Observationally based evaluation of NWP reanalyses in modeling cloud properties over the Southern Great Plains

Wei Wu,[1] Yangang Liu,[1] and Alan K. Betts[2]

[1]   This study evaluates three major Numerical-Weather-Prediction reanalyses (ERA-Interim, NCEP/NCAR Reanalysis I, and NCEP/DOE Reanalysis II) in modeling surface relative shortwave cloud forcing, cloud fraction, and cloud albedo. The observations used for this evaluation are decade-long surface-based continuous measurements of the U.S. Atmospheric Radiation Measurement (ARM) program from 03/25/1997 to 12/31/2008 over the Southern Great Plains site. These cloud properties from the reanalyses are evaluated at multiple temporal scales. Like the observations, all the reanalyses show a strong annual cycle, and relatively weak diurnal or inter-annual variations of the cloud properties. The reanalyses exhibit significant underestimation on the cloud properties, and the model biases in the cloud properties in general reveal a linear link to one another and are somewhat related to cloud fraction magnitude. Further examination shows that the cloud properties are strongly related to 2-m relative humidity, especially for the observations and ERA-Interim. However, the relationship between the cloud properties and 2-m temperature and specific humidity is much weaker. Also, the cloud fraction biases in the two NCEP reanalyses increase (decrease) with the relative humidity (temperature and specific humidity), but the cloud fraction biases in ERA-Interim show no (opposite) relationship with the relative humidity (temperature and specific humidity). The relative humidity biases have a positive (negative) linear relationship with the specific humidity (temperature) biases. A combined statistical analysis using the technique of Taylor diagrams and a newly developed metric "Relative Euclidean Distance" indicates that ERA-Interim and NCEP/NCAR reanalyses have the best and worst overall performance in modeling the cloud and meteorological properties examined, respectively, except that NCEP/DOE Reanalysis II ranks the best in modeling the monthly temperature and specific humidity.

1.   Introduction

[2]   Climate prediction depends on modeling, so there is a pressing need to quantify model uncertainties and reduce model biases. Among numerous model uncertainties, the representation of clouds and associated radiative processes has been recognized as one of the major factors in global climate models (GCMs), which limit the accuracy of climate prediction [*Intergovernmental Panel on Climate Change (IPCC)*, 2007]. As a consequence, model evaluations to identify deficiencies in the parameterization of clouds and associated processes remain a field of intensive research.

[3]   To address this long-standing issue, the U.S. Department of Energy's (DOE's) Earth System Modeling program funded a new model evaluation project in 2009: the FAst-physics System TEstbed and Research (FASTER) project. The main objective of this multi-institutional project is to utilize various long-term measurements collected by the DOE's Atmospheric Radiation Measurement (ARM) program over the ARM sites [*Stokes and Schwartz*, 1994; *Ackerman and Stokes*, 2003] to accelerate the evaluation and improvement of the parameterizations of cloud-related fast processes in large-scale climate models. This paper is an initial evaluation of cloud and radiative properties and their links to surface meteorology in three major Numerical-Weather-Prediction (NWP) reanalyses.

[4]   As representations of the state of the atmosphere, reanalyses are generated via a state-of-art analysis and forecast system assimilating data from a wide variety of observations including ships, satellites, ground stations and radar. The long-term, consistent global distributions of reanalyses make them particularly valuable in climate research, and have been widely used as a base-line in studying global climate change and climate modeling [*Lu et al.*, 2005;

[1]Atmospheric Sciences Division, Brookhaven National Laboratory, Upton, New York, USA.
[2]Atmospheric Research, Pittsford, Vermont, USA.

Corresponding author: W. Wu, Atmospheric Sciences Division, Brookhaven National Laboratory, 75 Rutherford Dr., Bldg. 815E, Upton, NY 11973, USA. (wwu@bnl.gov)

Anderson et al., 2008; *Haimberger et al.*, 2008; *Betts et al.*, 2009; *Simmons et al.*, 2010; *Rye et al.*, 2010]. Nonetheless, reanalysis data may have time-varying biases, which limit their value for characterizing long-term climate trends [e.g., *Thorne and Vose*, 2010; *Dee et al.*, 2011a]. Furthermore, cloud observations are not directly assimilated into current reanalyses, and thus cloud-related properties of reanalyses are expected to suffer from problems resulting from deficient model parameterizations, similar to GCMs. Accordingly, observationally based evaluations of widely used reanalyses in modeling cloud and associated radiative properties are crucial to both current climate research and future development of reanalysis.

[5] Observationally based evaluations of reanalyses in modeling cloud and radiative properties have been conducted in previous studies [*Jakob*, 1999; *Allan and Ringer*, 2003; *Chevallier et al.*, 2005; *Betts and Viterbo*, 2005; *Betts et al.*, 2006; *Bedacht et al.*, 2007; *Betts*, 2007; *Weidle and Wernli*, 2008; *Betts et al.*, 2009; *Xu*, 2009]. Most of them are done by using satellite-based observations. Evaluations using ARM surface-based cloud and radiation measurements are limited in both the number of studies and the scope of data evaluated. For example, *Walsh et al.* [2009] evaluated four reanalyses in modeling 1999–2006 monthly Arctic cloud fraction and radiative fluxes by using the ARM surface-based measurements at the North Slope of Alaska of cloud base height from ceilometers and shortwave/longwave flux from sky/ground radiometers. The four reanalyses were from National Centers for Environmental Prediction National Center for Atmospheric Research, 40-yr European Centre for Medium-Range Weather Forecasts Reanalysis, North American Regional Reanalysis, and the Japan Meteorological Agency and Central Research Institute of Electric Power Industry 25-yr Reanalysis. They found that the performance of the reanalyses in modeling radiative fluxes mainly depends on their performance in modeling cloud fraction, and the systematic errors of reanalysis cloud fractions are substantial. For the purpose of investigating the applicability for forcing single-column and cloud-resolving models, *Kennedy et al.* [2011] evaluated reanalysis data over the ARM Southern Great Plains (SGP) site from the Modern-Era Retrospective analysis for Research and Applications (MERRA) and the North American Regional Reanalysis (NARR) using 1999–2001 ARM continuous forcing product and surface-based sounding data from Climate Modeling Best Estimates (CMBE) data [*Xie et al.*, 2010]. They found that ARM continuous forcing and the reanalyses show good agreement with the CMBE sounding data, with biases being 0.5 K, 0.5 m s^{-1}, and 5% for temperature, wind, and relative humidity. However, larger disagreements occur in the upper troposphere for temperature, humidity, and zonal wind, in the boundary layer for meridional wind and humidity. They also found that the phase patterns of the seasonal cloud fraction from the 3-year MERRA and NARR are similar to those from ARM radar–lidar and Geostationary Operational Environmental Satellite, and MERRA has better agreement with ARM observations on surface shortwave and longwave fluxes than NARR.

[6] In this study we evaluate three major NWP reanalyses in modeling surface relative shortwave cloud forcing (see section 2.3.1 for the definition), cloud fraction, and cloud albedo. Multiscale (diurnal, annual and inter-annual) mean variations of the cloud properties are evaluated. We also

examine model-error propagation paths through the investigation of the model biases and their links to one another, and to the 2-m temperature and moisture (relative and specific humidity) fields. Decade-long (1997 to 2008) surface-based continuous ARM value-added products (VAP) over the SGP Central Facility site are used as a standard for the evaluation. The rest of the paper is organized as follows. Section 2 briefly introduces the data and methods used. Section 3 shows the multiscale mean variations of the cloud properties and discusses potential uncertainties. Section 4 analyzes model biases and their links. Section 5 evaluates the overall performance of the reanalyses in modeling the cloud properties and 2-m temperature/humidity. Section 6 summarizes this study.

2. Data and Methods

[7] This section briefly introduces the data (e.g., observations and reanalyses) and methods used in this study. Note that, to be consistent with shortwave radiation observations, this study only evaluates daytime (6 am to 6 pm) cloud properties and corresponding temperature/humidity.

2.1. Observations

[8] The observations used are the high-resolution ARM VAP (15-min SIRS data) derived from the well-calibrated surface-based Solar and Infrared Radiation System (SIRS) over the SGP Central Facility (262.51°E, 36.61°N) from 03/25/1997 to 12/31/2008 (http://www.arm.gov/instruments/sirs). Long and his coworkers [*Long and Ackerman*, 2000; *Long et al.*, 2006] produced a 15-min mean data set of all-sky and clear-sky-fit surface downwelling shortwave (SW) fluxes, and a fractional sky cover ("cloud fraction" hereafter) derived from an analysis of surface measurements of downwelling total and diffuse SW irradiance, carefully screened and corrected for optically thicker overcast cases. Here, the cloud fraction is retrieved based on clear-sky surface downwelling SW total/diffuse irradiance and the effect of clouds on the SW irradiance, estimated by using the methodology developed by *Long and Ackerman* [2000].

[9] The observed cloud fraction represents 15-min averages of 160° field-of-view (FOV) hemispheric fractional sky cover from the surface. *Long et al.* [2006] found good agreement (with root-mean square error <8%) between this estimate and hemispheric sky imager data. By definition this cloud fraction differs from the plane-parallel cloud fraction from a nadir view or from a climate model. *Liu et al.* [2011, Figure 2] found that the hourly averaged cloud fraction from the same 15-min hemispheric-view cloud fraction showed reasonable agreement (with correlation coefficient 0.86) with the hourly cloud fraction from the Geostationary Operational Environmental Satellite (GOES) satellite. More discussion on cloud fraction and associated uncertainty is deferred to Section 3.2.

[10] We then compute surface relative shortwave cloud forcing (SRCF) from the all-sky and clear-sky-fit SW fluxes, and cloud albedo is estimated using both SRCF and cloud fraction. In addition, we also use the 30-min averaged data streams of 2-m air temperature and relative humidity, and barometric pressure from the Surface Meteorological Observation System instruments from 04/01/2001 to 12/31/2008 for analyzing the links of the model biases in cloud properties to the 2-m meteorological conditions.

2.2. Reanalyses

[11] Three NWP reanalyses are evaluated in this study: ERA-Interim, NCEP/NCAR Reanalysis I (R1 hereafter), and NCEP/DOE Reanalysis II (R2 hereafter). The abbreviations "ERA," "NCEP," and "NCAR" denote "European Centre for Medium-Range Weather Forecasts (ECMWF) global atmospheric reanalysis," "National Center for Environmental Prediction," and "National Center for Atmospheric Research", respectively. A brief introduction is given below.

2.2.1. ERA-Interim

[12] ERA-Interim is the latest version of ECMWF's global atmospheric reanalysis available from 1989 to present [*Dee et al.*, 2011b]. This reanalysis was archived in a horizontal resolution of T255 spherical-harmonic representation for the basic dynamical fields or N128 reduced Gaussian grid with approximately uniform 79 km spacing for surface and other grid point fields, with 60 vertical levels (P. Berrisford et al., The ERA-Interim archive, August 2009, http://www.ecmwf.int/publications/library/do/references/list/782009). The global archive has 3-hourly (for surface parameters) or 6-hourly (for upper-air parameters) time resolution, but for selected points, including the ARM SGP site, hourly data were archived and used in this study. The major improvements in ERA-Interim include the representation of hydrological cycle, the quality of stratospheric circulation, and the handling of biases and changes in the observing system. One advantage of this reanalysis is its high spatial and temporal resolutions, better for studying regional diurnal variations. Detailed description about this reanalysis can be found at the ECMWF web site at http://www.ecmwf.int//research/era/do/get/ERA-Interim. We use: 1) hourly averaged clear-sky surface net SW flux, and all-sky surface net and surface downwelling SW fluxes, and total cloud cover, 2) hourly interval (i.e., instantaneous) 2-m air temperature and specific humidity, and surface pressure, from 03/25/1997 to 12/31/2008. The hourly data are the outputs from the first 0–12 h forecasts from twice-daily analysis. All the data used are from a Gaussian grid centered at (262.50°E, 36.84°N) over the ARM SGP site.

