A Comparison of MERRA and NARR Reanalyses with the DOE ARM SGP Data

AARON D. KENNEDY, XIQUAN DONG, AND BAIKE XI
Department of Atmospheric Sciences, University of North Dakota, Grand Forks, North Dakota

SHAOCHENG XIE AND YUNYAN ZHANG
Lawrence Livermore National Laboratory, Livermore, California

JUNYE CHEN
Earth System Science Interdisciplinary Center, University of Maryland, College Park, College Park, Maryland

(Manuscript received 5 August 2010, in final form 18 February 2011)

ABSTRACT

Atmospheric states from the Modern-Era Retrospective analysis for Research and Applications (MERRA) and the North American Regional Reanalysis (NARR) are compared with data from the Atmospheric Radiation Measurement Program (ARM) Southern Great Plains (SGP) site, including the ARM continuous forcing product and Cloud Modeling Best Estimate (CMBE) soundings, during the period 1999–2001 to understand their validity for single-column model (SCM) and cloud-resolving model (CRM) forcing datasets. Cloud fraction, precipitation, and radiation information are also compared to determine what errors exist within these reanalyses. For the atmospheric state, ARM continuous forcing and the reanalyses have good agreement with the CMBE sounding information, with biases generally within 0.5 K for temperature, 0.5 m s\(^{-1}\) for wind, and 5% for relative humidity. Larger disagreements occur in the upper troposphere (\(p < 300 \text{ hPa}\)) for temperature, humidity, and zonal wind, and in the boundary layer (\(p > 800 \text{ hPa}\)) for meridional wind and humidity. In these regions, larger errors may exist in derived forcing products. Significant differences exist for vertical pressure velocity, with the largest biases occurring during the spring upwelling and summer downwelling periods. Although NARR and MERRA share many resemblances to each other, ARM outperforms these reanalyses in terms of correlation with cloud fraction. Because the ARM forcing is constrained by observed precipitation that gives the adequate mass, heat, and moisture budgets, much of the precipitation (specifically during the late spring/early summer) is caused by smaller-scale forcing that is not captured by the reanalyses. While reanalysis-based forcing appears to be feasible for the majority of the year at this location, it may have limited usage during the late spring and early summer, when convection is common at the ARM SGP site. Both NARR and MERRA capture the seasonal variation of cloud fractions (CFs) observed by ARM radar–lidar and Geostationary Operational Environmental Satellite (GOES) with high correlations (0.92–0.78) but with negative biases of 14% and 3%, respectively. Compared to the ARM observations, MERRA shows better agreement for both shortwave (SW) and longwave (LW) fluxes except for LW-down (due to a negative bias in water vapor): NARR has significant positive bias for SW-down and negative bias for LW-down under clear-sky and all-sky conditions. The NARR biases result from a combination of too few clouds and a lack of sufficient extinction by aerosols and water vapor in the atmospheric column. The results presented here represent only one location for a limited period, and more comparisons at different locations and longer periods are needed.

1. Introduction

In the past decade, reanalysis datasets have become increasingly common to study a variety of meteorological and climatological questions. Reanalyses blend observation and model output to create a systematic long-term description of the climate system. While it is an excellent strategy to use model output to fill gaps in the observing systems and to diagnose variables unable to be measured directly, reanalyses are not error free because of the limitations of model and assimilation technology. Because the errors of reanalyses and their underlying models are
relatively unknown, their benefit for answering more complex questions involving the climate is questionable. For this reason, reanalyses have been used sparingly to generate forcing that provides initial and boundary conditions for single-column model (SCM) and cloud-resolving model (CRM) studies that can help develop improvements for general circulation models (GCMs).

To circumnavigate these issues, extensive work has been done to derive forcing using constrained variational analysis from observations during intensive observation periods (IOPs) at the Department of Energy (DOE) Atmospheric Radiation Measurement Program (ARM) sites (Zhang and Lin 1997; Zhang et al. 2001). More recently, Xie et al. (2003) evaluated the forcing datasets derived from the European Centre for Medium-Range Weather Forecasts (ECMWF) during three IOPs at the ARM Southern Great Plains (SGP) site. They found that although the two forcing datasets correlated well, the ECMWF-derived forcing was much weaker, owing to limitations in the model-predicted surface radiation and precipitation fields. Unfortunately, IOPs are expensive to run from a monetary and workload perspective. Continuously run models, however, offer long-term datasets that are valuable from a climate study perspective. To combine the benefits of long-term model results and high-quality IOP observations, Xie et al. (2004) developed a continuous forcing dataset using a combination of a model (atmospheric state variables, such as temperature, humidity, among others) from the Rapid Update Cycle 2 (RUC-2; Benjamin et al. 2004) and surface and top-of-atmosphere (TOA) observations at the ARM SGP site. Although this forcing dataset has shown good agreement with forcing developed during IOPs (Xie et al. 2004), a thorough comparison with long-term observed soundings has not been completed.

