Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1029/2018JD028805

Evaluation of Radiation and Clouds From Five Reanalysis Products in the Northeast Pacific Ocean

Lauren Schmeisser[1], Laura M. Hinkelman[2], and Thomas P. Ackerman[1,2]

[1]Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA, [2]Joint Institute for the Study of the Atmosphere and Ocean, University of Washington, Seattle, WA, USA

Key Points:
- An evaluation of radiative fluxes and total cloud fraction from five reanalysis products is performed using CERES EBAF satellite estimates
- Shortwave radiative flux is mostly overestimated in reanalysis data sets, while longwave radiative flux has both negative and positive biases, depending on the reanalysis
- There is no gold standard reanalysis—the most appropriate data set depends on variables, time period, and subregion of interest

Supporting Information:
- Supporting Information S1

Correspondence to:
L. Schmeisser,
schmeiss@uw.edu

Citation:
Schmeisser, L., Hinkelman, L. M., & Ackerman, T. P. (2018). Evaluation of radiation and clouds from five reanalysis products in the Northeast Pacific Ocean. *Journal of Geophysical Research: Atmospheres*, *123*, 7238–7253. https://doi.org/10.1029/2018JD028805

Received 9 APR 2018
Accepted 20 JUN 2018
Accepted article online 2 JUL 2018
Published online 20 JUL 2018

Abstract Atmospheric reanalyses are valuable tools for studying the atmosphere, as they provide temporally and spatially complete coverage of atmospheric variables. However, some regions are susceptible to large biases in reanalysis products due to the scarce data available to assimilate into the reanalyses. Consequently, evaluation of reanalyses using available measurements is essential for quantifying regional errors. Here we use NASA's CERES satellite estimates to evaluate surface radiative fluxes and total cloud fraction in the Northeast Pacific from five reanalysis products—ERA-Interim, MERRA2, JRA-55, NCEP2, and CFSR—from years 2001 to 2015. Results show that biases of surface incident shortwave radiative flux in reanalyses compared to satellite estimates range from 3.8 (CFSR) to 21.2 Wm-2 (NCEP2), with significant biases in JRA-55 and NCEP2. Mean surface downward longwave radiative flux in the reanalysis products is biased by −8.9 (MERRA2) to 3.9 Wm^{-2} (JRA-55), with significant biases in MERRA2 and NCEP2. Errors in the surface radiative fluxes are partially linked to differences in total cloud fraction in the satellite estimates and reanalyses, which show significant negative biases ranging from −8% (CFSR) to −21.7% (NCEP2). There is not one reanalysis that outperforms the rest in the NE Pacific. The most appropriate data set depends on the variables of interest, subregion of the NE Pacific being studied, time period of interest, and whether the reanalysis data will be used to study long-term or short-term climate processes. Using the errors presented for each reanalysis data set can help guide appropriate use and bound uncertainty for the five reanalysis products analyzed.

1. Introduction

Atmospheric reanalysis products are valuable tools for climate scientists because they provide long time series of most atmospheric variables with complete global spatial coverage. Reanalysis data can be useful for studying a variety of atmospheric and climate processes.

Reanalysis makes use of both observations and models to produce an optimized global set of atmospheric variables. This global data set is created using a single version of a forecast model and a series of atmospheric observations at high time resolution (subdaily). The data are assimilated into the model resulting in a model-generated data set that is global, uniformly produced, and as consistent with observations as possible (Betts et al., 2006). Even though observations of the atmosphere are not consistent in space or time, an assimilation model provides a forward interpolation that "fills in" these gaps and maintains dynamically and energetically consistent fields. Forecast models are imperfect representations of the atmosphere due to a combination of finite grid resolution, uncertainty in the parameterization of subgrid scale processes, and imprecision in initial conditions. However, the assimilation of observations continually adjusts the model fields to be more consistent with the observations (Betts et al., 2006). The resulting long-term data sets are the best available representation of long-term and large-scale atmospheric fields.

Despite the utility of reanalysis, past studies illustrate the potential for biases in reanalysis products (e.g., Betts et al., 2006; Bosilovich et al., 2008; Dee et al., 2011). Some atmospheric variables are better represented by reanalysis models than others. For example, variables for which we have large observational data sets, are smooth at all scales, and are well understood in terms of model physics (e.g., temperature) are generally reproduced well by reanalysis (e.g., Bao & Zhang, 2013; Betts et al., 1998, 2006; Donat et al., 2014; Dulière et al., 2011). There are exceptions to this, as some studies show that moisture is not well reproduced by reanalysis products (e.g., Jiang et al., 2015). On the other hand, variables for which we have fewer or less accurate observational data, are more variable at small scales, and depend on model parameterizations have much larger biases in reanalysis products (e.g., clouds and precipitation; Bosilovich et al., 2008; Lindsay et al.,

2014; Pfeifroth et al., 2013; Vey et al., 2010; Walsh et al., 2009; Zib et al., 2012). In addition, the quality of a reanalysis data set at a particular location and time depends on both the numerical model and the number and quality of observations input to the assimilation framework (Arakawa & Kitoh, 2004; Fujiwara et al., 2017), so biases may have a regional structure.

Numerous studies have specifically evaluated radiative fluxes and cloud fraction from reanalyses using observations, most found substantial biases in the reanalyses, many found a pattern of consistent overestimation of downward shortwave radiative flux and underestimation of cloud fraction. Zib et al. (2012) used surface radiative flux measurements to assess radiative fluxes and total cloud fraction at two locations in the Arctic region and found that all reanalyses had substantial biases in cloud fraction, especially during the winter. Radiative flux biases varied widely depending on the reanalysis. ERA-Interim and Climate Forecast System Reanalysis (CFSR) generally had the smallest biases, and National Centers for Environmental Prediction 2 (NCEP2) had the largest biases. Decker et al. (2012) used flux tower observations from all over the world to evaluate incoming shortwave radiation and found that, although the reanalyses captured the annual cycle very well, all reanalyses overestimated the incident shortwave radiative flux. NCEP had the largest biases, exceeding 25 Wm^{-2} at many of the sites analyzed, while ERA-Interim showed the smallest shortwave biases. Similarly, Zhang et al. (2016) found all reanalyses products overestimated incoming shortwave radiative fluxes, with the exception of ERA-Interim which had a small negative bias of -2.98 Wm^{-2} compared to CERES EBAF satellite data.

Most evaluations of reanalysis data sets have been performed on a global scale and tend to focus on investigating long-term annual averages (e.g., Dee et al., 2011). These analyses provide important assessments of reanalysis quality and usefulness. However, many atmospheric phenomena happen on smaller regional and temporal scales, and long-term global reanalysis evaluations do not provide information on the value of these data sets at smaller spatiotemporal scales. Regional evaluations give more detailed insight into reanalysis performance at finer resolutions, which is particularly important for "data-sparse" regions with limited surface and in situ observations. Surface and in situ observations provide quality controlled data with excellent temporal coverage that can help constrain both satellite and reanalysis data. While reanalysis products are arguably the best available atmospheric information for regions lacking surface and in situ measurements (Jakobson et al., 2012; Screen & Simmonds, 2011), due to their complete spatiotemporal coverage and comprehensive output of atmospheric variables, it is precisely in these data-sparse regions where reanalysis data sets are most susceptible to large uncertainties because there are less data to assimilate. The Arctic is an exemplar region exhibiting this discrepancy. While reanalysis data are essential tools for studying the Arctic atmosphere, reanalysis products perform poorly in this region, particularly with modeling temperature, radiative fluxes, precipitation, relative humidity, wind speed, and ice thickness (Bromwich et al., 2007; Jakobson et al., 2012; Lindsay et al., 2014; Lüpkes et al., 2010; Porter et al., 2011; Walsh et al., 2009). These case studies in the Arctic suggest that other data-sparse regions require reanalysis evaluations as well.

We are specifically interested in the performance of reanalysis products in the Northeast (NE) Pacific. There are a number of reasons to study climate and climate variability in the NE Pacific; for example, its atmosphere-ocean interactions affect weather downstream in North America and it is a sensitive and productive region for fisheries, including shellfish. Our attention was drawn specifically to this region by a recent large and persistent marine heat wave that occurred in this area from 2013 to 2016, colloquially referred to as "the Blob" (Bond et al., 2015). The NE Pacific region, however, has very few in situ observations: an array of National Data Buoy Center buoys near the coast (Meindl & Hamilton, 1992), a ship making observations between the coast and Ocean Station Papa at 50N, 145W, and only select variables available from satellite estimates (e.g., satisfactory turbulent heat fluxes are not available from satellite measurements). Thus, reanalysis products are an indispensable tool for studying the region. However, given potential uncertainty in reanalysis products, it is prudent to exercise caution when using them as a means to examine physical processes on small scales. In this study, we compare radiative fluxes and cloud fraction from reanalyses in the NE Pacific to satellite estimates in order to evaluate usefulness of these reanalysis quantities for regional studies.

We use NASA's Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) satellite estimates to evaluate the surface radiative fluxes and total cloud fraction from five reanalysis products. The reanalyses include the European Center for Medium-Range Weather Forecasts reanalysis

(ERA-Interim), NASA's Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA2), the Japanese Meteorological Agency 55-year reanalysis (JRA-55), the National Center for Environmental Prediction and Department of Energy reanalysis 2 (NCEP2), and NCEP CFSR. Three approaches are used to evaluate the reanalysis data sets, as summarized in the following research questions:

1. How well do reanalysis products replicate the seasonality of surface radiative fluxes and total cloud fraction in the NE Pacific?
2. Do reanalysis products adequately capture the spatial distribution of surface radiative fluxes and total cloud fraction in the NE Pacific?
3. How well do reanalysis products reproduce anomalies in surface radiative fluxes and total cloud fraction in the NE Pacific?

2. Background

The five reanalysis products assessed here have been evaluated in various capacities throughout many regions of the globe. Results show biases that vary widely by location, variable of interest, and reanalysis product (e.g., Bosilovich et al., 2008; Dee et al., 2011). Consequently, the choice of reanalysis product to use for climate analysis in a given region can yield very different results for the same diagnostic (Fujiwara et al., 2017).

