973.

Woodruff SD, Diaz HF, Worley SJ, Reynolds RW, Lubker SJ. 2005. Early ship observational data and ICOADS. *Climatic Change* **73**: 169–194.

Woodruff SD, Worley SJ, Lubker SJ, Ji Z, Freeman JE, Berry DI, Brohan P, Kent EC, Reynolds RW, Smith SR, Wilkinson C. 2010. ICOADS release 2.5: Extensions and enhancements to the surface marine meteorological archive. *Int. J. Climatol.* DOI: 10.1002/joc.2103 (in press).

Worley SJ, Woodruff SD, Reynolds RW, Lubker SJ, Lott N. 2005. ICOADS release 2.1 data and products. *Int. J. Climatol.* **25**: 823–842. DOI: 10.1002/joc.1166.

Yin X, Gleason BE, Compo GP, Matsui N, Vose RS. 2008. 'The International Surface Pressure Databank (ISPD) land component version 2.2'. National Climatic Data Center: Asheville, NC. Available from ftp://ftp.ncdc.noaa.gov/pub/data/ispd/doc/ISPD2_2.pdf