In recent years, several new reanalyses have been released, including the National Centers for Environmental Prediction (NCEP) North American Regional Reanalysis (NARR; Mesinger et al. 2006) and the Modern-Era Retrospective analysis for Research and Applications (MERRA) reanalysis (Rienecker et al. (2011)). Compared to their predecessors, these new reanalyses have been improved significantly. For example, NARR includes an assimilation of precipitation at a high resolution over North America and has shown improvements over the NCEP/Department of Energy Global Reanalysis 2 for a variety of variables (Mesinger et al. 2006). MERRA, conversely, assimilates numerous satellite datastreams to better constrain the earth’s energy and hydrologic budgets. Both reanalyses provide diagnostic outputs at high temporal and spatial resolutions that can be used to develop SCM/CRM forcing. To determine whether there is merit in developing SCM/CRM forcing from reanalyses, however, it is important to evaluate these reanalyses against both observations and the ARM continuous forcing data.

The primary purpose of this paper is to evaluate the NARR and the MERRA reanalysis at the ARM SGP site to determine their applicability for SCM/CRM forcing. This study will investigate whether these datasets can faithfully represent the atmospheric state and have limited errors due to their own model parameterizations that represent subgrid-scale processes. The similarities between the reanalyses and an existing forcing dataset will also be explored. To achieve this goal and to further validate the ARM continuous forcing, atmospheric soundings have been collected at the ARM SGP site during the period 1999–2001. Atmospheric state parameters, such as temperature, relative humidity (RH), and wind, will be compared between NARR, MERRA, and ARM continuous forcing with ARM Cloud Modeling Best Estimate (CMBE) soundings.

A secondary goal of this paper is to provide a much-needed benchmark of these reanalyses for other atmospheric variables. Additional observations at the ARM SGP site, including ground-based observed cloud, precipitation, and radiative fluxes, provide a unique dataset to evaluate these parameters from the reanalyses. The ARM SGP site is representative of a continental climate in the midlatitudes, and it has been used successfully in the past to evaluate a variety of model simulations, including the NCEPEta Model (Hinkelmann et al. 1999), ECMWF (Xie et al. 2004), and the NCEP Global Forecast System (GFS) (Yang et al. 2006). This 3-yr comparison will assist in determining whether parameterization issues exist in these recent reanalyses and aid other reanalysis users by providing error bars for commonly used variables. Studies of this nature have been encouraged by recent studies, such as Thorne and Vose (2010), that have sought to understand whether reanalyses can be used for diagnosing long-term trends.

This paper is formatted as follows: Section 2 gives a brief summary of the various datasets used in this study. In section 3, the atmospheric state is compared between ARM CMBE soundings, the reanalyses, and the ARM continuous forcing during the period 1999–2001. Cloud fraction, total precipitation, and radiative fluxes are then compared in section 4. A summary of findings and concluding remarks are provided in section 5.

2. Datasets

ARM CMBE (Xie et al. 2010) soundings, ARM continuous forcing, NARR, and the MERRA reanalysis have been collected at the ARM SGP site for the period 1999–2001. These 3 yr were chosen because the ARM
continuous forcing dataset is only available during this period. To have cloud information at the ARM SGP site, surface observations from a vertically pointing cloud radar and micropulse lidar pair have also been collected along with Geostationary Operational Environmental Satellite (GOES) observations. All forcing and reanalyses have been processed to identical temporal and spatial resolutions for comparison in sections 3 and 4. For example, the results from the two reanalyses are averaged in space to the domain of the ARM forcing, while the hourly continuous forcing is averaged in time to three-hourly increments to match the reanalyses.

a. ARM CMBE soundings

To provide vertical profiles of temperature, humidity, and winds, ARM has launched Vaisala RS-80 and RS-90 balloon soundings up to 4-times (0000, 0600, 1200, and 1800) daily at the ARM SGP Central Facility (SCF; 36.6°N, 97.5°W). In this study, this sounding information is provided by the ARM CMBE product, which uses the Microwave Radiometer-Scaled Sonde Profiles (LSSONDE) datastream (Turner et al. 1998). Because Vaisala RS-80 soundings have a known dry bias (Turner et al. 2003), this datastream uses an empirical technique to scale sounding moisture profiles to match the microwave radiometer–derived atmospheric precipitable water vapor.