[13] Note that, ERA-Interim hourly clear-sky surface downwelling SW flux (SW_{dn}^{clear}) is not available so it is calculated from equation (1) using the available clear-sky surface net SW flux (SW_{net}^{clear}), all-sky surface downwelling SW flux (SW_{dn}^{all}) and all-sky surface net SW flux (SW_{net}^{all}),

$$SW_{dn}^{clear} = \frac{SW_{net}^{clear} \times SW_{dn}^{all}}{SW_{net}^{all}}. \quad (1)$$

[14] Equation (1) is derived based on equation (2) in *Betts et al.* [2009] under the assumption that surface albedo is the same for clear-sky and cloudy-sky conditions. A small number of hourly data points were excluded from the analysis (25 out of the 12-year record) for reasons such as data truncation errors or unrealistically large derived surface albedos.

[15] Note also that, ERA-Interim hourly 2-m relative humidity is not available so it is calculated from the 2-m specific humidity using equation (2) [i.e., *Peixoto and Oort*, 1992, equation (3.62)],

$$\text{rh} = \frac{q}{q_s} = \frac{pq}{0.622 e_s} \quad (2)$$

where rh, q, q_s, p, e_s, and T represent relative humidity, specific humidity (kg/kg), saturation specific humidity (kg/kg), pressure of moist air (hPa), saturation vapor pressure (hPa), and temperature (K) of moist air. The saturation vapor pressure e_s is calculated using equation (3) of saturation vapor pressure, recommended by World Meteorological Organization (http://cires.colorado.edu/~voemel/vp.html),

$$\log_{10}(e_s) = 10.79574 \times \left(1 - \frac{273.16}{T}\right) - 5.02800$$
$$\times \log_{10}\left(\frac{T}{273.16}\right) + 1.50475 \times 10^{-4}$$
$$\times \left(1 - 10^{-8.2969 \times \left(\frac{T}{273.16} - 1\right)}\right) + 0.42873 \times 10^{-3}$$
$$\times \left(10^{4.76955 \times \left(1 - \frac{273.16}{T}\right)} - 1\right) + 0.78614. \quad (3)$$

[16] Modeled pressure in the reanalyses is very tightly constrained by the observations to about 1 hPa: the standard deviation of the difference between the hourly ERA-Interim and the observations equals 0.79 hPa, and that between the 6-hourly R1 (R2) and the observations equals 1.13 (1.28) hPa. Hence the ratio of specific humidity and saturated specific humidity accurately reflects the modeled relative humidity.

2.2.2. NCEP/NCAR Reanalysis I

[17] R1 [*Kalnay et al.*, 1996; *Kistler et al.*, 2001] provides data from 1948 to the present with 6-h temporal resolution. The spatial resolution archived in this reanalysis is T62 Gaussian grid (~210 km), or 2.5° latitude × 2.5° longitude non-Gaussian grid, with 28 vertical levels. This reanalysis was generated by the analysis and forecast system performing data assimilation using a wide variety of weather observations including ships, satellites, ground stations and radar. In this paper, we use: 1) 6-hourly averaged clear-sky and all-sky surface downwelling SW flux, and total cloud cover, 2) 6-hourly interval forecasted 2-m air temperature and specific humidity, and surface pressure from 03/25/1997 to 12/31/2008. Detailed description about the data can be found at http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.other_flux.html. All the data used are from a Gaussian grid centered at (262.50°E, 37.14°N) over the ARM SGP site. Note that the reanalysis forecasted data of 2-m air temperature (or specific humidity) and surface pressure are strongly impacted by the reanalysis model as stated in *Kalnay et al.* [1996]. These data are downloaded from the NOAA's Physical Science Division web site at http://www.esrl.noaa.gov/psd/data/gridded/reanalysis/. Note also that R1 6-hourly 2-m relative humidity is not available so it is calculated using equations (2) and (3).

2.2.3. NCEP/DOE Reanalysis II

[18] R2 is the second version of R1. It covers data from 1979 to present, available at the same web site as R1. R2 is believed to be an improved version of R1, with a number of errors fixed, updated parameterizations of physical processes (including a new SW radiation scheme which significantly reduced surface insolation by about 8%), and the addition of more observations [*Kanamitsu et al.*, 2002]. Similar to R1, we use: 1) R2 6-hourly averaged all-sky surface downwelling SW flux, and total cloud cover, 2) R2 6-hourly interval forecasted 2-m air temperature and specific humidity, and
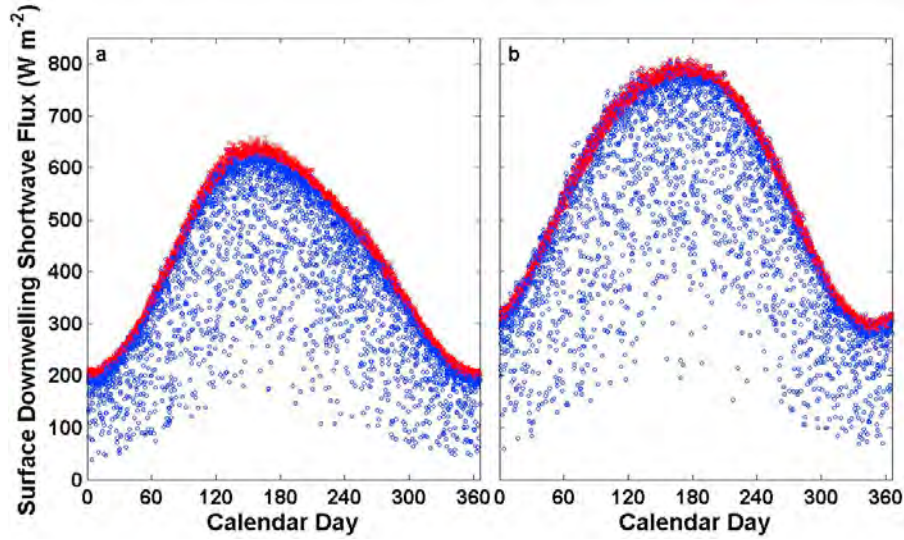
Figure 1. Annual distribution of R2 6-hourly (a) 6 am–12 pm LST and (b) 12 pm-6 pm LST surface downwelling shortwave flux over the ARM SGP Central Facility site from 03/25/1997 to 12/31/2008. The red cross represents R2 clear-sky surface downwelling shortwave flux, estimated by using R1 clear-sky surface downwelling shortwave flux. The blue circle represents available R2 all-sky surface downwelling shortwave flux.

surface pressure from 03/25/1997 to 12/31/2008 over the same Gaussian grid. Because R2 6-hourly 2-m relative humidity is not available, we calculated it using equations (2) and (3). Since the archive does not provide R2 clear-sky surface downwelling SW flux, we estimate it by using R1 clear-sky surface downwelling SW flux and the following expressions for the best curve fit to the upper envelop of the available R2 all-sky surface downwelling SW flux,

$$SW_{dn}^{clear}(t, \text{d}; \text{R2}) = SW_{dn}^{clear}(t, d; \text{R1}) \times \left[0.85 + \frac{(0.92 - 0.85)d}{60}\right]$$

$$\text{(for } d \leq 60) \tag{4a}$$

$$SW_{dn}^{clear}(t, \text{d}; \text{R2}) = SW_{dn}^{clear}(t, d; \text{R1}) \times 0.92$$

$$\text{(for } 60 < d < 120 \text{ or } 220 < d < 280) \tag{4b}$$

$$SW_{dn}^{clear}(t, \text{d}; \text{R2}) = SW_{dn}^{clear}(t, d; \text{R1})$$
$$\times \left[0.92 + 0.02 \times \sin\frac{\pi(d - 120)}{100}\right]$$
$$\text{(for } 120 \leq d \leq 220)) \tag{4c}$$

$$SW_{dn}^{clear}(t, \text{d}; \text{R2}) = SW_{dn}^{clear}(t, d; \text{R1})$$
$$\times \left[0.92 - \frac{(0.92 - 0.85)(d - 280)}{86}\right]$$
$$\text{(for } 280 < d \leq 366) \tag{4d}$$

where $SW_{dn}^{clear}(t, \text{d}; \text{R1})$ and $SW_{dn}^{clear}(t, \text{d}; \text{R2})$ denote R1 and R2 6-hourly averaged clear-sky surface downwelling SW fluxes, t denotes 6-hourly time of a day, and d denotes calendar day of a year. We found that the 8% surface insolation reduction [*Kanamitsu et al.*, 2002, Figure 8] was only valid over some calendar day ranges (4b), so we adjusted the

correction coefficient over calendar day ranges as given by the expressions: (4a) to (4d). This estimated R2 6-hourly averaged clear-sky surface downwelling SW flux is shown in Figure 1, together with the all-sky surface downwelling SW flux.

2.3. Methods

[19] The methods used in this study are detailed in this section. We first briefly introduce the calculations of SRCF and cloud albedo, and then give a detailed description of the procedures used in this study.

2.3.1. SRCF

[20] SRCF (also called "effective cloud albedo") was first proposed by *Betts and Viterbo* [2005] to quantify the impact of the cloud field on the surface radiative budget over a southwest basin of the Amazon. It is a non-dimensional measure of surface SW cloud forcing ($SW_{cld} = SW_{dn}^{all} - SW_{dn}^{clear}$, an upward flux), defined as

$$\alpha_{cld}^{SRF} = -\frac{SW_{cld}}{SW_{dn}^{clear}} = 1 - \frac{SW_{dn}^{all}}{SW_{dn}^{clear}}. \tag{5}$$

[21] Here, a positive value of the SW fluxes indicates a downward flux. Based on equation (5), we can calculate SRCF if all-sky and clear-sky surface downwelling SW fluxes are available.

[22] Furthermore, equation (5) indicates that SRCF represents the fraction of clear-sky incoming downward SW flux which is reflected and absorbed by clouds. This non-dimensional quantity offers an effective measure of surface SW cloud forcing and minimizes the influence from solar zenith angle and other non-cloud factors. More discussions and applications of SRCF can be found in previous studies [*Betts et al.*, 2006, *Betts*, 2009; *Betts et al.*, 2009; *Betts and Chiu*, 2010; *Liu et al.*, 2011].

2.3.2. Cloud Albedo

[23] *Liu et al.* [2011] derived an analytical expression that quantifies the relationship between SRCF, cloud fraction, and cloud albedo,

$$\alpha_{cld}^{SRF} = \alpha_r f \qquad (6)$$

Where f and α_r denote cloud fraction, and cloud albedo, respectively. *Liu et al.* [2011] demonstrated that the estimated cloud albedo by using the surface-based radiation and cloud fraction shows reasonable agreement (correlation coefficient 0.69) with that from satellite measurements. In this study, cloud albedo is estimated using the same procedure, and the same surface-based observations as in *Liu et al.* [2011]. Note that α_r implicitly includes cloud absorptance when cloud absorption cannot be ignored, esp., for strongly absorbing clouds.

2.3.3. Procedures of Evaluation

[24] Detailed procedures of the evaluation are described below. First, the 15-min all-sky/clear-sky surface downwelling SW flux and cloud fraction observations are averaged into hourly data. Here, only those with 4 valid 15-min data points within one hour are used. The valid 15-min data points refer to those with 15-min all-sky/clear-sky surface downwelling SW flux greater than zero and 15-min cloud fraction between 0 and 1. We use the hourly data (for example, $x_i(d, m, y)$, i, d, m, y represent hour, day, month, and year respectively) to calculate the mean variations of hourly all-sky/clear-sky surface downwelling SW flux and cloud fraction. Considering that the valid hourly observational data points are not evenly distributed temporally, we first calculate each-year seasonal mean values of each-hour variables $(\bar{x}_i(s, y) = \overline{x_i(d, m, y)}^s$, s represents season, the line on the top of the function represents mean, and the line with "s" represents seasonal mean), and then average into the overall mean values of each-hour variables. Those valid hourly all-sky/clear-sky surface downwelling SW flux and cloud fraction between 6 am and 6 pm (local standard time: LST) are further averaged into daytime-mean data. Daytime-mean (6 am–6 pm LST) SRCF is calculated using daytime-mean all-sky and clear-sky surface downwelling SW flux. Those daytime-mean SRCF (>−0.05) and cloud fraction are further averaged into monthly data.