We are aware of only one previous study that evaluates reanalysis data in the NE Pacific. Ladd and Bond (2002) compare NCEP1 (also referred to as NCEP/National Center for Atmosphere Research) reanalysis output with measurements from multiple moored buoys in the Bering Sea and one moored buoy (the aforementioned Ocean Station Papa) in the NE Pacific from 1995 to 2000. Wind, radiation, cloud cover, and sea level pressure were evaluated. The direction of 10 m winds and sea level pressure from the reanalysis product were satisfactory compared to buoy observations. Speed of 10 m winds, on the other hand, was biased high by about 5%. The shortwave radiative flux from NCEP1 showed a 70 Wm^{-2} bias in the Bering Sea and a 20 Wm^{-2} bias in the NE Pacific. The positive bias in shortwave radiative flux in the NE Pacific was present in both the summer and winter. Further analysis in Ladd and Bond (2002) suggests that radiative flux biases are likely due to the NCEP1 biases in clouds. In particular, they found the reanalysis model has the largest biases during fair weather events when low clouds are prevalent. Here we expand on this previous analysis by considering different meteorological quantities and five newer reanalysis data sets, all for a longer time period.

It should be noted that there are emerging efforts within the scientific community to systematically evaluate reanalysis products to provide comparisons of reanalysis data for all variables and regions of the globe. One such effort is the SPARC Reanalysis Intercomparison Project, which was formed to diagnose biases in reanalysis products, with a focus on the upper troposphere, stratosphere, and lower mesosphere (Fujiwara et al., 2017). Consequently, the SPARC Reanalysis Intercomparison Project will be less useful for those interested in the performance of reanalysis products at the surface, making assessments such as ours still necessary.

3. Data and Methods

3.1. NASA Satellite Estimates

We use the CERES Edition 4 EBAF satellite estimates (Kato et al., 2013, 2018; Wielicki et al., 1996) as our baseline for comparison with surface radiative fluxes from the five reanalyses. The CERES EBAF-Surface (Kato et al., 2013, 2018) product is based on the Edition 4 CERES synoptic 1° monthly averaged data (SYN1deg-Month; Rutan et al., 2015) radiative flux data set. The underlying SYN1deg-Hour fluxes are obtained by applying a radiative transfer model to atmospheric profiles of temperature and water vapor from GEOS5, cloud parameters from Moderate Resolution Imaging Spectroradiometer (MODIS), and geostationary satellite observations. The monthly averaged SYN1deg-Month data are then constrained to match the top of atmosphere (TOA) radiative fluxes from the CERES Ed. 4. EBAF-TOA (Loeb et al., 2009), which itself has been constrained to the ocean heat storage. The basic approach in this process is to vary the SYN1deg input values within their uncertainty bounds until a good match with the TOA fluxes is achieved. Adjustment of the cloud parameters takes advantage of the higher quality active cloud measurements from Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations and CloudSat (Kato et al., 2013, 2018). The final EBAF-Surface monthly mean downward and upward shortwave and longwave radiative fluxes at the surface are provided on a

Table 1
Reanalysis Products Evaluated

Product	Model grid resolution	Availability	Reference
ECMWF ERA-Interim	T255, 60 vertical levels	January 1979 to present	Dee et al. (2011)
NASA MERRA2	0.5°x0.625°, 72 vertical levels	January 1980 to present	Gelaro et al. (2017)
JRA-55	T319, 60 vertical levels	December 1957 to present	Kobayashi et al. (2015)
NCEP2	T62, 28 vertical levels	January 1979 to present	Kanamitsu et al. (2002)
CFSR	T382, 64 vertical levels	January 1979 to present	Saha et al. (2010) Saha et al. (2014)

1° × 1°global grid. Total cloud fraction data for comparison to the reanalyses are provided by SYN1deg-Month satellite estimates.

The satellite estimates are not without uncertainty. Since surface irradiance values are only available through a radiative transfer model that requires input of atmospheric variable measurements (e.g., aerosol and cloud), the accuracy of the satellite retrievals depends on the accuracy of these inputs (Kato, Loeb, Rose, et al., 2012). Estimates suggest that CERES EBAF monthly gridded estimates of surface upward shortwave radiative flux over the ocean are accurate within 8.3 Wm^{-2}, downward shortwave radiative flux within 6.9 Wm^{-2}, downward longwave radiative flux within 4.3 Wm^{-2}, and upward longwave radiative flux within 11.5 Wm^{-2} (scaled from Kato et al., 2018; see also Kato, Loeb, Rose, et al., 2012; Kato, Loeb, Rutan, et al., 2012; Kato et al., 2013). Cloud fraction uncertainty is estimated at 5%, per the CERES EBAF Surface Data Quality Summary (https://ceres.larc.nasa.gov/documents/DQ_summaries/CERES_EBAF-Surface_Ed4.0_DQS.pdf) We expect that these uncertainties are further reduced when spatially averaging over the region of the NE Pacific, so the uncertainty estimates here are conservative. The satellite uncertainty estimates are used here as thresholds for acceptance of agreement between satellite and reanalysis data. If the bias between reanalysis and satellite observations is less than the cited satellite uncertainty, the reanalysis and satellite observations are considered to be in adequate agreement. If the bias is larger than the satellite estimate uncertainty, the bias is significant and the reanalysis product is not in agreement with satellite estimates.

Clouds and the Earth's Radiant Energy System EBAF is one of the only long-term satellite data sets available for the entirety of the NE Pacific region. Other available satellite data sets include the International Satellite Cloud Climatology Project (ISCCP) Flux product (Zhang et al., 1995) and the NASA/Global Energy and Water Cycle Experiment Surface Radiation Budget product (Zhang et al., 2013). ISCCP is produced at 2.5° resolution—much lower than that of CERES EBAF. Furthermore, both ISCCP and NASA/Global Energy and Water Cycle Experiment estimates have similar or larger uncertainties than those for CERES EBAF, which is why CERES EBAF is used here for evaluation of reanalysis products.

Clouds and the Earth's Radiant Energy System satellite estimates of radiative fluxes have not been assimilated into ERA-Interim, MERRA2, JRA-55, NCEP2, or CFSR. Therefore, comparison to the CERES EBAF radiative fluxes can be considered an independent measure of the quality of the reanalysis radiative fluxes. However, CERES utilizes data from geostationary satellites, MODIS, and Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations to derive cloud variables and radiative fluxes, and some aspects of these products are assimilated into the reanalyses. For example, MODIS aerosol estimates are assimilated into MERRA2 and ERA-Interim.

3.2. Reanalysis Products

Monthly averages of radiative fluxes and total cloud fraction from five reanalysis products (ERA-Interim, MERRA2, JRA-55, NCEP2, and CFSR) are evaluated here. Although other reanalysis products do exist for this region, the five investigated here are (a) the most up to date versions of older available reanalysis models and (b) commonly used and highly cited in the literature. Table 1 outlines technical information about each of the reanalysis data sets. The rest of section 3 provides descriptions of each reanalysis product.

3.2.1. ECMWF ERA-Interim

European Center for Medium-Range Weather Forecasts reanalysis is the most recent long-term reanalysis product from ECMWF and is well documented in Dee et al. (2011). ERA-Interim data are available from January 1979 to December 2017 at subdaily, daily, and monthly time steps. The model uses a T255 grid

with 60 vertical levels, and data are available at 0.75° × 0.75° at 60 vertical levels. The ERA-Interim product improves upon the ERA-40 system (described in Uppala et al., 2005) by using a 4D-Var data assimilation system that more fully integrates observations from the satellite era and more completely incorporates measurements made between analysis times. ERA-Interim has an improved hydrologic cycle and stratospheric circulation compared to ERA-40 (Fujiwara et al., 2017).

3.2.2. NASA MERRA2

Modern-Era Retrospective analysis for Research and Applications version 2 is the most recent version of the reanalysis product from NASA's Global Modeling and Assimilation Office. Detailed information about MERRA2 can be found in Gelaro et al. (2017) and Bosilovich et al. (2015). MERRA2 has data available from January 1980 to November 2017 at subdaily, daily, and monthly time steps. The MERRA2 data are available at a resolution of 0.5° × 0.625° with 72 vertical levels. MERRA2 improves upon the former NASA MERRA reanalysis product by ingesting new data types and implementing improvements to the model, including closed balance between surface water fluxes and total atmospheric water, a modified gravity wave scheme, improved treatment of the upper atmosphere, and assimilation of aerosol optical depth measurements (Bosilovich et al., 2015; Fujiwara et al., 2017; Randles et al., 2017).

3.2.3. JMA JRA-55

The JRA-55 is the most recent version of the JMA reanalysis effort and is documented in Kobayashi et al. (2015). JRA-55 is available from December 1957 to July 2017 at subdaily and monthly time steps. The native and available model resolution is T319 with 60 levels. JRA-55 is an update of its predecessor, JRA-25 (Onogi et al., 2007), and assimilates satellite data as well as upper air data using a 4D-Var assimilation scheme (Fujiwara et al., 2017; Kobayashi et al., 2015).

3.2.4. NCEP/DOE Reanalysis 2

The NCEP/DOE Reanalysis 2 has data available from January 1979 to January 2018 at subdaily, daily, and monthly time steps. The native resolution of the NCEP2 reanalysis model is T62 with 28 levels, which is the lowest resolution of all the reanalysis products evaluated here. The available data resolution is 2.5° × 2.5° with 28 vertical levels. Detailed description of the NCEP2 and the corrections and updates implemented since its previous version NCEP/National Center for Atmosphere Research Reanalysis 1 (NCEP1; Kalnay et al., 1996) are found in Kanamitsu et al. (2002).

3.2.5. NCEP CFSR

The CFSR product is the most advanced reanalysis product produced by NOAA NCEP. CFSR is available from January 1979 to February 2011, and the operational model on which this reanalysis product is based was upgraded in March 2011 to the CFSR version 2 (CFSv2). Since the time period of interest in this study spans both the CFSR and CFSv2, the two reanalysis products were merged to get a continuous reanalysis data set, which is referred to throughout as CFSR. Detailed information on CFSR is found in Saha et al. (2010), while details on CFSv2 and how the model was improved compared to the previous version are found in Saha et al. (2014). Although there are small differences between the CFSR and CFSv2 reanalysis products that could affect working across the two data sets (Saha et al., 2014), we expect no problems with merging the two data sets here given the variables of interest, as many previous studies have done so without issue (Betts et al., 2006; Fujiwara et al., 2017).

The CFSR is unique from the other reanalyses studied here because it is a coupled atmosphere-ocean-land-sea ice system model. The model assimilates measurements of carbon dioxide, trace gases, and aerosols and should reflect changes in climate due to these variables. The CFSR is improved from the NCEP/DOE reanalysis in that it is a coupled model and assimilates more observations with a more sophisticated assimilation scheme. However, fewer evaluations of CFSR have been conducted compared to other reanalysis products, so its performance globally is not yet well established.