b. ARM continuous forcing

The ARM continuous forcing dataset developed for SCM/CRM applications is centered on the ARM SCF and is provided from January 1999 to December 2001. This forcing uses ARM surface and GOES-8 satellite observations as constraints to adjust atmospheric state variables to conserve the column-integrated mass, heat, and moisture through a variational analysis approach (Zhang and Lin 1997; Zhang et al. 2001). The forcing atmospheric state is provided by hourly RUC-2 (see Benjamin et al. 2004) analyses because of the lack of continuous sounding measurements (Xie et al. 2004). A comparison of the continuous forcing with selected IOPs by Xie et al. (2004) found root-mean-square errors (RMSEs) within 1 m s⁻¹ for horizontal wind, 0.5 K for temperature, and 0.5 g kg⁻¹ for moisture for the atmospheric column. The forcing represents an average over a circular area approximately 180 km in radius centered on the ARM SCF (an area bounded by a box of approximately 4.1° × 3.3° in longitude by latitude).

c. NARR

The NCEP NARR is a long-term (1979–present) climate dataset with 3-h temporal, 32-km horizontal, and 45-layer vertical resolutions over the North American domain (Mesinger et al. 2006). It contains outputs of many atmospheric variables and fluxes, and is nicely suited for diagnosis of synoptic and mesoscale conditions over the ARM SGP site. NARR uses the operational NCEP Eta Model and its three-dimensional variational data assimilation (3DVAR) technique on a wide variety of observation platforms, but it was principally developed to improve on NCEP reanalysis by assimilating precipitation accurately. Studies by Becker et al. (2009) and Bukovsky and Karoly (2007) found that this statement is generally true for NARR.

d. MERRA reanalysis

NASA released the MERRA reanalysis dataset in 2010 based on the Goddard Earth Observing System Data Analysis System, Version 5 (GEOS-5 DAS; Rienecker et al. 2011). This global reanalysis covers the same period as NARR (1979–present). MERRA takes advantage of a variety of recent satellite datastreams—for example, the observations from the National Aeronautics and Space Administration (NASA)’s Earth Observing System (EOS)—to improve the representation of the earth’s energy and water cycles. GEOS-5 includes the GEOS-5 AGCM and the gridpoint statistical interpolation (GSI) atmospheric analysis developed jointly with the National Oceanic and Atmospheric Administration (NOAA)’s NCEP Environmental Modeling Center (EMC). Incremental analysis update (IAU) technique (Bloom et al. 1996) is incorporated into the GEOS-5 to minimize the 6-hourly shock from the observation input. The model has a native spatial resolution of 72 layers in the vertical and 1/2° × 1/2° in the horizontal. In addition to the 6-hourly 3D analyses at the native spatial resolution, MERRA also provides 1-hourly 2D diagnostics at 1/8° × 1/8° resolution and 3-hourly 3D diagnostics at 1.25° × 1.25° resolution on 42 vertical levels.

e. Cloud observations

For several portions of the study, cloud information is used to determine relationships with the atmospheric state and clear-sky radiative fluxes. Cloud information comes from two sources. Ground-based observations from the ARM 35-GHz millimeter wavelength cloud radar (MMCR; Moran et al. 1998) are combined with a Belfort laser ceilometer and Micro Pulse Lidar (MPL) to determine cloud bases, tops, and vertical distributions. While information is collected at 5-min intervals, it has been binned to 1-h cloud fractions (CFs) at the resolution of the forcing in a fashion identical to that described in Xi et al. (2010) and Kennedy et al. (2010). This cloud product is similar to the Active Remotely Sensed Cloud Locations (ARSCL; Clothiaux et al. 2000)
cloud product except the original datastream is the Mace principal investigator product (Mace et al. 2006), which merges the original radar modes differently. Considering cloud information is only used at a 1–3-hourly resolution, the differences between the two products is negligible.

The second source of cloud information is total cloud fractions derived from Visible Infrared (IR) Solar-Infrared Split-Window Technique (VISST)-retrieved satellite cloud products (Minnis et al. 2010) using algorithms developed for the NASA Clouds and the Earth’s Radiant Energy System (CERES) project. Cloud properties are retrieved from half-hourly 4-km 0.65, 3.9, 10.8 (IR), and 12.0-µm radiances taken by GOES-8. Cloudy pixels are identified using an adaptation of the method described by Minnis et al. (2008). The areal fraction of clouds [or the amount when present, (AWP)] is the ratio of the number of pixels classified as cloudy to the total number of pixels within a specified area. Cloud fraction is then calculated at the resolution of the forcing by considering the quantity of 0.5° × 0.5° grid boxes contained within the area of interest. Once again, this methodology is consistent with that used in the Xi et al. (2010) and Kennedy et al. (2010) studies. The reader is referred to these publications for additional details on the process.