[25] The mean value of overall hourly cloud albedo is calculated using the mean values of overall hourly SRCF and cloud fraction. The daytime-mean cloud albedo is calculated using daytime-mean SRCF and cloud fraction for those with daytime-mean cloud fraction greater than 0.05. We use this filter because cloud albedo is not well described by equation (6) when cloud fraction is small [*Liu et al.*, 2011]. The monthly cloud albedo is calculated by using monthly SRCF and cloud fraction. The mean variations of monthly/yearly cloud properties are the averages of the monthly cloud properties.

[26] Next, the cloud properties from hourly ERA-Interim and 6-hourly R1 and R2 are used to calculate the mean values of daytime hourly and 6-hourly cloud properties. The calculation procedures are the same as for the observations. Here, only those hourly/6-hourly reanalysis data concurrent to those valid hourly observations are used. The concurrent 6-hourly reanalysis data refer to those: within those concurrent 6 h the hourly observations have valid data. Further, those concurrent hourly/6-hourly reanalyses are averaged into daytime-mean, and then those daytime-mean reanalysis data concurrent to those valid daytime-mean observations are averaged into monthly data. The yearly data are the averages of monthly data. The diurnal/annual/interannual cloud properties from the reanalyses are then evaluated based on the observations.

[27] After that, for diagnosing the path of model-error propagation, the model biases (model minus observation) in the cloud properties and their relationships are analyzed. Further examined are the relationship between the cloud properties and 2-m temperature/humidity. To do so, we first aggregate the observed 30-min averaged 2-m air temperature, relative humidity and surface pressure into hourly data, and then use hourly temperature, relative humidity and surface pressure to calculate hourly specific humidity by equations (2) and (3). After that, using the same method as for the cloud properties, we use concurrent valid hourly data to generate daytime-mean, monthly, yearly temperature/humidity. And, the three reanalyses' hourly/6-hourly cloud properties and the meteorological variables with concurrency to the valid hourly observations are averaged into daytime-mean and then monthly data. The relationship between the daytime-mean/monthly cloud properties and temperature/humidity from all the data sets are first examined. Then, we compare the multiscale mean variations of 2-m temperature/humidity and corresponding model biases. After that, the relationship between relative humidity (or cloud fraction) biases and the temperature/humidity is examined.

[28] It is noted that R1 and R2 6-hourly 2-m temperature and relative/specific humidity have three daytime data points at 6 A.M., 12 P.M. and 6 P.M. (LST). Thus, the calculation of the daytime mean is not as straight-forward as the hourly/6-hourly averaged data or the hourly interval data (considering that variations within one hour are small). We examine three common ways to obtain the daytime mean from the three data points using the hourly ERA-Interim data: [(6 am + 12 pm)/2 + (12 pm + 6 pm)/2]/2, [(6 am + 12 pm + 6 pm)/3], and [6 am + 4 × (12 pm) + 6 pm]/6. The last formula is from the Simpson's rule for parabolas. The daytime mean from the three methods are compared to that from the mean of the 13 daytime hourly points. The results are shown in Figure 2, indicating that the first method has the smallest difference (standard deviation) from the 13-point averaged value. Based on this analysis, we chose the first method to calculate R1 and R2 daytime-mean temperature/humidity.

[29] Finally, for evaluating the overall performance of the reanalyses in modeling the cloud properties and the meteorological variables, we employ the widely used technique of the Taylor diagram [*Taylor*, 2001], and also develop a new metric "Relative Euclidean Distance" as a supplement to the Taylor diagram. The Taylor diagram reveals concise and easy-to-visualize second-order statistical differences between two (or more) different time series. It is especially useful for evaluating a model's performance in phase and amplitude of variations (as measured by the correlation coefficient r and standard deviation σ, respectively), and a model's "centered root-mean-square error" E ("RMS error" hereafter). This technique has been widely used in climate researches and IPCC assessment [e.g., *IPCC*, 2001; *Anderson et al.*, 2004; *Martin et al.*, 2006; *Miller et al.*, 2006; *Bosilovich et al.*, 2008; *Gleckler et al.*, 2008; *Pincus et al.*, 2008]. Briefly,
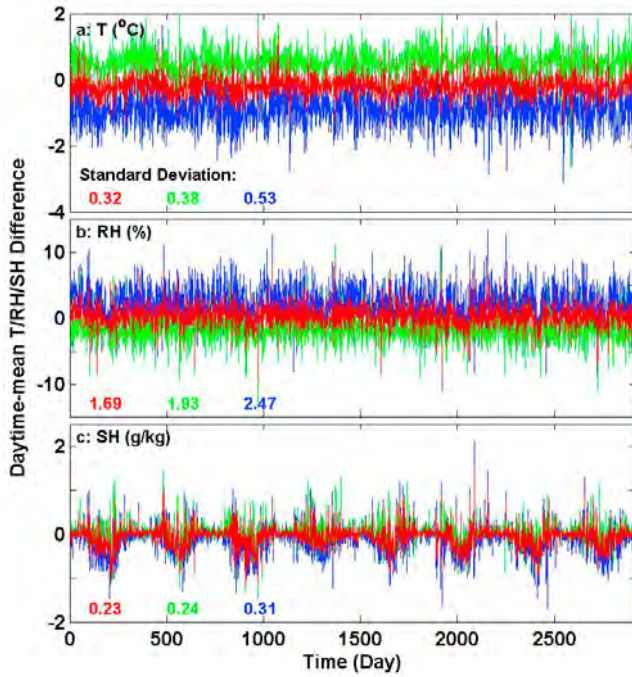
Figure 2. Comparison of the differences between the day-time-mean values from 3 daytime 6-hourly data points using the three methods (red, green, and blue) and those from 13 daytime hourly data points (from 6 am to 6 pm), based on the hourly ERA-Interim data. Red: [(6 am + 12 pm)/2 + (12 pm + 6 pm)/2]/2, green: [6 am + 4 × (12 pm) + 6 pm]/6, and blue: [(6 am + 12 pm + 6 pm)/3]. T, RH, and SH represent 2-m temperature, relative humidity, and specific humidity.

the expressions for calculating r, σ, and E are shown below [*Taylor*, 2001],

$$r = \frac{\frac{1}{N}\sum_{n=1}^{N}(M_n - \bar{M})(O_n - \bar{O})}{\sigma_M \sigma_O} \qquad (7)$$

$$E^2 = \frac{1}{N}\sum_{n=1}^{N}[(M_n - \bar{M}) - (O_n - \bar{O})]^2 \qquad (8)$$

$$\sigma_M^2 = \frac{1}{N}\sum_{n=1}^{N}(M_n - \bar{M})^2 \qquad (9)$$

$$\sigma_O^2 = \frac{1}{N}\sum_{n=1}^{N}(O_n - \bar{O})^2 \qquad (10)$$

with

$$\bar{M} = \frac{1}{N}\sum_{n=1}^{N}M_n \qquad (11)$$

$$\bar{O} = \frac{1}{N}\sum_{n=1}^{N}O_n \qquad (12)$$

where "M_n" or "O_n" denote a modeled or observed variable, defined at N discrete temporal (or spatial) points; and the subscript "M" or "O" denote "model" or "observation." Note that, the correlation coefficient r, standard deviation σ and RMS error E are calculated without removing the periodic signals of the time series.

[30] As a supplement to the Taylor diagram, we develop a new metric "Relative Euclidean Distance" (D), based on Euclidean-Distance technique [e.g., *Elmore and Richman*, 2001] and first- and second -order statistics,

$$D = \sqrt{\left[\frac{(\bar{M} - \bar{O})}{\bar{O}}\right]^2 + \left[\frac{(\sigma_M - \sigma_O)}{\sigma_O}\right]^2 + (1 - r)^2}. \qquad (13)$$

[31] As can be seen from the expression (13), D measures the overall model performance. A perfect agreement corresponds to D = 0, and the overall model performance degrades as D increases.

3. Evaluation of Multiscale Cloud Properties and a Discussion on Potential Uncertainty

[32] This section first evaluates the cloud properties (i.e., SRCF, cloud fraction, and cloud albedo) from the reanalyses at diurnal, annual, and inter-annual temporal scales. Then, cloud fraction and potential uncertainty are further discussed.

3.1. Evaluation of Multiscale Cloud Properties

[33] Figure 3 shows the multiscale mean variations of SRCF from the reanalyses and the observations (Figures 3a, 3b, and 3c), and corresponding relative model biases [i.e., (model minus observation) divided by observation] (Figures 3d, 3e, and 3f). The daytime each-hour standard deviation is calculated for each specific hour using seasonal-mean hourly SRCF after removing seasonal cycles (i.e., each-season mean value). The SRCF standard deviation in early morning (or late afternoon) shows slightly larger value than those at other daytime hours. The each-month standard deviations are calculated for each specific month using monthly SRCF. The SRCF shows larger standard deviation in fall/winter months than in other months. The standard deviation for annually averaged SRCF is calculated using monthly SRCF after removing seasonal cycles (i.e., each-month mean value). The cause of the large standard deviation in early morning (or late afternoon) or in fall/winter months is probably associated with the accuracy of the measurement and retrieval methods.

[34] It is evident that the SRCF from the observations and ERA-Interim has a strong annual cycle, with amplitude about 0.16 for the observations and about 0.10 for ERA-Interim. The SRCF peaks in March and reaches its lowest value in July. R2 exhibits a similar phase pattern with weaker amplitude of 0.05. It is noteworthy that despite having a variation amplitude (0.06) comparable to R2, R1 shows an opposite phase pattern, suggesting that R1 is the most problematic in modeling the observed phase of SRCF annual cycle, although it shows the smallest biases from April to September. The diurnal (inter-annual) variations of SRCF are weaker than the annual cycle, with amplitude ranging from 0.01 to 0.05 (from 0.04 to 0.07). For the diurnal cycle, the observations and R2 show larger (smaller) SRCF in the morning
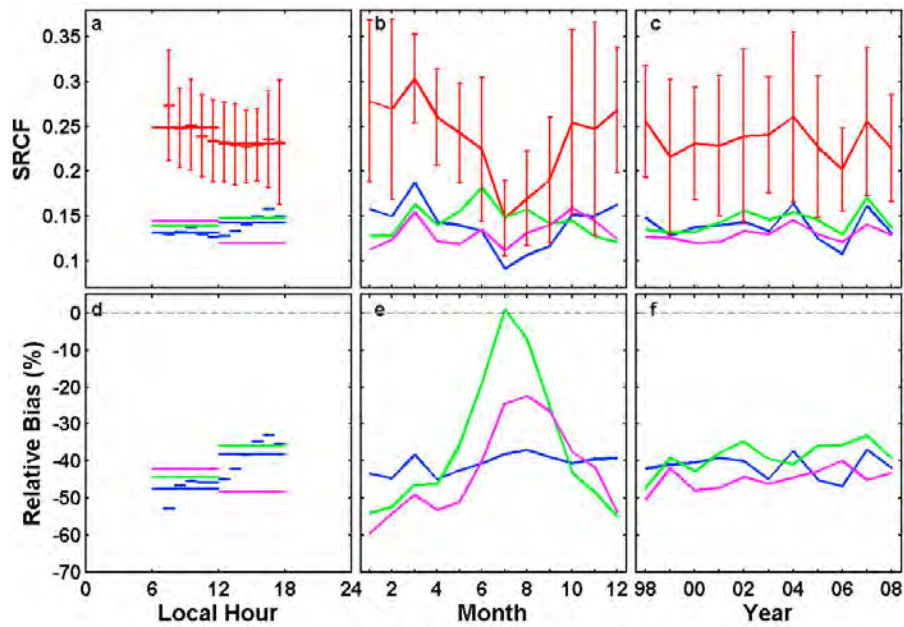
Figure 3. (a, b, and c) Multiscale mean variations of surface relative shortwave cloud fraction (SRCF) and (d, e, and f) corresponding relative model biases, defined as [(model minus observation) divided by observation]. Red, blue, green, and purple represent the observations, ERA-Interim, R1, and R2, respectively. The data used are daytime (6 am to 6 pm LST) data from 03/25/1997 to 12/31/2008 over ARM SGP Central Facility site. The red bar represents ± one standard deviation of the observations. The relative model biases show the significance of the model biases compared to the observations.

