Assessment of Arctic Cloud Cover Anomalies in Atmospheric Reanalysis Products Using Satellite Data

YINGHUI LIU

Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin–Madison, Madison, Wisconsin

JEFFREY R. KEY

Center for Satellite Applications and Research, NOAA/NESDIS, Madison, Wisconsin

(Manuscript received 4 December 2015, in final form 16 May 2016)

ABSTRACT

Cloud cover is one of the largest uncertainties in model predictions of the future Arctic climate. Previous studies have shown that cloud amounts in global climate models and atmospheric reanalyses vary widely and may have large biases. However, many climate studies are based on anomalies rather than absolute values, for which biases are less important. This study examines the performance of five atmospheric reanalysis products—ERA-Interim, MERRA, MERRA-2, NCEP R1, and NCEP R2—in depicting monthly mean Arctic cloud amount anomalies against Moderate Resolution Imaging Spectroradiometer (MODIS) satellite observations from 2000 to 2014 and against *Cloud–Aerosol Lidar and Infrared Pathfinder Satellite Observation* (*CALIPSO*) observations from 2006 to 2014. All five reanalysis products exhibit biases in the mean cloud amount, especially in winter. The Gerrity skill score (GSS) and correlation analysis are used to quantify their performance in terms of interannual variations. Results show that ERA-Interim, MERRA, MERRA-2, and NCEP R2 perform similarly, with annual mean GSSs of 0.36/0.22, 0.31/0.24, 0.32/0.23, and 0.32/0.23 and annual mean correlation coefficients of 0.50/0.51, 0.43/0.54, 0.44/0.53, and 0.50/0.52 against MODIS/*CALIPSO*, indicating that the reanalysis datasets do exhibit some capability for depicting the monthly mean cloud amount anomalies. There are no significant differences in the overall performance of reanalysis products. They all perform best in July, August, and September and worst in November, December, and January. All reanalysis datasets have better performance over land than over ocean. This study identifies the magnitudes of errors in Arctic mean cloud amounts and anomalies and provides a useful tool for evaluating future improvements in the cloud schemes of reanalysis products.

1. Introduction

The Arctic has changed dramatically over the past few decades. It is warming at a higher rate than any other region in the world, a phenomenon known as polar amplification (Serreze and Francis 2006; Holland and Bitz 2003). Arctic sea ice extent and thickness have been decreasing dramatically (Maslanik et al. 2007; Kwok and Untersteiner 2011; Stroeve et al. 2012), and there have been important changes to the large-scale atmospheric circulation (Zhang et al. 2008; Wu and Zhang 2010; Wu et al. 2014). It has become increasingly clear that the Arctic is a highly variable and sensitive region in the global climate system (Walsh et al. 2002; Francis et al. 2009).

Changes in cloud cover and cloud properties affect the surface energy budget and contribute to sea ice growth and melt, which in turn feeds back to cloud formation (e.g., Kay et al. 2008; Liu and Key 2014; Liu et al. 2012a; Schweiger et al. 2008). Unlike clouds on a global scale, Arctic clouds warm the surface most of the year except for a short period in the summer (Intrieri et al. 2002; Wang and Key 2005; Schweiger and Key 1994). The influence of clouds on sea ice is particularly relevant for studies of Arctic climate. The interaction between Arctic clouds and sea ice has been examined by a number of investigators recently, primarily with regard to

Denotes Open Access content.

Corresponding author address: Yinghui Liu, CIMSS, 1225 West Dayton Street, Madison, WI 53706.
E-mail: yinghui.liu@ssec.wisc.edu

DOI: 10.1175/JCLI-D-15-0861.1

record minimum ice extents in 2007 and 2012. Liu and Key (2014), Taylor et al. (2015), and Letterly et al. (2016) provide a review of the relevant literature.

Arctic cloud formation and dissipation mechanisms are complex (Curry et al. 1996; Vavrus and Waliser 2008; Beesley and Moritz 1999; Walsh et al. 2009), and clouds are one of the main sources of uncertainty in modeling the Arctic climate (Solomon et al. 2007; Boucher et al. 2013). There are large discrepancies in modeled clouds, the reasons for which are discussed elsewhere (Randall et al. 1998; Sandvik et al. 2007; Walsh et al. 2002, 2005, 2009; Vavrus 2004; Inoue et al. 2006; Birch et al. 2009; Wyser et al. 2008; de Boer et al. 2012), and radiative fluxes in both regional models (Aas et al. 2015; Klaus et al. 2012; Paquin-Ricard et al. 2010; Simjanovski et al. 2011; Tjernström et al. 2008; Wilson et al. 2012; Wyser et al. 2008) and global climate models (Du et al. 2011; Vavrus et al. 2009; Walsh et al. 2005, 2002; Xie et al. 2013; Zhang et al. 2005; Zhao and Wang 2010).

Arctic clouds and surface radiation have been examined in some of the major atmospheric reanalysis products (Bromwich et al. 2007; Chaudhuri et al. 2014; Chernokulsky and Mokhov 2012; Clark and Walsh 2010). Lindsay et al. (2014) compared Arctic surface temperature, radiative fluxes, precipitation, and wind speed in seven reanalysis products to surface observations and found that three were most consistent with observations: the Climate Forecast System Reanalysis (CFSR), the National Aeronautics and Space Administration (NASA) Modern-Era Retrospective Analysis for Research and Applications (MERRA), and the European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis (ERA-Interim). Zib et al. (2012) used Baseline Surface Radiation Network (BSRN) observations from Barrow and Ny-Alesund to intercompare cloud fraction and surface radiative fluxes in MERRA, CFSR, ERA-Interim, the NOAA Twentieth Century Reanalysis Project (20CR), and the National Centers for Environmental Prediction (NCEP) U.S. Department of Energy (DOE) Reanalysis-2 (NCEP R2). They found that all of the reanalyses showed large biases in cloud fraction, especially in winter, and that the cloud fraction biases lead to surface radiation biases. Walsh et al. (2009, 2002) examined cloud fraction and radiative fluxes in the NCEP–National Center for Atmospheric Research (NCAR) Reanalysis-1 (NCEP R1), the 40-yr ECMWF Re-Analysis (ERA-40), the NCEP–NCAR North American Regional Reanalysis (NARR), and the Japan Meteorological Agency and Central Research Institute of Electric Power Industry Japanese 25-year Reanalysis (JRA-25) through comparisons to surface data from Barrow, Alaska. They found that the reanalyses simulate the radiative fluxes well if and when the cloud fraction is simulated correctly, but that the systematic errors are substantial. Cloud fraction and radiation biases showed considerable variability in the annual mean and seasonal cycle. ERA-40 was found to perform best for shortwave and longwave fluxes because its cloud cover was the most realistic.

Satellite data have been employed, though not widely, in evaluating clouds and radiation in models. Zhang et al. (2005) and Kay et al. (2012) are two examples of using satellite data to examine clouds in general circulation models. For the Arctic, Liu et al. (2005) used two satellite datasets and surface measurements to evaluate Arctic surface radiation fluxes in NCEP R2 and ERA-40. They found that one of the satellite products and NCEP R2 were most similar to in situ measurements of surface temperature, that both satellite products were more accurate for downwelling shortwave radiation, and that ERA-40 was best for downwelling longwave fluxes. Zygmuntowska et al. (2012) compared Arctic cloud amount and surface radiation in ERA-Interim to surface measurements and cloud cover detected with active and passive satellite instruments. They found that the total cloud cover in ERA-Interim agreed well with surface observations, but that cloud amount and surface radiation in the two satellite products and ERA-Interim were very different, particularly in winter.

In data-sparse regions, such as the Arctic, reanalysis products are important for studies of climate variability and change because they synthesize the generally sparse observational data. But with large uncertainties and inconsistencies in modeled cloud amount (Zib et al. 2012; Walsh et al. 2002, 2009; Zygmuntowska et al. 2012), can reanalyses produce reasonable and consistent interannual variations in clouds and other variables that are related to clouds, such as surface radiation? Furthermore, if cloud cover in a reanalysis product exhibits a bias, does that preclude its use in climate studies that are based on anomalies rather than absolute cloud amount? For example, Liu and Key (2014) examined the relationship between wintertime cloud anomalies and summertime sea ice anomalies in 2013. Letterly et al. (2016) extended that work to the period 1983–2013. In both studies, a bias in the absolute cloud amount was not relevant because the anomalies were found to be realistic.

This study examines the performance of five reanalysis products in depicting the interannual cloud amount variations in the Arctic. The monthly mean cloud amount (fractional cloud cover) anomalies are evaluated using satellite-derived anomalies from the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument on the NASA *Terra* and *Aqua* satellites and the Cloud–Aerosol Lidar with Orthogonal Polarization (CALIOP) instrument on the *Cloud–Aerosol Lidar and Infrared Pathfinder Satellite Observation* (*CALIPSO*) satellite. The satellite-derived cloud information from MODIS is not necessarily considered as "truth", but rather as the most spatially complete observational dataset available and, therefore, as a useful reference point. CALIOP, being an active sensor, has superior cloud detection capability in the polar regions, especially at night, when information in visible channels is not available. The period of study is restricted to the periods of data availability: 2000–14 for MODIS and 2006–14 for CALIOP/*CALIPSO* (here called *CALIPSO*). The five reanalysis products are NCEP R1, NCEP R2, ERA-Interim, MERRA, and the second version of MERRA (MERRA-2). This work extends previous studies in that satellite data provide more complete spatial coverage than in situ point measurements, five major reanalysis products are examined, and their skill in depicting both cloud cover (amount) and monthly anomalies is assessed.

We first compare the monthly mean cloud amounts from MODIS, *CALIPSO*, and reanalysis products to show their differences and similarities. We then present cases of monthly mean cloud amount anomalies in January 2013 and June 2013 to show how each reanalysis product performs in describing the negative and positive anomalies. Finally, we assess the overall performance of these five reanalysis products in depicting the monthly mean cloud amount anomalies using the Gerrity skill score (GSS; Gerrity 1992; Livezey 2011), and correlation analysis for the entire Arctic (60°–90°N), over the Arctic Ocean and over Arctic land.

2. Data and methods

The MODIS instrument on NASA's *Terra* and *Aqua* satellites measures radiances at 36 wavelengths, including infrared and solar channels, at spatial resolutions of 250 m, 500 m, and 1 km, depending on the channel. Improvements in the cloud detection algorithm have been made over recent years (Ackerman et al. 1998; Liu et al. 2004; Frey et al. 2008). During the day (sunlight), 2.7% of the cloudy cases identified by surface radar/lidar are misidentified as clear in the MODIS cloud mask, and 6.9% of the clear cases identified by surface radar/lidar are misidentified as cloud; at night, the misidentification rates of cloud as clear and clear as cloud are 16.3% and 8.6% for MODIS in the Arctic (Liu et al. 2004). Liu et al. (2010) used active satellite sensors to evaluate errors in the MODIS cloud detection over the Arctic. The results revealed a dependence of MODIS cloud amount on sea ice concentration such that the differences between MODIS and spaceborne radar–lidar cloud amount are −0.1% per 1% sea ice concentration (Liu et al. 2010).

The MODIS level-2 cloud mask (MOD35_L2 for *Terra* and MYD35_L2 for *Aqua*) provides cloud cover with four confidence levels: confident clear, probably clear, uncertain/probably cloudy, and cloudy. All level-2 MODIS cloud mask granules are used to calculate the daily statistics, and the monthly level-3 product is computed from all the daily files in a particular month on an equal-angle, 1° latitude by 1° longitude grid. In each grid cell, cloud fraction is the percentage of cloudy and uncertain/probably cloudy L2 pixels in all available level-2 pixels (Hubanks et al. 2015). Here, we use the monthly mean cloud fraction (cloud amount) from 2000 to 2014. The mean and standard deviation of monthly mean cloud amount from 2000 to 2014 are calculated at each grid point, as are the monthly mean anomalies.