3.3. Methods

For our purposes, the NE Pacific region is defined from 40°–60°N and 180°–120°W. In order to compare reanalysis and satellite products, all data sets were regridded to match the data set with the lowest resolution—in this case, the NCEP2 2.5° × 2.5° grid. Regridding was performed using bilinear interpolation. The time period from 2001 to 2015 was used in order to maximize overlap of reanalysis products with currently available satellite data (the processing of CERES data to produce surface fluxes lags a half year or more behind acquisition of the satellite data).

Table 2
Statistical Comparison of Monthly Radiative Fluxes and Total Cloud Fraction From Reanalysis Data and CERES EBAF Satellite Estimates, Spatially Averaged Over the Domain From 40°–60°N and 180°–120°W

Data product	Measures	Downward shortwave radiative flux (Wm^{-2})	Upward shortwave radiative flux (Wm^{-2})	Downward longwave radiative flux (Wm^{-2})	Upward longwave radiative flux (Wm^{-2})	Total cloud fraction (%)
CERES EBAF	Mean	119.4	9.09	313.5	356.2	88.4
	Uncertainty	6.9	8.3	4.3	11.5	5.0
ERA-Interim	Mean	124.9	10.1	313.9	356.9	78.9
	Bias (MAD)	5.5 **(7.7)**	1.0 (1.0)	0.41 (2.1)	0.71 (0.72)	**−9.5 (9.5)**
	Relative bias (%)	4.4	10.0	0.13	0.20	12.0
	RMSD	10.5	1.4	2.6	0.86	9.6
	Correlation coeff.	0.99	0.99	0.99	0.99	0.67
MERRA2	Mean	122.4	10.7	304.6	353.5	70.8
	Bias (MAD)	3.0 **(7.1)**	1.7 (1.7)	**−8.9 (8.9)**	−2.7 (2.7)	**−17.5 (17.5)**
	Relative bias (%)	2.4	15.5	2.9	0.78	24.7
	RMSD	10.1	1.7	9.2	2.8	17.7
	Correlation coeff.	0.99	0.99	0.99	0.99	0.74
JRA-55	Mean	131.7	11.2	309.6	358.7	66.7
	Bias (MAD)	**12.3 (12.3)**	2.1 (2.1)	−3.8 (3.9)	2.5 (2.5)	**−21.6 (21.6)**
	Relative bias (%)	9.4	18.6	1.3	0.69	32.4
	RMSD	15.1	2.1	4.7	2.5	21.7
	Correlation coeff.	0.99	0.99	0.99	0.99	0.41
NCEP2	Mean	140.6	12.7	307.6	357.4	66.7
	Bias (MAD)	**21.2 (21.2)**	3.6 (3.6)	**−5.8 (5.9)**	1.2 (2.1)	**−21.7 (21.7)**
	Relative bias (%)	15.1	28.6	1.9	0.35	32.6
	RMSD	26.0	4.0	7.7	2.4	22.3
	Correlation coeff.	0.99	0.99	0.96	0.99	0.43
CFSR	Mean	120.3	11.1	317.1	357.4	80.4
	Bias (MAD)	0.86 (3.8)	2.0 (2.0)	3.6 (3.6)	1.3 (1.3)	**−8.0 (8.0)**
	Relative bias (%)	0.7	18.0	1.1	0.36	9.9
	RMSD	4.6	2.2	3.9	1.4	8.1
	Correlation coeff.	0.99	0.99	0.99	0.99	0.79

Note. Statistics are provided for the mean, bias, mean absolute deviation (MAD), relative bias, root-mean-square deviation (RMSD) and correlation coefficient. Biases and MADs are bolded if they are larger than the uncertainty in satellite estimates, meaning bias is significant.

Surface all-sky radiative fluxes, including downward shortwave, upward shortwave, downward longwave, and upward longwave, are evaluated here. Not all of the reanalysis data sets provide clear-sky radiative fluxes, and thus, analysis of clear-sky variables is not included here. For the entirety of this manuscript, reference to radiative fluxes assumes surface, all-sky values unless otherwise indicated. Total cloud fraction is analyzed instead of cloud fraction at different levels, because the satellite and reanalysis products do not provide cloud fraction information at the same vertical levels.

We examine how reanalysis products reproduce seasonality, spatial distribution, and anomalies of radiative fluxes and total cloud fraction in the NE Pacific compared to satellite observations, which we treat as the objective standard. Monthly means of variables are utilized throughout. Most of the evaluation presented here, particularly the statistics and plots presented in sections 4.1 and 4.3, are spatially averaged over a box from 40–60°N and 180–120°W, unless indicated otherwise.

Monthly climatologies of the variables are computed by averaging the monthly mean values for each month of each year of the time series from 2001 to 2015 to yield a "typical" annual cycle. Five measures are utilized throughout to quantify reanalysis performance: bias, mean absolute deviation, relative bias (as percent of the mean), root-mean-square deviation and correlation coefficient. All differences are computed as model minus satellite values. The mean absolute deviation is computed by averaging the absolute value of the difference at every point in a time series, which avoids underreporting of the magnitude of differences due to averaging out of positive and negative biases. Relative bias reports the bias as a percentage of the multiyear annual mean of that variable.

Mean spatial distributions (as in Figures 3 and 4) were computed by averaging the variable annual means from 2001 to 2015 in each grid box. The contour plots show the smoothed long-term annual means at

each point in the domain. Spatial distributions of reanalysis bias were computed by subtracting annual averaged satellite estimates from annual averaged reanalysis data, then plotted at each point in the domain.

Anomalies were computed as monthly values minus long-term monthly climatology. Standardized anomalies were computed by dividing the anomaly by the standard deviation of the time series. Standardized anomalies take into account differences in the reanalysis time series and help visually compare magnitudes of anomalies between reanalyses and satellite estimates. To evaluate more intense anomalous events for each variable, all time steps where the CERES estimates exceeded one standard deviation of the mean were considered. At each of those time steps, anomaly biases were computed as reanalysis value minus satellite value at that time step.

4. Results

4.1. Monthly Climatology and Seasonality of Radiative Fluxes and Total Cloud Fraction

Monthly mean values of the radiative fluxes and total cloud fraction are computed for reanalysis products and CERES EBAF for the period from 2001 to 2015 and are summarized in Table 2. The table also provides statistical comparison of the reanalyses to observations of radiative fluxes and total cloud fraction over the entire NE Pacific domain. In the left column in Figure 1 are plots of the monthly climatology of radiative fluxes from satellite observations in black and the reanalyses in the colors. In the right column are reanalysis biases compared to CERES EBAF radiative fluxes.

Annual average downward shortwave radiative flux at the surface is overestimated by every reanalysis data set for most months of the year. Annual average positive biases range from 0.7 Wm^{-2} (CFSR) to 21.2 Wm^{-2} (NCEP2). The annual average of the absolute values of the differences range from 3.8 Wm^{-2} (CFSR) to 21.2 Wm^{-2} (NCEP2), though JRA-55 and NCEP2 are the only products with biases that exceed the uncertainty of the satellite estimate and are thus considered significant. The average NCEP2 shortwave bias is similar to the NCEP1 bias of 20 Wm^{-2} in the NE Pacific reported by Ladd and Bond (2002). The high bias in the NCEP2 reanalysis constitutes 15% of the average downward shortwave radiative flux from that product; but the summer bias (47 Wm^{-2}) is much greater than the winter bias and constitutes closer to 20% of the average downward shortwave radiative flux in the summer. Reanalyses have the largest biases in the summer, which is expected since summer also has the largest magnitude of shortwave radiative flux at the surface. However, every reanalysis product, with the exception of MERRA2, gets the seasonal cycle of downward shortwave radiative flux correct. This finding is not surprising given the pronounced annual cycle of solar radiation in the north Pacific region; it serves, however, as a zero-order evaluation of radiative fluxes in the reanalyses. Zhang et al. (2016) found that globally MERRA, JRA-55, NCEP2, and CFSR overestimated downward shortwave radiative flux at the surface, while ERA-Interim underestimated downward shortwave radiative flux. They found that global mean biases ranged from -3 to 22 Wm^{-2}, which is similar to biases found here, with the exception of ERA-Interim, which showed that biases of the opposite sign. CFSR also had the smallest global biases in downward shortwave radiative flux according to Zhang et al. (2016).

Annual average upward shortwave radiative flux at the surface is overestimated by all reanalysis products, with biases ranging from 1.0 Wm^{-2} (ERA-Interim) to 3.6 Wm^{-2} (NCEP2), corresponding to 10 to 28.6% of average upward shortwave radiative flux values (see Table 2). These biases are not significant, however, because they are all within the bounds of uncertainty of the satellite estimates. Biases in upward shortwave radiative flux in the reanalysis data set over the NE Pacific are largely attributable to the biases in downward shortwave radiative flux, since the domain of interest is mostly ocean surface (little to no change in surface albedo). Despite biases, correlation coefficients for both upward and downward shortwave radiative fluxes are 0.97 or higher for all reanalysis data sets. This is not surprising, given that the seasonal cycle of these variables in each data set is adequate and dominates the correlation (Zhang et al., 2016). Yang et al. (1999) evaluated the older versions of the ECMWF and NCEP reanalysis products, ERA-40 and NCEP1, respectively, and found consistent overestimation of reflected shortwave radiative flux throughout the globe compared to Earth Radiation Budget Experiment. Though the Yang et al. (1999) study is not directly comparable to results shown here, due to the difference in observations used, reanalysis products, time periods (Yang et al., 1999 used years 1985 and 1986), and analysis location (they looked at the top of the atmosphere), this does highlight a consistent problem in overestimation of upward shortwave radiative flux across reanalyses.