3. Atmospheric state

a. Comparison with CMBE soundings

ARM CMBE soundings are used as ground truth to evaluate the atmospheric states of the reanalyses and ARM continuous forcing (ARM). There are approximately 3700 temperature and wind soundings and 2400 humidity soundings available during the 3-yr period. The actual number of samples, however, varies by pressure level because of balloons malfunctioning at different heights. Note that although the soundings are used as ground truth, they represent point observations that vary in location by height and are not area averaged, such as the forcing and reanalysis datasets. For any given launch, a sounding may have a positive or negative anomaly as compared to the other datasets because of inhomogeneity in both space and time. After averaging many samples, however, these point observations can be directly compared with those for the region encompassed by the ARM forcing and reanalyses. Kennedy et al. (2010) and Xi et al. (2010) show that this is a valid comparison for the point observations (MMCR) of cloud fraction at the ARM SGP cloud fraction (CF) versus model simulations and satellite observations of cloud fraction, which cover a 2.5° × 2.0° region.

Averaged vertical profiles of bias and RMSEs during the 3-yr period are given in Fig. 1. Temperature is characterized by having biases <0.5 K within the low–midtroposphere for all three datasets (Fig. 1a). Biases become larger in the upper troposphere at 300–100 hPa. Although the reanalyses are generally within 0.5 K of the soundings in this region, ARM has a large bimodal bias that peaks at ~1.5 K, corresponding with RMSE values nearly double of the reanalyses at this level (Fig. 1e). This large discrepancy appears to be associated with the tropopause height, which varies by season (not shown). The large RMSEs near the ground are likely related to the aforementioned issues of spatial and temporal inhomogeneity.

Zonal and meridional wind biases are provided in Figs. 1b and 1c, respectively. These values are slightly negative and typically within 0.5 m s⁻¹ of the observed except for two cases. First, ARM has a larger negative bias (and corresponding increase in RMSE, Fig. 1f) in the upper troposphere with a peak bias at ~2 m s⁻¹. Second, all three datasets have weaker meridional winds within the boundary layer (~1.7 m s⁻¹). Histograms for these variables reveal that these biases are primarily caused by a lack of stronger winds in both cases (not shown).

Relative humidity biases have several interesting features (Fig. 1d). Near the ground, biases are within 3% of the soundings for ARM and NARR, albeit of opposite sign; ARM has a moist bias, while NARR is drier than the soundings. These RH biases correspond to errors in the mixing ratio of less than 0.5 g kg⁻¹. MERRA also has a dry bias that is much larger than the other two datasets (~6%, 1 g kg⁻¹). Near the level of non-divergence (~400–500 hPa), all biases change in sign from negative to positive. The MERRA bias has a peak of 8% near 300 hPa and then decreases toward 0% at 100 hPa; ARM and NARR have positive biases that peak around 15%–20% at 100–200 hPa. Although not shown, humidity biases vary by season. Most notably, the boundary layer moist bias for ARM increases in depth and magnitude during the summer. The MERRA moist bias in the upper troposphere is also larger during the summer months and doubles during periods of precipitation.

To better understand these humidity biases, histograms were calculated at 925 and 200 hPa (Fig. 2), which represent the boundary layer and near the tropopause, respectively. All four datasets show gammalike distributions for 925-hPa RH that peak at values between 38% and 55% (Fig. 2a). The moist bias for ARM can be seen as values greater than the soundings between 35% and 75%. At values greater than 75%, the sounding frequency is greater than for all three datasets. This feature is likely a function of the sounding being a point observation versus area average. Despite this issue,
FIG. 1. Biases of ARM continuous forcing (solid), NARR (dotted), and MERRA (dashed) to ARM CMBE sounding profiles for the period 1999–2001 for (a) temperature, (b) zonal wind, (c) meridional wind, and (d) RH. (e)–(h) As in (a)–(d), but for the RMSE.
however, Fig. 2a clearly shows that MERRA is dry, as its distribution is shifted approximately 5%–10% to the left of the other datasets.

At 200 hPa, the RH histograms span a variety of distributions. The frequencies of the sounding and MERRA have near-linear distributions that gradually decrease from 0.15 at RH = 5% to zero at RH = 60%. ARM’s distribution is quasi normal, with a peak at RH ~ 30%. NARR has a drastically different distribution: it is narrower with a peak frequency at RH ~ 40%, a factor of 2 higher than other distributions. This issue only occurs at pressure levels near or above the tropopause.