Monthly mean cloud amounts from NCEP R1, NCEP R2, MERRA, MERRA-2, and ERA-Interim over the same period are used. Other reanalysis products are not included because they did not cover the entire time period when MODIS data were available (2000–14) when we started this study. Monthly mean cloud amounts at spectral T62 (210 km) model resolution from NCEP R1 and NCEP R2, 0.66° longitude by 0.5° latitude from MERRA, 0.625° longitude by 0.5° latitude from MERRA-2, and 1° longitude by 1° latitude from ERA-Interim are employed. All monthly means are interpolated to a 1° × 1° latitude–longitude grid using bilinear interpolation to match the MODIS data. The mean and standard deviation of monthly mean cloud amount from 2000 to 2014 are calculated for each reanalysis product at each grid point, as are the monthly mean anomalies.

The *CALIPSO* vertical feature mask (VFM) from 2006 to 2014 at 5-km resolution (Vaughan et al. 2009) from the Atmospheric Science Data Center of the NASA Langley Research Center is used to calculate a monthly mean cloud amount at 10° longitude by 3° latitude resolution using the approach in Liu et al. (2012b). It is important to note, however, that while an active instrument like CALIOP may be more sensitive to clouds than a passive instrument like MODIS, its limited sampling introduces some uncertainty. Liu (2015) estimated the error in monthly mean cloud amount due to the limited, nadir-view sampling for sensors like CALIOP and found that the errors are less than 6.5% (11.5%), with a probability of 80% (95%) for a 100-km resolution Equal-Area Scalable Earth Grid (EASE-Grid) cell with a sample size of 1000 and that errors decrease with increasing sample size. In this study, the *CALIPSO* monthly mean cloud cover is calculated at 10° longitude by 3° latitude with sample sizes larger than 1000 to minimize the sampling error. The mean and standard deviation of monthly mean cloud amount from 2006 to 2014 are calculated at each grid point, as are the monthly mean anomalies.

To match the *CALIPSO* data in time and space, the monthly mean cloud amounts from MODIS and all reanalysis products from 2006 to 2014 are averaged at

10° longitude by 3° latitude weighted by the square root of the cosine of latitude. Standard deviations of monthly mean cloud amount and monthly anomalies from 2006 to 2014 are then calculated.

NCEP R1 (Kalnay et al. 1996; Kistler et al. 2001) uses the NCEP global spectral model for assimilation, including a simplified Arakawa–Schubert convective parameterization scheme (Pan and Wu 1994; Grell 1993) and a diagnostic cloud scheme (Campana et al. 1994) with a parameterized relative humidity–cloud cover relationship. NCEP R2 (Kanamitsu et al. 2002) uses the same cloud schemes with a more realistic cloud cover simulation as a result of further tuning of the schemes. MERRA uses the Goddard Earth Observing System (GEOS), version 5 (Rienecker et al. 2011), which includes a prognostic cloud scheme (Bacmeister et al. 2006) and a modified version of the relaxed Arakawa–Schubert convective scheme (Moorthi and Suarez 1992). MERRA-2 (GMAO 2015) uses an updated GEOS-5 data assimilation system; major changes are summarized in Reichle et al. (2015). ERA-Interim (Dee et al. 2011) uses the ECMWF Integrated Forecast System (IFS), which includes a full prognostic approach for condensate cloud fraction.

The GSS (Gerrity 1992; Livezey 2011) is used to assess the capability of the reanalysis products in depicting cloud amount interannual variation. GSS is a score recommended by the World Meteorological Organization (WMO) for overall assessment of forecast quality, using two or more categories, when compared to independent observations (WMO 2010). The GSS is commonly used to measure categorical forecast quality using independent observations, but it can also be applied for assessing ranges of continuous variables: for example, to evaluate model performance in forecasting precipitation (Lee and Seo 2013) and 2-m air temperature (Stefanova et al. 2012). One advantage of the GSS is that it has reasonable reward matrices for successful forecasting of less likely events (Livezey 2011).

In this study, GSS is applied to assess how the reanalyses perform in depicting the below-normal, normal, and above-normal monthly mean cloud anomalies. The ''forecasts'' are Arctic cloud amount monthly anomalies from a reanalysis product at each grid cell. The independent observations are the MODIS cloud amount monthly anomalies. There are three categories: below normal, normal, and above normal, defined as the anomalies more than one standard deviation below the mean, between one standard deviation below and above the mean, and more than one standard deviation above the mean. Means are calculated over the period 2000–14. The GSS is calculated as

$$\mathrm{GSS} = \sum_{i=1}^{3}\sum_{j=1}^{3} p_{ij}s_{ij},$$

where p_{ij} is a contingency table with the relative frequency of forecast category i and observed category j, weighted by the cosine of the latitude at each grid cell:

$$p_{ij} = \frac{\sum_{g=1}^{N}\cos(l_g)_{ij}}{\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{g=1}^{N}\cos(l_g)_{ij}},$$

where l is latitude, g is a grid cell, and N is the number of grid cells. The scoring matrix s_{ij} contains the reward/penalty for every forecast–observation outcome, given by

$$s_{ij} = \begin{Bmatrix} \frac{1}{2}\left(\sum_{r=1}^{i-1}a_r^{-1} + \sum_{r=i}^{2}a_r\right), & i=j \\ \frac{1}{2}\left[\sum_{r=1}^{i-1}a_r^{-1} - (j-i) + \sum_{r=j}^{2}a_r\right], & 1\le i<j\le 3 \\ \frac{1}{2}\left[\sum_{r=1}^{j-1}a_r^{-1} - (i-j) + \sum_{r=i}^{2}a_r\right], & 1\le j<i\le 3 \end{Bmatrix},$$

where

$$a_i = \frac{1-\sum_{r=1}^{i}p_r}{\sum_{r=1}^{i}p_r}$$

and

$$p_r = \sum_{i=1}^{3}p_{ir}.$$

The value of GSS is 1 for a perfect forecast and 0 for random and constant forecasts. The GSS can be calculated using data from all grids cells and can be done separately for different surface types (e.g., land and ocean).

A Monte Carlo bootstrap approach is used to test the statistical significance of the calculated GSS. The GSS calculation is repeated 1000 times with shuffled time series of monthly anomalies of the reanalysis products and fixed time series of MODIS observations. The significance level of the GSS is determined as the percentage of the 1000 values smaller than the calculated GSS.

Correlations between monthly mean cloud amount anomalies from two different datasets are also calculated. Statistical significance is determined by the two-tailed

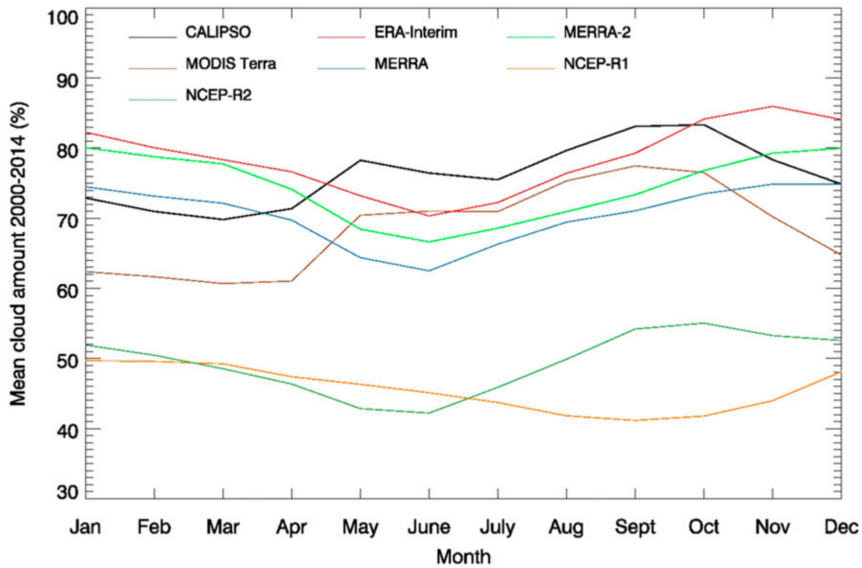

FIG. 1. Mean cloud amount over the Arctic (60°–90°N) from *Terra* MODIS, ERA-Interim, MERRA, MERRA-2, NCEP R1, and NCEP R2 during 2000–14 and from *CALIPSO* during 2006–14.

Student's *t* test. The correlation coefficients measure the linear association between two monthly mean cloud anomalies. Having both the correlation analysis and the GSS may provide a better assessment than either one alone.

3. Results

a. *Mean and standard deviation of cloud amount*

A comparison of monthly mean cloud amount from MODIS and reanalysis datasets shows that there is a broad range in the monthly mean cloud amount in the reanalysis products (Fig. 1). The annual mean cloud amounts of ERA-Interim, MERRA, MERRA-2, NCEP R2, and NCEP R1 from 2000 to 2014 are 78.6%, 70.5%, 74.6%, 49.4%, and 45.7%, respectively, compared to 68.5% for *Terra* MODIS from 2000 to 2014, and 76.2% for *CALIPSO* from 2006 to 2014. The annual mean cloud amount over ocean is higher than that over land for ERA-Interim, MERRA, and MERRA-2 (not shown). The difference between the ocean and land cloud amounts is large for ERA-Interim and MERRA-2, over 11% in the annual mean. For MERRA, the ocean–land difference is approximately 6% for the nonsummer months but somewhat larger from June to August. The monthly mean cloud amount of NCEP R2 is only slightly higher—around 2%—over ocean than land in every month. The only reanalysis with higher monthly mean cloud amount over land than over ocean is NCEP R1, which is about 14% higher throughout the year. For *Terra* MODIS the difference between ocean and land is larger from May to September, over 10%, than for other months.

In addition to overall differences in the monthly mean cloud amounts, the annual cycles of cloud amount in the reanalysis datasets do not resemble that from MODIS. The MODIS annual cycle has cloud amounts over 70% from May to November and cloud amounts less than 70% from December to April. This is consistent with annual cycles from surface observations (Hahn et al. 1995), other passive satellite datasets (Wang and Key 2005), and active lidar–radar satellite datasets (Liu et al. 2012b). In fact, the *CALIPSO* annual cycle from 2006 to 2014 is uniformly larger than the MODIS by 7.7%, and the correlation between these two cloud amount annual cycles is 0.96.

Except for NCEP R1, the reanalysis products generally have the highest cloud amounts from November to January, a decrease to the minimum in June, and then an increase (Fig. 1). Four of the five reanalysis products exhibit similar annual cycles, with the correlation coefficients of MERRA, MERRA-2, and NCEP R2 with ERA-Interim of 0.95, 0.91, and 0.89, respectively. In contrast, the correlation coefficients of monthly mean cloud amount annual cycles from ERA-Interim, MERRA, MERRA-2, and NCEP R2 with *Terra* MODIS and *CALIPSO* are −0.10 and 0.11, −0.27 and −0.10, −0.48 and −0.29, and 0.18 and 0.34, respectively. The annual cycle from NCEP R1 is different in shape and magnitude from the other annual cycles.

