Evaluation of Seven Different Atmospheric Reanalysis Products in the Arctic*

R. LINDSAY, M. WENNAHAN, A. SCHWEIGER, AND J. ZHANG

Polar Science Center, Applied Physics Laboratory, University of Washington, Seattle, Washington

(Manuscript received 27 December 2012, in final form 1 October 2013)

ABSTRACT

Atmospheric reanalyses depend on a mix of observations and model forecasts. In data-sparse regions such as the Arctic, the reanalysis solution is more dependent on the model structure, assumptions, and data assimilation methods than in data-rich regions. Applications such as the forcing of ice–ocean models are sensitive to the errors in reanalyses. Seven reanalysis datasets for the Arctic region are compared over the 30-yr period 1981–2010: National Centers for Environmental Prediction (NCEP)–National Center for Atmospheric Research Reanalysis 1 (NCEP-R1) and NCEP–U.S. Department of Energy Reanalysis 2 (NCEP-R2), Climate Forecast System Reanalysis (CFSR), Twentieth-Century Reanalysis (20CR), Modern-Era Retrospective Analysis for Research and Applications (MERRA), ECMWF Interim Re-Analysis (ERA-Interim), and Japanese 25-year Reanalysis Project (JRA-25). Emphasis is placed on variables not observed directly including surface fluxes and precipitation and their trends. The monthly averaged surface temperatures, radiative fluxes, precipitation, and wind speed are compared to observed values to assess how well the reanalysis data solutions capture the seasonal cycles. Three models stand out as being more consistent with independent observations: CFSR, MERRA, and ERA-Interim. A coupled ice–ocean model is forced with four of the datasets to determine how estimates of the ice thickness compare to observed values for each forcing and how the total ice volume differs among the simulations. Significant differences in the correlation of the simulated ice thickness with submarine measurements were found, with the MERRA products giving the best correlation ($R = 0.82$). The trend in the total ice volume in September is greatest with MERRA ($-4.1 \times 10^3 \text{ km}^3 \text{ decade}^{-1}$) and least with CFSR ($-2.7 \times 10^3 \text{ km}^3 \text{ decade}^{-1}$).

1. Introduction

Atmospheric reanalyses in the Arctic provide estimates of the state of the atmosphere in a region of large climatic changes as the sea ice cover diminishes, glacial and ice sheet melt increases, and polar amplification of surface temperature changes is observed. Atmospheric reanalyses are critical tools in our attempts to document and understand these changes. One particularly pertinent application is the use of atmospheric reanalyses to force ice–ocean models. This class of models has a coupled representation of the ice–ocean systems but does not have an atmosphere. Reanalysis datasets are used to

provide surface atmospheric variables to link the historic state of the atmosphere to the dynamically evolving ice–ocean state in the model. A notable application of this class of models is the production of retrospective time series of sea ice thickness and volume (e.g., Zhang and Rothrock 2003; Hunke and Holland 2007; Lindsay et al. 2009; Schweiger et al. 2011; Notz et al. 2013). Observations for these ice variables are typically too sparse in time and space to allow an analysis for climate purposes and the use of atmospheric reanalysis data allows the reconstructions of the ice and ocean state. Because the ice–ocean environment is strongly forced by the atmosphere, ice–ocean model simulations are sensitive to the forcing data used and uncertainties in the reanalysis datasets should be considered.

Kalnay et al. (1996) make a useful distinction in the reliability of reanalysis output variables. Output variables are classified depending on the degree to which they are influenced by the observations or by the model. For example, class A indicates that a variable is strongly influenced by observed data (e.g., sea level pressure or upper air temperatures) whereas class C variables (e.g.,

* Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/JCLI-D-13-00014.s1>.

Corresponding author address: Ron Lindsay, Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, WA 98105.
E-mail: lindsay@apl.washington.edu

precipitation and surface fluxes) are completely determined by the model and subject to the largest uncertainties. Unfortunately, these variables tend to also be the ones for which validation data are absent or limited in time and spatial coverage so that uncertainties need to be established through a combination of comparisons with observations and intercomparisons between the candidate reanalysis datasets. Earlier studies have compared reanalysis products to observations in the Arctic region and some have considered the application to sea ice model forcing. For example, excellent agreement for sea level pressures and a relatively good correlation for surface winds are reported by Makshtas et al. (2007) in a study that compared the National Centers for Environmental Prediction (NCEP)–National Center for Atmospheric Research (NCAR) Reanalysis 1 (hereafter NCEP-R1) products to observations from the North Pole drifting stations for the period 1954–2006. The observed temperature is in good agreement with reanalysis data only in winter. Liu et al. (2008) compared surface air temperatures from the NCEP–U.S. Department of Energy (DOE) Reanalysis 2 (hereafter NCEP-R2) and 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) to drifting sea ice buoy observations and found that the reanalyses have warm annual-mean biases and underestimate the observed interannual variability in summer.

To determine the impact of different forcing data for sea ice models, Hunke and Holland (2007) used three different forcing datasets (NCEP-R1 was the only global reanalysis) and found significant differences in the ice thickness and ocean circulation. Additionally, Notz et al. (2013) found much different mean sea ice volume and volume trends in the Max Planck Institute Earth System Model using NCEP-R1 or ECMWF Interim Re-Analysis (ERA-Interim) forcing. These studies illustrate how important it is to consider the relative merits of reanalysis datasets for forcing an ice–ocean model.

Broader reviews beyond the application to sea ice model forcing examine other variables and combinations of datasets for the Arctic. For example, specific humidity and cloudiness at the stations are not reproduced well by the reanalysis. Bromwich et al. (2007) report the largest differences in three products are related to clouds and their associated radiation impacts; ERA-40 captures the cloud variability better than NCEP-R1 and the Japanese 25-yr Reanalysis Project (JRA-25), but the ERA-40 and JRA-25 clouds are too optically thin for shortwave radiation. In a comparison of the seasonal climatology for clouds of eight different reanalyses to different satellite and surface observations,

Chernokulsky and Mokhov (2012) report that NCEP-R1, NCEP-R2, and JRA-25 have less total cloud fraction (TCF) than observations during the whole year; other reanalyses are in close agreement with observations during summer and have noticeably higher TCF values than observations during winter. Zib et al. (2012) evaluate cloud fraction and radiative fluxes compared to surface observations for five models and also find strong biases.

Serreze et al. (2012) report that the National Aeronautics and Space Administration (NASA) Modern-Era Retrospective Analysis for Research and Applications (MERRA), Climate Forecast System Reanalysis (CFSR), and ERA-Interim have positive cold season humidity and temperature biases below 850 hPa based on comparisons with radiosonde data; these reanalyses also do not capture well the observed low-level humidity and temperature inversions. MERRA has the smallest biases. Simmons et al. (2004) compared global surface air temperature trends from NCEP-R1 and ERA-40 and report generally good agreements since 1979. Finally, two studies have compared NCEP-R1 and ERA-40 in the Southern Hemisphere: Bromwich and Fogt (2004) investigate skill trends and Hines et al. (2000) investigate surface pressure trends and they both find a strong dependence of the trends on the reanalysis model where observations are sparse.