Scatterplots of RH at 200 hPa reveal the nature of the different distributions (Fig. 3). ARM has a linear relationship with the sounding data (Fig. 3a), with positive biases during periods of drier conditions (RH ~ 0%–20%). NARR’s scatterplot has a logarithmic curve that quickly rises under drier conditions and levels off for other RH values (Fig. 3b). Of the three datasets, MERRA shows the best agreement with the ARM CMBE soundings but with noticeable scatter (Fig. 3c).

b. **ARM forcing and reanalyses intercomparison**

Both humidity and vertical pressure velocity are crucial for developing accurate forcing required by SCM/CRM applications. For example, the RH biases in the boundary layer and near the tropopause, as shown in Fig. 1, will directly translate to biases in the modeled cloud simulations because RH is often considered as a trigger to generate clouds in the stratiform cloud parameterizations of these models. For this reason and for the fact that vertical velocities are difficult to measure directly, these two variables warrant additional investigation. In doing so, it may be possible to determine whether the reanalyses have issues within their own parameterizations. Sampling issues and the inability of soundings to measure vertical velocity limit this investigation to only the ARM continuous forcing and reanalyses.

The seasonal variations of RH and vertical pressure velocity (omega) derived from the ARM continuous forcing and the NARR and MERRA reanalyses over
the ARM SGP site during the period 1999–2001 are provided in Fig. 4. As illustrated in Figs. 4a and 4b, the RH values derived from ARM and NARR are in good agreement and have a bimodal distribution with peaks in the boundary layer and in the upper troposphere. This is consistent with the seasonal variation of radar-lidar-derived cloud fraction at the ARM SGP site (Kennedy et al. 2010). The decrease in RH during the late summer (August–September) is primarily due to the influence of large-scale ridging and a lack of baroclinic wave activity over Oklahoma. The RH differences between ARM and NARR occur near the top of the troposphere during the summer and in the boundary layer throughout the year, which is consistent with Fig. 1. The former of these two differences suggests that there may be an issue with ARM/RUC-2 because there is no physical explanation for higher humidity at this level during the summer months. Despite these differences, monthly maximums are present in both datasets during January and March. MERRA captures the general shape of RH at the ARM SGP site (Fig. 4c) but with a ~5% negative bias throughout the year in the upper troposphere—except during late spring and early summer, when convection is most common at the ARM SGP site. During this period, MERRA has a considerable positive bias (~10%–15%) compared to ARM and NARR. Seasonal RMSE plots (not shown) demonstrate that the largest disagreement between MERRA and ARM continuous forcing for the mixing ratio occurs during the spring [March–May (MAM)] and summer seasons [June–August (JJA)] in the boundary layer and upper troposphere. The maximum RH for MERRA occurs during June, when boundary layer humidity is highest. As will be shown later, cloud fraction in MERRA also peaks in June, suggesting that this may be a by-product of the convective parameterization used in the AGCM. This is also supported by the fact that the RH bias in the upper troposphere doubles during periods of precipitation in the summer months. Like ARM and NARR, additional peaks occur during January and March. It is concluded that the seasonal cycle of RH from the three different datasets generally agree during this 3-yr period, except for the upper troposphere during the summer months.

Contrary to the RH comparison, significant differences exist for the omega field, as shown in Figs. 4d–f. As illustrated in Fig. 4d, there are two periods of upwelling (cool colors) for the ARM dataset: one during the late spring, from May to June, peaking at ~1.75 mb h⁻¹ and the other in the early fall, during September–October, with weak upward motion. Downwelling branches occur during the late fall/early winter and the late summer in the lower troposphere. Although NARR and MERRA omega values are similar to each other, they differ considerably from the ARM dataset. NARR is characterized by capturing the seasonal pattern of omega, however, with much different amplitudes than ARM. For the upwelling motion, the largest upward motion in NARR occurs during March (Fig. 4e) instead of the late spring (May–June) as shown in Fig. 4d. The upward motion during the early fall is also much weaker, whereas the downwelling motion is notably stronger than ARM, with maximum values around ~1 mb h⁻¹. This is most evident during the summer months, when the downwelling branch extends throughout the atmospheric column. MERRA (Fig. 4f) shares many resemblances with NARR, especially with regard to the weaker spring upwelling and the stronger summer downwelling. Perhaps the most unique feature with MERRA is that the upward motion is largest near the surface, just above the PBL.