The standard deviations of the monthly mean cloud amount over the period 2000–14 (2006–14 for *CALIPSO*)

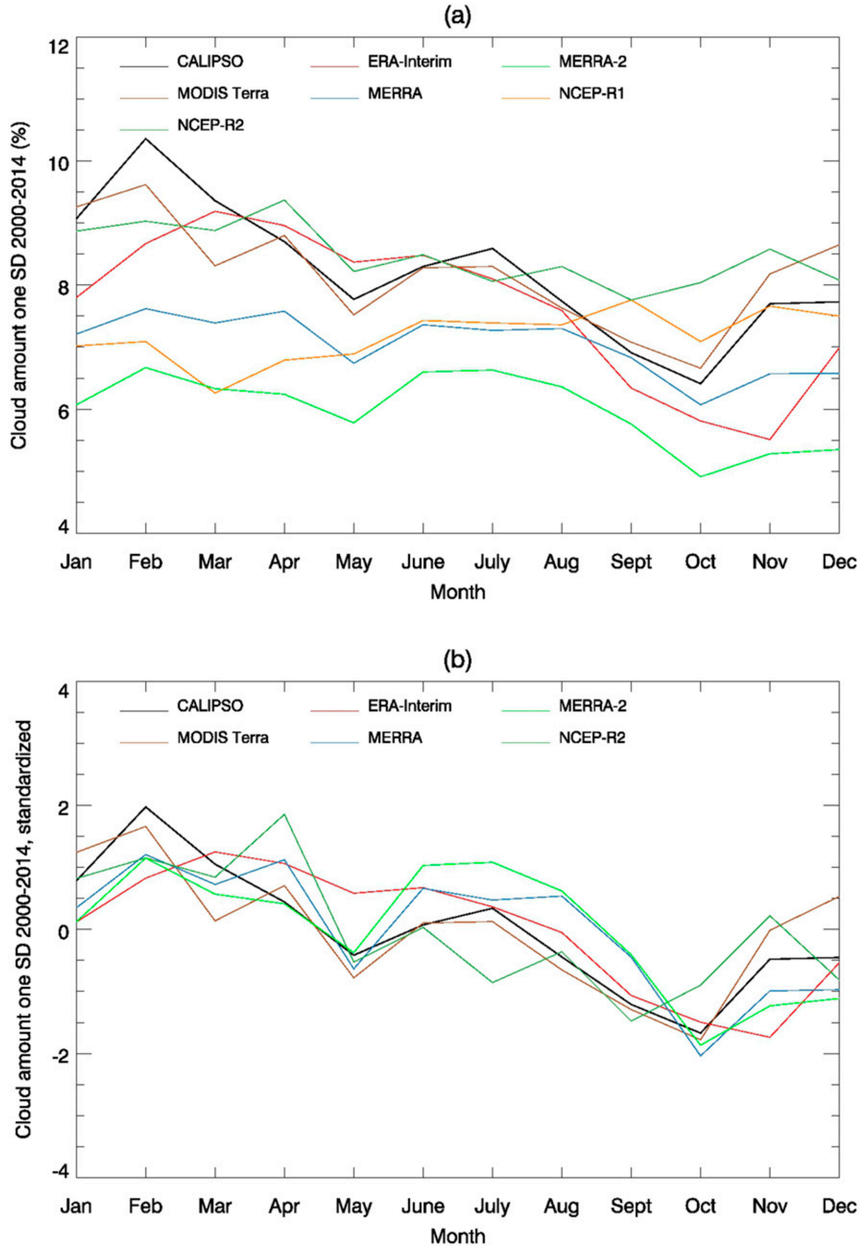

FIG. 2. (a) One standard deviation (SD) of cloud amount over the Arctic (60°–90°N) from *Terra* MODIS, ERA-Interim, MERRA, MERRA-2, NCEP R1, and NCEP R2 during 2000–14 and from *CALIPSO* during 2006–14; (b) as in (a), but standardized and without NCEP R1.

(Fig. 2a) exhibit the following characteristics: First, the standard deviations from MODIS, *CALIPSO*, and most reanalysis datasets except NCEP R1 have maximum values from January to March. They then decrease to a minimum between September and November, with annual mean standard deviations of 7.65, 7.04, 6.00, 7.19, 8.47, 8.19, and 8.22 for ERA-Interim, MERRA, MERRA-2, NCEP R1, NCEP R2, MODIS, and *CALIPSO*. The correlations between the monthly mean standard deviations of ERA-Interim, MERRA, MERRA-2, NCEP R1, and NCEP R2 with MODIS and *CALIPSO* are 0.54 and 0.76, 0.66 and 0.82, 0.53 and 0.74, −0.22 and −0.49, and 0.72 and 0.75, respectively, which indicates that their variabilities are similar over the annual cycle. This is illustrated by the standardized cloud amount standard deviations shown in Fig. 2b. In the figure, the monthly standard deviations are standardized by subtracting their annual mean and dividing by their annual standard deviation. Second, the differences

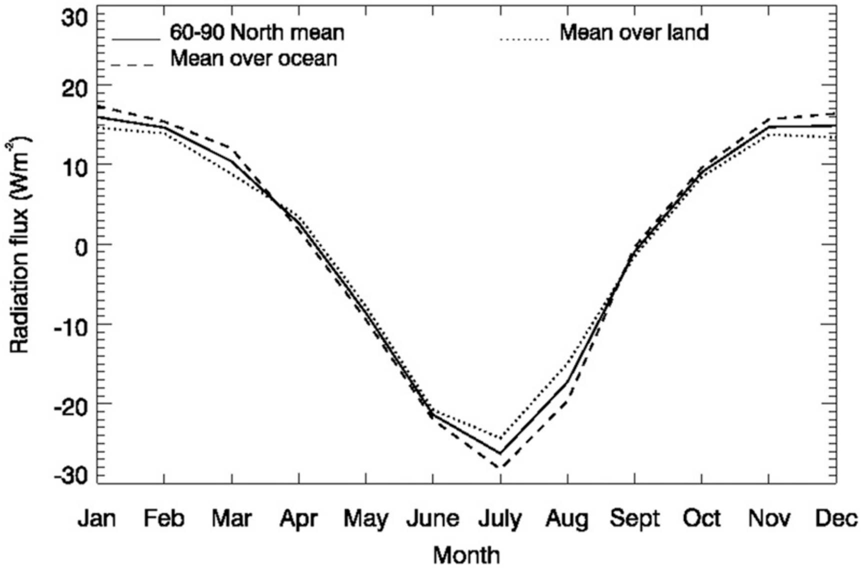

FIG. 3. The difference in cloud radiative forcing from MERRA calculated for cloud amount one standard deviation above and below the monthly mean, averaged over the entire Arctic, the Arctic Ocean, and Arctic land areas.

between the standard deviations of the reanalysis datasets and *Terra* MODIS are relatively small except for MERRA-2. MERRA-2 shows larger differences in standard deviations compared to *Terra* MODIS. Mean differences are −0.54, −1.15, −2.19, −1.0, and 0.28 of ERA-Interim, MERRA, MERRA-2, NCEP R1, and NCEP R2 compared to MODIS, with root-mean-square errors (RMSEs) of 1.20, 1.36, 2.40, 1.47, and 0.68. Third, the standard deviations over land are larger than those over ocean in most months for ERA-Interim, MERRA, MERRA-2, MODIS, and *CALIPSO* (not shown).

b. *Cloud amount anomaly*

To illustrate the importance of accurate cloud amount anomalies, anomalies in the cloud radiative effect at the surface, or "forcing," are examined. Cloud radiative forcing is calculated as the difference between the all-sky and clear-sky net radiative fluxes at the surface. The surface cloud forcing was calculated for every month in every year at each grid point. The surface cloud forcing was then averaged over the entire Arctic, over Arctic land, and over the Arctic Ocean, but only for monthly cloud amount anomalies larger than one standard deviation above the mean, and again for cloud amount anomalies less than one standard deviation below the mean. The differences between cloud forcing based on these anomalously high and low cloud amounts calculated from MERRA are shown in Fig. 3. The differences are positive in winter, with a maximum value for the entire Arctic of $16\,\mathrm{W\,m^{-2}}$. The differences decrease after January and become negative in April–May, reaching a minimum of $-26\,\mathrm{W\,m^{-2}}$ in August. The differences then start to increase and become positive in September–October. The differences are larger over ocean in the daytime (sunlit) portion of the year and more negative over ocean during nighttime (dark) than those over land. The differences between cloud forcing based on anomalously high and low cloud amounts calculated from MERRA-2 and the ERA-Interim are shown in Fig. 4. The results from MERRA-2 show similar characteristics as MERRA, except that the differences are closer over ocean and over land. The largest positive and negative cloud forcings for ERA-Interim are 14 and $-23\,\mathrm{W\,m^{-2}}$ in December and July, respectively; the differences are larger over land in the daylight portion of the year and more negative over land during the nighttime than those over ocean.

Why is this important? One example is the influence of cloud forcing on sea ice growth. Theoretically, a $1\,\mathrm{W\,m^{-2}}$ radiative flux anomaly would produce a 0.85-cm sea ice growth anomaly in one month (Liu and Key 2014). Therefore, a $-16\,\mathrm{W\,m^{-2}}$ radiative flux anomaly would produce a +14-cm sea ice growth anomaly in a month, and a $26\,\mathrm{W\,m^{-2}}$ radiative flux anomaly would produce a −22-cm sea ice growth anomaly (i.e., less ice growth). Accurate estimates of the cloud amount and cloud forcing anomalies are critical for studying feedbacks and interactions in Arctic climate.

Previous studies have shown that the reanalysis products have some skill in estimating monthly mean Arctic cloud

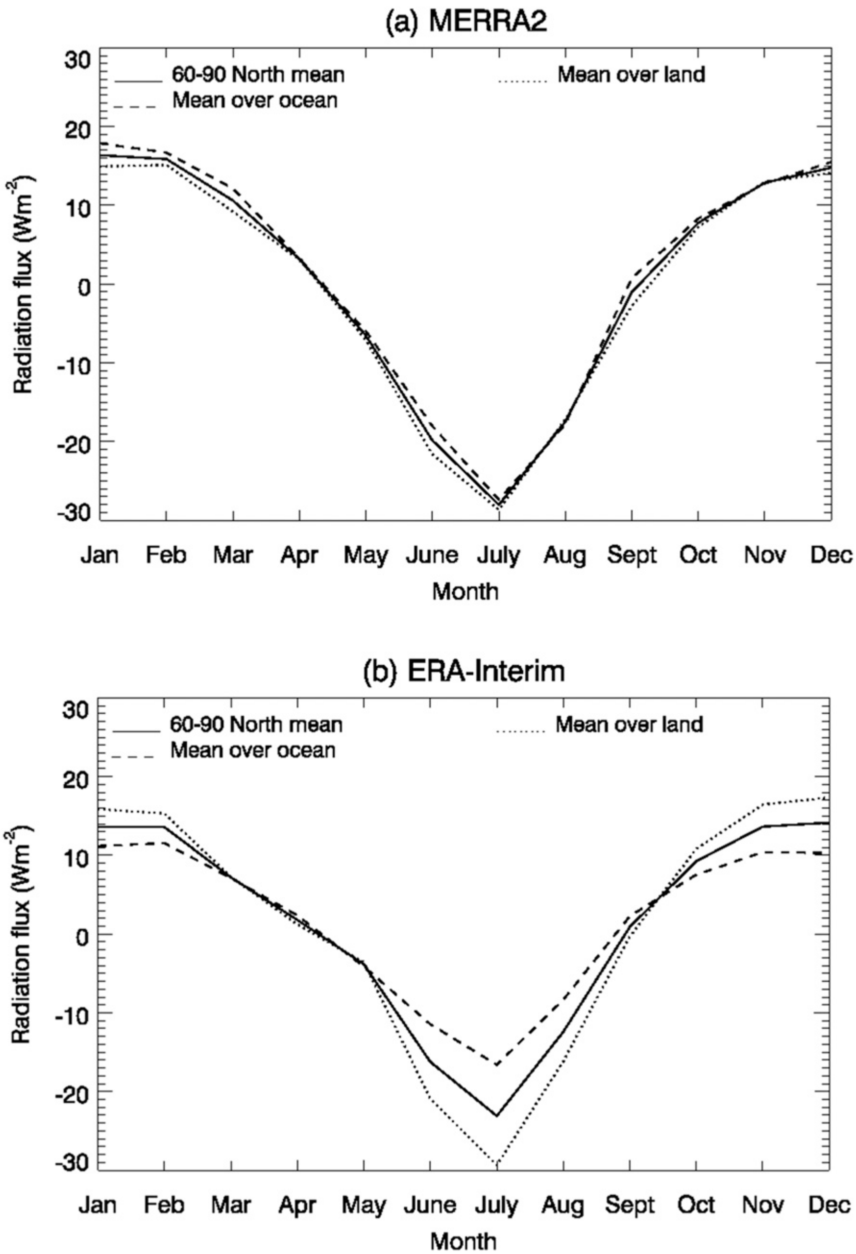

FIG. 4. As in Fig. 3, but for (a) MERRA-2 and (b) ERA-Interim.