None of these previous studies provides a comprehensive analysis of all the variables required for ice–ocean model forcing in the Arctic from all of the modern global reanalysis projects. This paper attempts to fill this gap by providing an intercomparison of sea level pressure, near-surface air temperatures, surface shortwave and longwave radiative fluxes, precipitation, and wind speed from seven projects: NCEP-R1 and NCEP-R2, CFSR, Twentieth-Century Reanalysis (20CR), MERRA, ERA-Interim, and JRA-25. Four products are from different branches of the National Oceanic and Atmospheric Administration (NOAA), one from NASA, one from ECMWF, and one from the Japanese Meteorological Agency (JMA). We consider here the period 1980–2009 because it spans the satellite era, all of the reanalysis products cover the period, and a 30-yr period is the standard for climatological means.

We provide comparisons with in situ observations where quality observations are available using only observations that are not assimilated. We then compare the fields from the different projects by referencing each dataset to the median of the seven reanalyses. We examine how well the different analyses agree with each other and determine if there are significant regional or seasonal variations in the discrepancies between the models. Finally we drive our Pan-Arctic Ice Ocean

TABLE 1. Reanalysis products, sponsoring agencies, and websites for documentation and data.

Reanalysis	Sponsoring agencies
NCEP-R1	National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) http://rda.ucar.edu/datasets/ds090.0/
NCEP-R2	National Centers for Environmental Prediction–U.S. Department of Energy (NCEP–DOE) http://rda.ucar.edu/datasets/ds091.0/
CFSR	National Centers for Environmental Prediction (NCEP) http://cfs.ncep.noaa.gov/cfsr/ http://rda.ucar.edu/datasets/ds093.2/
20CR	National Oceanic and Atmospheric Administration (NOAA) Earth System Research Laboratory (ESRL)–Cooperative Institute for Research in Environmental Sciences (CIRES) http://rda.ucar.edu/datasets/ds131.1/ http://www.esrl.noaa.gov/psd/data/gridded/data.20thC_ReanV2.html
MERRA	National Aeronautics and Space Administration (NASA) Global Modeling and Assimilation Office (GMAO) http://gmao.gsfc.nasa.gov/merra/
ERA-Interim	European Centre for Medium-Range Weather Forecasts (ECMWF) http://rda.ucar.edu/datasets/ds627.0/ http://data-portal.ecmwf.int/data/d/interim_moda/
JRA-25	Japanese Meteorological Agency (JMA) http://jra.kishou.go.jp/JRA-25/index_en.html http://rda.ucar.edu/datasets/ds625.0/

Modeling and Assimilation System (PIOMAS) with four forcing datasets and examine the impact of the selection on the ice volume time series from 1979 to 2009.

The paper is organized as follows: The main discriminating features for each reanalysis product are described in section 2, a summary table of the data sources is shown in Table 1, each model's key characteristics are shown in Table 2, and the variables considered in this study are given in Table 3. Observations of air temperature, radiative fluxes, precipitation, and wind speed that have not been assimilated by the different models are compared to the simulations in section 3. Three-month seasonal averages of the different models are then compared to each other in section 4. In section 5 trends in the surface temperature, radiative fluxes, and precipitation from the different datasets are compared. Section 6 examines the differences in the simulated ice volume with the use of four of the datasets to force an ice–ocean model. In the supplemental material, a table of the seasonal medians and the deviations of each of the models from the medians for ocean areas north of 70°N

and for land areas north of 65°N for each of the variables in Table 3 as well as a figure illustrating these deviations from the median are given. Figures of the seasonal 2-m air temperature trends are also found in the supplemental material.

2. The reanalysis products

a. NCEP-R1

The NCEP–NCAR Reanalysis 1 project (Kalnay et al. 1996; Kistler et al. 2001) was the first major reanalysis effort and has been used in a wide array of studies. NCEP and NCAR collaborated to produce a global analysis of atmospheric fields beginning in 1948. This first reanalysis effort uses a forecast model with T62 (210 km) model resolution and 28 vertical levels.

b. NCEP-R2

The main objective of the NCEP–DOE Reanalysis 2 project is to correct known errors in NCEP-R1 and update the parameterizations of the physical processes (Kanamitsu et al. 2002). The resolution of the NCEP-R2 model is the same as NCEP-R1, T62 (210 km) with 28 vertical sigma levels. NCEP-R2 starts in 1979. A key difference in the polar regions is the way sea ice cover is specified for R2, which follows Atmospheric Model Intercomparison Project phase 2 (AMIP-II) sea ice specifications provided by the Program for Climate Model Diagnosis and Intercomparison.

c. CFSR

The Climate Forecast System Reanalysis, operated by NCEP, is a global coupled atmosphere–ocean–land surface–sea ice system designed to provide the best estimate of the state of these domains since 1979 (Saha et al. 2010). The CFSR includes coupling of the atmosphere and ocean during the generation of 9-h forecast fields, an interactive sea ice model, and assimilation of satellite radiances over the entire period. The CFSR atmosphere resolution is ~38 km (T382) with 64 levels extending from the surface to 0.26 hPa. The ocean resolution is 0.25° at the equator, extending to 0.5° beyond the tropics, with 40 levels to a depth of 4737 m. The land surface model has four soil levels and the sea ice model has three levels. Satellite observations were used in radiance form and were bias corrected.

The sea ice model is from the Geophysical Fluid Dynamics Laboratory (GFDL) Sea Ice Simulator. It has three vertical layers, including two equal layers of sea ice and one layer of snow. There are five categories of possible sea ice thicknesses. Sea ice dynamics are based on the elastic–viscous–plastic technique (Hunke and

TABLE 2. Reanalysis product characteristics.

Reanalysis	NCEP-R1	NCEP-R2	CFSR	20CR	MERRA	ERA-Interim	JRA-25
Time interval	1948–present	1979–present	1979–present	1871–2008	1979–present	1979–present	1979–present
Type	Spectral	Spectral	Spectral	Spectral	Finite volume	Spectral	Spectral
Spatial resolution	T62 (210 km)	T62 (210 km)	T382 (38 km)	T62 (210 km)	$0.5^{\circ} \times 0.66^{\circ}$ (65 km)	T255 (79 km)	T106 (120 km)
Vertical levels (top pressure)	Sigma, 28 (3 hPa)	Hybrid sigma, 28 (3 hPa)	Hybrid sigma, 64 (0.26 hPa)	Hybrid sigma, 28 (0.2 hPa)	Hybrid sigma, 72 (0.1 hPa)	Hybrid sigma, 60 (0.1 hPa)	Hybrid sigma, 40 (0.4 hPa)
Data assimilation method	3D-Var	3D-Var	3D-Var	Ensemble Kalman filter	3D-Var/IAU	4D-Var	3D-Var
Sea ice and SST	Various prescribed	AMIP-II prescribed	Various interactive	HadISST prescribed	Reynolds prescribed	NCEP* prescribed	SMMR and SSM/I prescribed

* NCEP refers here to a changing suite of operational sources from the National Centers for Environmental Prediction.

Dukowicz 1997) to calculate ice internal stress and ice thermodynamics are based on Winton (2000).

d. 20CR

The Twentieth-Century Reanalysis is an effort led by the NOAA Earth System Research Laboratory, Physical Sciences Division, and the Cooperative Institute for Research in Environmental Sciences (CIRES) Climate Diagnostics Center to produce a reanalysis dataset spanning the entire twentieth century, assimilating only surface observations of air pressure and using the observed monthly sea surface temperature and sea ice concentration as lower boundary conditions (Compo et al. 2006, 2011; Whitaker et al. 2004). The dataset provides estimates of the atmospheric variability from 1871 to 2010. This reanalysis has the advantage of not being subject to gross changes in the type or quantity of data assimilated in the modern era or to unknown changes in the bias of radiosonde or satellite observations.