To further investigate the omega differences between the three datasets, the histograms of 3-hourly omega at 300 hPa for all and nonprecipitating (dry hours) periods are presented in Fig. 5. Despite having a large positive bias compared to ARM, NARR occasionally produces larger upward motions (Fig. 5a). These upward motions, however, disappear under the dry period (Fig. 5b), indicating that these upward motions are associated with precipitation. It is believed that these large upward velocities result from spurious grid-scale precipitation (SGSP), as first documented by West et al. (2007). In brief, the mismatch between assimilated and Eta modeled precipitation used in NARR introduces spurious latent heating, which in turn causes unreasonable upward velocities, usually during times of convection. Given this only occurred several dozen times during the 3-yr period, this study agrees with the West et al. (2007) finding that “SGSP will probably have little or no effect on long-term hydrometeorological averages prior to 2003.” This phenomenon is a nonissue in MERRA, which has a smaller tail for upward velocities. Figures 5a and 5b demonstrate that both NARR and MERRA have more downward motion than ARM at the 300-hPa level, which is consistent with the results in Fig. 4.

Determining which dataset is closer to the atmospheric “truth” is a difficult question to answer without direct measurements of vertical velocity. Therefore, it is necessary to find other observed parameters that may be related to vertical velocity to evaluate the three datasets during the 3-yr period. In this study, it is hypothesized that a more accurate large-scale relative humidity and vertical motion field will have a stronger relationship with observed cloud fraction. This has the added benefit of accessing the validity of cloud parameterizations that use these variables to predict cloud fraction.

Correlations were calculated between 3-h mean RH, omega, and cloud fraction as determined by the ARM
Fig. 4. Monthly means of RH over the ARM SGP domain for the period 1999–2001 for (a) ARM continuous forcing, (b) NARR, and (c) MERRA. (d)–(f) As in (a)–(c), but for the omega field.
MMCR/MPL data at the ARM SGP site during the 3-yr period (Fig. 6). RH correlations were calculated for matching layers of RH and cloud fraction (i.e., 925-hPa RH correlated with 925-hPa cloud fraction). For omega, correlations were calculated at an observed CF pressure level against 300-hPa omega. These correlations (Fig. 6b) were higher than those calculated at each level (i.e., 925-hPa CF correlated with 925-hPa omega) because the vertical motion is typically small and more turbulent at lower levels. The results are qualitatively the same as choosing another upper-troposphere pressure level, such as 500 hPa. The 300-hPa level was chosen on the basis that correlations were slightly higher than at other levels. Since these RH and omega correlations are calculated from a point observation (CF derived from ARM radar lidar) and a forcing domain-averaged mean (RH and omega), these correlations may be lower than reality because clouds might occur elsewhere in the forcing domain but were not observed by ARM radar–lidar. As illustrated in Fig. 6a, the vertical distributions of the CF and RH correlations for the three datasets are nearly identical, although values are slightly higher for ARM. Overall, RH has a moderate correlation with CF and is characterized by being bimodal, with peak values of 0.5–0.6 at the top of the boundary layer and the upper troposphere. A larger value at the lowest levels for MERRA is a result of fewer samples at the first level; unlike NARR, MERRA does not calculate variables below ground level (i.e., surface pressure less than the pressure level). Correlations for omega (Fig. 6b) are highest (~0.45) and peak at a level of 450 hPa. MERRA falls between ARM and NARR, with a peak value of approximately ~0.4, and it has a similar vertical distribution to those of ARM and NARR, although it is slightly bimodal. In the upper troposphere, however, the rate of change in the MERRA correlation is much smaller, which results in higher correlations than those of ARM and NARR. This is most likely caused by a sampling issue because the vertical resolution of MERRA is less than those from NARR and ARM above 300 hPa (50 versus 25 hPa).

To understand the seasonal variation of the RH/omega relationship with cloud fraction, Fig. 7 is produced. The RH correlations from the three datasets have similar seasonal variations with a large range. Correlations are highest from late fall to early spring, when clouds are more closely linked to baroclinic wave activity. Correlations then decrease until they reach their lowest level.
(<0.2) during the months of July and August, suggesting that cloud parameterizations that are dependent on RH to trigger clouds may perform poorly during this season and need to be improved in the future.