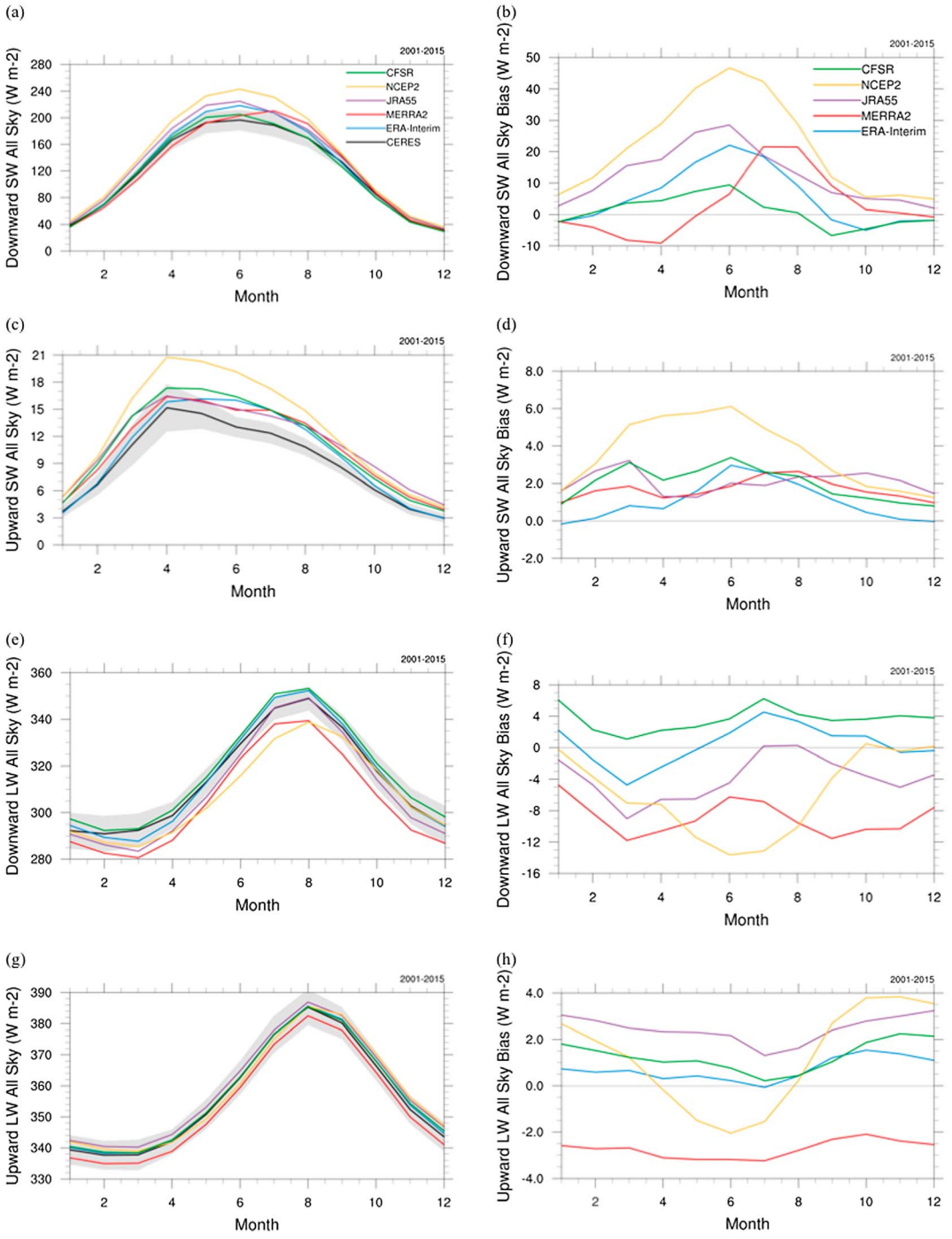

Figure 1. (left column) Monthly climatology and (right column) bias of radiative fluxes. The black line in plots (a), (c), (e), and (g) represents Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) climatology, and the gray shading is standard deviation of CERES EBAF variables. The green line is Climate Forecast System Reanalysis, the yellow line is NCEP/DOE R2, the purple line is JRA55, the red line is MERRA2, and the blue line is ERA-Interim in all plots. (a and b) Downward shortwave, (c and d) upward shortwave, (e and f) downward longwave, and (g and h) upward longwave radiative flux.

The reanalysis data sets have annual average biases in downward longwave radiative flux ranging from −8.9 (MERRA2) to 3.6 Wm^{-2} (CFSR; Figures 1e and 1f and Table 2). Downward longwave flux biases from ERA-Interim, JRA-55, and CFSR are in reasonable agreement with the satellite estimates, while biases in MERRA2 and NCEP2 are significant. Few studies have investigated reanalysis bias of downward longwave radiative

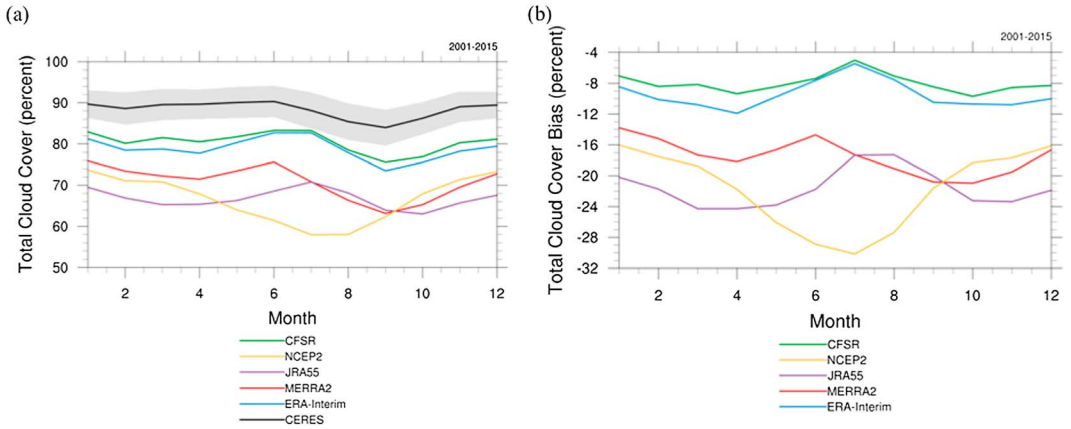

Figure 2. (left column) Monthly climatology and (right column) bias of total cloud cover. The black line in plot (a) represents Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) climatology, and the gray shading is standard deviation of CERES EBAF variables. The green line is Climate Forecast System Reanalysis, the yellow line is NCEP/DOE R2, the purple line is JRA55, the red line is MERRA2, and the blue line is ERA-Interim in all plots.

flux, and we are not aware of any studies that provide a detailed analysis of downward longwave radiative flux in the NE Pacific specifically. Wang and Dickinson (2013) evaluated downward longwave radiative flux from reanalysis against in situ measurements and found high correlation coefficients (0.96–0.98) and globally averaged negative biases of -0.19 Wm^{-2} (CFSR), -2.31 (ERA-Interim), and -14.51 (MERRA). The biases in MERRA2 in the NE Pacific reported here are similar to those Wang and Dickinson (2013) report for MERRA globally. The contour maps in Figure 16 in Wang and Dickinson (2013) show biases between CERES SYN satellite estimates and reanalysis in the NE Pacific region that are very similar to those reported here. Biases in longwave radiative flux are most likely attributable to difficulty in accurately modeling clouds (Wang & Dickinson, 2013; Yang et al., 1999) but could also be attributed to differences in temperature and water vapor profiles.

Reanalysis biases in upward longwave radiative flux are shown in Figures 1g and 1h. ERA-Interim, JRA55, and CFSR all show positive biases throughout the year, with slightly smaller biases in summer compared to winter. MERRA2 shows consistently negative biases in upward longwave radiative flux throughout the year. NCEP2 has positive biases in the winter and negative biases in the summer. However, since all biases are smaller than uncertainties in the satellite estimates, the biases are not significant. Since the reanalysis products (with the exception of CFSR) do not interactively assimilate sea surface temperatures (in other words, reanalysis products read in SST observations which stay fixed and do not change based on radiation and cloud changes in the reanalysis model), the biases in upward longwave radiative flux are likely a combination of differences in SST and differences in downward longwave radiative flux reaching the surface. Differences in upward longwave radiative flux in CFSR could be due in part to differences in sea surface temperatures compared to CERES (CFSR has a SST bias of 0.22 K, not shown).

Figure 2a shows monthly climatologies of total cloud cover from satellite estimates and reanalysis data, with the corresponding biases in Figure 2b. All reanalysis products underestimate total cloud cover, with multiyear mean negative biases ranging from -8% (CFSR) to -21.7% (NCEP2) total cloud fraction. Cloud biases in all reanalysis products are significant. NCEP2 most grossly underestimates total cloud cover, sometimes by ~30%, particularly in the summer months, which is likely part of the reason for the large positive biases in downward shortwave radiative flux in NCEP2. While cloud fraction bias and shortwave bias tend to be negatively correlated in the reanalyses, sometimes the relationship is weaker than expected, suggesting that other factors like different in cloud optical thickness, cloud height, or, possibly, aerosols also contribute (see supplemental materials).

There are no studies that evaluate total cloud fraction in the NE Pacific in particular; however, many other studies point out discrepancies in cloud fraction estimates by reanalysis products. Arguably, the most comparable studies to our region of interest analyze the Arctic. Walsh et al. (2009) find that NCEP1 and JRA25

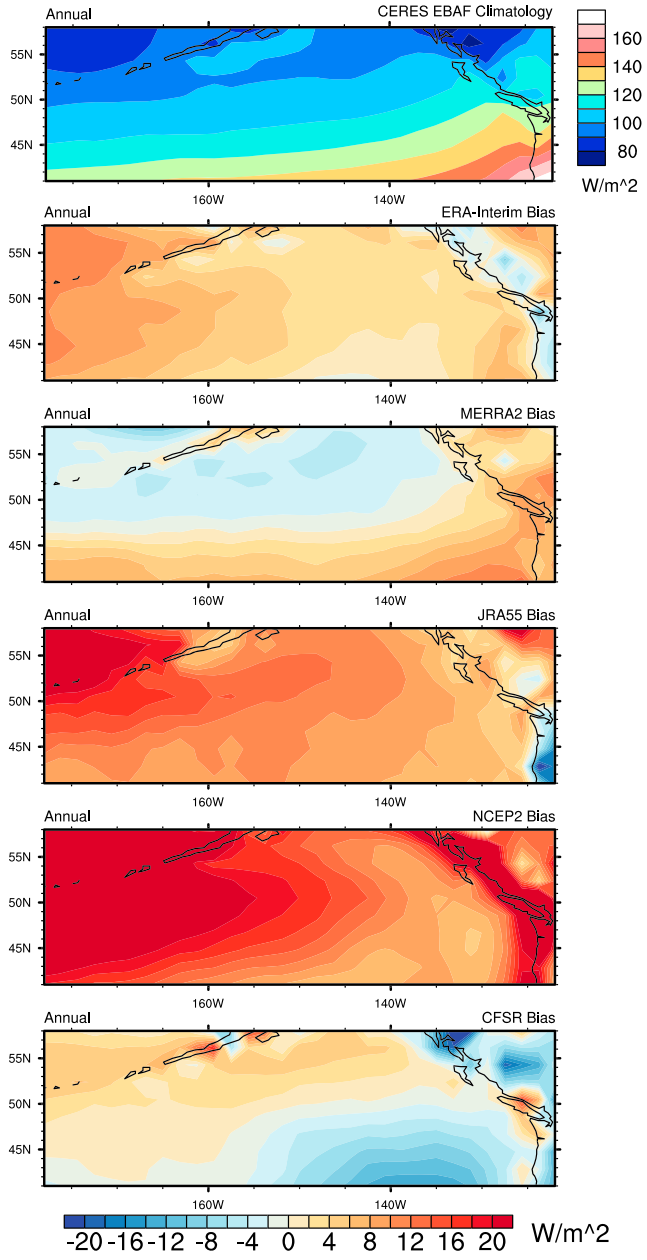

Figure 3. (top) Annual mean climatology of net shortwave radiative flux from Clouds and the Earth's Radiant Energy System Energy Balanced and Filled satellite estimates in the NE Pacific. Subsequent panels show spatial distribution of biases in reanalysis data set net shortwave radiative flux for ERA-Interim, MERRA2, JRA55, NCEP2, and CFSR, respectively.

consistently underestimate cloud fraction at the Barrow, Alaska site, while ERA-40 represents cloud fraction better; furthermore, they note that radiative flues are well represented in reanalysis data sets if/when the cloud fraction is also well represented.