amount anomalies (Liu and Key 2014; Letterly et al. 2016). Figure 5 shows the monthly mean cloud amount anomalies in January 2013 and June 2013 from *CALIPSO*, MODIS, ERA-Interim, MERRA, MERRA-2, NCEP R1, and NCEP R2. In January 2013, both *CALIPSO* and *Terra* MODIS show negative cloud amount anomalies over most of the Beaufort Sea, Chukchi Sea, Canada Basin, the Laptev Sea, northern Greenland, and part of north central Russia, with maximum negative anomalies of −20% or lower. Of the five reanalysis products, MERRA and NCEP R2 show similar spatial distributions and magnitudes. The cloud amount anomaly from ERA-Interim is negative over part of the Arctic Ocean, and MERRA-2 resembles ERA-Interim without the warming pattern over the Canada Basin. NCEP R1 shows very little similarity to other reanalyses and to *Terra* MODIS. In June 2013, both *CALIPSO* and *Terra* MODIS show positive cloud amount anomalies over the Canada Basin, Beaufort Sea, and Chukchi Sea and negative cloud amount anomalies over northern Europe. NCEP R2 is most similar to MODIS and *CALIPSO*, with both positive and negative anomalies over the ocean and land. ERA-Interim, MERRA, and

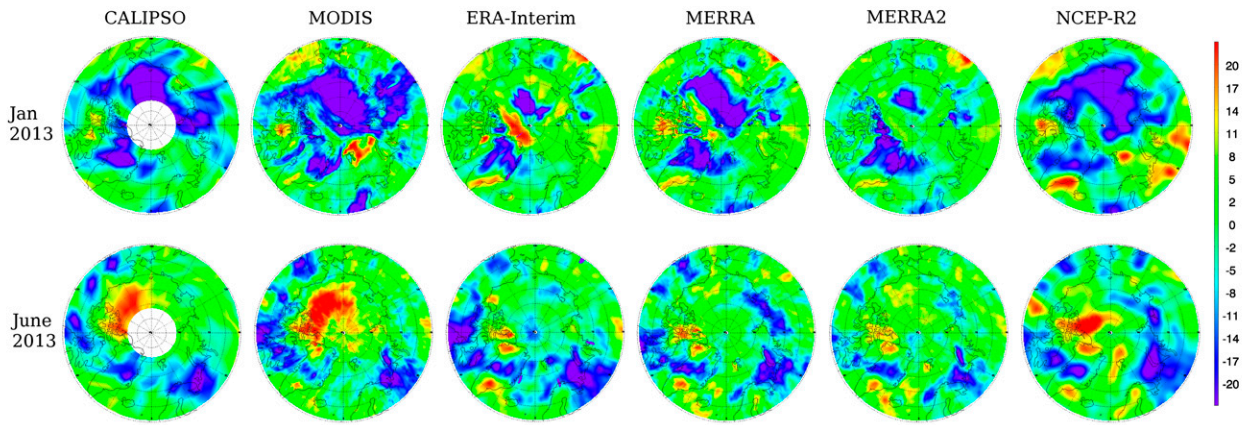

FIG. 5. Cloud amount anomalies (%) in (top) January 2013 and (bottom) June 2013 from (left)–(right) *CALIPSO* relative to the 2006–14 mean, and *Terra* MODIS, ERA-Interim, MERRA, MERRA-2, and NCEP R2 relative to the 2000–14 mean.

MERRA-2 only show negative anomalies over northern Europe. As in January, NCEP R1 bears little resemblance to MODIS and to the other reanalyses.

To assess the overall performance of the reanalysis products, the GSS and correlation coefficients are calculated for each reanalysis as the ''forecast'' with MODIS (2000–14) as the independent observation for the Arctic, Arctic land, and Arctic Ocean. The GSS and correlation coefficients are also calculated between MERRA and ERA-Interim, with MERRA as the forecast and ERA-Interim as the independent observation, and between *Aqua* MODIS and *Terra* MODIS, with *Aqua* MODIS as the forecast and *Terra* MODIS as the observation. All the GSS values and correlation coefficients are given in Table 1.

1) GSS WITH MODIS 2000–14 AND *CALIPSO* 2006–14

Figure 6 shows the GSS for each month in the entire Arctic (Fig. 6a), over the Arctic Ocean (Fig. 6b), and over Arctic land (Fig. 6c). The figure reveals the following features: First, the GSSs for *Aqua* MODIS against *Terra* MODIS are between 0.65 and 0.8 throughout the year, with average scores of 0.71, 0.68, and 0.74 in the Arctic, over the Arctic Ocean, and over Arctic land, respectively. A perfect score would be expected for these two satellite products if the two satellites viewed the same clouds and if the cloud detection schemes were the same. The spectral characteristics and cloud detection capabilities of *Aqua* and *Terra* MODIS are nearly identical, so the departures from perfect scores can be attributed to internal variability in cloud cover on the synoptic scale and diurnal variations, as *Terra* and *Aqua* have different equator crossing times [see Shupe et al. (2011) for a discussion of Arctic cloud diurnal cycles].

Second, all the scores are statistically significant at the 95% confidence level except for NCEP R1, as determined using the Monte Carlo test. This indicates that the reanalysis products have skill in estimating cloud amount anomalies relative to MODIS. The annual mean scores are close for all reanalysis datasets except NCEP R1, with average scores of 0.36, 0.31, 0.32, and 0.32 for ERA-Interim, MERRA, MERRA-2, and NCEP R2. There is no statistically significant difference between

TABLE 1. GSSs and correlation coefficients of reanalysis for the Arctic overall, over land, and over ocean.

	GSS		Correlation coefficient	
	Terra MODIS	*CALIPSO*	*Terra* MODIS	*CALIPSO*
ERA-Interim	0.36, 0.45, 0.26	0.22, 0.29, 0.16	0.50, 0.70, 0.35	0.51, 0.61, 0.39
MERRA	0.31, 0.40, 0.22	0.24, 0.29, 0.18	0.43. 0.65, 0.30	0.54, 0.65, 0.43
MERRA-2	0.32, 0.40, 0.23	0.23, 0.29, 0.16	0.44, 0.64, 0.31	0.53, 0.66, 0.40
NCEP R1	0.0, 0.0, 0.0	0.0, 0.0, 0.0	0.0, 0.0, 0.0	−0.02, 0.01, −0.04
NCEP R2	0.32, 0.37, 0.26	0.23, 0.25, 0.19	0.50, 0.62, 0.44	0.52, 0.58, 0.47
ERA-Interim/MERRA	0.49, 0.56, 0.42		0.56, 0.75, 0.44	
Terra MODIS		0.33, 0.35, 0.30		0.49, 0.70, 0.35
Aqua MODIS	0.71, 0.74, 0.68	0.35, 0.36, 0.32	0.92, 0.94, 0.90	0.72, 0.75, 0.69

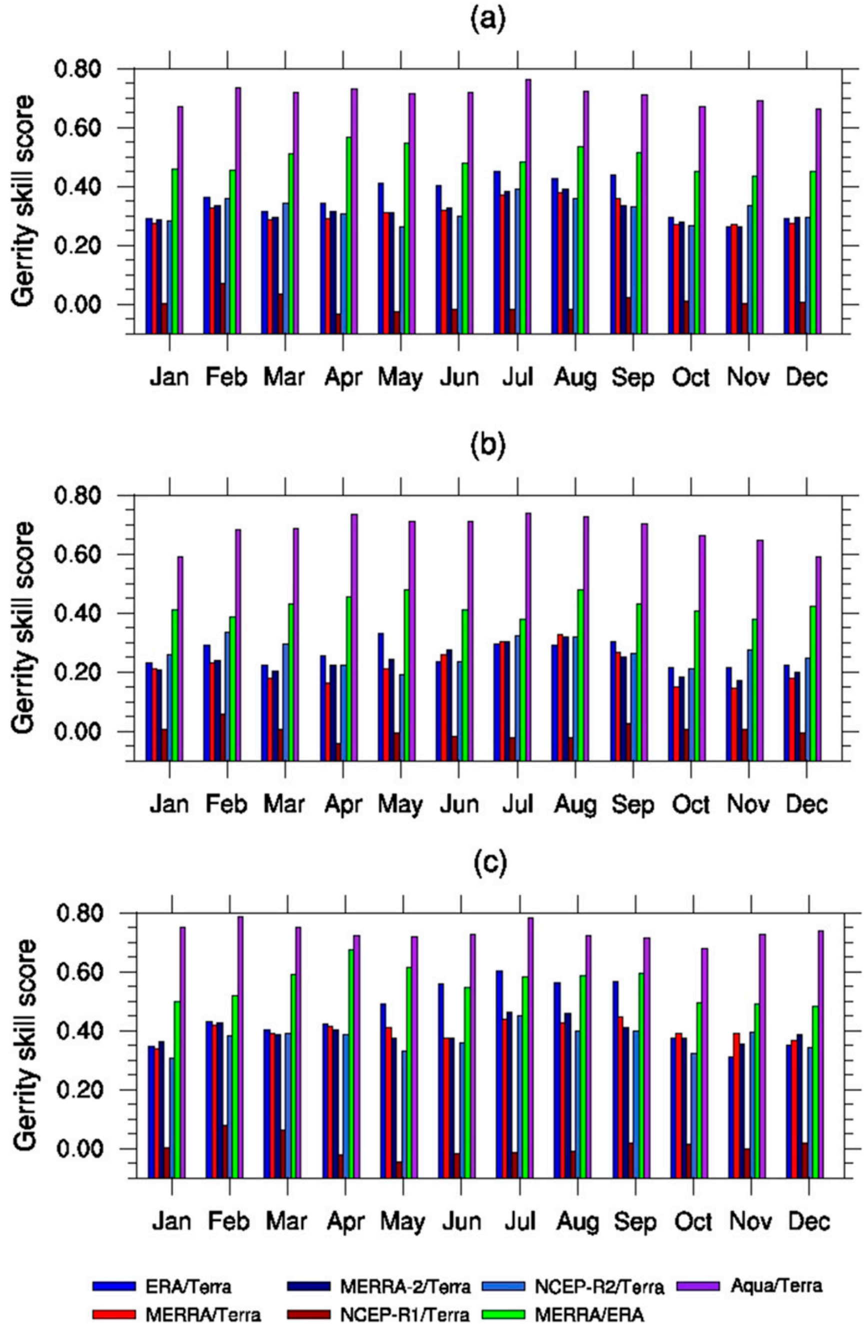

FIG. 6. GSSs of ERA-Interim, MERRA, MERRA-2, NCEP R1, and NCEP R2 cloud amount anomalies assessed by *Terra* MODIS cloud amount anomalies for the period 2000–14 over (a) the Arctic, (b) the Arctic Ocean, and (c) Arctic land.

these annual mean scores at the 99% confidence level. At the 95% confidence level, ERA-Interim has more skill than MERRA; the NCEP R2 has more skill than MERRA over ocean; and ERA-Interim and MERRA have more skill than NCEP R2 over land. These scores are lower than a perfect score, which suggests that future improvements are needed. All reanalyses show the best performance in July, August, and September and the worst performance in November, December, and January. It should be emphasized that ERA-Interim shows somewhat better performance with respect to MODIS from May to September.