(afternoon); but ERA-Interim and R1 show an opposite phase pattern. This suggests that ERA-Interim and R1 do not model the observed phase of the SRCF diurnal cycle correctly. For the inter-annual variations, the observed SRCF reveals an increase of 0.04 from 1999 to 2004 followed by a drop of 0.06 until 2006, and the three reanalyses show similar phase pattern. One notable phenomenon is that the observed SRCF is systematically much larger than the reanalyses at all the temporal scales, suggesting that the reanalyses significantly underestimate the observed SRCF. This fact can also be seen from Figures 3d, 3e, and 3f, where the relative biases reaches up to −50% (−60%) for the diurnal/interannual (annual) variations.

[35] Similarly, Figures 4 and 5 show the comparisons of the multiscale mean variations of cloud fraction and cloud albedo, respectively. Variations of the standard deviations from the observations in general follow those of SRCF. As shown in Figure 4, the multiscale mean variations of cloud fraction look remarkably similar to those of SRCF, with a strong annual cycle (amplitude up to 0.22) and relatively weak diurnal/inter-annual variations (amplitude up to 0.06/0.10). The phase patterns of the multiscale mean variations of cloud fraction look similar to those of SRCF, suggesting that large cloud fraction in general corresponds to large SRCF. Similar to SRCF, ERA-Interim and R1 do not capture diurnal-cycle phase pattern of the cloud-fraction observations, and R1 suffers from evident deficiency in modeling the annual-cycle variations. The modeled CF biases are the largest in summer for ERA-Interim, similar to the results from 3-year MERRA and NARR by *Kennedy et al.* [2011]. But, this is not the case for R1 and R2. The underestimation of the modeled cloud fraction is also significant at

all the temporal scales, ranging from −20% to −40% for the majority. Figure 5 reveals a strong annual cycle (amplitude up to 0.22) and slightly weak diurnal/inter-annual variations (amplitude up to 0.16/0.12) for the modeled and observed cloud albedo. The phase patterns of the multiscale mean variations of cloud albedo mainly follow those of SRCF or cloud fraction, except for the annual cycle of the modeled cloud albedo which has the lowest value in winter season, opposite to other annual-cycle patterns. Similar to SRCF and cloud fraction, ERA-Interim and R1 show a problem in modeling the phase pattern of the observed cloud-albedo diurnal cycle. It is noted that the phase pattern of R1 cloud-albedo annual cycle agrees with the observations better than that of SRCF and cloud fraction. Note also that, cloud fraction is only one of the many variables (e.g., cloud thickness and liquid water path) affecting SRCF and cloud albedo. The differences in the phase patterns may imply different cloud types, although more research is needed to ascertain this. For all the models the underestimation of cloud albedo is small from May to September, but increases to −25% to −33% in the cold season.

3.2. Further Discussion on Cloud Fraction and Potential Uncertainty

[36] The observed cloud fraction shown in Section 3.1 represents 15-min averages of 160°FOV hemispheric fractional sky cover from the surface [*Long et al.*, 2006]. Although the hemispheric fractional sky cover is consistent with the surface radiation measurements evaluated in this paper and the hemispheric fractional cover has been shown to agree with the GOES satellite measurements reasonably well [*Liu et al.*, 2011], it is desirable to examine the effect of
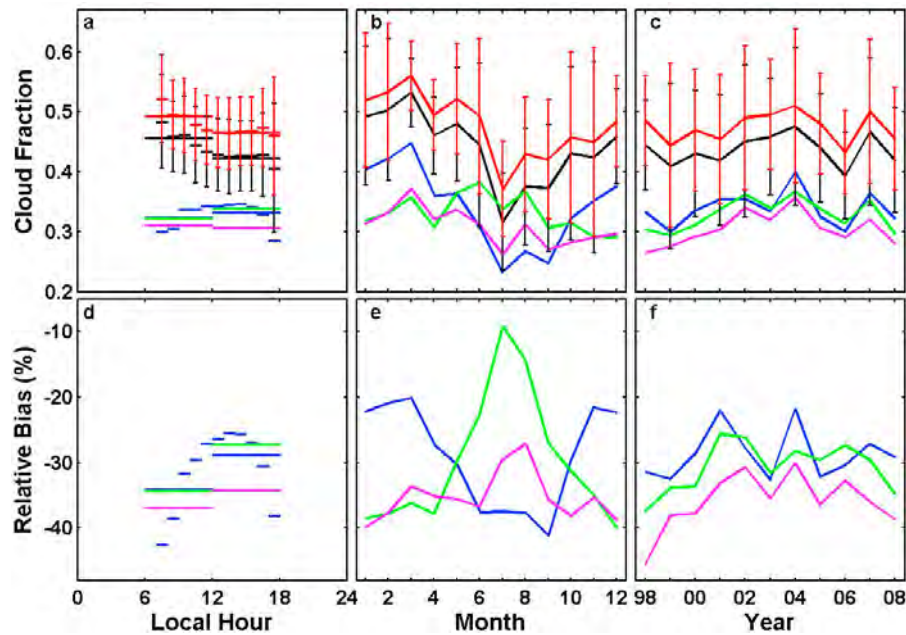
Figure 4. The same as Figure 3 but for cloud fraction instead of surface relative shortwave cloud fraction (SRCF). The black curves represent the multiscale mean variations of the cloud fraction generated by using the FOV-adjusted hourly cloud fraction CF(60°) instead of CF(160°).

different FOV angles on cloud fraction. *Kassianov et al.* [2005], using Monte Carlo simulations of surface-based hemispherical measurements based on four-dimensional cloud fields produced by a large-eddy simulation model, analyzed the differences in hemispheric cloud fraction estimates depending on averaging time, FOV and cloud spatial structure. They found that 15 min is an appropriate averaging time and that hemispheric-view cloud fractions are almost unbiased for FOV ≤ 100°. However, for the same cloud field, hemispheric-view cloud fractions increase as FOV increases,

and the difference between the hemispheric-view cloud fractions corresponding to different FOVs increases as cloud fraction decreases. *Kassianov et al.* [2005] further provided the following linear fit between the 160°FOV hemispheric fractional sky cover, CF(160°), and the 60°FOV value, CF(60°),

$$CF(160°) = 0.17 + 0.815 \times CF(60°) \qquad (14)$$

Because CF(60°) appears to be more representative of the plane-parallel cloud fraction of the reanalyses, here we explore
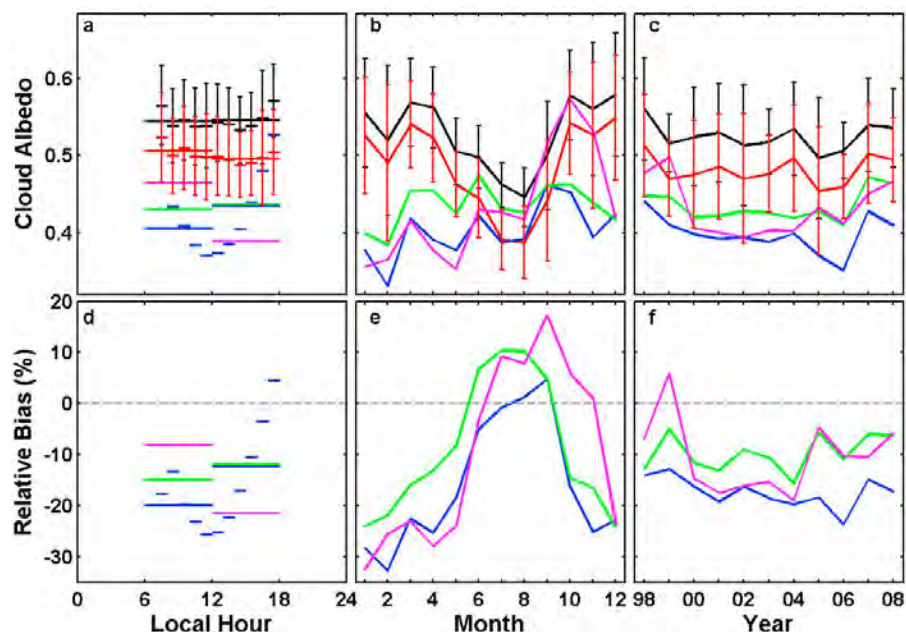

Figure 5. The same as Figures 4 but for cloud albedo instead of cloud fraction.
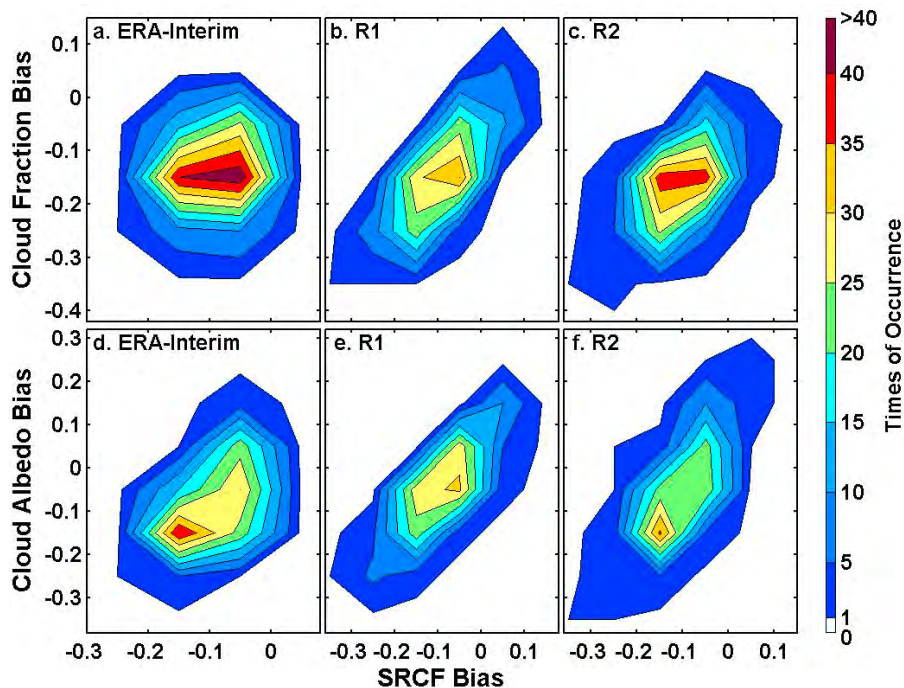
Figure 6. Contour plots showing the range and relationship of the monthly model biases between the cloud properties in each of the three reanalyses.

the correction of the CF(160°) data based on the inverse of equation (14),

$$CF(60°) = 1.227 \times CF(160°) - 0.2086. \tag{15}$$

[37] Equation (15) is used as a correction for 0.25 < = CF(160°) < = 0.918. The upper limit is chosen where the correction goes to zero, and for CF(160°) > 0.918 we set CF(60°) = CF(160°). For CF(160°) < 0.25, we reduced the correction linearly to zero, as CF(60°) = (1 − 0.1519/0.25) × CF(160°).

[38] The black curves in Figures 4 and 5 represent the multiscale mean variations of the cloud fraction and cloud albedo generated by using the hourly cloud fractions CF(60°) instead of CF(160°). As can be seen, the cloud fractions from CF(60°) are systematically lower than those from CF(160°) in a range from 0.02 to 0.06. The corresponding cloud albedo from CF(60°) are systematically higher than those derived from CF(160°) in a range from 0.03 to 0.07.

[39] From Figure 4, we see that using as a reference the derived cloud fraction, CF(60°), would reduce but not remove the low bias of the reanalyses. From Figure 5 we see that using as a reference cloud albedo derived from CF(60°) would increase the low bias of the reanalyses. However, there are other measurement comparison issues here which need further study. For example, correction (15) is based on *Kassianov et al.* [2005] for cloud fraction. However, our observed SRCF is also derived from hemispheric radiometers which see the cloud field over all viewing angles and solar zenith angles. This is not the same geometry as the plane-parallel assumption of the reanalyses, averaged over a horizontal grid-cell. The geometry-induced difference in SRCF may make the difference in cloud albedo spuriously larger. An ultimate solution requires considering SRCF, cloud fraction and cloud

albedo simultaneously and consistently; however, it is beyond the scope of this paper. Nevertheless, the comparison of the results from CF(60°) and CF(160°) gives an estimate of possible uncertainties resulting from different FOV angles.