4.2. Spatial Distribution of Radiative Fluxes and Total Cloud Fraction

Evaluation of the domain-averaged parameter climatologies in section 4.1 obscures spatial variations in the satellite measurements and reanalyses, which we explore in this section. For brevity, we discuss only spatial distribution of biases in net shortwave and net longwave radiative fluxes at the surface. It is useful to keep in mind that the climatological biases discussed in section 4.1 show that net shortwave biases are dominated by biases in downward shortwave radiative flux and net longwave biases are generally dominated by biases in downward longwave radiative flux. There is very little spatial variation in total cloud cover biases so we do not discuss them further.

Most of the reanalysis products analyzed here exhibit an east-west gradient in radiative flux biases. The top panel of Figure 3 shows the annual mean climatological distribution of CERES EBAF net shortwave radiative flux, while the following panels of Figure 3 show the long-term mean annual differences between the satellite estimates and the reanalysis products. ERA-Interim, JRA55, NCEP2, and CFSR all show larger positive net shortwave biases in the west portion of the domain, with smaller positive or even negative (CFSR) net shortwave biases in the eastern portion of the domain nearer to the coast. MERRA2 is the only reanalysis product that has a north-south gradient in net shortwave radiative flux bias, with positive biases in the south, corresponding to larger average climatological net shortwave fluxes at the surface. Betts et al. (2006) compared NCEP2 and ERA-40 reanalysis data to ISLSCP-II data from 1986 to 1995 and present spatial maps of biases in their Figures 10 and 11 for downward shortwave radiative flux. Not much spatial variation is apparent in the NE Pacific in winter, though summer shows higher downward shortwave biases nearest the coast for ERA-40 reanalysis (which is different than what is found here for ERA-Interim) and highest biases off the coast of Washington and Oregon and then south of the Gulf of Alaska for NCEP2. Large biases are also apparent in NCEP2 near the coast here, though Betts et al. (2006) do not report the large biases also seen here in the western part of the NE Pacific domain. Lack of agreement in Betts et al. (2006) results and those shown here could be due to a variety of factors, including the different time periods analyzed, the different observations used for comparison, or the different versions of the ECMWF reanalysis products evaluated.

Patterns in the spatial distribution of net longwave radiative flux biases in reanalysis data sets are less consistent (Figure 4). ERA-Interim has positive biases in the open ocean in the south of the domain and negative biases in the north and along the coast. MERRA2 has negative biases throughout the domain, with largest biases in the zone of coastal upwelling. JRA55 has little spatial variation, although largest negative biases occur to the north. NCEP2 also has large negative biases in the north and along the coast. Finally, CFSR biases vary from west to east, with larger positive biases for the ocean in the west and negative biases occurring along the coast. Figure 14 in Betts et al. (2006) show that the largest biases in summertime downward longwave radiative flux in the NE Pacific occur near the coast for both NCEP2 and ERA-40. These results are not consistent with those presented here, likely for similar reasons mentioned previously.

Another subregion of interest within the domain is the mountain chain along the coast, ranging from Oregon and Washington up into British Columbia. There are clearly larger biases along these mountain ranges

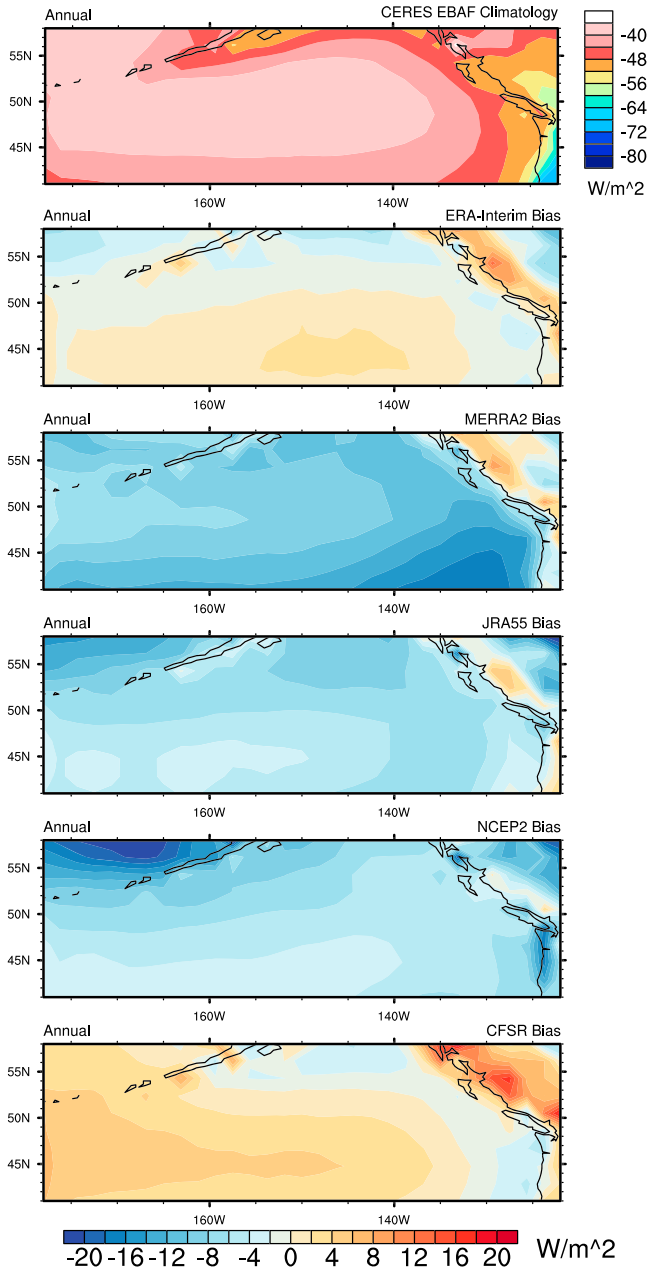

Figure 4. (top) Annual mean climatology of net longwave radiative flux from Clouds and the Earth's Radiant Energy System Energy Balanced and Filled satellite estimates in the NE Pacific. Subsequent panels show spatial distribution of biases in reanalysis data set net longwave radiative flux for ERA-Interim, MERRA2, JRA55, NCEP2, and CFSR, respectively.

compared to surrounding areas. Biases in reanalyses at higher altitudes are documented in other studies (e.g., Dulière et al., 2011) so will not be discussed in detail here.

4.3. Anomalies in Radiative Fluxes and Total Cloud Fraction

Just as global reanalysis evaluations do not supply detail of reanalysis performance at small spatial scales, annual climatological reanalysis evaluations do not supply detail of reanalysis performance on short temporal scales. Anomalous events are of particular interest to climate scientists as they provide insight into the climate and interactive processes when the system is pushed toward extremes. For the reasons mentioned in the introduction, reanalysis products can be important tools for studying anomalous climatic events, but it is essential to understand the performance of reanalysis models during the anomalous periods. Some studies have evaluated representation of temperature and precipitation extremes in reanalysis products (Donat et al., 2014; Dulière et al., 2011), though we are unaware of studies that evaluate anomalies in radiative fluxes and cloud fraction.

Time series of anomalies of radiative parameters and total cloud fraction from CERES EBAF satellite estimates are compared to those from the reanalysis data, spatially averaged over 40°–60°N and 180°–120°W. Table 3 presents statistics both for comparison of all anomalies, as well as comparison of just anomalous events that exceed 1 standard deviation (σ) of the mean.

Figure 5a shows the time series of standardized anomalies of downward shortwave radiative flux from 2001 to 2015 for CERES EBAF and the five reanalysis data sets. On average, the mean absolute deviations of the anomalies range from 2.2 Wm^{-2} (CFSR) to 3.5 Wm^{-2} (MERRA2; Table 3), which are all small enough to be in acceptable agreement with the satellite estimates. For anomalous events that exceed 1σ, those mean absolute deviations are much larger, ranging from 6.3 Wm^{-2} (CFSR) to 34.1 Wm^{-2} (NCEP2). The deviations from ERA-Interim, JRA-55, and NCEP2 are significant. The time series in Figure 5a make it apparent that CFSR best captures the anomalous changes in downward shortwave radiative flux that are documented in the CERES EBAF data sets, particularly in the last few years of the time period. The anomaly comparison statistics indicate that on the whole, the reanalysis products are not reproducing extreme anomalous events (≥1σ) well, even when they are adequately reproducing the climatology of downward shortwave radiative flux in the NE Pacific.

Figure 5b shows the time series of standardized anomalies of downward longwave radiative flux from 2001 to 2015 for CERES EBAF and the five reanalysis data sets. On average the mean absolute deviations of anomalies range from 1.8 Wm^{-2} (ERA-Interim) to 2.8 Wm^{-2} (CFSR). These deviations are not significant. For anomalous events that exceed 1σ, those mean absolute deviations are larger, ranging from 2.8 Wm^{-2} (ERA-Interim) to 9.3 Wm^{-2} (MERRA2), again suggesting that reanalyses do less well in capturing the magnitudes of large anomalies compared to smaller anomalies. The deviations in MERRA2, JRA-55, and NCEP2 are significant. The larger deviations for larger anomalies are likely at least in part due to the larger magnitudes associated with more extreme anomalous events.

For brevity, anomaly biases for total cloud fraction are not discussed here, though comparison statistics are presented in Table 2. All cloud fraction biases and MADs for anomalous events exceeding 1σ are significant.