Third, the performance of reanalysis products is better over land than over ocean. The average scores are 0.45,

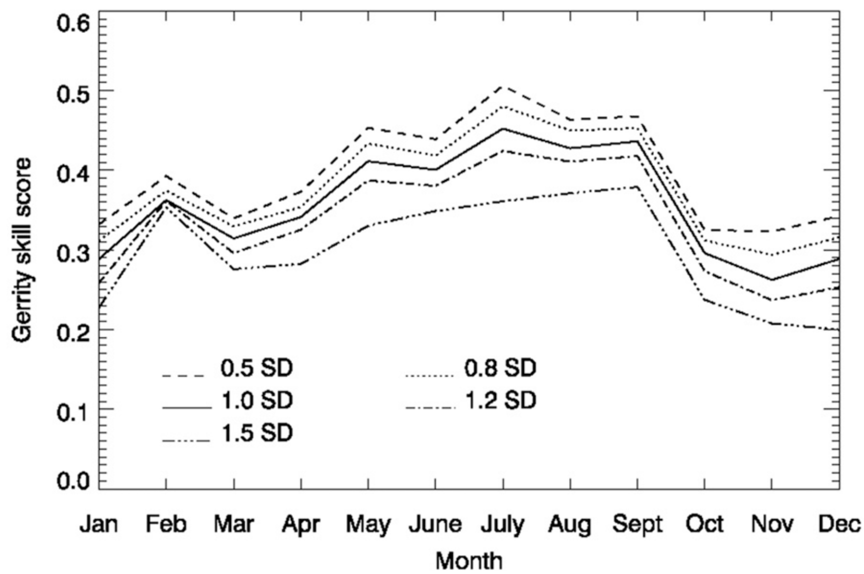

FIG. 7. Dependence of ERA-Interim cloud amount anomaly GSSs on the standard deviation threshold. GSS is relative to *Terra* MODIS for the period 2000–14.

0.40, 0.40, and 0.37 over land and 0.26, 0.22, 0.23, and 0.26 over ocean for ERA-Interim, MERRA, MERRA-2, and NCEP R2. All are statistically significant at the 99% confidence level.

Fourth, the scores of MERRA against ERA-Interim are 0.49, 0.56, and 0.42 for the Arctic overall, Arctic land, and Arctic Ocean. The score over land is significantly better than over ocean at the 99% confidence level, and the scores are significantly better than those against MODIS.

Fourth, the score depends on how an anomaly is defined. The results shown are for the normal defined as between one and minus one standard deviation of the monthly mean cloud amounts over the period 2000–14. When the standard deviation threshold used to define the normal is varied over the range 0.5–1.5, the scores decrease for ERA-Interim with increasing threshold (Fig. 7), as well as for MERRA, MERRA-2, and NCEP R2. This implies that the capability of reanalysis products to depict cloud amount anomalies weakens for more extreme anomalies. It should be emphasized that the significance level of GSS is higher than 95% even when the standard deviation threshold is set to 1.5.

GSS is also calculated for each reanalysis product as the "forecast" with *CALIPSO* (2006–14) as the independent observation to assess the reanalysis performance. The results are shown in Fig. 8. The GSSs for *Aqua* (*Terra*) MODIS against *CALIPSO* have average scores of 0.35 (0.33), 0.36 (0.35), and 0.32 (0.30) in the Arctic, over the Arctic land, and over the Arctic Ocean, respectively, with no significant differences at the 95% confidence level. The GSSs are higher from May to September than in other months, which implies better daytime MODIS cloud detection. The annual mean (range) GSSs of ERA-Interim, MERRA, MERRA-2, NCEP R1, and NCEP R2 against *CALIPSO* are 0.22 (0.29, 0.16), 0.24 (0.29, 0.18), 0.23 (0.29, 0.16), 0.00 (0.00, 0.00), and 0.23 (0.25, 0.19) in the Arctic, over Arctic land, and over the Arctic Ocean. All scores are statistically significant at the 95% level except for NCEP R1, as determined using the Monte Carlo test. There is no statistically significant difference between these values at the 95% level. As with the scores against MODIS, all reanalyses also demonstrate the best performance in July, August, and September and the worst performance in November, December, and January against *CALIPSO*. The performance of all reanalysis datasets is significantly better over land than over ocean at the 95% level. All features are consistent with those seen with MODIS.

The GSSs of reanalyses against *CALIPSO* are significantly lower than those against MODIS. These two datasets have different temporal coverage and spatial resolutions, where *CALIPSO* covers 2006–14 with 10° longitude by 3° latitude resolution, and MODIS covers 2000–14 with 1° longitude and 1° latitude resolution. GSSs of the reanalyses against MODIS for the same time period and spatial resolution as *CALIPSO* were also calculated. They are not significantly different than those for the original MODIS analysis. This suggests that the GSSs are sensitive to the validation datasets used but may not be very sensitive to the spatial resolution or the time period of the products.

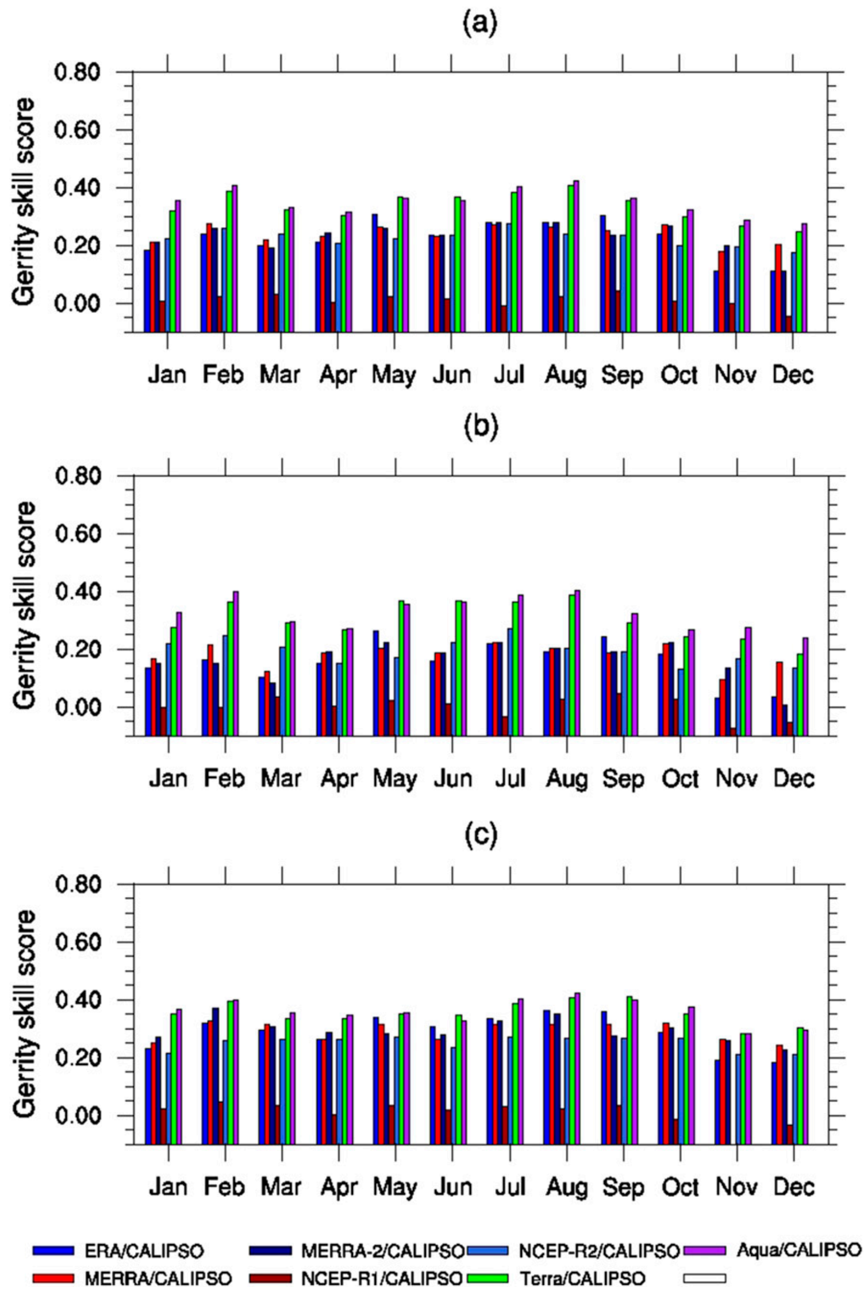

FIG. 8. GSSs of ERA-Interim, MERRA, MERRA-2, NCEP R1, and NCEP R2 cloud amount anomalies assessed by *CALIPSO* cloud amount anomalies for the period 2006–14 over (a) the Arctic, (b) the Arctic Ocean, and (c) Arctic land.

2) Correlation analysis with MODIS 2000–14 and *CALIPSO* 2006–14

The correlation analysis is presented in the same way as that for GSSs. The conclusions are also similar. Figure 9 shows the correlation coefficients for each month over the entire Arctic (Fig. 9a), over the Arctic Ocean (Fig. 9b), and over Arctic land (Fig. 9c). The correlation coefficients for *Aqua* MODIS against *Terra* MODIS are high throughout the year, with average scores of 0.92, 0.94, and 0.90 in the Arctic, over Arctic land, and over the Arctic Ocean, respectively. The annual mean correlation coefficients are similar for all reanalysis datasets except NCEP R1, with average values of 0.50, 0.43, 0.44, and 0.50 for ERA-Interim, MERRA, MERRA-2, and NCEP R2. All the correlations are statistically significant at the 95%

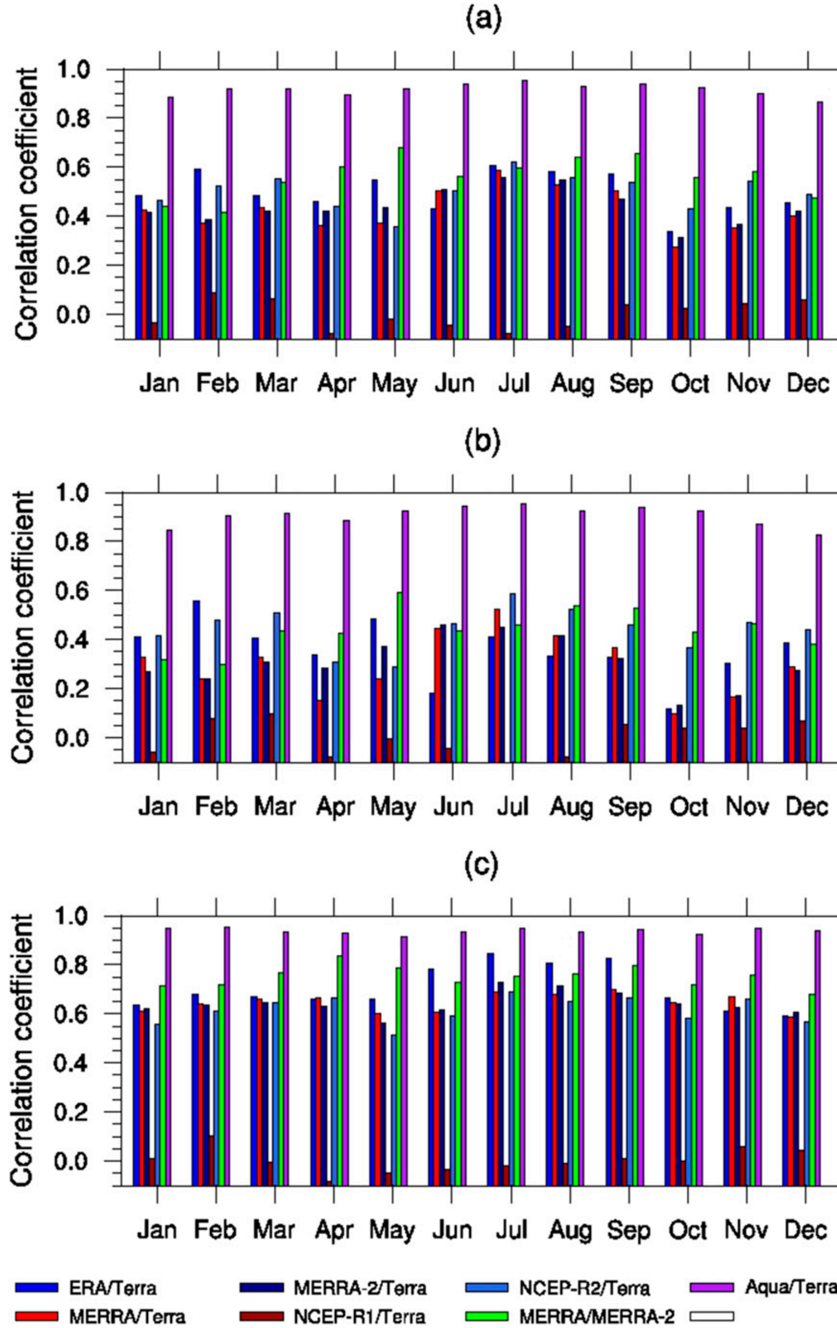

FIG. 9. Correlation coefficients of ERA-Interim, MERRA, MERRA-2, NCEP R1, and NCEP R2 cloud amount anomalies assessed against *Terra* MODIS cloud amount anomalies for the period 2000–14 over (a) the Arctic, (b) the Arctic Ocean, (c) and Arctic land.