4. Model Biases and Their Links

[40] The previous section shows that, except for warm season cloud albedo, the three reanalyses in general significantly underestimate all the three cloud properties. To further explore the model-error propagation path, this section examines the range of the model biases in the cloud properties, their mutual relationships, and the relationship of the cloud properties to 2-m temperature and relative/specific humidity. The analysis has been performed for both daytime-mean and monthly data; but only the monthly results are presented below. The daytime-mean results are similar except for a larger range of their biases.

4.1. Model Biases in the Cloud Properties and Their Mutual Relationships

[41] Figure 6 are the contour plots showing the range and relationship of the monthly model biases in the cloud properties. As can be seen, the model biases of the cloud properties in the reanalyses mainly fall within a centered region, ranging from −0.20 to 0 for SRCF, from −0.25 to 0 for cloud fraction, and from −0.20 to 0.10 for cloud albedo. The points within the centered regions exhibit a positive linear relationship to one another, except for the model biases of ERA-Interim cloud fraction and SRCF. The strong positive correlation is evident in the corresponding scatterplots (Figure 7) as well, where the overlapped relationships from different reanalyses and the strength of the relationships are clearly shown.
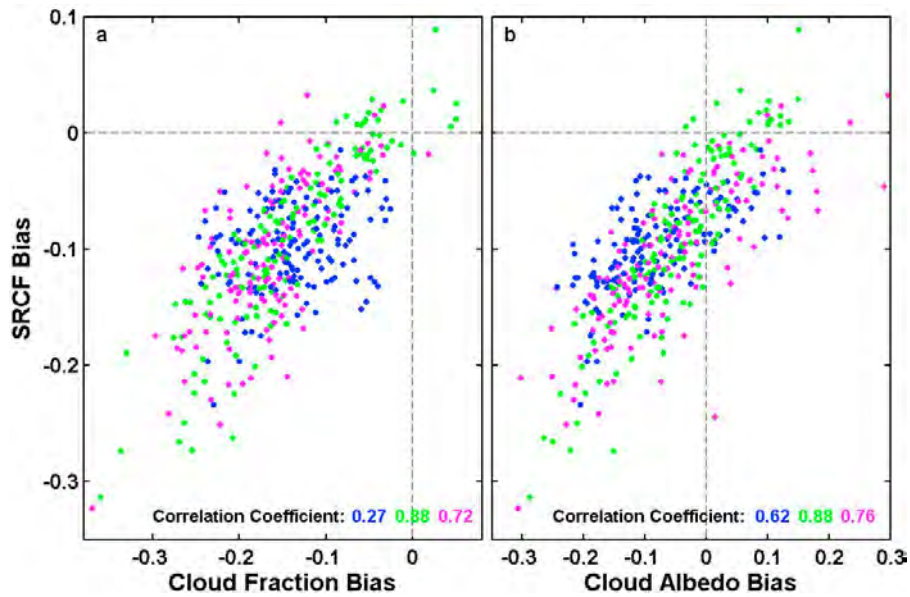
Figure 7. Scatterplots showing the overlapped relationships from different reanalyses and the correlations (the strength of the relationships). Blue, green, and purple represent ERA-Interim, R1, and R2, respectively. The colored numbers represent correlation coefficients from the same colored dots.

[42] The relationship between the model biases in the cloud properties and the observed (a, b, and c) or modeled (d, e, and f) cloud fraction are shown in Figure 8. It is evident that the model biases in the cloud properties are large when observed cloud fractions are large, except for ERA-interim with small correlation coefficients for cloud fraction biases (−0.25) and cloud albedo biases (−0.18). There is no clear relationship between cloud fraction biases and the modeled cloud fraction. The model biases in SRCF (or cloud albedo)
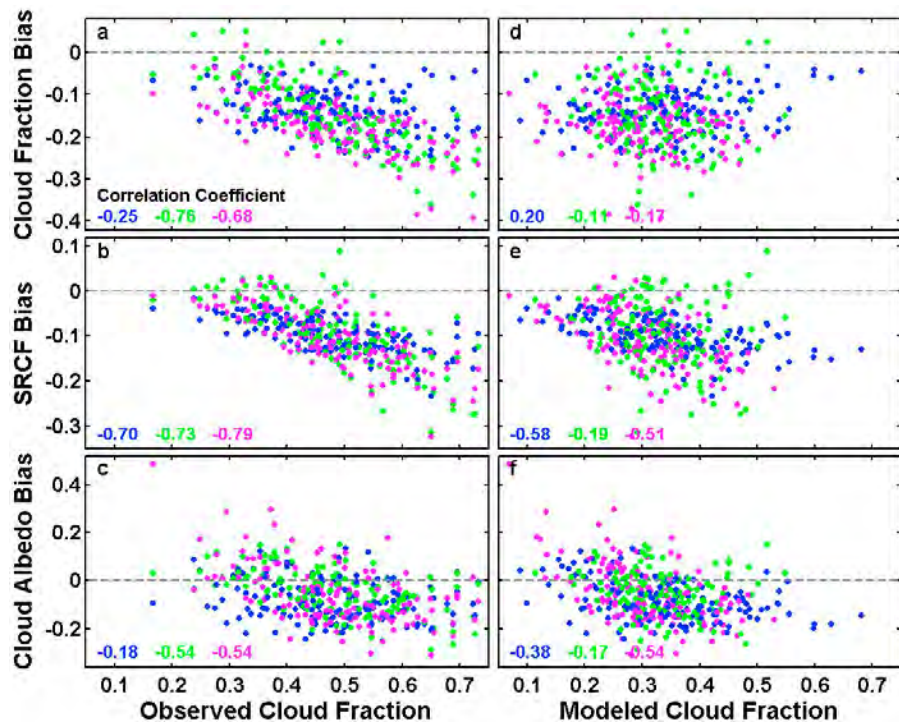

Figure 8. Scatterplots showing the relationship between (a, b, and c) observed or (d, e, and f) modeled cloud fraction and the model biases in the cloud properties. Blue, green, and purple represent ERA-Interim, R1, and R2, respectively. The colored numbers represent correlation coefficients from the same colored dots.
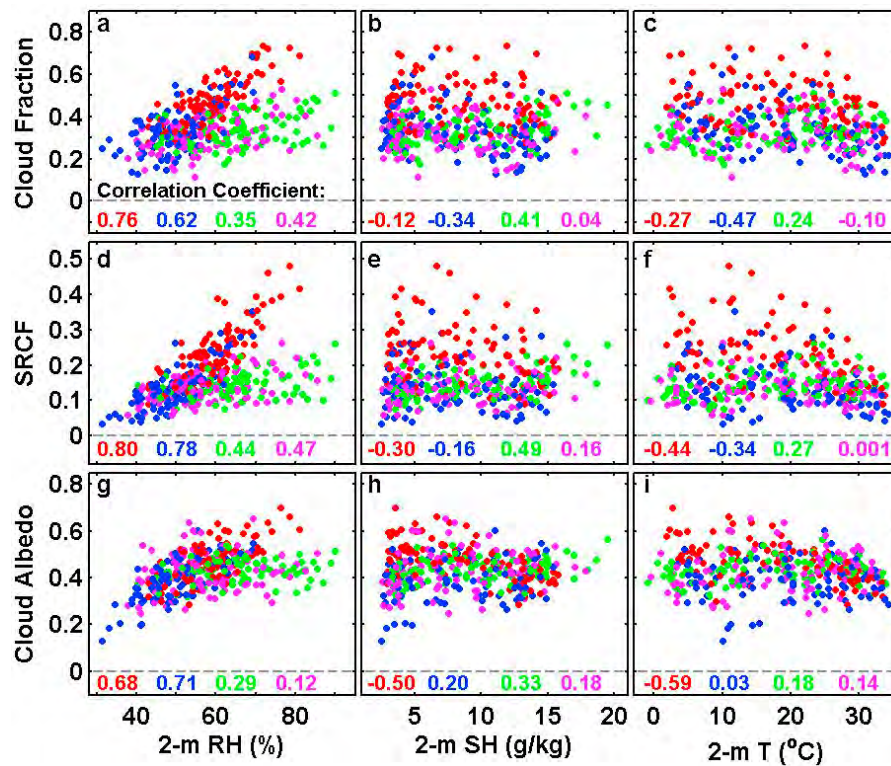
Figure 9. Relationship between the cloud properties and the 2-m temperature/humidity.

are large when modeled cloud fraction is large, except for R1 with small correlation coefficients.

4.2. Relationship of Cloud Fraction Biases to 2-m Temperature and Humidity

[43] It is expected that cloud properties are related to 2-m meteorological variables, especially for relative humidity, at least for clouds with roots in the boundary layer. Furthermore, the modeled 2-m air temperature/humidity are strongly impacted by the model used in the reanalysis [*Kalnay et al.*, 1996]. This section examines the relationship between the cloud properties and the temperature/humidity, the multiscale mean variations of the temperature/humidity from all the data sets and the corresponding model biases, and the relationship between the model biases in the relative humidity (or cloud fraction) and the temperature/humidity. Results from concurrent monthly data are presented.

[44] Figure 9 shows the cloud properties as a function of temperature and humidity. The cloud properties are strongly related to the relative humidity for all the data sets, especially for the observations and ERA-Interim with correlation coefficients between 0.62 and 0.80. There is some relationship between the cloud properties and the temperature (or specific humidity), but the relationship is not as strong as that with the relative humidity. The ERA-Interim appears to describe the observed relationships better than the two NCEP reanalyses in general. One exception is that cloud albedo decreases with temperature (and specific humidity) in the observations but not in the reanalyses. These relationships are likely related to the processes that couple boundary layer clouds with the land surface, and to the parameterizations

used in the models. The parameterization details and their effects on the results warrant further investigation.

[45] Figure 10 shows the multiscale mean variations of 2-m temperature from all the data sets and corresponding model biases. In general, the model biases are small, ranging from −2.10°C to 1.65°C, which is comparable to the standard deviation of the observations. ERA-Interim (R1) shows slightly warmer (colder) than the observations at all the temporal scales, except that R1 has warm biases in April/May. Relatively, R2 shows the smallest overall biases, as evident by the annual cycle and interannual variations. The phase patterns of the multiscale mean variations from the reanalyses agree with the observations for the three temporal scales. The diurnal (annual) cycle peaks at 3pm (July) and shows the coldest temperature at 6 am (January).

[46] Figure 11 shows the multiscale mean variations of 2-m specific humidity from all the data sets and corresponding model biases. Like the temperature, the model biases in the specific humidity are small (from −0.94 to 1.47 g/kg), comparable to the standard deviations of the observations. The specific humidity in ERA-Interim/R2 (R1) is smaller (larger) than the observations, except for R2 monthly specific humidity in January/August and annually averaged specific humidity in 2007/2008 with an opposite behavior. The multiscale mean specific-humidity variations from all the data sets have similar phase patterns, indicating that the reanalyses are capable of modeling the observed phase patterns.