The omega comparison basically follows that for RH, except for a few important features. In particular, ARM correlations (Fig. 7d) have maxima during the months of February, April, and June. Although NARR and MERRA (Figs. 7e and 7f) capture the peaks for the winter and early spring months, they do not have a maximum during June. This warrants further investigation. Given that the ARM forcing is constrained by
precipitation, this suggests that during the late spring and early summer, precipitation is more likely caused by local forcing [i.e., isolated thunderstorms developing along weak boundaries with weak synoptic-scale support (Dong et al. 2010)] that cannot be captured by the reanalyses. Like the RH comparison, ARM correlations are slightly higher (0.1–0.2) than those of NARR and MERRA at any given time and height. In other words, ARM, NARR, and MERRA all agree on the hour-to-hour variation of vertical velocity and its relationship to cloud occurrence.

c. Feasibility of reanalysis-based forcing

The results from this section paint a mixed picture on how useful reanalysis-based forcing may be at the ARM SGP site. While biases against ARM CMBE soundings are quite small and even better than the ARM continuous forcing in some cases, there is significant disagreement for vertical pressure velocity. The reanalyses capture the relationships between RH and omega with clouds and their seasonal variations, except during the month of June. During this month, convection is common at the ARM SGP site and either the reanalyses cannot capture the forcing mechanism for the precipitation or convection is too weak. When MERRA simulates convection during this month, however, it transports too much water vapor to the upper troposphere, resulting in a positive bias of RH. This is a perfect example of a model parameterization issue that will significantly affect the SCM/CRM simulations.

It is concluded that reanalysis-based forcing is a worthy endeavor despite the noted issues. Forcing should have many similarities to the ARM continuous forcing during the majority of the year. Efforts are already underway to produce forcing based on the NARR at this location. Based on the results of this study, the forcing will be developed both with and without postprocessed vertical velocity information. It is believed that this forcing will provide a good resource when either is used in tandem with the ARM continuous forcing, or at locations that do not frequently experience convection.

4. Precipitation, cloud fraction, and surface radiation

In this section, the precipitation, cloud fraction, and surface radiation fields from both NARR and MERRA are evaluated with observations at the DOE ARM SGP site during the period 1999–2001. As shown in Fig. 8, ARM and NARR precipitation have excellent agreement with each other, capturing the monthly variability in precipitation during this period, which should be expected, given the design of NARR to assimilate observed precipitation. This is certainly not a new finding because it has been documented in Becker et al. (2009) and Bukovsky and Karoly (2007). The largest precipitation amounts occur during the month of June, followed by the earlier spring, and fall months. For many months, the two lines are nearly indistinguishable. Conversely, MERRA appears to have a negative bias for most of the 3-yr period. Despite this bias, however, it does capture the monthly variability of precipitation. Figure 9 shows the scatterplots of the monthly and daily total precipitation for the three datasets. As demonstrated in Figs. 8 and 9a, NARR monthly total precipitation has excellent agreement with ARM forcing, with a correlation of 0.99 and bias of −2.8 mm. MERRA monthly total precipitation (Fig. 9b), however, has a larger bias of −22.2 mm. Despite this bias, there is still a linear trend with a strong correlation of 0.86. Precipitation is also over simulated on occasion during low-precipitation months (<50 mm), hence the intercept of 15.66 mm.
Reducing precipitation to daily totals leads toward more disagreement between ARM and reanalyses, as noted by the smaller values of slope and correlation. For NARR (Fig. 9c), the slope is reduced from 0.96 to 0.86 and the correlation from 0.99 to 0.91. Overall, there is an approximate -0.1 mm bias per day. This panel is similar to the “Great Plains” panel in Fig. 2 from Becker et al. (2009). The more significant scattering and values at 0 for one dataset suggest that the assimilation process might introduce some uncertainty into the original observations, either in time and/or location. Becker et al. (2009) found that, in general, NARR has less intensity and higher-frequency precipitation than the observations, so some care should be taken in analyzing individual cases. Daily precipitation correlation for MERRA (Fig. 9d) is reduced to 0.69 with a bias of -0.73 mm.

Figure 10 shows the CF comparison between ARM radar lidar, GOES, NARR, and MERRA at the ARM SGP site during the period 1999–2001. The monthly CF difference between ARM radar lidar and GOES observations may be due to the spatial-scale difference (point versus a 2.5° × 2° grid box) and remote sensing method (active vs passive). The annual mean CF difference between ARM radar lidar and GOES observations is within 1% (43% vs 44%) for the entire 3-yr period. This result is consistent with the findings in the Xi et al. (2010) and Kennedy et al. (2010) studies. Cloud fraction is characterized by having maximum values during the late winter and spring (peaking in March) and then having another local maximum during June, when precipitation and upward motion peaks. CF then decreases to a minimum during the summer, when Oklahoma is typically under large-scale ridging. Both NARR and MERRA reanalyses capture the same seasonal variations as the ARM radar lidar and GOES observations but with negative biases. Of the two, however, MERRA has better agreement with a larger maximum during June and is overall within 3%–4% of observations. Correlations and RMSEs between the reanalyses and observations are also calculated based on a total of 36 monthly means and are summarized in Table 1.
has a larger RMSE against both ARM and GOES observations than MERRA, its correlations are higher, indicating that NARR captures month-to-month variability better. Note that the CF correlation between ARM and GOES is 0.91 and the RMSE is 5.8%. While the CF correlation is higher for NARR than ARM, the correlation between GOES and MERRA is nearly the same as that between GOES and NARR, and the RMSE values for MERRA are much smaller than those for NARR. This may be a matter of MERRA incorporating GOES data into its assimilation process.