The analysis is performed with an ensemble Kalman filter as described in Compo et al. (2011). Each 6-hourly analysis is the most likely state of the global atmosphere and the uncertainty in that analysis is estimated as well. The short-term forecast ensemble is generated in parallel from 56 9-h integrations of the atmospheric component of the Climate Forecast System (CFS) model (Saha et al. 2006). The model has a spatial resolution of about 200 km on an irregular Gaussian grid (T62). There are 28 vertical levels and the model top is at 0.2 hPa. The monthly sea surface temperature and sea ice fields are from Hadley Centre Sea Ice and Sea Surface Temperature (HadISST) data obtained from the Met Office.

There is a problem with how sea ice is treated in the 20CR model, particularly in coastal regions where the sea ice concentration is often much less than observed. Compo et al. (2011) acknowledge the problem and report that it influences the lower tropospheric temperature structure in both polar regions, creating a warm bias

compared to other reanalysis products during the cold seasons. As a result the 20CR is the most notable outlier for many of the variables considered but for others, particularly in the summer when the temperature difference is minimal, the ice concentration error is less significant and meaningful comparisons can be made.

e. MERRA

The Modern-Era Retrospective Analysis for Research and Application is a NASA reanalysis that uses the Goddard Earth Observing System Data Assimilation System, version 5 (GEOS-5) (Rienecker et al. 2008; Bosilovich et al. 2008). A particular focus of the data assimilation in the model is to simulate the hydrological cycle correctly. The MERRA GEOS-5 data assimilation system uses three-dimensional variational data assimilation (3D-Var) as the assimilation framework and the incremental analysis updates (IAU) procedure to slowly adjust the model state toward the observed state. The water cycle is improved as spindown is reduced. In addition, the model physical parameterizations were tested and evaluated with data assimilation present, which reduces the shock of adjusting the model system

TABLE 3. Variables considered in this study.

Short name	Long name
SLP	Sea level pressure
T2M	2-m air temperature
SHF	Sensible heat flux
LHF	Latent heat flux
DWSWRFB	Surface downwelling shortwave flux
DWLWRFB	Surface downwelling longwave flux
UWSWRFB	Surface upwelling shortwave flux
UWLWRFB	Surface upwelling longwave flux
DWSWRFT	Top downwelling shortwave flux
UWSWRFT	Top upwelling shortwave flux
UWLWRFT	Top upwelling longwave flux
PRECIP	Precipitation
AVG_SPD	Average 10-m wind speed

to new data. Sea surface temperature and sea ice concentration boundary conditions are derived from the weekly 1° sea surface temperature product of Reynolds et al. (2002). Serreze et al. (2012) report that the MERRA record in particular shows evidence of artifacts in the lower tropospheric temperature and humidity in the region north of 70°N, likely introduced by changes in assimilation data.

f. ERA-Interim

ERA-Interim represents an undertaking by the ECMWF to produce a reanalysis with an improved atmospheric model and assimilation system that replaces those used in ERA-40 (Dee et al. 2011). ERA-Interim uses four-dimensional variational data assimilation (4D-Var) rather than 3D-Var as in ERA-40. Horizontal resolution is increased from T159 (nominally 1.125°) for ERA-40 to T255 (nominally 0.70°) for ERA-Interim. It retains the same 60 model levels used for ERA-40 with the highest level being 0.1 hPa. In addition, data assimilation of ERA-Interim benefits from more extensive use of radiances with an improved fast radiative transfer model. Dee et al. (2011) note the data assimilation excludes all Television Infrared Observation Satellite (TIROS) Operational Vertical Sounder (TOVS) and scatterometer data over sea ice.

The earlier reanalysis product, ERA-40, is not considered here because it ends in 2002. It has a significant discontinuity in the polar air temperatures below about 500 mb starting in 1997 when processing of High-Resolution Infrared Radiation Sounder (HIRS) satellite data was improved (Bromwich et al. 2007; Screen and Simmonds 2011). This discontinuity distorts the computed trends in the air temperatures (Bitz and Fu 2008; Grant et al. 2008; Thorne 2008). The seasonal trends in the new product are broadly similar to those seen in the other models.

g. JRA-25

The Japanese 25-yr Reanalysis represents the first long-term global atmospheric reanalysis undertaken by the Japanese Meteorological Agency (Onogi et al. 2007). It uses the JMA numerical assimilation and forecast system and observational and satellite data from many sources including the ECMWF, the National Climatic Data Center (NCDC), and the Meteorological Research Institute of JMA.

In JRA-25, 3D-Var data assimilation and a global spectral model are employed to produce 6-hourly atmospheric analysis and forecast cycles. The global spectral model is based on a 320×160 ($\sim 1.125^\circ$) Gaussian grid with T106 truncation. The vertical grid uses a hybrid sigma-pressure coordinate scheme utilizing 40 levels up

to 0.4 hPa. Daily sea ice concentration was obtained using the NASA team algorithm (Cavalieri et al. 1984) based on Special Sensor Microwave Imager (SSM/I) and Scanning Multichannel Microwave Radiometer (SMMR) brightness temperatures (Matsumoto et al. 2006). The bias between SMMR and SSM/I was corrected with the Cavalieri et al. (1999) method.

h. A note on sea ice

Table 2 lists how sea ice is prescribed or computed for each model. While the different reanalysis projects used different sources to specify the ice extent, the differences are minor except in the aforementioned case of 20CR. For example, sea ice concentration data assimilated by CFSR changes depending on the time period: from 1979 to 1996 it is from the National Snow and Ice Data Center (NSIDC; Cavalieri et al. 1996); from 1997 to February 2000 it is from the NCEP operational analysis (Grumbine 1996); from March 2000 to October 2007 it is from the newer NCEP sea ice analysis system; and from November 2007 to the present it is from the operational NCEP passive microwave analysis. Other reanalysis projects have similar mixes of data sources for ice concentration. CFSR is the only product with a modeled ice thickness, although satellite-based observations of the ice concentration are assimilated. Comparison maps of the ice concentration for the different projects (not shown) verify that minor differences mostly occur in coastal zones where there are modest differences in the land masks.

3. Comparison of reanalysis products to independent observations

a. Near-surface air temperatures

The 2-m air temperatures are compared to the monthly average land station data from the Climate Research Unit at the University of East Anglia (Brohan et al. 2006). Figure 1 shows the station locations, the seasonal-mean bias, and the anomaly correlations for each of the reanalysis models for which the monthly mean of each station and each model at the station location is first removed. A total of 449 stations north of 60°N during the period 1979–2009 are used, yielding 150 000 station-months. Most of the models estimate the 2-m air temperature by interpolating between the skin temperature and the lowest model temperature. However, the ERA-Interim uses the observed 2-m temperature and background temperatures from the previous analysis time step in an optimal interpolation scheme (Dee et al. 2011). As a consequence, its bias compared to the observations is smallest, and its correlation is highest. A large winter

FIG. 1. Comparison of the T2M measured at 449 land stations to estimates from seven reanalysis products representing 150 000 station-months. (a) Station locations, (b) mean difference for all stations by month, and (c) anomaly correlations for which the spatial variability is removed. The data are from the Climate Research Unit at the University of East Anglia (Brohan et al. 2006).

bias in the 20CR is apparent (Compo et al. 2011) and can be attributed to the bad sea ice specification mentioned earlier, though it is largely absent in summer. The other models have a small warm bias in the winter of up to 2°C and a slightly larger cold bias in the summer. The MERRA product has a very small bias and high correlation and CFSR and JRA-25 are nearly as good.

b. Surface radiative fluxes

The surface radiative fluxes are highly dependent on the model parameterization of cloud processes, so large differences are expected among the reanalyses as reported by Walsh et al. (2009). The surface fluxes are not assimilated by any of the models. We use the monthly averaged surface radiative fluxes from the NASA Clouds and the Earth's Radiant Energy System (CERES)–Atmospheric Radiation Measurement Program (ARM) Validation Experiment (CAVE). CAVE is a dataset containing radiation and meteorological data for sites having CERES top-of-atmosphere (TOA) broadband observations collocated with surface broadband flux measurements from the Baseline Surface Radiation Network (BSRN). There are two Arctic sites: Barrow, Alaska (1992–2009; operated by ARM), and Ny-Ålesund (on Spitsbergen Island, 1998–2009; operated by the Alfred Wegener Institute).