confidence level except for NCEP R1. There is no statistically significant difference between these annual means at the 99% level. For the 95% confidence level, the ERA-Interim and NCEP R2 are better (higher, significant correlations) than MERRA and MERRA-2; NCEP R2 is better than other reanalyses over ocean; and ERA-Interim is better than other reanalyses over land. All reanalyses show overall better performance in July, August, and September than that in December, and January. The performance of reanalysis products is better over land than over ocean. The average scores are 0.70, 0.65, 0.64, and 0.62 over land and 0.35, 0.30,

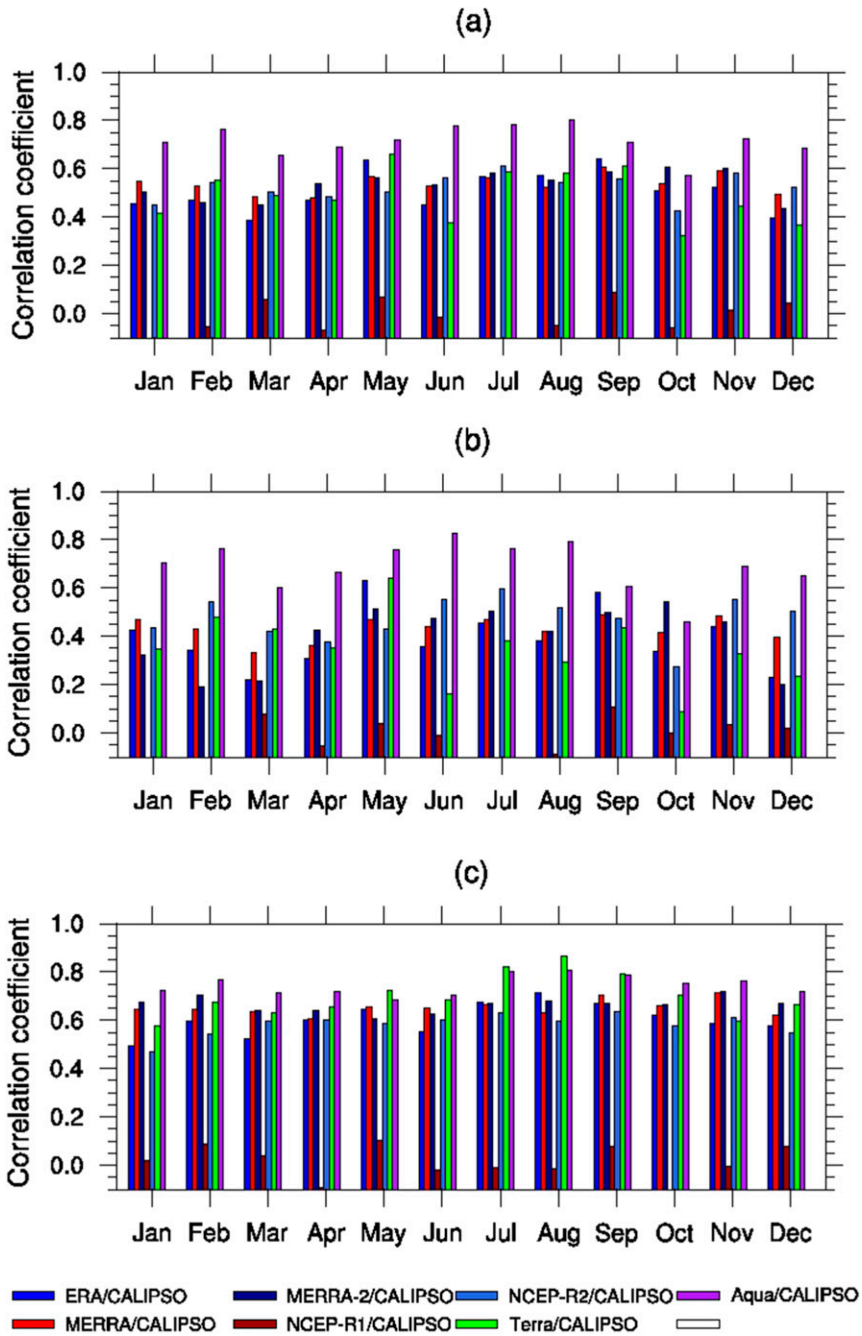

FIG. 10. Correlation coefficients of ERA-Interim, MERRA, MERRA-2, NCEP R1, and NCEP R2 cloud amount anomalies assessed against *CALIPSO* cloud amount anomalies for the period 2006–14 over (a) the Arctic, (b) the Arctic Ocean, and (c) Arctic land.

0.31, and 0.44 over ocean for ERA-Interim, MERRA, MERRA-2, and NCEP R2. All are statistically significant at the 99% level. The agreement of MERRA and ERA-Interim (0.56) is significantly better than that between each reanalysis and MODIS, and it is significantly better over land than over ocean, with average scores of 0.75 and 0.44. The findings of the correlation analysis are consistent with those of the GSS analysis.

Correlation analysis is also carried out for each reanalysis dataset with *CALIPSO* (Fig. 10). The correlation coefficients for *Aqua* (*Terra*) MODIS against *CALIPSO* have average scores of 0.72 (0.49), 0.75 (0.70), and

0.69 (0.35) in the Arctic, over Arctic land, and over the Arctic Ocean, respectively, with significantly higher values for *Aqua* MODIS over the Arctic and the Arctic Ocean. The best performance is from July to September, possibly as a result of more daylight in the *Aqua* data and therefore better cloud detection. The annual mean (range) correlation coefficients of ERA-Interim, MERRA, MERRA-2, NCEP R1, and NCEP R2 against *CALIPSO* are 0.51 (0.61, 0.39), 0.54 (0.65, 0.43), 0.53 (0.66, 0.40), −0.02 (0.01, −0.04), and 0.52 (0.58, 0.47) in the Arctic (over Arctic land and over the Arctic Ocean). All coefficients are statistically significant at the 95% level except for NCEP R1. There is no statistically significant difference between these annual means at the 95% level over the entire Arctic. Over land, the coefficients of MERRA and MERRA-2 are significantly higher than those of ERA-Interim and NCEP R2. Over ocean the coefficients are higher for NCEP R2 than for ERA-Interim. The seasonal differences in the coefficients are not apparent. The performance of all reanalysis datasets is significantly better over land than over ocean. All the features are consistent with those for the MODIS analysis.

In contrast to the GSS analysis, the correlation coefficients of the reanalyses against *CALIPSO* are slightly higher than those against MODIS for MERRA and MERRA-2, with no significant differences for ERA-Interim or NCEP R2. Correlation coefficients for the reanalyses against MODIS for the *CALIPSO* time period and spatial resolution were also calculated, with no significant differences. As with the GSS analysis, this suggests that the correlation analysis may not be sensitive to the spatial resolution and time period of the validation datasets. The sensitivity to different validation datasets is much lower than that of the GSS.

4. Discussion and conclusions

There are significant differences in the monthly mean Arctic cloud amount in reanalysis products, and none of the reanalysis products examined here resembles the annual cycle of cloud amount from MODIS or *CALIPSO* satellite products. Despite differences in the mean cloud amount, the reanalysis datasets do exhibit some capability for depicting the monthly mean cloud amount anomalies, as demonstrated for the period 2000–14. The GSS and correlation coefficients of all reanalysis products included in this study except NCEP R1 are statistically significant, which demonstrates that there is skill in estimating the cloud amount anomalies with reanalysis products. While cloud schemes in the reanalysis models have been periodically updated and improved (Kanamitsu et al. 2002; Dee et al. 2011; Rienecker et al. 2011), the results of this study indicate that there is no significant difference between the performance of the prognostic and diagnostic schemes in simulating the monthly mean cloud cover anomalies, and further improvement may be needed. The GSS and correlation coefficient values provide a baseline for the current performance and can be used to evaluate future improvements.

Cloud detection in the polar regions using passive satellite instruments, such as MODIS, is difficult given the similarity in temperature and reflectance of the ubiquitous low-level clouds and the surface. This is particularly true in the wintertime, when shortwave (visible through solar infrared) data are not available and when stronger temperature inversions exist (Zygmuntowska et al. 2012). In addition to MODIS, there are other options for reference cloud amount and anomaly climatologies, such as active spaceborne radar–lidar and surface observations. *CALIPSO* cloud amount data from 2006 to 2014 are therefore employed in this study as a complement to the MODIS cloud data. However, while active sensors may be more effective in cloud detection than passive sensors, there are challenges in their use, including difficulties in detecting the low-level clouds (Zygmuntowska et al. 2012; Liu et al. 2012b) and uncertainties due to limited spatial sampling (Liu 2015). Sampling error is a potential problem in the analysis of (Wyser et al. 2008), particularly for comparisons over a short time period, such as daily averages. For the evaluation of monthly means, the uncertainty introduced by sampling error is much smaller (Guan et al. 2013). Here, the *CALIPSO* cloud amount is averaged over a large area in this study to minimize the sampling error. Regarding in situ observations, their spatial coverage is very limited in the polar regions. It would be a difficult task to optimally combine all these observations to produce a high-quality cloud amount product.

The annual cycles of Arctic cloud cover from the reanalyses do not resemble those from MODIS and *CALIPSO*. This is consistent with results of other studies (Wyser et al. 2008; Wyser and Jones 2005). MERRA, MERRA-2, and ERA-Interim show higher cloud amount in the winter than MODIS and *CALIPSO*. It has been suggested that an optical thickness threshold should apply to modeled cloud amount so that clouds with thicknesses lower than the threshold would be removed (Wyser et al. 2008; Wyser and Jones 2005; Karlsson and Svensson 2011). Such an approach results in a better agreement between the modeled and observed annual cycle. This approach, however, will not work for NCEP R2 or NCEP R1, which have lower cloud amounts than MODIS throughout the year.

The reanalysis products perform better in depicting the monthly mean cloud amount anomalies over land than over ocean. This may be a result of the fact that the reanalysis products assimilate many available atmospheric and surface state variables directly (e.g., air temperature, pressure, and moisture), and such measurements are dense over land but sparse over the ocean. The land–ocean difference is apparent in the GSS.

Satellite simulation software has been developed to evaluate the cloud parameterization of numerical weather prediction and climate models using satellite retrievals (Bodas-Salcedo et al. 2011). Cloud outputs generated with the Cloud Feedback Model Intercomparison Project (CFMIP) Observation Simulator Package (COSP) simulators have been used to understand and quantify climate model cloud biases (Kay et al. 2012). Such information would be useful, but it is not currently available (Naud et al. 2014).