[47] Figure 12 shows the multiscale mean variations of 2-m relative humidity from all the data sets and corresponding model biases. The model biases in relative humidity range from −12.52% to 14.93%, slightly larger than the standard
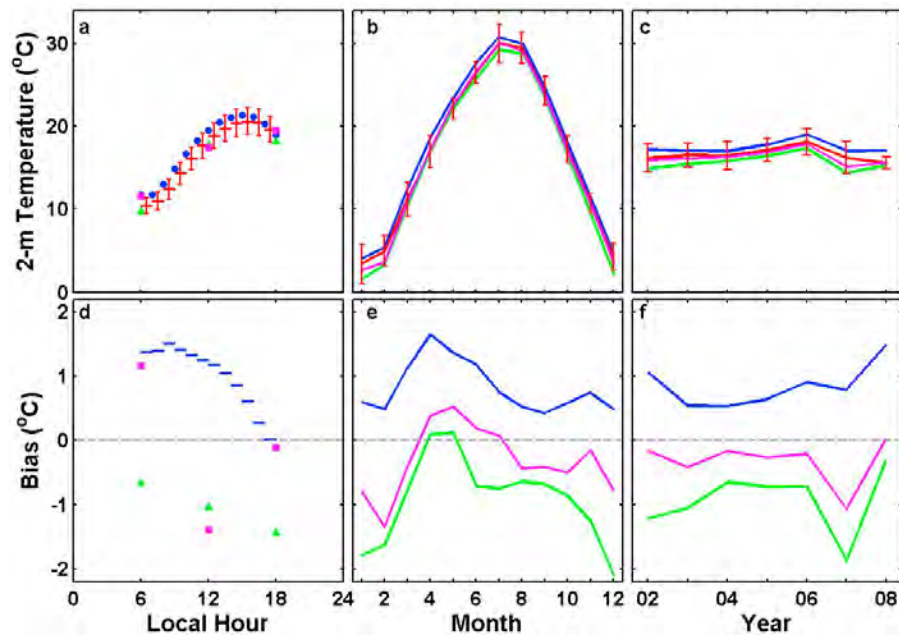
Figure 10. (a, b, and c) Multiscale mean variations of 2-m air temperature and (d, e, and f) corresponding model biases, defined as (model minus observation). Red, blue, green, and purple represent the observations, ERA-Interim, R1, and R2, respectively. The data used are daytime (6 am to 6 pm LST) data from 04/01/2001 to 12/31/2008 over ARM SGP Central Facility site. The red bar represents ± one standard deviation of the observations.

deviations of the observations. It is evident that ERA-Interim (R1) relative humidity is systematically smaller (larger) than the observations at all the temporal scales. The underestimation (overestimation) in ERA-Interim (R1) is especially large, up to −12.52% in early morning (late afternoon) or from late Fall to early Spring. R2 relative humidity matches the observations the best among the three reanalyses, although smaller than the observations at 6 am (−7.69%) or in April and May (−7.01%). The phase pattern from all the reanalyses agrees well with the observations, with a peak in early morning, then decreases until 3pm, and then increases again (note that the 6-hourly R1 and R2 cannot show the detailed hourly pattern). The relative humidity reaches the smallest value in July, and then increases from July to
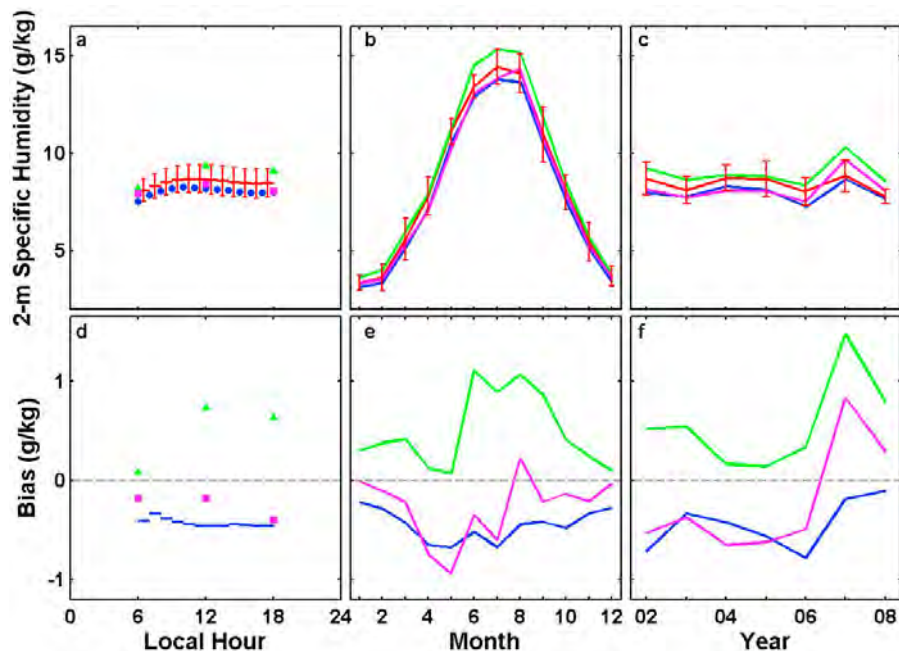

Figure 11. The same as Figure 10 but for 2-m specific humidity instead of 2-m air temperature.
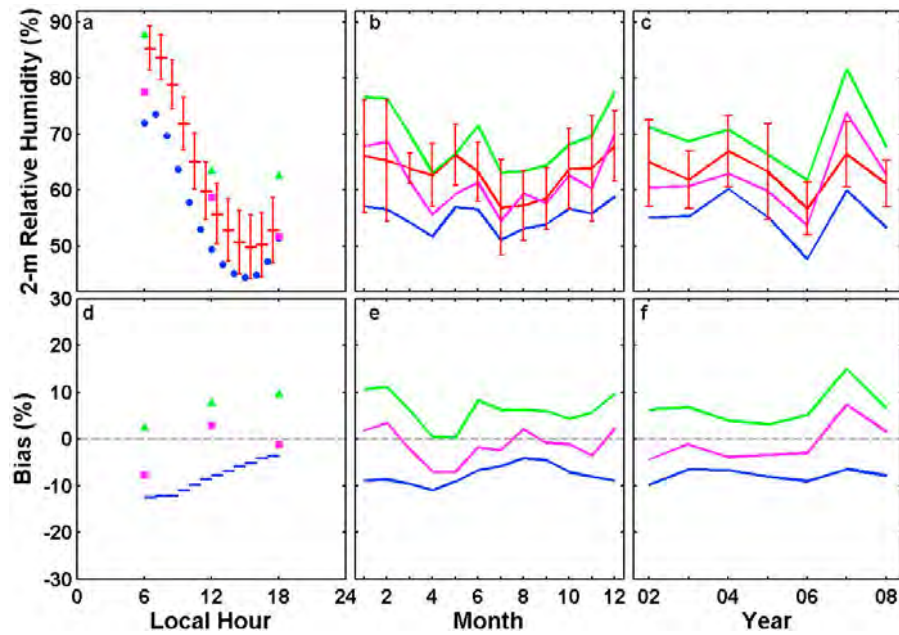
Figure 12. The same as Figure 10 but for 2-m relative humidity instead of 2-m air temperature.

December, followed by a decrease until April, and then an increase from April to May (from April to June for R1 and R2), then a drop to July. A notable phenomenon of the annually averaged relative humidity is the large drop (−12.51%) from 2004 to the hot year 2006, followed by the

large increase (20.04%) to 2007, and then a drop about −13.54% to 2008.

[48] The relationship between the relative humidity biases and the temperature/humidity is shown in Figure 13. As shown in Figures 13a and 13b, the relative humidity biases in
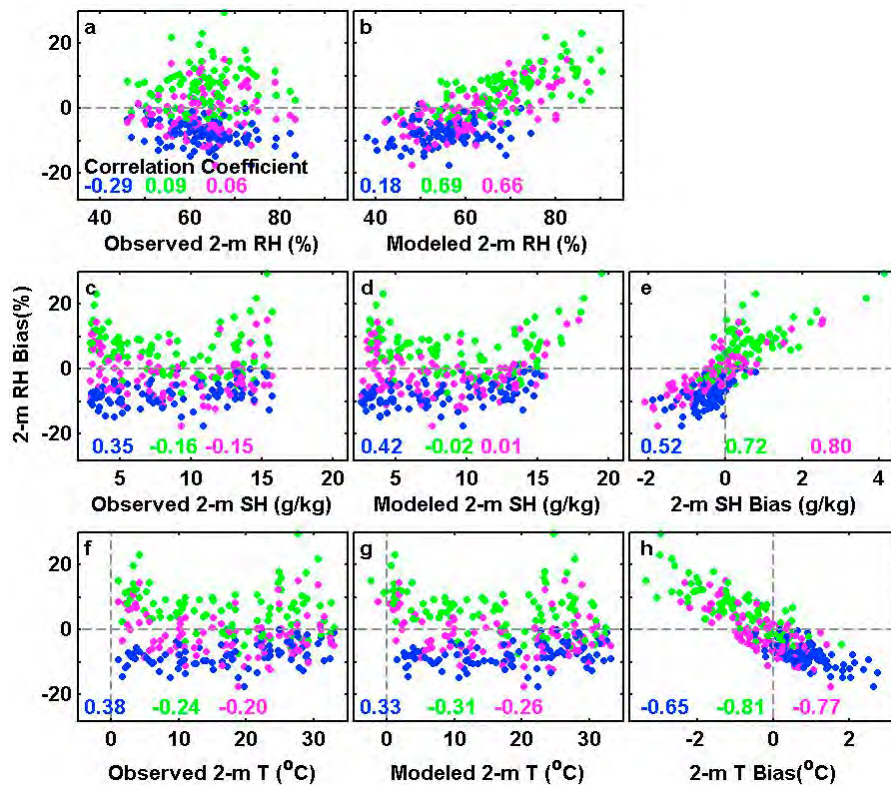

Figure 13. Scatterplots showing the relationships between 2-m relative humidity (RH) biases and 2-m temperature (T) [or specific humidity (SH), or relative humidity (RH)]. Blue, green, and purple represent ERA-Interim, R1, and R2, respectively. The colored numbers represent correlation coefficients from the same colored dots.

Figure 14. Scatterplots showing the relationships between cloud fraction biases and 2-m temperature (T) [or specific humidity (SH), or relative humidity (RH)]. Blue, green, and purple represent ERA-Interim, R1, and R2, respectively. The colored numbers represent correlation coefficients from the same colored dots.
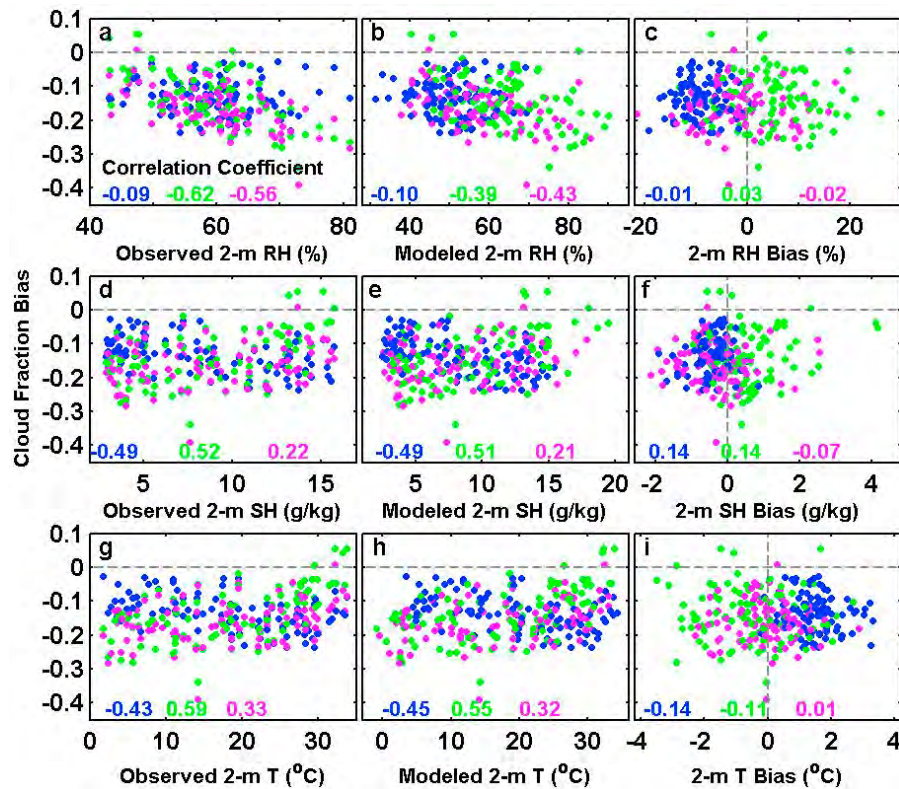
R1 and R2 are strongly positively related to the modeled relative humidity, with correlation coefficients greater than 0.66. The relationship between ERA-Interim relative humidity and its biases is much weaker, with correlation coefficient only 0.18. The dependence of relative humidity biases on the observed relative humidity, observed/modeled temperature and specific humidity is also weak, with correlation coefficient smaller than 0.42. Figure 13 further shows that the relative humidity biases have a positive (negative) linear relationship with the specific humidity (temperature) biases, suggesting that an overestimated (underestimated) relative humidity results from combined effects of an overestimated (underestimated) specific humidity and underestimated (overestimated) temperature.