Comparisons of the downwelling surface longwave and shortwave radiative fluxes are shown in Fig. 2. The longwave bias is large and positive in the winter for 20CR and large and negative nearly all year for the NCEP-R1 and NCEP-R2 products; the smallest biases

are in the ERA-Interim and CFSR products. The anomaly correlation, for which the monthly interannual mean of each station and each model is first removed, of all of the models is poor in the summer when the cloud fraction is high and there is little variability. The bias in the shortwave radiation is severe for the NCEP-R1 and NCEP-R2 products, rising to almost 100 W m^{-2} too high, as noted by other researchers (Zib et al. 2012). Cullather and Bosilovich (2012) compared the MERRA downwelling shortwave flux to observations and report a negative bias of 12 W m^{-2} for the year, consistent with the comparisons shown here. They also note that the MERRA parameterized albedo for sea ice is fixed at 0.60, lower than measured at the Surface Heat Budget of the Arctic Ocean (SHEBA), which contributes to an underestimate in the upwelling shortwave flux. For the CFSR and the ERA-Interim products the bias in both fluxes is small for the entire year. The relatively low biases in ERA-Interim and CFSR indicate that the annual cycle in these properties is likely well represented. Low anomaly correlations for radiative fluxes points to difficulties in simulating the variability in the cloud coverage and properties.

c. Precipitation

Comparing point precipitation measurements to the precipitation fields from the reanalyses is difficult because of a strong influence of local conditions on precipitation measurement and the fact that a substantial fraction of the precipitation falls as snow, which is notoriously difficult to measure (Serreze et al. 2005). Here, we compare the monthly mean values to those computed

FIG. 2. Comparison of the downwelling (left) longwave and (right) shortwave radiative fluxes measured at two stations to the estimates from the models. Mean difference for all stations by month (top); anomaly correlations for which the monthly mean of each station and each model is first removed (bottom). The data are from the two BSRN measurement sites at Barrow (71.32°N, 156.61°W) and Ny-Ålesund (78.93°N, 11.95°E).

by the gridded Global Precipitation Analysis Products of the Global Precipitation Climatology Centre (GPCC) (Rudolf and Schneider 2005; Rudolf et al. 2010). The analyses are based on automatic and intensive manual quality control of the observed precipitation data from land stations. This analysis includes the monthly mean precipitation from all stations within a $0.5^\circ \times 1.0^\circ$ latitude by longitude grid box. Approximately 800 grid boxes per month contain at least one quality-controlled monthly mean observation over the period 1989–2009. The number of stations in each grid box ranges from 1 to 22, with an average of 2 (Fig. 3a). The GPCC grid boxes containing at least one station are not uniformly distributed over the landmass.

There is a very large bias in the spring for the 20CR and CFSR models (Fig. 3b). The JRA-25 model has the lowest mean bias. The anomaly correlations are better than expected given the poor showing for the radiative fluxes. Correlations are best in the ERA-Interim product (Fig. 3c).

Cullather and Bosilovich (2011) compare MERRA arctic data with long-term station measurements and some field experiment sites. For example, a time series of MERRA precipitation compared to observations

made at a North Pole drifting ice station shows discrepancies in the magnitude of a few events, but the comparison (correlations of 0.74) is quite reasonable given the uncertainties of precipitation measurements at high latitudes.

d. Wind speed

The 10-m wind speed is compared to that measured at the North Pole drifting ice stations of the former Soviet Union (Lindsay 1998). The observed daily average speed for 14 station-years is computed from the average vector magnitude of 3-hourly values of the meridional and zonal wind components. The daily average wind speeds for the reanalysis products are computed from either daily or 6-hourly vector magnitudes. Figure 4 shows the monthly mean difference in the wind speed (model – observed) for six of the reanalysis models (daily values for JRA-25 are not readily available). The bias in the two NCEP products are of opposite sign: NCEP-R1 is too low and NCEP-R2 too high. CFSR, MERRA, and ERA-Interim biases are mostly less than 0.5 m s^{-1} , with CFSR having the smallest mean bias. Makshtas et al. (2007) compared wind speed data from the North Pole

FIG. 3. Comparison of the monthly precipitation to the model estimates. (a) Map of the station locations, (b) monthly precipitation bias of each model expressed as a fraction of the observed precipitation, and (c) anomaly correlation in which the monthly means for the observations and the models have been removed for each location. The data are from the GPCC Full Data Reanalysis, version 5 (<http://gpcc.dwd.de>), and the model monthly mean values are linearly interpolated to the station locations for each month for which there are observations.

stations for the period 1954–2006 to NCEP-R1 and found that the average wind speed of NCEP-R1 was nearly the same in winter and $0.5\text{--}1.0\text{ m s}^{-1}$ too low in summer. The model with the best correlation in wind speed is ERA-Interim. Variations in the bias and correlations between the models may be related to how the models represent the lower atmospheric stratification and the translation of the geostrophic wind to surface wind. Model resolution may also play a role in that the models with the highest resolution,

ERA-Interim, CFSR, and MERRA, also have the highest correlations.

4. Comparisons of products to each other

Each of the variables from each of the reanalysis products is regridded with linear interpolation to a common 100-km grid [the Equal-Area Scalable Earth (EASE) grid from National Snow and Ice Data Center] that is centered on the pole and extends to south of 50°N . To

FIG. 4. Comparison of the daily average 10-m wind speed from the North Pole drifting ice stations of the former Soviet Union with five model estimates. Thirteen station-years are included between the years of 1981 and 1990. Daily or 6-hourly wind magnitudes are used to compute the daily averages for the reanalysis products. (a) Map of the station locations, (b) monthly wind speed bias of each model (model – observed; mean annual wind speed is 4.1 m s^{-1}), and (c) correlation of the daily wind speeds by month. The observational data are from the Arctic and Antarctic Research Institute in Leningrad and are archived at the NSIDC.

compare the products to each other, we need a common point of comparison. We use gridpoint local median value as the common reference because it represents a consensus estimate that is not influenced by outliers. This consensus may not be as close to the observed values as one or more of the individual models but the median reference helps identify commonalities and outliers. The comparisons are made on a seasonal basis: March–May (MAM), June–August (JJA), September–November (SON), and December–February (DJF). A table and figure in the supplemental material provides the seasonal medians and deviations from the median of each model for ocean areas north of 70°N and for land areas north of 65°N for each of the variables in Table 3.