Walsh et al. (2005) advocates that both cloud amount and radiative fluxes should be assessed in evaluating modeled clouds in the Arctic. The reason is that errors in cloud amount and cloud microphysical properties can offset each other and result in similar radiative fluxes, and it is the radiative effect of clouds that controls the energy budget (de Boer et al. 2014; Cox et al. 2014). In this study, only cloud amount was evaluated against satellite data, as satellite-derived radiative flux estimates have large uncertainties because they depend on cloud amount, other cloud properties, surface temperature and albedo, and atmospheric temperature and water vapor. Errors in each of these quantities propagate through the calculation of radiative fluxes and can collectively be substantial (Key et al. 1997). When a reliable satellite-derived radiative flux product is available, radiative fluxes in the reanalyses should be evaluated using the same approach employed here for cloud amount.

Acknowledgments. This work was supported by the NOAA/National Climatic Data Center (NA10NES4400013) and the Joint Polar Satellite System (JPSS) Program Office. MODIS data were obtained from the Atmosphere Archive and Distribution System of the NASA Goddard Space Flight Center (https://ladsweb.nascom.nasa.gov/). NCEP reanalysis data were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado (http://www.esrl.noaa.gov/psd/). The ERA-Interim data were provided by the European Centre for Medium-Range Weather Forecasts (ECMWF; http://apps.ecmwf.int/datasets/). MERRA data were provided by the Goddard Earth Sciences Data and Information Services Center (http://disc.sci.gsfc.nasa.gov/mdisc/). MERRA-2 data were provided by the Goddard Earth Sciences Data and Information Services Center (http://disc.sci.gsfc.nasa.gov/datacollection/M2TMNXRAD_V5.12.4.shtml). The *CALIPSO* data were obtained from the NASA Langley Research Center Atmospheric Science Data Center. The views, opinions, and findings contained in this report are those of the author(s) and should not be construed as an official National Oceanic and Atmospheric Administration or U.S. government position, policy, or decision.

REFERENCES

Aas, K. S., and Coauthors, 2015: A comparison between simulated and observed surface energy balance at the Svalbard Archipelago. *J. Appl. Meteor. Climatol.*, **54**, 1102–1119, doi:10.1175/JAMC-D-14-0080.1.

Ackerman, S., K. Strabala, W. Menzel, R. Frey, C. Moeller, and L. Gumley, 1998: Discriminating clear sky from clouds with MODIS. *J. Geophys. Res.*, **103**, 32 141–32 157, doi:10.1029/1998JD200032.

Bacmeister, J. T., M. J. Suarez, and F. R. Robertson, 2006: Rain reevaporation, boundary layer–convection interactions, and Pacific rainfall patterns in an AGCM. *J. Atmos. Sci.*, **63**, 3383–3403, doi:10.1175/JAS3791.1.

Beesley, J., and R. Moritz, 1999: Toward an explanation of the annual cycle of cloudiness over the Arctic Ocean. *J. Climate*, **12**, 395–415, doi:10.1175/1520-0442(1999)012<0395:TAEOTA>2.0.CO;2.

Birch, C. E., I. M. Brooks, M. Tjernstrom, S. F. Milton, P. Earnshaw, S. Soderberg, and P. O. G. Persson, 2009: The performance of a global and mesoscale model over the central Arctic Ocean during late summer. *J. Geophys. Res.*, **114**, D13104, doi:10.1029/2008JD010790.

Bodas-Salcedo, A., and Coauthors, 2011: COSP: Satellite simulation software for model assessment. *Bull. Amer. Meteor. Soc.*, **92**, 1023–1043, doi:10.1175/2011BAMS2856.1.

Boucher, O., and Coauthors, 2013: Clouds and aerosols. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 571–657, doi:10.1017/CBO9781107415324.016.

Bromwich, D. H., R. L. Fogt, K. I. Hodges, and J. E. Walsh, 2007: A tropospheric assessment of the ERA-40, NCEP, and JRA-25 global reanalyses in the polar regions. *J. Geophys. Res.*, **112**, D10111, doi:10.1029/2006JD007859.

Campana, K. A., Y.-T. Hou, K. E. Mitchell, S.-K. Yang, and R. Cullather, 1994: Improved diagnostic cloud parameterization in NMC's global model. Preprints, *10th Conf. on Numerical Weather Prediction*, Portland, OR, Amer. Meteor. Soc., 324–325.

Chaudhuri, A. H., R. M. Ponte, and A. T. Nguyen, 2014: A comparison of atmospheric reanalysis products for the Arctic Ocean and implications for uncertainties in air–sea fluxes. *J. Climate*, **27**, 5411–5421, doi:10.1175/JCLI-D-13-00424.1.

Chernokulsky, A., and I. 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.

Clark, J. V., and J. E. Walsh, 2010: Observed and reanalysis cloud fraction. *J. Geophys. Res.*, **115**, D23121, doi:10.1029/2009JD013235.

Cox, C. J., V. P. Walden, G. P. Compo, P. M. Rowe, M. D. Shupe, and K. Steffen, 2014: Downwelling longwave flux over Summit, Greenland, 2010–2012: Analysis of surface-based observations and evaluation of ERA-Interim using wavelets. *J. Geophys. Res. Atmos.*, **119**, 12 317–12 337, doi:10.1002/2014JD021975.

Curry, J. A., W. B. Rossow, D. Randall, and J. L. Schramm, 1996: Overview of Arctic cloud and radiation characteristics. *J. Climate*, **9**, 1731–1764, doi:10.1175/1520-0442(1996)009<1731:OOACAR>2.0.CO;2.

de Boer, G., W. Chapman, J. E. Kay, B. Medeiros, M. D. Shupe, S. Vavrus, and J. Walsh, 2012: A characterization of the present-day Arctic atmosphere in CCSM4. *J. Climate*, **25**, 2676–2695, doi:10.1175/JCLI-D-11-00228.1.

——, and Coauthors, 2014: Near-surface meteorology during the Arctic Summer Cloud Ocean Study (ASCOS): Evaluation of reanalyses and global climate models. *Atmos. Chem. Phys.*, **14**, 427–445, doi:10.5194/acp-14-427-2014.

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.

Du, P., E. Girard, A. K. Bertram, and M. D. Shupe, 2011: Modeling of the cloud and radiation processes observed during SHEBA. *Atmos. Res.*, **101**, 911–927, doi:10.1016/j.atmosres.2011.05.018.

Francis, J. A., D. M. White, J. J. Cassano, W. J. Gutowski Jr., L. D. Hinzman, M. M. Holland, M. A. Steele, and C. J. Vörösmarty, 2009: An Arctic hydrologic system in transition: Feedbacks and impacts on terrestrial, marine, and human life. *J. Geophys. Res.*, **114**, G04019, doi:10.1029/2008JG000902.

Frey, R., S. Ackerman, Y. Liu, K. Strabala, H. Zhang, J. Key, and X. Wang, 2008: Cloud detection with MODIS. Part I: Improvements in the MODIS cloud mask for collection 5. *J. Atmos. Oceanic Technol.*, **25**, 1057–1072, doi:10.1175/2008JTECHA1052.1.

Gerrity, J. P., 1992: A note on Gandin and Murphy's equitable skill score. *Mon. Wea. Rev.*, **120**, 2709–2712, doi:10.1175/1520-0493(1992)120<2709:ANOGAM>2.0.CO;2.

GMAO, 2015: MERRA-2 tavgM_2d_rad_Nx: 2d, Monthly mean, time-averaged, single-level, assimilation, radiation diagnostics, version 5.12.4. Goddard Space Flight Center Distributed Active Archive Center, accessed 1 December 2015, doi:10.5067/OU3HJDS973O0.

Grell, G. A., 1993: Prognostic evaluation of assumptions used by cumulus parameterizations. *Mon. Wea. Rev.*, **121**, 764–787, doi:10.1175/1520-0493(1993)121<0764:PEOAUB>2.0.CO;2.

Guan, B., D. E. Waliser, J.-L. F. Li, and A. da Silva, 2013: Evaluating the impact of orbital sampling on satellite-climate model comparisons. *J. Geophys. Res. Atmos.*, **118**, 355–369, doi:10.1029/2012JD018590.

Hahn, C., S. Warren, and J. London, 1995: The effect of moonlight on observation of cloud cover at night, and application to cloud climatology. *J. Climate*, **8**, 1429–1446, doi:10.1175/1520-0442(1995)008<1429:TEOMOO>2.0.CO;2.

Holland, M. M., and C. M. Bitz, 2003: Polar amplification of climate change in coupled models. *Climate Dyn.*, **21**, 221–232, doi:10.1007/s00382-003-0332-6.

Hubanks, P., S. Platnick, M. King, and B. Ridgway, 2015: MODIS Atmosphere L3 gridded product algorithm theoretical basis document and user's guide, version 4.1. MODIS Algorithm Theoretical Basis Doc. ATBD-MOD-30, 96 pp.

Inoue, J., J. Liu, J. O. Pinto, and J. A. Curry, 2006: Intercomparison of Arctic regional climate models: Modeling clouds and radiation for SHEBA in May 1998. *J. Climate*, **19**, 4167–4178, doi:10.1175/JCLI3854.1.

Intrieri, J. M., M. D. Shupe, T. Uttal, and B. J. McCarty, 2002: An annual cycle of Arctic cloud characteristics observed by radar and lidar at SHEBA. *J. Geophys. Res.*, **107**, 8030, doi:10.1029/2000JC000423.

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

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, doi:10.1175/BAMS-83-11-1631.

Karlsson, J., and G. Svensson, 2011: The simulation of Arctic clouds and their influence on the winter surface temperature in present-day climate in the CMIP3 multi-model dataset. *Climate Dyn.*, **36**, 623–635, doi:10.1007/s00382-010-0758-6.

Kay, J. E., T. L'Ecuyer, A. Gettelman, G. Stephens, and C. O'Dell, 2008: The contribution of cloud and radiation anomalies to the 2007 Arctic sea ice extent minimum. *Geophys. Res. Lett.*, **35**, L08503, doi:10.1029/2008GL033451.

——, and Coauthors, 2012: Exposing global cloud biases in the Community Atmosphere Model (CAM) using satellite observations and their corresponding instrument simulators. *J. Climate*, **25**, 5190–5207, doi:10.1175/JCLI-D-11-00469.1.

Key, J. R., A. J. Schweiger, and R. S. Stone, 1997: Expected uncertainty in satellite-derived estimates of the surface radiation budget at high latitudes. *J. Geophys. Res.*, **102**, 15 837–15 847, doi:10.1029/97JC00478.

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

Klaus, D., W. Dorn, K. Dethloff, A. Rinke, and M. Mielke, 2012: Evaluation of two cloud parameterizations and their possible adaptation to Arctic climate conditions. *Atmosphere*, **3**, 419–450, doi:10.3390/atmos3030419.

Kwok, R., and N. Untersteiner, 2011: The thinning of Arctic sea ice. *Phys. Today*, **64**, 36–41, doi:10.1063/1.3580491.

Lee, S.-E., and K.-H. Seo, 2013: The development of a statistical forecast model for Changma. *Wea. Forecasting*, **28**, 1304–1321, doi:10.1175/WAF-D-13-00003.1.

Letterly, A., J. Key, and Y. Liu, 2016: The influence of winter cloud on summer sea ice in the Arctic, 1983–2013. *J. Geophys. Res. Atmos.*, **121**, 2178–2187, doi:10.1002/2015JD024316.

Lindsay, R., M. Wensnahan, A. Schweiger, and J. Zhang, 2014: Evaluation of seven different atmospheric reanalysis products in the Arctic. *J. Climate*, **27**, 2588–2606, doi:10.1175/JCLI-D-13-00014.1.

Liu, J. P., J. A. Curry, W. B. Rossow, J. R. Key, and X. J. Wang, 2005: Comparison of surface radiative flux data sets over the Arctic Ocean. *J. Geophys. Res.*, **110**, C02015, doi:10.1029/2004JC002381.

Liu, Y., 2015: Estimating errors in cloud amount and cloud optical thickness due to limited spatial sampling using a satellite imager as a proxy for nadir-view sensors. *J. Geophys. Res. Atmos.*, **120**, 6980–6991, doi:10.1002/2015JD023507.