[49] Figure 14 shows the relationship between cloud fraction biases and the temperature/humidity. From Figures 14a, 14b, and 14c, the cloud fraction biases in R1 and R2 increase with observed and modeled relative humidity, except that ERA-Interim cloud fraction biases show no clear relationship with relative humidity. Figures 14d–14i also shows that the cloud fraction biases in R1 and R2 decrease with observed and modeled temperature/specific humidity, while ERA-Interim has an opposite behavior. There is no clear relationship between cloud fraction biases and the temperature/humidity biases.

5. Evaluation on the Overall Performance of the Reanalyses

[50] This section evaluates the overall performance of the reanalyses in modeling the cloud properties and the

meteorological variables. The evaluation is conducted using concurrent daytime-mean and monthly data.

[51] Figure 15 shows the Taylor diagrams of the cloud properties from daytime-mean (a) and monthly (b) data. The radial distance represents the amplitude of the variations, normalized by the amplitude of the observationally based variations. The cosine of azimuthal angle of each point gives the correlation between the reanalyses and the observations. The distance between each point and the reference point "Obs" represents the RMS error, normalized by the amplitude of the observationally based variations. As this distance approaches to zero, the modeled variations approach the observations.

[52] For daytime-mean cloud properties, ERA-Interim exhibits the best performance in modeling the phase of the cloud properties and the magnitude of the SRCF variations, although it slightly overestimates the magnitude of the cloud-fraction and cloud-albedo variations. R2 exhibits the best performance in modeling the magnitude of the cloud-fraction and cloud-albedo variations, although it significantly underestimates the magnitude of the SRCF variations. R2 shows slightly better phase pattern of the cloud properties than R1. R1 significantly underestimates the magnitude of the cloud properties and also shows the worst phase similarity to the observations. ERA-Interim has the smallest RMS errors of all the cloud properties while R1 (or R2) exhibits the largest RMS errors of SRCF (or cloud fraction and cloud albedo).

[53] For daytime-mean meteorological variables, ERA-Interim exhibits the best performance in modeling the phase of the temperature/humidity and the magnitude of the
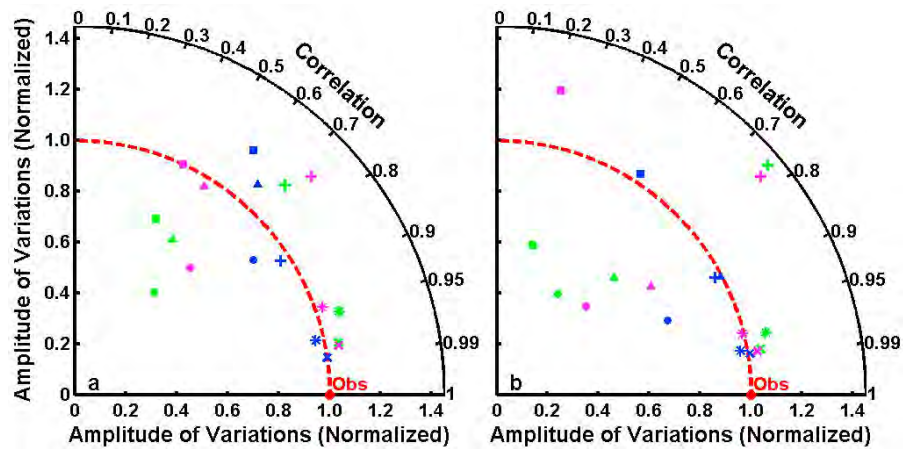
Figure 15. Taylor diagrams of the cloud properties and the 2-m temperature/humidity from (left) daytime-mean or (right) monthly data. Red, blue, green, and purple represent the observations, ERA-Interim, R1, and R2, respectively. The symbols "circle," "triangle," "square," "x-mark," "plus," and "star" denote surface relative shortwave cloud fraction (SRCF), cloud fraction, cloud albedo, temperature, relative humidity, and specific humidity, respectively.

temperature variations, although it slightly underestimates the magnitude of the relative/specific humidity variations. R1 (R2) exhibits the best performance in modeling the magnitude of the relative (specific) humidity variations, and both overestimate the magnitude of all the daytime-mean meteorological variables. R1 shows slightly better phase pattern than R2 for specific humidity whereas the opposite is true for temperature and relative humidity.

[54] For monthly cloud properties, ERA-Interim exhibits the best performance in modeling the phase and magnitude and the smallest RMS error of the cloud properties, even

though ERA-Interim significantly underestimates the observed monthly SRCF. R2 significantly overestimates (underestimates) the magnitude of the cloud-albedo (SRCF/cloud fraction) variations, with the largest RMS error of cloud albedo. R1 significantly underestimates all the cloud properties, showing the worst in modeling the phase and magnitude and the largest RMS errors of cloud fraction and SRCF.

[55] For monthly meteorological variables, ERA-Interim exhibits the best performance in modeling the phase and magnitude and the smallest RMS error of the temperature/humidity variations, except for R2 showing the best in
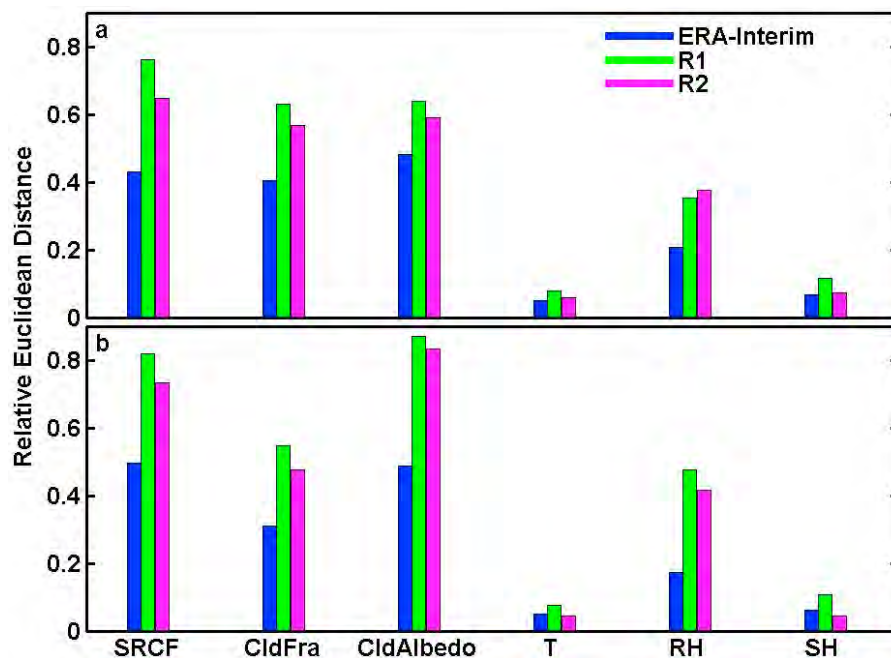

Figure 16. "Relative Euclidean Distance" of the cloud properties and the 2-m temperature/humidity from (a) daytime-mean and (b) monthly data. The shorter the Euclidean distance, the better the model performance is.

modeling the magnitude of specific humidity variations. R1 (R2) shows slightly better performance in simulating the phase pattern of the specific humidity (temperature and relative humidity). R1 significantly (slightly) overestimates the magnitude of the observed monthly relative humidity (temperature, and specific humidity) variations. R2 significantly (slightly) overestimates the magnitude of the observed monthly relative humidity (temperature), with a slightly small RMS error than R1 for all the meteorological variables.

[56] Figure 16 further compares the "Relative Euclidean Distance" D values of the reanalyses in modeling the daytime-mean (Figure 16a) or monthly (Figure 16b) cloud properties and the meteorological variables. As can be seen, ERA-Interim (R1) has the smallest (largest) D values for all the cloud properties and the meteorological variables at both daytime-mean and monthly temporal scales. The only exception is that R2 has the best overall performance in modeling the monthly temperature and specific humidity. This suggests that ERA-Interim (R1) ranks the best (worst) overall performance in modeling the cloud properties and the meteorological variables among the three reanalyses, except that R2 ranks the best in modeling the monthly temperature and specific humidity.

6. Summary

[57] This study evaluates three major reanalyses (ERA-Interim, NCEP/NCAR Reanalysis I, NCEP/DOE Reanalysis II) in modeling surface relative shortwave cloud forcing (SRCF), cloud fraction, and cloud albedo. The cloud properties at diurnal, annual, and inter-annual temporal scales from all the reanalyses are first evaluated. Then, the model biases are quantified and their links to one another or to 2-m temperature/humidity are examined. The overall performance of the reanalyses in modeling the cloud properties and the 2-m temperature/humidity is evaluated using a combined statistical method (i.e., the technique of Taylor diagrams and a newly developed metric "Relative Euclidean Distance"). Decade-long (1997 to 2008) surface-based continuous ARM measurements over the Southern Great Plains (SGP) Central Facility site are used as a standard for this evaluation.

[58] Results show that the reanalyses significantly underestimate the cloud properties, with relative biases ranging mainly from −60% to −30% for SRCF, from −40% to −20% for cloud fraction, and from −30% to 10% for cloud albedo. The annual cycle of the models' relative biases is substantial. For SRCF, ERA-Interim has a uniform negative bias throughout the year whereas NCEP/NCAR Reanalysis I and NCEP/DOE Reanalysis II have much reduced biases in summer than in winter/spring. The annual cycles of the cloud-fraction relative biases differ between the three reanalyses. But for the derived cloud albedo, all the models show a small bias from May to September, increasing in the cold season to a large negative bias of −25% to −33%. NCEP/DOE Reanalysis II is the only reanalysis having similar diurnal phase pattern to the observed cloud properties, and NCEP/NCAR Reanalysis I is the only reanalysis showing an opposite phase pattern to the observed annual-cycle cloud fraction or SRCF.

[59] The model biases of the cloud properties predominantly range from −0.20 to 0 for SRCF, from −0.25 to 0 for cloud fraction, and from −0.20 to 0.10 for cloud albedo. The

model biases in general exhibit a positive linear relationship to one another. The model biases in the cloud properties increase with the cloud fraction observations, except for ERA-interim model biases in cloud fraction (or cloud albedo).

[60] In analyzing the relationship between cloud fraction and 2-m temperature/humidity, we found that the cloud fraction biases in the two NCEP reanalyses increase (decrease) with their relative humidity (temperature and specific humidity), but the cloud fraction biases in ERA-Interim show no (opposite) relationship with relative humidity (temperature and specific humidity). There is no clear relationship between cloud fraction biases and the temperature (or specific humidity) biases. We also found that the reanalyses have small (large) biases in modeling 2-m temperature/specific humidity (relative humidity), and the relative humidity biases have a positive (negative) linear relationship with the specific humidity (temperature) biases. These findings suggest that the model biases in the cloud properties are closely linked to the near-surface temperature/humidity fields, likely through the parameterizations used to couple boundary layer clouds with the land surface.

[61] A combined analysis of the Taylor diagram and the relative Euclidean distance, based on first- and second- order statistics, indicates that ERA-Interim and NCEP/NCAR Reanalysis I have the best and worst overall performance in modeling the cloud properties and the meteorological variables, respectively, except that R2 ranks the best in modeling the monthly temperature and specific humidity. The findings from this study highlight the significant underestimation of the cloud properties over the ARM SGP site in the three major NWP reanalyses, and the model biases of the cloud properties are closely linked to one another, and also linked to 2-m temperature/humidity. This suggests that caution must be taken when using the reanalyses as a standard (e.g., a substitute of observations) for evaluating climate models, especially considering the large biases in the forecasted 2-m relative humidity. Furthermore, the underestimation of the cloud properties is a crucial issue in climate modeling, since it could substantially influence the estimation of the Earth's surface and atmospheric energy budget, hydrological cycle and general circulation.