a. Sea level pressure

The sea level pressure is assimilated by all of the models, so very good agreement among them is expected. This is generally the case, though the winter season has the largest discrepancy among the models (see the table in the supplemental material). The 20CR solution has a large positive bias over Asia (not shown). This is somewhat surprising because surface pressure is the sole parameter assimilated by 20CR. Other models, such as NCEP-R1, MERRA, and JRA-25, have large differences over Greenland and, in the case of MERRA, over other mountainous areas. This is likely due to different methods of reducing the model surface pressure to sea level.

b. Wind speed

The 10-m wind speed is of particular importance for sea ice modeling because it is used to determine the surface wind stress, an important factor for ice dynamics. The strongest winds and the largest differences among the models are in the winter months (see the table in the supplemental material). The greatest variability between the models is over the land surfaces, and the greatest outlier is the NCEP-R2 product. The other significant outlier is the NCEP-R1 product, which has winds lower than the median over much of the Arctic Ocean. The comparisons to observations from the North Pole drifting ice stations also show that the NCEP-R1 model has lower winds than the others over the Arctic Ocean and that they are closer to the observed values. In this case, NCEP-R1 appears to be a better estimate than the median.

c. Near-surface air temperature

The median and deviations from the median of each model for the 2-m air temperature estimates for winter and summer are shown in Fig. 5. There are large

discrepancies in winter for 20CR related to the error in ice concentration. It has a smaller warm bias over land in the summer. NCEP-R1 is relatively cold in the winter over the Arctic Ocean. All of the models are in good agreement over the Arctic Ocean during the summer as the surface is melting uniformly, at least until recent years. Over land the model with the smallest bias is MERRA, which is close to the median in most land areas in summer but 1.6° colder than the median in winter.

d. Radiative fluxes

The downwelling radiative fluxes are much more dependent on the model parameterizations of physical processes, particularly clouds, than the sea level pressure, air temperature, or height fields. It is no surprise that there are large variations among the models, as seen in the comparisons of models to observations. Median surface downwelling longwave radiative fluxes for winter and summer for all models and each model's deviation from the median are shown in Fig. 6. In winter the downwelling longwave bias is smallest for MERRA, amounting to about 10 W m^{-2} at the two ARM stations. It averages 7.5 W m^{-2} more than the median over the oceans and 2.0 W m^{-2} over land, so the median may be a generally good representation of the observed fluxes in this case. In the summer both ERA-Interim or CFSR have a very low bias, less than 5 W m^{-2} . However these two models have a substantial positive bias compared to the median, $12\text{--}19 \text{ W m}^{-2}$ for ERA-Interim, so the maps should be interpreted with the understanding that these two model fields give the best estimate, not the median.

In the winter, the largest deviations in downwelling longwave fluxes are with the 20CR products, which are very large in the regions where the sea ice concentration is wrongly assigned. The JRA-25 has much lower and the two NCEP models have somewhat lower fluxes than the median, consistent with the comparisons of models to observations. In the summer there are positive anomalies over the ocean in the CFSR and ERA-Interim products and negative anomalies in the NCEP-R1, NCEP-R2, and JRA-25 products. ERA-Interim and CFSR appear to have the closest match with the observations during the summer and their anomaly patterns are similar. This is one case in which the median is not a good representation of the flux because of the large bias in the comparisons to observations in four of the products: NCEP-R1 and NCEP-R2, MERRA, and JRA-25.

The largest variations, in terms of mean energy flux, are in the downwelling shortwave radiative fluxes in the

FIG. 5. T2M in (top) summer and (bottom) winter. The median of the seasonal means for the seven models is in the upper left of each eight-map set and the deviation from the median for each of the models is in the subsequent maps.

summer (Fig. 7), when the mean difference between the median and the individual model fields reaches 60 W m^{-2} on average over the ocean (see supplemental material). The NCEP-R1 and NCEP-R2 products differ the most from the median with much higher fluxes

arising from much lower cloudiness. The 20CR value is the closest to the median field, although the comparisons to the observations suggest the median of all the models is 33 W m^{-2} too high in the summer months at the locations of the two BSRN stations.

FIG. 6. As in Fig. 5, but for surface downwelling longwave radiative flux in the (top) summer and (bottom) winter.

The incoming shortwave flux, a parameterized choice of the modeling group, is consistent across models, as expected, but there is a small deviation for the ERA-Interim product, which has a flux about 0.8% (3 W m^{-2}) higher than the other models in the summer (not shown).

JRA-25 shows the largest deviation from the median in the TOA upwelling longwave flux with an average of 12 W m^{-2} more than the median over land (see supplemental material). The TOA upwelling summer shortwave flux is much higher than the median in NCEP-R2

FIG. 7. As in Fig. 5, but for surface downwelling shortwave radiative flux in the summer.

for locations where sea ice is present, which suggests that the cloud cover is too low, as also seen in the downwelling surface radiative fluxes.

e. Precipitation

The annual-mean precipitation fields are shown in Fig. 8. The NCEP-R1 field is striking in the regularity of the anomalies, a “ringing” that has been noted before for this model in many of the fields related to clouds (Trenberth and Guillemot 1998). The CFSR has heavier precipitation than the others over the North Atlantic and North Pacific. As with the comparisons to observations, aside from CFSR and 20CR, the winter bias in precipitation is small over land.

f. Turbulent fluxes

The surface sensible and latent heat turbulent fluxes are critical boundary conditions for the atmosphere and are sensitive to the land (or ice) surface model, the surface radiative fluxes, and the atmospheric state. They depend strongly on the model physics and data assimilation procedures and hence we expect large differences between the models. In a comparison of winter sensible heat flux and the summer latent heat flux, the winter heat fluxes are greatest over the North Atlantic and North Pacific regions and in the Barents and Bering Seas. They are generally negative over the Arctic Ocean and over land (see supplemental material). The largest deviations with respect to the median are in NCEP-R2,

particularly over land. MERRA has positive deviations over the Arctic Ocean and to a lesser extent over land. In the summer, the latent heat flux is strongly positive over the land and small or slightly negative over the oceans. Again the largest deviations are in NCEP-R2 over land.

5. Comparison of reanalysis trends

a. Near-surface air temperature

All datasets show significant positive annual-mean 2-m air temperature trends (Fig. 9). All show significant trends in the Pacific sector of the Arctic Ocean, where substantial ice loss has occurred in the late summer and early fall. Other regions showing a significant trend in most of the models are the Barents Sea and the eastern portion of Canada, although the strength and spatial patterns of the trends differ among the models with the two NCEP models having generally stronger trends. The ERA-Interim model uses the observed 2-m temperature in postprocessing so this model is expected to be most representative of the true trends. The North Pacific and the west coast of North America have no significant trend as does the region north of Canada where the ice is generally thicker. The trends are strongly positive in all of the models in regions where the sea ice has diminished most, north of Asia and Alaska, but other regions also show positive trends.

FIG. 8. As in Fig. 5, but for annual precipitation.

The annual trends from the reanalysis models are generally lower in the Arctic than those computed by the Goddard Institute for Space Studies (GISS) Surface Temperature Analysis (GISTEMP) dataset (Hansen et al. 2010). Figure 10 shows the annual-mean 2-m air temperature trend from GISTEMP 250-km smoothing radius product over the same 1980–2009 period used in this study. In many land areas, the GISTEMP trends are more than twice the median air temperature trends from the reanalyses. All of the reanalyses appear to be underestimating the annual trend as estimated by GISTEMP.

An analysis of seasonal temperature trends (see the figures in the supplemental material) shows that in the spring the trends are strongly positive over the Arctic Ocean; trends over the land surface, however, are not generally significantly different from zero at the 95% significance level. All reanalyses have warming over the Asian landmass and cooling over the North American landmass but the trends are not generally significant at the 95% level. The strongest trends in the median field are in the East Siberian, Barents, and Labrador Seas.