Table 3
Statistical Comparisons of Radiative Flux and Total Cloud Fraction Monthly Mean Anomalies From Reanalysis Data and CERES EBAF Satellite Estimates, Spatially Averaged Over the Domain From 40–60°N and 180–240°E

Data product	Measures	Downward shortwave radiative flux (Wm^{-2})	Upward shortwave radiative flux (Wm^{-2})	Downward longwave radiative flux (Wm^{-2})	Upward longwave radiative flux (Wm^{-2})	Total cloud fraction (%)
ERA-Interim	MAD	2.6	0.28	1.8	0.51	1.4
	RMSD	3.5	0.39	2.5	0.71	1.7
	Correlation coeff.	0.62	0.76	0.83	0.97	0.56
	Bias (≥1σ (MAD)	**12.9 (13.5)**	1.0 (1.1)	−0.86 (2.8)	0.46 (0.94)	**−9.6 (9.6)**
	RMSD (≥1σ)	15.6	1.4	3.6	1.1	9.9
	Correlation coeff. (≥1σ)	0.99	0.97	0.99	0.99	0.73
MERRA2	MAD	3.5	0.27	1.9	0.48	1.8
	RMSD	4.9	0.36	2.5	0.62	2.2
	Correlation coeff.	0.48	0.82	0.80	0.98	0.37
	Bias (≥1σ) (MAD)	2.4 (9.4)	**1.5 (1.5)**	**−9.3 (9.3)**	−2.9 (2.9)	**−17.4 (17.4)**
	RMSD (≥1σ)	11.6	1.6	9.8	3.1	17.6
	Correlation coeff. (≥1σ)	0.96	0.98	0.98	0.99	0.75
JRA-55	MAD	2.4	0.26	1.8	0.27	1.5
	RMSD	3.2	0.36	2.5	0.36	2.0
	Correlation coeff.	0.67	0.81	0.82	0.99	0.47
	Bias (≥1σ) (MAD)	**20.3 (20.3)**	**2.0 (2.0)**	**−5.1 (5.2)**	2.4 (2.4)	**−22.1 (22.1)**
	RMSD (≥1σ)	21.8	2.2	6.1	2.4	22.3
	Correlation coeff. (≥1σ)	0.99	0.95	0.99	0.99	0.56
NCEP2	MAD	3.1	0.51	2.1	0.37	1.7
	RMSD	4.4	0.81	2.8	0.49	2.2
	Correlation coeff.	0.40	0.83	0.75	0.99	0.45
	Bias (≥1σ) (MAD)	**34.1 (34.1)**	**5.1 (5.1)**	**−4.4 (5.4)**	1.0 (2.1)	**−21.6 (21.6)**
	RMSD (≥1σ)	36.4	5.4	7.3	2.4	22.3
	Correlation coeff. (≥1σ)	0.98	0.96	0.95	0.99	0.43
CFSR	MAD	2.2	0.38	2.8	0.36	1.4
	RMSD	2.9	0.52	3.4	0.46	1.8
	Correlation coeff.	0.82	0.78	0.60	0.99	0.75
	Bias (≥1σ) (MAD)	5.1 (6.3)	**2.5 (2.5)**	1.5 (3.5)	1.2 (1.3)	**−8.2 (8.2)**
	RMSD (≥1σ)	7.2	2.7	4.2	1.5	8.5
	Correlation coeff. (≥1σ)	0.99	0.97	0.98	0.99	0.89

Note. Statistics include mean absolute deviation (MAD), root mean square deviation (RMSD), and correlation coefficient comparing all anomalies in CERES data and reanalysis data, as well as bias, mean absolute deviation (MAD), RMSD and correlation coefficient for anomalous events exceeding 1 standard deviation (≥1σ). Biases and MADs are bolded if they are larger than the uncertainty in satellite estimates, meaning bias is significant.

These biases in radiative fluxes and cloud fraction have implications for the use of reanalysis products when studying anomalous events. For example, reanalysis products are not informative tools to use to study a climate extreme event in which the signal falls within the biases presented here.

5. Discussion

Our analysis evaluates radiative fluxes and total cloud fraction from reanalysis models relative to CERES EBAF satellite estimates. One interpretation of these biases is that they provide bounds of uncertainty for parameters in reanalysis data sets. In this case, the uncertainty in regional mean downward shortwave radiative flux can be interpreted as ±7.7 Wm^{-2} (ERA-Interim), ±7.1 Wm^{-2} (MERRA2), ±12.3 Wm^{-2} (JRA-55), ±21.2 Wm^{-2} (NCEP2), and ±3.8 Wm^{-2} (CFSR). Uncertainties for the remaining variables can be specified from the mean absolute deviations presented in Table 2. These uncertainties are similar or greater for anomalous events greater than 1σ (see Table 2). This means that extra caution should be taken when using reanalyses to analyze anomalous events in the climate system in the NE Pacific.

For most of the fields evaluated here, CFSR has the smallest differences compared to satellite estimates. The low biases, particularly in total cloud fraction, found in CFSR suggest the importance of air-ocean-sea ice interactions, specifically how SSTs are important for determination of clouds and radiative fluxes and vice versa, at least in the region of the NE Pacific. Studies have shown that SSTs and low cloud are negatively correlated in the subtropical and midlatitude North Pacific (e.g., Klein et al., 1995; Norris & Leovy, 1994);

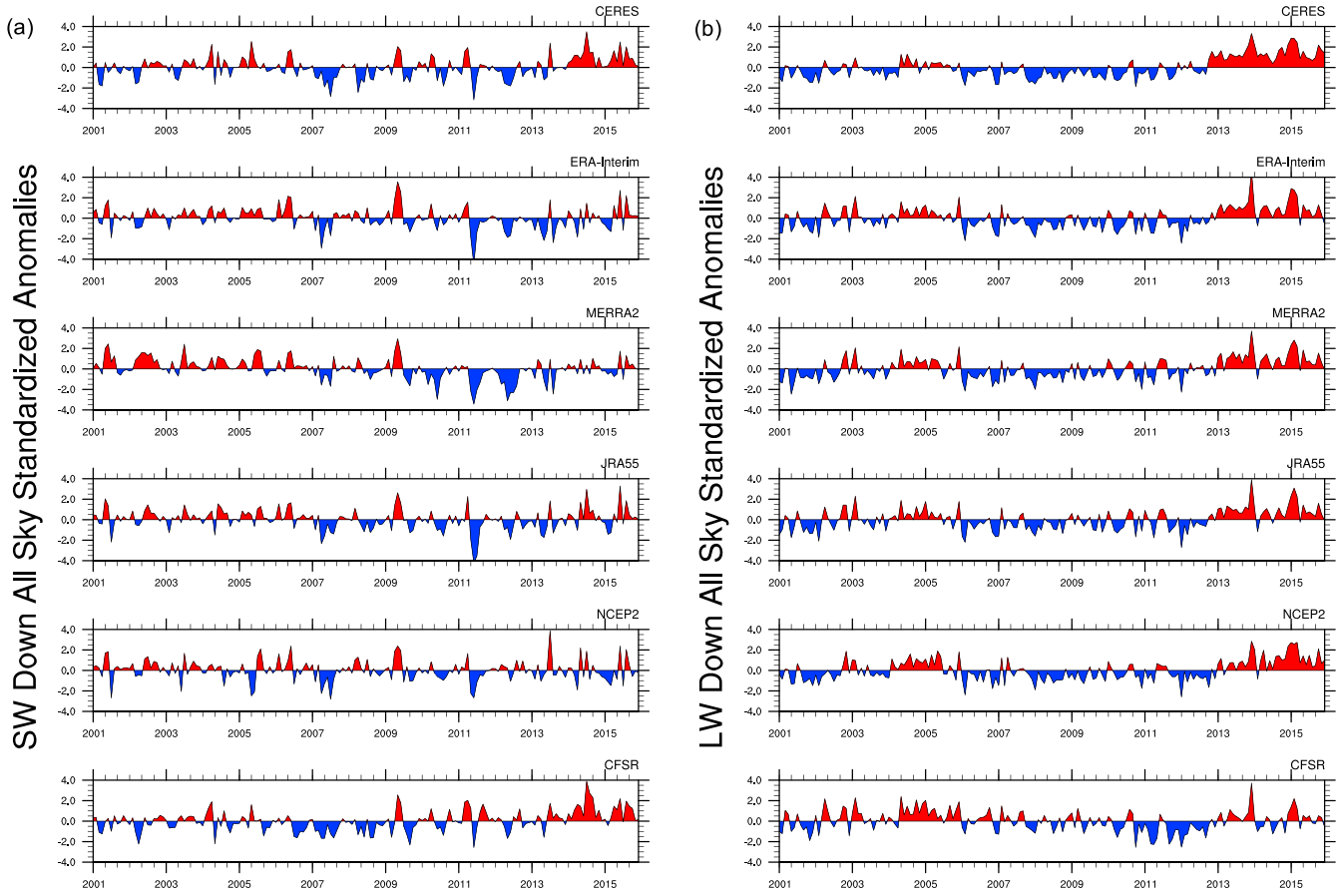

Figure 5. Time series from 2001 to 2015 of monthly standardized anomalies of all-sky downward (a) shortwave radiative flux and (b) longwave radiative flux from Clouds and the Earth's Radiant Energy System, ERA-Interim, MERRA2, JRA55, NCEP2, and CFSR. The positive anomalies are shaded red, and the negative anomalies are shaded blue.

therefore, it makes sense that using a coupled ocean-atmosphere model in this region could be important for replicating observed cloud cover. NCEP2 often has the largest biases in radiative fluxes and total cloud fraction compared to satellite estimates, which suggests very large uncertainties in the reanalysis data. Though CFSR and NCEP2 both come from the same modeling center, the models and resolutions used in each are quite different. Furthermore, the difference in performance between the coupled reanalysis product (CFSR) and noncoupled reanalysis product (NCEP2) is stark. The evidence that the coupled CFSR outperforms NCEP2, and the other noncoupled reanalysis in most variables suggests that the ocean-atmosphere coupling could be partially responsible for improved reproduction of atmospheric variables. This could be especially pertinent in the NE Pacific midlatitudes where atmosphere-ocean interactions are important, and coupled models should be utilized when possible.