[62] Several points are noteworthy. First, we have shown that there is a potential uncertainty in comparing the ARM hemispheric observations with the spatially averaged model data from the reanalyses which use a plane-parallel assumption. A systematic investigation to quantify the uncertainty of cloud fraction observations and to develop new approaches to reduce the observational uncertainty [e.g., *Min et al.*, 2008; *Hogan et al.*, 2009] is needed to further assess observational effects discussed briefly in Section 3.2. Second, the ARM measurements and three reanalyses have different spatial and temporal resolutions, which could partially impact the analyzed results. Although aggregating large number of data points as done in this study tends to minimize the resolution effect, more work is needed to pin down the causes of the model biases shown in this study. Third, as pointed out in *Liu et al.* [2011], the current observational data of cloud fraction and cloud albedo are derived separately from surface-based radiation measurements. Such a separate treatment could lead to potentially compensating errors, especially when different FOVs are involved (Section 3.2). A retrieval approach to

simultaneously obtain cloud fraction and cloud albedo is desirable. Fourth, in this study we examined the relationships between the model biases in the cloud properties and the 2-m temperature/humidity. Detailed vertical profiles of temperature, humidity, vertical velocity and wind are crucial to examining the cause of the large model biases found in this study and further investigation along this line is under way. Finally, it is expected that the identified cloud-property biases in the reanalyses, their connections to near-surface meteorological variables, and the disparities between the three reanalyses arise largely from inadequate and different parameterizations of related processes/quantities in the models used to generate the corresponding reanalyses. Further relating the identified biases to specific parameterization deficiencies and improving the parameterizations call for in-depth investigation of model parameterizations, which are the challenges to the climate and NWP community in general, and to the reanalysis community in particular.

References

Ackerman, T. P., and G. M. Stokes (2003), The Atmospheric Radiation Measurement program, *Phys. Today*, *56*, 39–44.

Allan, R. P., and M. A. Ringer (2003), Inconsistencies between satellite estimates of longwave cloud forcing and dynamical fields from reanalyses, *Geophys. Res. Lett.*, *30*(9), 1491, doi:10.1029/2003GL017019.

Anderson, B. T., G. Salvucci, A. C. Ruane, J. O. Roads, and M. Kanamitsu (2008), A new metric for estimating the influence of evaporation on seasonal precipitation rates, *J. Hydrometeorol.*, *9*, 576–588, doi:10.1175/2007JHM968.1.

Anderson, J. L., et al. (2004), The new GFDL global atmosphere and land model AM2–LM2: Evaluation with prescribed SST simulations, *J. Clim.*, *17*, 4641–4673, doi:10.1175/JCLI-3223.1.

Bedacht, E., S. K. Gulev, and A. Macke (2007), Intercomparison of global cloud cover fields over oceans from the VOS observations and NCEP/NCAR reanalysis, *Int. J. Climatol.*, *27*, 1707–1719, doi:10.1002/joc.1490.

Betts, A. K. (2007), Coupling of water vapor convergence, clouds, precipitation, and land-surface processes, *J. Geophys. Res.*, *112*, D10108, doi:10.1029/2006JD008191.

Betts, A. K. (2009), Land-surface-atmosphere coupling in observations and models, *J. Adv. Model Earth Syst.*, *1*, art. #4, 18 pp., doi:10.3894/JAMES.2009.1.4.

Betts, A. K., and J. C. Chiu (2010), Idealized model for changes in equilibrium temperature, mixed layer depth, and boundary layer cloud over land in a doubled CO_2 climate, *J. Geophys. Res.*, *115*, D19108, doi:10.1029/2009JD012888.

Betts, A. K., and P. Viterbo (2005), Land-surface, boundary layer and cloud-field coupling over the south-western Amazon in ERA-40, *J. Geophys. Res.*, *110*, D14108, doi:10.1029/2004JD005702.

Betts, A. K., J. Ball, A. Barr, T. A. Black, J. H. McCaughey, and P. Viterbo (2006), Assessing land-surface-atmosphere coupling in the ERA-40 reanalysis with boreal forest data, *Agric. For. Meteorol.*, *140*, 365–382, doi:10.1016/j.agrformet.2006.08.009.

Betts, A. K., M. Köhler, and Y. Zhang (2009), Comparison of river basin hydrometeorology in ERA-Interim and ERA-40 reanalyses with observations, *J. Geophys. Res.*, *114*, D02101, doi:10.1029/2008JD010761.

Bosilovich, M. G., J. Chen, F. R. Robertson, and R. F. Adler (2008), Evaluation of global precipitation in reanalyses, *J. Appl. Meteorol. Climatol.*, *47*, 2279–2299, doi:10.1175/2008JAMC1921.1.

Chevallier, F., G. Kelley, A. J. Simmons, S. Uppala, and A. Hernandez (2005), High clouds over oceans in the ECMWF 15- and 45-yr reanalysis, *J. Clim.*, *18*, 2647–2661, doi:10.1175/JCLI3429.1.

Dee, D. P., E. Källén, A. J. Simmons, and L. Haimberger (2011a), Comments on "Reanalyses suitable for characterizing long-term trends," *Bull. Am. Meteorol. Soc.*, *92*, 65–70, doi:10.1175/2010BAMS3070.1.

Dee, D. P., et al. (2011b), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, *137*, 553–597, doi:10.1002/qj.828.

Elmore, K. L., and M. B. Richman (2001), Euclidean distance as a similarity metric for principal component analysis, *Mon. Weather Rev.*, *129*, 540–549, doi:10.1175/1520-0493(2001)129<0540:EDAASM>2.0.CO;2.

Gleckler, P. J., K. E. Taylor, and C. Doutriaux (2008), Performance metrics for climate models, *J. Geophys. Res.*, *113*, D06104, doi:10.1029/2007JD008972.

Haimberger, L., C. Tavolato, and S. Sperka (2008), Toward elimination of the warm bias in historic radiosonde temperature records—Some new results from a comprehensive intercomparison of upper-air data, *J. Clim.*, *21*, 4587–4606, doi:10.1175/2008JCLI1929.1.

Hogan, R. J., E. J. O'Connor, and A. J. Illingworth (2009), Verification of cloud-fraction forecasts, *Q. J. R. Meteorol. Soc.*, *135*, 1494–1511, doi:10.1002/qj.481.

Intergovernmental Panel on Climate Change (IPCC) (2001), *Climate Change 2001: The Scientific Basis, Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, edited by J. T. Houghton et al., 881 pp., Cambridge Univ. Press, New York.

Intergovernmental Panel on Climate Change (IPCC) (2007), *Climate Change 2007: The Physical Science Basis*, edited by S. Solomon et al., Cambridge Univ. Press, New York.

Jakob, C. (1999), Cloud cover in the ECMWF reanalysis, *J. Clim.*, *12*, 947–959, doi:10.1175/1520-0442(1999)012<0947:CCITER>2.0.CO;2.

Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. 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. Florino, and G. L. Potter (2002), NCEP–DOE AMIP–II reanalysis (R-2), *Bull. Am. Meteorol. Soc.*, *83*, 1631–1643, doi:10.1175/BAMS-83-11-1631.

Kassianov, E., C. N. Long, and M. Ovtchinnikov (2005), Cloud sky cover versus cloud fraction: Whole-sky simulations and observations, *J. Appl. Meteorol.*, *44*, 86–98, doi:10.1175/JAM-2184.1.

Kennedy, A. D., X. Dong, B. Xi, S. Xie, Y. Zhang, and J. Chen (2011), A comparison of MERRA and NARR reanalyses with the DOE ARM SGP data, *J. Clim.*, *24*, 4541–4557, doi:http://dx.doi.org/10.1175/2011JCLI3978.1

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

Liu, Y., W. Wu, M. P. Jensen, and T. Toto (2011), Relationship between cloud radiative forcing, cloud fraction and cloud albedo, and new surface-based approach for determining cloud albedo, *Atmos. Chem. Phys.*, *11*, 7155–7170, doi:10.5194/acp-11-7155-2011.

Long, C. N., and T. P. Ackerman (2000), Identification of clear skies from broadband pyranometer measurements and calculation of downwelling shortwave cloud effects, *J. Geophys. Res.*, *105*(D12), 15,609–15,626, doi:10.1029/2000JD900077.

Long, C. N., T. P. Ackerman, K. L. Gaustad, and J. N. S. Cole (2006), Estimation of fractional sky cover from broadband shortwave radiometer measurements, *J. Geophys. Res.*, *111*, D11204, doi:10.1029/2005JD006475.

Lu, C.-H., M. Kanamitsu, J. O. Roads, W. Ebisuzaki, K. E. Mitchell, and D. Lohmann (2005), Evaluation of soil moisture in the NCEP–NCAR and NCEP–DOE global reanalyses, *J. Hydrometeorol.*, *6*, 391–408, doi:10.1175/JHM427.1.

Martin, G. M., M. A. Ringer, V. D. Pope, A. Jones, C. Dearden, and T. J. Hinton (2006), The physical properties of the atmosphere in the new Hadley Centre Global Environmental Model (HadGEM1). Part I: Model description and global climatology, *J. Clim.*, *19*, 1274–1301, doi:10.1175/JCLI3636.1.

Miller, R. L., G. A. Schmidt, and D. T. Shindell (2006), Forced annular variations in the 20th century Intergovernmental Panel on Climate Change Fourth Assessment Report models, *J. Geophys. Res.*, *111*, D18101, doi:10.1029/2005JD006323.

Min, Q., T. Wang, C. N. Long, and M. Duan (2008), Estimating fractional sky cover from spectral measurements, *J. Geophys. Res.*, *113*, D20208, doi:10.1029/2008JD010278.

Peixoto, J. P., and A. H. Oort (1992), *Physics of Climate*, 520 pp., Springer, New York.

Pincus, R., C. P. Batstone, R. J. P. Hofmann, K. E. Taylor, and P. J. Glecker (2008), Evaluating the present-day simulation of clouds, precipitation, and radiation in climate models, *J. Geophys. Res.*, *113*, D14209, doi:10.1029/2007JD009334.

Rye, C. J., N. S. Arnold, I. C. Willis, and J. Kohler (2010), Modeling the surface mass balance of a high Arctic glacier using the ERA-40 reanalysis, *J. Geophys. Res.*, *115*, F02014, doi:10.1029/2009JF001364.

Simmons, A. J., K. M. Willett, P. D. Jones, P. W. Thorne, and D. P. Dee (2010), Low-frequency variations in surface atmospheric humidity, temperature and precipitation: Inferences from reanalyses and monthly gridded observational data sets, *J. Geophys. Res.*, *115*, D01110, doi:10.1029/2009JD012442.

Stokes, G. M., and S. E. Schwartz (1994), The Atmospheric Radiation Measurement (ARM) program: Programmatic background and design of the cloud and radiation test bed, *Bull. Am. Meteorol. Soc.*, *75*, 1201–1221, doi:10.1175/1520-0477(1994)075<1201:TARMPP>2.0.CO;2.

Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, *106*, 7183–7192, doi:10.1029/2000JD900719.

Thorne, P. W., and R. S. Vose (2010), Reanalyses suitable for characterizing long-term trends: Are they really achievable?, *Bull. Am. Meteorol. Soc.*, *91*, 353–361, doi:10.1175/2009BAMS2858.1.

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

Weidle, F., and H. Wernli (2008), Comparison of ERA40 cloud top phase with POLDER-1 observations, *J. Geophys. Res.*, *113*, D05209, doi:10.1029/2007JD009234.

Xie, S., et al. (2010), CLOUDS AND MORE: ARM climate modeling best estimate data, *Bull. Am. Meteorol. Soc.*, *91*, 13–20, doi:10.1175/2009BAMS2891.1.

Xu, K.-M. (2009), Evaluation of cloud physical properties of ECMWF analysis and re-analysis (ERA) against CERES tropical deep convective cloud object observations, *Mon. Weather Rev.*, *137*, 207–223. doi:http://dx.doi.org/10.1175/2008MWR2633.1