Most of the summer trends show little or no warming over the Arctic Ocean, as would be expected in a melting ice bath, but show significant warming over Europe and central Asia. The exception is NCEP-R2, which has significant warming over much of the Arctic Ocean. NCEP-R1 is the only model that has very strong positive trends over Alaska; all models have significant positive trends in the North Atlantic.

In the fall there is strong warming over the Chukchi Sea and to the east in the Arctic Ocean, where there have been significant reductions in sea ice extent but lesser warming than over most of the landmasses. Eastern North America and the North Atlantic have smaller trends, but they are generally significant at the 95% level. The patterns of warming are generally similar in the different models, although the two NCEP models have more widespread warming over the Arctic Ocean. In the winter, the warming is generally less than in the other seasons and there is slight cooling over the Asian landmass. The only region with significant warming in all models is in the Barents Sea.

b. Radiative fluxes

The downwelling longwave radiative fluxes are linked closely to the near-surface air temperatures so the trends in these fluxes have similar patterns to those of the 2-m air temperature. The downwelling shortwave fluxes depend on the model cloud and aerosol properties and have strongly differing patterns (Fig. 11). There are significant negative trends for the majority of the models in the Canadian Archipelago and the Barents Sea indicating increasing cloud cover in these regions. An area in eastern Asia has an increasing trend. The spatial patterns in the trends are otherwise highly variable. The most notable outlier is the MERRA model, which has very strong negative trends over the oceans, most evident in the North Pacific, again indicative of a strong positive trend in cloudiness.

FIG. 9. Comparisons of annual T2M trends. All trends are actual, not anomalies. The trend is plotted only for regions exceeding 95% confidence that the trend is nonzero. The top left map is the median trend of all seven models.

c. Precipitation

The spatial patterns for the trend in the annual precipitation are in general agreement for the different models, despite differences in climatology for clouds, precipitable water, and annual precipitation. Figure 12 shows the decadal trends as a fraction of the mean annual precipitation. The only region with a significant trend (downward) in most of the models is in eastern Asia. Otherwise, much of the region does not have significant trends in precipitation. The spatial patterns in the trends for the three best models for precipitation (MERRA, ERA-Interim, and JRA-25) are not consistent. Also note that the ringing in the NCEP-R1 precipitation fields is not apparent in the trends.

6. Forcing an ice–ocean model

Returning to one of the motivating questions, that of how the datasets compare for use in simulating past sea ice conditions, a coupled ice–ocean model was forced with winds, temperatures, and downwelling radiative fluxes from four of the datasets (NCEP-R1, CFSR, MERRA, and ERA-Interim) for the period 1980–2009. Sea ice volume for the entire Arctic is calculated using PIOMAS (Zhang and Rothrock 2003). To allow us to assess the full impact of different forcing datasets, data assimilation in PIOMAS is turned off for this study. To

account for mean differences in the forcing fields, sea ice model parameters for each dataset are adjusted to minimize the mean difference between the estimated ice draft and that measured by submarines. The two

FIG. 10. Trends in the annual-mean air temperature from GISTEMP for the period 1980–2009 with a 250-km smoothing radius. (Data are from data.giss.nasa.gov/gistemp.)

FIG. 11. As in Fig. 8, but for trends in the (top) summer surface downwelling shortwave radiative flux and (bottom) winter surface downwelling longwave radiative flux.

adjustable parameters in this study are the summer ice albedo and the sea ice aerodynamic roughness. Submarine sea ice draft data from the Unified Sea Ice Thickness Climate Data Record are used for comparisons to model estimates (1979–2005; 50-km averages,

$N = 1647$) (Lindsay 2010; Lindsay 2013). After adjustments, the mean bias in draft ranged between -0.02 and 0.30 m. The temporal and spatial variability of the model ice thickness, however, is not easily improved by simple parameter adjustments. Following model tuning,

FIG. 12. As in Fig. 8, but for annual precipitation trends expressed as a percent of the mean annual total precipitation at each location.

the correlation between the model ice draft and the measured ice draft for the four models is 0.76, 0.79, 0.80, and 0.82 for ERA-Interim, CFSR, NCEP-R1, and MERRA, respectively. These correlations are for 50-km ice draft measurements taken within the data release area for the submarines from different months, but primarily in the spring and fall. Thus, MERRA does a little better in simulating the ice draft variability than the other products and the difference in the correlation from the next best, NCEP-R1, is significant at a p value of 0.98.

There are also significant differences in the time series of the simulated total ice volume estimates (Fig. 13), particularly in March in the first half of the record, and the resulting ice volume trends differ. The trends in March are -2.5 , -2.9 , -3.6 , and $-3.4 \pm 0.03 \times 10^3 \text{ km}^3 \text{ decade}^{-1}$ for CFSR, ERA-Interim, MERRA, and NCEP-R1, respectively (1980–2009). Thus, the trend computed using MERRA and NCEP-R1 forcings is significantly stronger than for the other two. In September the trends are -2.7 , -3.4 , -4.2 , and $-4.1 \pm 0.3 \times 10^3 \text{ km}^3 \text{ decade}^{-1}$ in the same order. Again the trend computed using MERRA and NCEP-R1 forcings is significantly stronger than for the other two.

These experiments show that there are important differences in the results of ice simulations depending on which forcing dataset is used. This sensitivity of coupled ice–ocean models to the atmospheric forcing dataset was

also found by both Hunke and Holland (2007) and Notz et al. (2013). In the Notz et al. (2013) study, the simulated September ice volume was less than half as much using ERA-Interim forcing compared to using NCEP-R1 forcing in part because the model was not fully adjusted to the different biases of each dataset. As a result the trend in September ice volume was also very different,

FIG. 13. Total Arctic sea ice volume computed by PIOMAS for (top) March and (bottom) September using forcing data from four different reanalysis datasets.

$-2.6 \times 10^3 \text{ km}^3 \text{ decade}^{-1}$ for the NCEP-R1 forcing (1979–2007) compared to a much weaker trend in the run with thinner ice, $-1.6 \times 10^3 \text{ km}^3 \text{ decade}^{-1}$ for the ERA-Interim forcing. In contrast, our range of differences in mean state and volume trends using different forcing datasets is substantially smaller and in line with uncertainties estimated through independent ice thickness validation (Schweiger et al. 2011). The ability of this class of models to reproduce the mean state is important for assessing trend sensitivities of the models to different forcing datasets.

7. Discussion and conclusions

In a comparison of the monthly averaged products from seven different reanalysis efforts for the period 1980–2009, some of the fields that are related directly to observations differ little among products, while the fields that are less closely related to observations in some cases differ substantially.

In comparison to observations the 2-m air temperature for four models were generally closely correlated to the observations and showed small biases: CFSR, MERRA, ERA-Interim, and JRA-25. For the longwave radiative fluxes, the models with the smallest bias and best correlations are CFSR, MERRA, and ERA-Interim, although in the summer the correlations were all rather poor, below 0.7. The shortwave radiative flux bias was also smallest for CFSR, MERRA, and ERA-Interim. The summer correlations were also generally poor, near 0.7 or less. However, this may simply be due to persistent large cloud fractions during summer and the relatively small interannual variability in the longwave flux yielding a small signal-to-noise ratio. Precipitation for land stations has the smallest bias and largest correlation in three models: MERRA, ERA-Interim, and JRA-25. Large biases were found in the CFSR and 20CR products. Compared to observations from drifting ice camps, the smallest bias for wind speed is found with the CFSR model, but the biases of the MERRA and ERA-Interim models are not much larger. The best correlation for wind speed is for the ERA-Interim model in all seasons, rising to $R = 0.93$ in June. In general CFSR, MERRA, and ERA-Interim all perform well compared to the observations, but ERA-Interim has more consistently good scores for the observations we analyzed.