Disregarding data set biases when using reanalysis products can produce substantial problems. Consider an analysis of the ocean surface radiative budget in the NE Pacific using data from the poorest performing reanalysis product, NCEP2, in the summer months, when NCEP2 typically has its largest biases. With summer downward shortwave biases on the order of 40 Wm^{-2}, upward shortwave biases on the order of 6 Wm^{-2}, downward longwave biases on the order of -12 Wm^{-2}, and upward longwave biases on the order of -2 Wm^{-2}, the net radiative flux biases would be on the order of 32 Wm^{-2}. Assuming a mixed layer ocean depth of 10 m (a reasonable summer mixed layer depth in this region), these net radiation biases would mean that estimations of temperature for 1 m^2 of the ocean mixed layer would be off by ~2.1 °C (~3.8 °F) over 1 month. This "worst case scenario" of SST error is well outside of typical standard deviation of SSTs in this region.

Our analysis also highlights the need for more in-depth evaluation of reanalysis products before their application to climate studies. A thorough understanding of reanalysis performance is necessary in order to (a) choose the reanalysis product most appropriate for the region and variables of interest and (b) quantify uncertainties in results reached by using these reanalyses. Though there are some in-depth, observationally based, regional evaluations of reanalyses (e.g., Eastern Himalayas, Jin-Huan et al., 2014; North Atlantic, Josey, 2001; Josey et al., 2002; tropical Pacific, Newman et al., 2000; Kumar & Hu, 2012; China, Ma et al., 2008; Antarctica, Yu et al., 2010; Ming et al., 2013), more projects that systematically evaluate and compare reanalysis products for many regions around the globe would help address this need.

Our study suggests that there is no one "gold standard" reanalysis that is best for all climate studies. The most appropriate reanalysis product depends on the atmospheric variables of interest, the subregion of the NE Pacific domain that is being studied, the time period of interest, and whether the reanalysis products are needed to represent climatology or anomalies. Generally, reanalysis products perform best at large spatial and temporal scales, while the biases get much larger when looking at smaller spatial scales and subannual time scales. This study, for example, shows that for the region of the NE Pacific, CFSR may be the most appropriate reanalysis product to use if studying downward shortwave radiative flux at the surface, since this data set has the smallest biases (Table 2). On the other hand, ERA-Interim has the smallest biases for downward longwave radiative flux at the surface, which means the ECMWF product may be best for studies that are particularly sensitive to longwave radiative flux accuracy (Table 2). A study of the coastal NE Pacific might avoid using NCEP2 and MERRA2, which show particularly large biases in shortwave and longwave radiative fluxes near the coast compared to the open ocean (Figures 3 and 4). Furthermore, studies focused on anomalous events might choose to use CFSR data, which has the smallest biases during anomalous events defined to be outside 1σ of climatology (Figure 5).

6. Conclusions

This study evaluates the quality of radiative fluxes and total cloud fraction from five reanalysis products compared to CERES EBAF satellite estimates in the NE Pacific (40°–60°N, 180°–120°W). The analysis quantifies uncertainties in the variables that can help guide appropriate use of reanalyses in this region. Biases in climatological annual cycles of downward shortwave radiation ranged from 0.86 Wm^{-2} (CFSR) to 21.2 Wm^{-2} (NCEP2), while biases in climatological mean of upward shortwave radiative flux ranged from 1.0 Wm^{-2} (ERA-Interim) to 3.6 Wm^{-2}. Biases in climatological annual cycles of downward longwave radiation ranged from -8.9 Wm^{-2} (MERRA2) to 3.6 Wm^{-2} (CFSR), while biases in climatological annual cycles of upward longwave radiation ranged from -2.7 Wm^{-2} (MERRA2) to 2.5 Wm^{-2} (JRA-55). In all cases, biases between reanalysis and satellite estimates are larger during anomalous events that exceed 1σ compared to all anomalous deviations. Furthermore, deviations during anomalous events that exceed 1σ are similar to or much greater than climatological deviations. These results indicate that there are limitations to application of reanalysis products. Using reanalysis data to study climate processes with a signal within the bounds of errors cited in this manuscript may not be useful. Caution should be taken to fully understand the uncertainties within reanalysis products before applying them to climate analysis, since results presented here show large variability in performance between reanalyses, variables within the data sets, and subregions of the study domain.

Acknowledgments
The lead author was funded by the National Science Foundation IGERT Program on Ocean Change at the University of Washington. The authors have no real or perceived financial conflicts of interest. The senior author acknowledges institutional support from JISAO. The authors are grateful to the reviewers for providing comments that enhanced the quality of this paper. All data used in this study are available publicly. CERES EBAF satellite estimates can be downloaded from the NASA CERES website at https://ceres.larc.nasa.gov/order_data.php. ECMWF ERA-Interim reanalysis is available from the ECMWF website at https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim. MERRA2 reanalysis data sets are available at https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/data_access/. JRA-55 reanalysis data can be downloaded from http://jra.kishou.go.jp/JRA-55/index_en.html. NCEP2 reanalysis data are available from NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, and can be downloaded from https://www.esrl.noaa.gov/psd/data. Finally, CFSR data are available from https://rda.ucar.edu/pub/cfsr.html and https://rda.ucar.edu/datasets/ds094.2/ (for CFSRv2).

References

Arakawa, O., & Kitoh, A. (2004). Comparison of local precipitation-SST relationship between the observation and a reanalysis dataset. *Geophysical Research Letters*, *31*, L12206. https://doi.org/10.1029/2004GL020283

Bao, X., & Zhang, F. (2013). Evaluation of NCEP-CFSR, NCEP-NCAR, ERA-Interim, and ERA-40 reanalysis datasets against independent sounding observations over the Tibetan Plateau. *Journal of Climate*, *26*(1), 206–214. https://doi.org/10.1175/JCLI-D-12-00056.1

Betts, A. K., Viterbo, P., Beljaars, A., Pan, H. L., Hong, S. Y., Goulden, M., & Wofsy, S. (1998). Evaluation of land-surface interaction in ECMWF and NCEP/NCAR reanalysis models over grassland (FIFE) and boreal forest (BOREAS). *Journal of Geophysical Research*, *103*(D18), 23,079–23,085. https://doi.org/10.1029/98JD02023

Betts, A. K., Zhao, M., Dirmeyer, P. A., & Beljaars, A. C. M. (2006). Comparison of ERA40 and NCEP/DOE near-surface data sets with other ISLSCP-II data sets. *Journal of Geophysical Research*, *111*, D22S04. https://doi.org/10.1029/2006JD007174

Bond, N. A., Cronin, M. F., Freeland, H., & Mantua, N. (2015). Causes and impacts of the 2014 warm anomaly in the NE Pacific. *Geophysical Research Letters*, *42*(9), 3414–3420. https://doi.org/10.1002/2015GL063306

Bosilovich, M., Akella, S., Coy, L., Cullather, R., Draper, C., Gelaro, R., et al. (2015). MERRA-2: Initial evaluation of the climate. NASA Tech. Rep. Series on Global Modeling and Data Assimilation (NASA/TM-2015-104606, Vol. 43). Greenbelt, MD: NASA Goddard Space Flight Center.

Bosilovich, M. G., Chen, J., Robertson, F. R., & Adler, R. F. (2008). Evaluation of global precipitation in reanalyses. *Journal of Applied Meteorology and Climatology*, *47*(9), 2279–2299. https://doi.org/10.1175/2008JAMC1921.1

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

Decker, M., Brunke, M. A., Wang, Z., Sakaguchi, K., Zeng, X., & Bosilovich, M. G. (2012). Evaluation of the reanalysis products from GSFC, NCEP, and ECMWF using flux tower observations. *Journal of Climate*, *25*(6), 1916–1944. https://doi.org/10.1175/JCLI-D-11-00004.1

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

Donat, M. G., Sillmann, J., Wild, S., Alexander, L. V., Lippmann, T., & Zwiers, F. W. (2014). Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis datasets. *Journal of Climate*, *27*(13), 5019–5035. https://doi.org/10.1175/JCLI-D-13-00405.1

Dulière, V., Zhang, Y., & Salathé, E. P. Jr. (2011). Extreme precipitation and temperature over the US Pacific Northwest: A comparison between observations, reanalysis data, and regional models. *Journal of Climate*, *24*(7), 1950–1964. https://doi.org/10.1175/2010JCLI3224.1

Fujiwara, M., Wright, J. S., Manney, G. L., Gray, L. J., Anstey, J., Birner, T., et al. (2017). Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems. *Atmospheric Chemistry and Physics*, *17*(2), 1417–1452. https://doi.org/10.5194/acp-17-1417-2017

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Jakobson, E., Vihma, T., Palo, T., Jakobson, L., Keernik, H., & Jaagus, J. (2012). Validation of atmospheric reanalyses over the Central Arctic Ocean. *Geophysical Research Letters*, *39*, L10802. https://doi.org/10.1029/2012GL051591

Jiang, J. H., Su, H., Zhai, C., Wu, L., Minschwaner, K., Molod, A. M., & Tompkins, A. M. (2015). An assessment of upper-troposphere and lower-stratosphere water vapor in MERRA, MERRA2 and ECMWF reanalyses using Aura MLS observations. *Journal of Geophysical Research: Atmospheres*, *120*, 11,468–11,485. https://doi.org/10.1002/2015JD023752

Jin-Huan, Z., Shu-Po, M., Han, Z., Li-Bo, Z., & Peng, L. (2014). Evaluation of reanalysis products with in situ GPS sounding observations in the Eastern Himalayas. *Atmospheric and Oceanic Science Letters*, *7*(1), 17–22. https://doi.org/10.3878/j.issn.1674-2834.13.0050

Josey, S. A. (2001). A comparison of ECMWF, NCEP-NCAR, and SOC heat fluxes with moored buoy measurements in the subduction region of the Northeast Atlantic. *Journal of Climate*, *13*, 1780–1789. https://doi.org/10.1175/1520-0442(2001)014%3C1780:ACOENN%3E2.0.CO;2

Josey, S. A., Kent, E. C., & Taylor, P. K. (2002). Wind stress forcing of the ocean in the SOC climatology: Comparisons with the NCEP-NCAR, ECMWF, UWM/COADS, and Hellerman and Rosenstein datasets. *Journal of Physical Oceanography*, *32*(7), 1993–2019. https://doi.org/10.1175/1520-0485(2002)032%3C1993:WSFOTO%3E2.0.CO;2

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, *77*(3), 437–471. https://doi.org/10.1175/1520-0477(1996)077%3C0437:TNYRP%3E2.0.CO;2

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S. K., Hnilo, J. J., Fiorino, M., & Potter, G. L. (2002). NCEP-DOE AMIP-II reanalysis (R-2). *Bulletin of the American Meteorological Society*, *83*(11), 1631–1644. https://doi.org/10.1175/BAMS-83-11-1631