——, and J. R. Key, 2014: Less winter cloud aids summer 2013 Arctic sea ice return from 2012 minimum. *Environ. Res. Lett.*, **9**, 044002, doi:10.1088/1748-9326/9/4/044002.

——, ——, R. A. Frey, S. A. Ackerman, and W. P. Menzel, 2004: Nighttime polar cloud detection with MODIS. *Remote Sens. Environ.*, **92**, 181–194, doi:10.1016/j.rse.2004.06.004.

——, S. A. Ackerman, B. C. Maddux, J. R. Key, and R. A. Frey, 2010: Errors in cloud detection over the Arctic using a satellite imager and implications for observing feedback mechanisms. *J. Climate*, **23**, 1894–1907, doi:10.1175/2009JCLI3386.1.

——, J. R. Key, Z. Liu, X. Wang, and S. Vavrus, 2012a: A cloudier Arctic expected with diminishing sea ice. *Geophys. Res. Lett.*, **39**, L05705, doi:10.1029/2012GL051251.

——, ——, S. A. Ackerman, G. G. Mace, and Q. Zhang, 2012b: Arctic cloud macrophysical characteristics from *CloudSat* and *CALIPSO*. *Remote Sens. Environ.*, **124**, 159–173, doi:10.1016/j.rse.2012.05.006.

Livezey, R. E., 2011: Deterministic forecasts of multi-category events. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, 2nd ed. I. T. Jolliffe and D. B. Stephenson, Eds., John Wiley & Sons, 61–75, doi:10.1002/9781119960003.ch4.

Maslanik, J. A., C. Fowler, J. Stroeve, S. Drobot, J. Zwally, D. Yi, and W. Emery, 2007: A younger, thinner Arctic ice cover: Increased potential for rapid, extensive sea-ice loss. *Geophys. Res. Lett.*, **34**, L24501, doi:10.1029/2007GL032043.

Moorthi, S., and M. J. Suarez, 1992: Relaxed Arakawa–Schubert: A parameterization of moist convection for general circulation models. *Mon. Wea. Rev.*, **120**, 978–1002, doi:10.1175/1520-0493(1992)120<0978:RASAPO>2.0.CO;2.

Naud, C. M., J. F. Booth, and A. D. Del Genio, 2014: Evaluation of ERA-Interim and MERRA cloudiness in the Southern Ocean. *J. Climate*, **27**, 2109–2124, doi:10.1175/JCLI-D-13-00432.1.

Pan, H.-L., and W.-S. Wu, 1994: Implementing a mass flux convective parameterization package for the NMC medium-range forecast model. Preprints, *10th Conf. on Numerical Weather Prediction*, Portland, OR, Amer. Meteor. Soc., 96–98.

Paquin-Ricard, D., C. Jones, and P. A. Vaillancourt, 2010: Using ARM observations to evaluate cloud and clear-sky radiation processes as simulated by the Canadian regional climate model GEM. *Mon. Wea. Rev.*, **138**, 818–838, doi:10.1175/2009MWR2745.1.

Randall, D., and Coauthors, 1998: Status of and outlook for large-scale modeling of atmosphere–ice–ocean interactions in the Arctic. *Bull. Amer. Meteor. Soc.*, **79**, 197–219, doi:10.1175/1520-0477(1998)079<0197:SOAOFL>2.0.CO;2.

Reichle, R. H., R. A. Lucchesi, J. V. Ardizzone, G.-K. Kim, E. B. Smith, and B. H. Weiss, 2015: Soil Moisture Active Passive (SMAP) Mission Level 4 Surface and Root Zone Soil Moisture (L4_SM) Product Specification Document. GMAO Office Note 10, 82 pp. [Available online at http://gmao.gsfc.nasa.gov/pubs/office_notes.]

Rienecker, M. M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648, doi:10.1175/JCLI-D-11-00015.1.

Sandvik, A., M. Biryulina, N. G. Kvamsto, J. J. Stamnes, and K. Stamnes, 2007: Observed and simulated microphysical composition of arctic clouds: Data properties and model validation. *J. Geophys. Res.*, **112**, D05205, doi:10.1029/2006JD007351.

Schweiger, A. J., and J. R. Key, 1994: Arctic Ocean radiative fluxes and cloud forcing estimated from the ISCCP C2 cloud dataset, 1983–1990. *J. Appl. Meteor.*, **33**, 948–963, doi:10.1175/1520-0450(1994)033<0948:AORFAC>2.0.CO;2.

——, J. Zhang, R. W. Lindsay, and M. Steele, 2008: Did unusually sunny skies help drive the record sea ice minimum of 2007? *Geophys. Res. Lett.*, **35**, L10503, doi:10.1029/2008GL033463.

Serreze, M. C., and J. A. Francis, 2006: The Arctic amplification debate. *Climatic Change*, **76**, 241–264, doi:10.1007/s10584-005-9017-y.

Shupe, M. D., V. P. Walden, E. Eloranta, T. Uttal, J. R. Campbell, S. M. Starkweather, and M. Shiobara, 2011: Clouds at Arctic atmospheric observatories. Part I: Occurrence and macrophysical properties. *J. Appl. Meteor. Climatol.*, **50**, 626–644, doi:10.1175/2010JAMC2467.1.

Simjanovski, D., E. Girard, and P. Du, 2011: An evaluation of Arctic cloud and radiation processes simulated by the limited-area version of the Global Multiscale Environmental Model (GEM-LAM). *Atmos.–Ocean*, **49**, 219–234, doi:10.1080/07055900.2011.604266.

Solomon, S., and Coauthors, 2007: *Climate Change 2007: Synthesis Report*. Cambridge University Press, 22 pp.

Stefanova, L., V. Misra, J. J. O'Brien, E. P. Chassignet, and S. Hameed, 2012: Hindcast skill and predictability for precipitation and two-meter air temperature anomalies in global circulation models over the Southeast United States. *Climate Dyn.*, **38**, 161–173, doi:10.1007/s00382-010-0988-7.

Stroeve, J. C., M. C. Serreze, M. M. Holland, J. E. Kay, J. Malanik, and A. P. Barrett, 2012: The Arctic's rapidly shrinking sea ice cover: A research synthesis. *Climatic Change*, **110**, 1005–1027, doi:10.1007/s10584-011-0101-1.

Taylor, P. C., S. Kato, K.-M. Xu, and M. C. J. D. Cai, 2015: Covariance between Arctic sea ice and clouds within atmospheric state regimes at the satellite footprint level. *J. Geophys. Res.*, **120**, 12 656–12 678, doi:10.1002/2015JD023520.

Tjernström, M., J. Sedlar, and M. D. Shupe, 2008: How well do regional climate models reproduce radiation and clouds in the Arctic? An evaluation of ARCMIP simulations. *J. Appl. Meteor. Climatol.*, **47**, 2405–2422, doi:10.1175/2008JAMC1845.1.

Vaughan, M. A., and Coauthors, 2009: Fully automated detection of cloud and aerosol layers in the *CALIPSO* lidar measurements. *J. Atmos. Oceanic Tech.*, **26**, 2034–2050, doi:10.1175/2009JTECHA1228.1.

Vavrus, S., 2004: The impact of cloud feedbacks on Arctic climate under greenhouse forcing. *J. Climate*, **17**, 603–615, doi:10.1175/1520-0442(2004)017<0603:TIOCFO>2.0.CO;2.

——, and D. Waliser, 2008: An improved parametrization for simulating Arctic cloud amount in the CCSM3 climate model. *J. Climate*, **21**, 5673–5687, doi:10.1175/2008JCLI2299.1.

——, ——, A. Schweiger, and J. Francis, 2009: Simulations of 20th and 21st century Arctic cloud amount in the global climate models assessed in the IPCC AR4. *Climate Dyn.*, **33**, 1099–1115, doi:10.1007/s00382-008-0475-6.

Walsh, J. E., V. M. Kattsov, W. L. Chapman, V. Govorkova, and T. Pavlova, 2002: Comparison of Arctic climate simulations by uncoupled and coupled global models. *J. Climate*, **15**, 1429–1446, doi:10.1175/1520-0442(2002)015<1429:COACSB>2.0.CO;2.

——, S. J. Vavrus, and W. L. Chapman, 2005: Workshop on modeling of the Arctic atmosphere. *Bull. Amer. Meteor. Soc.*, **86**, 845–852, doi:10.1175/BAMS-86-6-845.

——, W. L. Chapman, and D. Portis, 2009: Arctic cloud fraction and radiative fluxes in atmospheric reanalyses. *J. Climate*, **22**, 2316–2334, doi:10.1175/2008JCLI2213.1.

Wang, X. J., and J. R. Key, 2005: Arctic surface, cloud, and radiation properties based on the AVHRR Polar Pathfinder dataset. Part I: Spatial and temporal characteristics. *J. Climate*, **18**, 2558–2574, doi:10.1175/JCLI3438.1.

Wilson, A. B., D. H. Bromwich, and K. M. Hines, 2012: Evaluation of Polar WRF forecasts on the Arctic System Reanalysis Domain: 2. Atmospheric hydrologic cycle. *J. Geophys. Res.*, **117**, D04107, doi:10.1029/2011JD016765.

WMO, 2010: Appendix II. Manual on the global data-processing and forecasting system, World Meteorological Organization Doc. WMO-485, II.8-1–II.8-17.

Wu, Q., and X. Zhang, 2010: Observed forcing-feedback processes between Northern Hemisphere atmospheric circulation and Arctic sea ice coverage. *J. Geophys. Res.*, **115**, D14119, doi:10.1029/2009JD013574.

——, J. Zhang, X. Zhang, and W. Tao, 2014: Interannual variability and long-term changes of atmospheric circulation over the Chukchi and Beaufort Seas. *J. Climate*, **27**, 4871–4889, doi:10.1175/JCLI-D-13-00610.1.

Wyser, K., and C. G. Jones, 2005: Modeled and observed clouds during Surface Heat Budget of the Arctic Ocean (SHEBA). *J. Geophys. Res.*, **110**, D09207, doi:10.1029/2004JD004751.

——, and Coauthors, 2008: An evaluation of Arctic cloud and radiation processes during the SHEBA year: Simulation results from eight Arctic regional climate models. *Climate Dyn.*, **30**, 203–223, doi:10.1007/s00382-007-0286-1.

Xie, S., X. Liu, C. Zhao, and Y. Zhang, 2013: Sensitivity of CAM5-simulated Arctic clouds and radiation to ice nucleation parameterization. *J. Climate*, **26**, 5981–5999, doi:10.1175/JCLI-D-12-00517.1.

Zhang, M. H., and Coauthors, 2005: Comparing clouds and their seasonal variations in 10 atmospheric general circulation models with satellite measurements. *J. Geophys. Res.*, **110**, D15S02, doi:10.1029/2004JD005021.

Zhang, X., A. Sorteberg, Z. Jing, R. Gerdes, and J. C. Comiso, 2008: Recent radical shifts of atmospheric circulations and rapid changes in Arctic climate system. *Geophys. Res. Lett.*, **35**, L22701, doi:10.1029/2008GL035607.

Zhao, M., and Z. Wang, 2010: Comparison of Arctic clouds between European Center for Medium-Range Weather Forecasts simulations and Atmospheric Radiation Measurement Climate Research Facility long-term observations at the North Slope of Alaska Barrow site. *J. Geophys. Res.*, **115**, D23202, doi:10.1029/2010JD014285.

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, doi:10.1175/JCLI-D-11-00147.1.

Zygmuntowska, M., T. Mauritsen, J. Quaas, and L. Kaleschke, 2012: Arctic clouds and surface radiation—A critical comparison of satellite retrievals and the ERA-Interim reanalysis. *Atmos. Chem. Phys.*, **12**, 6667–6677, doi:10.5194/acp-12-6667-2012.