Reanalysis seasonal-mean fields are compared by first computing the median value from the seven models and then determining the deviation from the median for each model. As expected, sea level pressure fields differed very little, except for the 20CR model. The winter-mean wind speeds were in general agreement except the NCEP-R1 model is lower than the median over the

oceans, consistent with the comparisons to the North Pole drifting ice station observations; the NCEP-R2 model has higher wind speeds over land. Summer 2-m air temperatures are in agreement for the seven different models but there is greater variability among them in winter. The 20CR product is much too warm and NCEP-R1 is colder than the others over sea ice. There is large variability among the models for the downwelling radiation. The two NCEP models produce much less longwave flux and much more shortwave flux in the summer than the other models, consistent with the comparisons to observations and the likely too small representation of cloud fraction in these two models. The TOA upwelling longwave flux is much greater in the JRA-25 model than in the others. There is large variability among the models in the summer TOA upwelling shortwave flux, also reflecting the large variations in model cloud coverage results. Total annual-mean precipitation is markedly higher in CFSR and 20CR than the others; these two products share the same atmospheric model. The turbulent sensible and latent heat fluxes have the greatest variability over land and the variations from the median are large, exceeding 50 W m^{-2} in some locations. The recent study by Bourassa et al. (2013) also found large variations in the surface fluxes in reanalysis and other gridded surface flux fields.

Kistler et al. (2001) point out that agreement among reanalyses in the trend is an important and necessary, but not sufficient, condition for confidence in the trends because of the changing nature of the observational basis of the reanalyses. Here we show there is large variability between the different models and, in the case of the 2-m air temperature, between the models and a global quality-controlled air temperature dataset, GISTEMP. The spatial patterns of the seasonal trends, and indeed the signs of the trends, differ between the models in some locations. Only in the fall do all the models agree; there is a significant warming trend in the Pacific sector of the Arctic Ocean, where sea ice decline is most pronounced. These observations further reinforce the notion that trends in the reanalysis products should be used with caution and that uncertainties at the regional scale are substantial.

Sea ice thickness and total ice volume were computed with the PIOMAS coupled ice–ocean model using forcing from four of the reanalysis projects. The results are broadly similar but in comparison to submarine ice draft measurements forcing fields from one of the product sets show a better correlation, the MERRA model. The bias of the different forcing sets is accounted for by adjusting parameters in the ice model to minimize the bias. Ice volume trends are different in the different simulations with CFSR forcing producing the smallest

trend and NCEP-R1 the largest, 50% larger than that of CFSR. Determining the specific mechanisms responsible for producing these differences is beyond the scope of this study but would be an interesting line of research.

Global atmospheric and oceanic reanalysis is a rapidly developing field of research. Advances are being made in observational systems and quality control, data assimilation, and the core forecast models. New major efforts are anticipated from several groups. The current status of various reanalysis projects can be found at the Reanalysis Intercomparison and Observations web page (reanalyses.org) or the Atmospheric Reanalysis section of the NCAR Climate Data Guide (<https://climatedataguide.ucar.edu/reanalysis/atmospheric-reanalysis-overview-comparison-tables>).

Acknowledgments. This study was supported by the NOAA Climate Program Office Modeling Analysis Prediction and Projection Program, award NA11OAR4310085, and the NASA Cryospheric Sciences Program. The NCEP-R1 data are from the Research Data Archive (RDA), which is maintained by the Computational and Information Systems Laboratory (CISL) at the National Center for Atmospheric Research (NCAR). NCAR is sponsored by the National Science Foundation (NSF). The original data are available from the RDA (<http://rda.ucar.edu>) in dataset number ds090.0. NCEP-R2 data are provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their website (<http://www.esrl.noaa.gov/psd/>). The 20th Century Reanalysis V2 data are provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their website (<http://www.esrl.noaa.gov/psd/>). The CFSR data were developed by NOAA's National Centers for Environmental Prediction (NCEP). The data for this study are from NOAA's National Operational Model Archive and Distribution System (NOMADS), which is maintained at NOAA's National Climatic Data Center (NCDC). The MERRA data were provided by the Global Modeling and Assimilation Office (GMAO) and the GES DISC for the dissemination of MERRA. The ERA-Interim data are from the NCEP Research Data Archive (RDA) (rda.ucar.edu) in dataset number ds627.0. The JR-25 datasets used for this study are provided from the cooperative research project of the JRA-25 long-term reanalysis by the Japan Meteorological Agency (JMA) and the Central Research Institute of Electric Power Industry (CRIEPI). ARM data are made available through the U.S. Department of Energy as part of the Atmospheric Radiation Measurement Program. The authors thank the reviewers for helpful and constructive comments.

REFERENCES

- Bitz, C. M., and Q. Fu, 2008: Arctic warming aloft is data set dependent. *Nature*, **455**, E3–E4, doi:10.1038/nature07258.
- Bosilovich, M. L. G., J. Chen, F. R. Robertson, and R. F. Adler, 2008: Evaluation of global precipitation in reanalyses. *J. Appl. Meteor. Climatol.*, **47**, 2279–2299.
- Bourassa, M. A., and Coauthors, 2013: High-latitude ocean and sea ice surface fluxes: Challenges for climate research. *Bull. Amer. Meteor. Soc.*, **94**, 403–423.
- Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones, 2006: Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J. Geophys. Res.*, **111**, D12106, doi:10.1029/2005JD006548.
- Bromwich, D. H., and R. L. Fogt, 2004: Strong trends in the skill of the ERA-40 and NCEP–NCAR reanalyses in the high and midlatitudes of the Southern Hemisphere, 1958–2001. *J. Climate*, **17**, 4603–4619.
- , —, K. I. Hodges, and J. E. Walsh, 2007: A tropospheric assessment of the ERA-40, NCEP, and JRA-25 global reanalyses in the polar regions. *J. Geophys. Res.*, **112**, D10111, doi:10.1029/2006JD007859.
- Cavalieri, D. J., P. Gloersen, and W. J. Campbell, 1984: Determination of sea ice parameters with the NIMBUS 7 SMMR. *J. Geophys. Res.*, **89** (D4), 5355–5369.
- , C. L. Parkinson, P. Gloersen, and H. J. Zwally, 1996 (updated yearly): Sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS passive microwave data. National Snow and Ice Data Center Boulder, CO, digital media. [Available online at <http://nsidc.org/data/nsidc-0051.html>.]
- , —, —, J. C. Comiso, and H. J. Zwally, 1999: Deriving long-term time series of sea ice cover from satellite passive-microwave multisensor data sets. *J. Geophys. Res.*, **104** (C7), 15 803–15 814.
- Chernokulsky, A., and I. Mokhov, 2012: Climatology of total cloudiness in the Arctic: An intercomparison of observations and reanalyses. *Adv. Meteor.*, **2012**, 542093, doi:10.1155/2012/542093.
- Compo, G. P., J. S. Whitaker, and P. D. Sardeshmukh, 2006: Feasibility of a 100-year reanalysis using only surface pressure data. *Bull. Amer. Meteor. Soc.*, **87**, 175–190.
- , and Coauthors, 2011: The Twentieth Century Reanalysis Project. *Quart. J. Roy. Meteor. Soc.*, **137**, 1–28.
- Cullather, R. L., and M. Bosilovich, 2011: The moisture budget of the polar atmosphere in MERRA. *J. Climate*, **24**, 2861–2879.
- , and —, 2012: The energy budget of the polar atmosphere in MERRA. *J. Climate*, **25**, 5–24.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.
- Grant, A. N., S. Brönnimann, and L. Haimberger, 2008: Recent Arctic warming vertical structure contested. *Nature*, **455**, E2–E3, doi:10.1038/nature07257.
- Grumbine, R. W., 1996: Automated passive microwave sea ice concentration analysis at NCEP. OMB Tech. Note 120, NCEP, 13 pp.
- Hansen, J., R. Ruedy, M. Sato, and K. Lo, 2010: Global surface temperature change. *Rev. Geophys.*, **48**, RG4004, doi:10.1029/2010RG000345.
- Hines, K. M., D. H. Bromwich, and G. J. Marshall, 2000: Artificial surface pressure trends in the NCEP–NCAR reanalysis over the Southern Ocean and Antarctica. *J. Climate*, **13**, 3940–3952.