Kato, S., Loeb, N. G., Rose, F. G., Doelling, D. R., Rutan, D. A., & Caldwell, T. E. (2012). Modeled surface irradiances consistent with CERES-derived top-of-atmosphere shortwave and longwave irradiances. *Journal of Climate*, *26*, 2719–2740. https://doi.org/10.1175/JCLI-D-12-00436.1

Kato, S., Loeb, N. G., Rose, F. G., Doelling, D. R., Rutan, D. A., Caldwell, T. E., et al. (2013). Surface irradiances consistent with CERES-derived top-of-atmosphere shortwave and longwave irradiances. *Journal of Climate*, *26*(9), 2719–2740. https://doi.org/10.1175/JCLI-D-12-00436.1

Kato, S., Loeb, N. G., Rutan, D. A., Rose, F. G., Sun-Mack, S., Miller, W. F., & Chen, Y. (2012). Uncertainty estimate of surface irradiances computed with MODIS-, CALIPSO-, and CloudSat-derived cloud and aerosol properties. *Surveys in Geophysics*, *33*(3–4), 395–412. https://doi.org/10.1007/s10712-012-9179-x

Kato, S., Rose, F. G., Rutan, D. A., Thorsen, T. J., Loeb, N. G., Doelling, D. R., et al. (2018). Surface irradiances of edition 4.0 Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) data product. *Journal of Climate*, *31*(11), 4501–4527. https://doi.org/10.1175/JCLI-D-17-0523.1

Klein, S. A., Hartmann, D. L., & Norris, J. R. (1995). On the relationships among low-cloud structure, sea surface temperature, and atmospheric circulation in the summertime Northeast Pacific. *Journal of Climate*, *8*(5), 1140–1155. https://doi.org/10.1175/1520-0442(1995)008%3C1140:OTRALC%3E2.0.CO;2

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., et al. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, *93*(1), 5–48. https://doi.org/10.2151/jmsj/2015-001

Kumar, A., & Hu, Z.-Z. (2012). Uncertainty in the ocean-atmosphere feedbacks associated with ENSO in the reanalysis products. *Climate Dynamics*, *39*(3-4), 575–588. https://doi.org/10.1007/s00382-011-1104-3

Ladd, C., & Bond, N. A. (2002). Evaluation of the NCEP/NCAR reanalysis in the NE Pacific and the Bering Sea. *Journal of Geophysical Research*, *107*(C10), 3158. https://doi.org/10.1029/2001JC001157

Lindsay, R., Wensnahan, M., Schweiger, A., & Zhang, J. (2014). Evaluation of seven different atmospheric reanalysis products in the Arctic. *Journal of Climate*, *27*(7), 2588–2606. https://doi.org/10.1175/JCLI-D-13-00014.1

Loeb, N. G., Wielicki, B. A., Doelling, D. R., Smith, G. L., Keyes, D. F., Kato, S., et al. (2009). Toward optimal closure of the Earth's top-of-atmosphere radiation budget. *Journal of Climate*, *22*(3), 748–766. https://doi.org/10.1175/2008JCLI2637.1

Lüpkes, C., Vihma, T., Jakobson, E., König-Langlo, G., & Tetzlaff, A. (2010). Meteorological observations from ship cruises during summer to the Central Arctic: A comparison with reanalysis data. *Geophysical Research Letters*, *37*, L09810. https://doi.org/10.1029/2010GL042724

Ma, L., Zhang, T., Li, Q., Frauenfeld, O. W., & Qin, D. (2008). Evaluation of ERA-40, NCEP-1, and NCEP-2 reanalysis air temperatures with ground-based measurements in China. *Journal of Geophysical Research*, *113*, D15115. https://doi.org/10.1029/2007JD009549

Meindl, E. A., & Hamilton, G. D. (1992). Programs of the National Data Buoy Center. *Bulletin of the American Meteorological Society*, *73*(7), 985–993. https://doi.org/10.1175/1520-0477(1992)073%3C0985:POTNDB%3E2.0.CO;2

Ming, L., Qinghua, Y., Jiechen, Z., Lin, Z., Chunhua, L., & Shang, M. (2013). Evaluation of reanalysis and satellite-based sea surface winds using in situ measurements from Chinese Antarctic expeditions. *Chinese Journal of Polar Research*, *24*(3-English), 147–152. https://doi.org/10.3724/SP.J.1085.2013.00147

Newman, M., Sardeshmukh, P. D., & Bergman, J. W. (2000). An assessment of the NCEP, NASA, and ECMWF reanalyses over the tropical West Pacific warm pool. *Bulletin of the American Meteorological Society*, *81*(1), 41–48. https://doi.org/10.1175/1520-0477(2000)081%3C0041:AAOTNN%3E2.3.CO;2

Norris, J. R., & Leovy, C. B. (1994). Interannual variability in stratiform cloudiness and sea surface temperature. *Journal of Climate*, *7*(12), 1915–1925. https://doi.org/10.1175/1520-0442(1994)007%3C1915:IVISCA%3E2.0.CO;2

Onogi, K., Tsutsui, J., Koide, H., Sakamoto, M., Kobayashi, S., Hatsushika, H., et al. (2007). The JRA-25 reanalysis. *Journal of the Meteorological Society of Japan. Ser. II*, *85*(3), 369–432. https://doi.org/10.2151/jmsj.85.369

Pfeifroth, U., Mueller, R., & Ahrens, B. (2013). Evaluation of satellite-based and reanalysis precipitation data in the tropical Pacific. *Journal of Applied Meteorology and Climatology*, *52*(3), 634–644. https://doi.org/10.1175/JAMC-D-12-049.1

Porter, D. F., Cassano, J. J., & Serreze, M. C. (2011). Analysis of the Arctic atmospheric energy budget in WRF: A comparison with reanalyses and satellite observations. *Journal of Geophysical Research*, *116*, D22108. https://doi.org/10.1029/2011JD016622

Randles, C. A., Da Silva, A. M., Buchard, V., Colarco, P. R., Darmenov, A., Govidaraju, R., et al. (2017). The MERRA-2 aerosol reanalysis, 1980 onward. Part I: System description and data assimilation evaluation. *Journal of Climate*, *30*(17), 6823–6850. https://doi.org/10.1175/JCLI-D-16-0609.1

Rutan, D. A., Kato, S., Doelling, D. R., Rose, F. G., Nguyen, L. T., Caldwell, T. E., & Loeb, N. G. (2015). CERES synoptic product: Methodology and validation of surface radiant flux. *Journal of Atmospheric and Oceanic Technology*, *32*(6), 1121–1143. https://doi.org/10.1175/JTECH-D-14-00165.1

Saha, S., Moorthi, S., Pan, H. L., Wu, X., Wang, J., Nadiga, S., et al. (2010). The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, *91*(8), 1015–1058. https://doi.org/10.1175/2010BAMS3001.1

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., et al. (2014). The NCEP climate forecast system version 2. *Journal of Climate*, *27*(6), 2185–2208. https://doi.org/10.1175/JCLI-D-12-00823.1

Screen, J. A., & Simmonds, I. (2011). Erroneous Arctic temperature trends in the ERA-40 reanalysis: A closer look. *Journal of Climate*, *24*(10), 2620–2627. https://doi.org/10.1175/2010JCLI4054.1

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D., Fiorino, M., et al. (2005). The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, *131*(612), 2961–3012. https://doi.org/10.1256/qj.04.176

Vey, S., Dietrich, R., Rülke, A., Fritsche, M., Steigenberger, P., & Rothacher, M. (2010). Validation of precipitable water vapor within the NCEP/DOE reanalysis using global GPS observations from one decade. *Journal of Climate*, *23*(7), 1675–1695. https://doi.org/10.1175/2009JCLI2787.1

Walsh, J. E., Chapman, W. L., & Portis, D. H. (2009). Arctic cloud fraction and radiative fluxes in atmospheric reanalyses. *Journal of Climate*, *22*(9), 2316–2334. https://doi.org/10.1175/2008JCLI2213.1

Wang, K., & Dickinson, R. E. (2013). Global atmospheric downward longwave radiation at the surface from ground-based observations, satellite retrievals, and reanalyses. *Reviews of Geophysics*, *51*, 150–185. https://doi.org/10.1002/rog.20009

Wielicki, B. A., Barkstrom, B. R., Harrison, E. F., Lee, R. B., Louis Smith, G., & Cooper, J. E. (1996). Clouds and the Earth's Radiant Energy System (CERES): An Earth observing system experiment. *Bulletin of the American Meteorological Society*, *77*, 853–868. https://doi.org/10.1175/1520-0477(1996)077%3C0853:CATERE%3E2.0.CO;2

Yang, S. K., Hou, Y. T., Miller, A. J., & Campana, K. A. (1999). Evaluation of the Earth radiation budget in NCEP-NCAR reanalysis with ERBE. *Journal of Climate*, *12*(2), 477–493. https://doi.org/10.1175/1520-0442(1999)012%3C0477:EOTERB%3E2.0.CO;2

Yu, L., Zhang, Z., Zhou, M., Zhong, S., Lenschow, D., Hsu, H., et al. (2010). Validation of ECMWF and NCEP-NCAR reanalysis data in Antarctica. *Advances in Atmospheric Sciences*, *27*(5), 1151–1168. https://doi.org/10.1007/s00376-010-9140-1

Zhang, T., Stackhouse, P. W. Jr., Gupta, S. K., Cox, S. J., Mikovitz, J. C., & Hinkelman, L. M. (2013). The validation of the GEWEX SRB surface shortwave flux data products using BSRN measurements: A systematic quality control, production and application approach. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *122*, 127–140. https://doi.org/10.1016/j.jqsrt.2012.10.004

Zhang, X., Liang, S., Wang, G., Yao, Y., Jiang, B., & Cheng, J. (2016). Evaluation of the reanalysis surface incident shortwave radiation products from NCEP, ECMWF, GSFC, and JMA using satellite and surface observations. *Remote Sensing*, *8*(3), –225. https://doi.org/10.3390/rs8030225

Zhang, Y. C., Rossow, W. B., & Lacis, A. A. (1995). Calculation of surface and top of atmosphere radiative fluxes from physical quantities based on ISCCP data sets: 1. Method and sensitivity to input data uncertainties. *Journal of Geophysical Research*, *100*(D1), 1149–1165. https://doi.org/10.1029/94JD02747

Zib, B. J., Dong, X., Xi, B., & Kennedy, A. (2012). Evaluation and intercomparison of cloud fraction and radiative fluxes in recent reanalyses over the Arctic using BSRN surface observations. *Journal of Climate*, *25*(7), 2291–2305. https://doi.org/10.1175/JCLI-D-11-00147.1