- Hunke, E. C., and J. K. Dukowicz, 1997: An elastic-viscous-plastic model for sea ice dynamics. *J. Phys. Oceanogr.*, **27**, 1849–1867.
- , and M. M. Holland, 2007: Global atmospheric forcing data for Arctic ice–ocean modeling. *J. Geophys. Res.*, **112**, C04S14, doi:10.1029/2006JC003640.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP–DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, **83**, 1631–1643.
- Kistler, R., and Coauthors, 2001: The NCEP–NCAR 50-Year Reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.*, **82**, 247–267.
- Lindsay, R. W., 1998: Temporal variability of the energy balance of thick Arctic pack ice. *J. Climate*, **11**, 313–333.
- , 2010: New unified sea ice thickness climate data record. *Eos, Trans. Amer. Geophys. Union*, **91**, 405–406.
- , J. Zhang, A. J. Schweiger, M. A. Steele, and H. Stern, 2009: Arctic sea ice retreat in 2007 follows thinning trend. *J. Climate*, **22**, 165–176.
- , 2013: Unified sea ice thickness climate data record collection spanning 1947–2012. National Snow and Ice Data Center, Boulder, CO, digital media. [Available online at <http://nsidc.org/data/G10006>.]
- Liu, J., Z. Zhang, Y. Hu, L. Chen, Y. Dai, and X. Ren, 2008: Assessment of surface air temperature over the Arctic Ocean in reanalysis and IPCC AR4 model simulations with IABP/POLES observations. *J. Geophys. Res.*, **113**, D10105, doi:10.1029/2007JD009380.
- Makstas, A., D. Atkinson, M. Kulakov, S. Shutin, R. Krishfield, and A. Proshutinsky, 2007: Atmospheric forcing validation for modeling the central Arctic. *Geophys. Res. Lett.*, **34**, L20706, doi:10.1029/2007GL031378.
- Matsumoto, T., M. Ishii, Y. Fukuda, and S. Hirahara, 2006: Sea ice data derived from microwave radiometer for climate monitoring. *Extended Abstracts, 14th Conf. on Satellite Meteorology and Oceanography*, Atlanta, GA, Amer. Meteor. Soc., P2.21. [Available online at https://ams.confex.com/ams/Annual2006/techprogram/paper_101105.htm.]
- Notz, D., F. A. Haumann, H. Haak, J. H. Jungclauss, and J. Marotzke, 2013: Arctic sea-ice evolution as modeled by Max Planck Institute for Meteorology's Earth system model. *J. Adv. Model. Earth Syst.*, **5**, 173–194, doi:10.1002/jame.20016.
- Onogi, K., and Coauthors, 2007: The JRA-25 Reanalysis. *J. Meteor. Soc. Japan*, **85**, 369–432.
- Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate*, **15**, 1609–1625.
- Rienecker, M. M., and Coauthors, 2008: The GEOS-5 Data Assimilation System—Documentation of versions 5.0.1, 5.1.0, and 5.2.0. NASA Technical Report Series on Global Modeling and Data Assimilation, Vol. 27, NASA/TM-2008-105606, 118 pp.
- Rudolf, B., and U. Schneider, 2005: Calculation of gridded precipitation data for the global land-surface using in-situ gauge observations. *Proc. Second Workshop of the International Precipitation Working Group*, Monterey, CA, EUMETSAT, 231–247.
- , A. Becker, U. Schneider, A. Meyer-Christoffer, and M. Ziese, 2010: GPCC status report December 2010: On the most recent gridded global dataset issued in fall 2010 by the Global Precipitation Climatology Centre (GPCC). GPCC, 7 pp.
- Saha, S., and Coauthors, 2006: The NCEP Climate Forecast System. *J. Climate*, **19**, 3483–3517.
- , and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057.
- Schweiger, A., R. Lindsay, J. Zhang, M. Steele, H. Stern, and R. Kwok, 2011: Uncertainty in modeled Arctic sea ice volume. *J. Geophys. Res.*, **116**, C00D06, doi:10.1029/2011JC007084.
- Screen, J. A., and I. Simmonds, 2011: Erroneous Arctic temperature trends in the ERA-40 reanalysis: A closer look. *J. Climate*, **24**, 2620–2627.
- Serreze, M. C., A. P. Barrett, and F. Lo, 2005: Northern high-latitude precipitation as depicted by atmospheric reanalyses and satellite retrievals. *Mon. Wea. Rev.*, **133**, 3407–3430.
- , —, and J. Stroeve, 2012: Recent changes in tropospheric water vapor over the Arctic as assessed from radiosondes and atmospheric reanalyses. *J. Geophys. Res.*, **117**, D10104, doi:10.1029/2011JD017421.
- Simmons, A. J., and Coauthors, 2004: Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature. *J. Geophys. Res.*, **109**, D24115, doi:10.1029/2004JD005306.
- Thorne, P. W., 2008: Arctic tropospheric warming amplification? *Nature*, **455**, E1–E2, doi:10.1038/nature07256.
- Trenberth, K., and C. Guillemot, 1998: Evaluation of the atmospheric moisture and hydrological cycle in the NCEP reanalyses. *Climate Dyn.*, **14**, 213–231.
- Walsh, J. E., W. L. Chapman, and D. H. Portis, 2009: Arctic cloud fraction and radiative fluxes in atmospheric reanalyses. *J. Climate*, **22**, 2316–2334.
- Whitaker, J. S., G. P. Compo, X. Wei, and T. M. Hamill, 2004: Reanalysis without radiosondes using ensemble data assimilation. *Mon. Wea. Rev.*, **132**, 1190–1200.
- Winton, M., 2000: A reformulated three-layer sea ice model. *J. Atmos. Oceanic Technol.*, **17**, 525–531.
- Zhang, J. L., and D. A. Rothrock, 2003: Modeling global sea ice with a thickness and enthalpy distribution model in generalized curvilinear coordinates. *Mon. Wea. Rev.*, **131**, 845–861.
- Zib, B. J., X. Dong, B. Xi, and A. Kennedy, 2012: Evaluation and intercomparison of cloud fraction and radiative fluxes in recent reanalyses over the Arctic using BSRN surface observations. *J. Climate*, **25**, 2291–2305.