Comparisons of monthly-mean surface fluxes for clear-sky and all-sky conditions from the three datasets are shown in Fig. 11 and summarized in Table 2. For a detailed discussion, refer to the Dong et al. (2006) study, which investigated the seasonal variations of CF and surface radiative fluxes at the ARM SGP during the period 1997–2002. Despite the slightly longer period in the Dong et al. (2006) study, the differences between this study (ARM results) and Dong et al. (2006) are within a few watts per square meter, as listed in Table 2.

Overall, the reanalyses capture the seasonal variability seen in ARM quite well, albeit with biases (Table 2). These biases are smallest for periods of clear-sky, which is expected; surface fluxes in reanalyses are dependent not only on their parameterizations for surface radiation but also clouds. Compared to the all-sky ARM results, the NARR SW-down (SWDN) is significantly higher (47 W m$^{-2}$) and LW-down (LWDN) is lower ($-9$ W m$^{-2}$), which is consistent with the negative bias of cloud fraction found in Fig. 10. Markovic et al. (2009) found similar results for NARR analyzed at six surface sites within the United States and suggested that high biases in mean annual all-sky SW-down ($-40$ W m$^{-2}$) were attributed to a negative bias of CF. The clear-sky comparisons are nearly the same as their all-sky counterparts; that is, SW-down is 23 W m$^{-2}$ higher and LW-down is $-7$ W m$^{-2}$ lower, suggesting that the impacts of water vapor and aerosols on radiative transfer in NARR also need to be improved. Given that NARR is based on the NCEP Eta Model, this is consistent with Hinkelman et al. (1999), who found that the Eta Model had an average excess of 50 W m$^{-2}$ for SW-down, with approximately half of this bias attributed to deficient extinction.

The comparisons between MERRA and ARM agree much better than those between NARR and ARM, as shown in Fig. 11 and listed in Table 2. However, there are a few exceptions. MERRA has larger biases than NARR for LW-down under both clear-sky and all-sky conditions ($-20$ and $-19$ W m$^{-2}$). Compared to ARM and NARR, these negative biases are consistent with the drier conditions in MERRA, as demonstrated in Figs. 1, 2, and 4, and the seasonal variations of precipitable water vapor (not shown). Atmospheric water vapor is extremely important for LW-down fluxes under both clear-sky and all-sky conditions (Dong et al. 2006) and is supported by the fact that these biases are largest during the warm season.

Finally, comparisons of monthly-mean TOA fluxes for clear-sky and all-sky conditions are given in Fig. 12 and are summarized in Table 3. Reanalysis fluxes under clear-sky condition have small positive biases within 5 W m$^{-2}$ of ARM (GOES) observations. As expected, TOA SW-up (SWUP) fluxes for all-sky condition are highest during months with high cloud fraction, and the differences between reanalyses and ARM are related to their CF differences. For example, NARR TOA flux biases [negative for SW-up and positive for LW-up (LWUP)] are consistent with the year-round negative CF bias. MERRA biases vary by season depending on the amount of cloud cover produced. The peak in SW-up and minimum in LW-up during June are strongly associated with the peak of CF during that month. Despite this disagreement, biases in MERRA are noticeably smaller than those of NARR, as listed in Table 3.

5. Summary and conclusions

The atmospheric state, precipitation, total cloud fraction, and surface radiative fluxes from MERRA and
NARR reanalyses were collected and compared with the ARM SGP continuous forcing and ARM CMBE sounding datasets during the period 1999–2001. The key findings are summarized below.

1) For atmospheric state, biases between the sounding profiles and ARM forcing, and reanalyses are typically within 0.5 K for temperature, 0.5 m s\(^{-1}\) for wind, and 5% for RH. Several areas of disagreement

FIG. 11. Monthly-mean clear-sky (a) SW-down, (b) LW-down, (c) SW-up, and (d) LW-up fluxes measured by Eppley Precision Spectral Pyranometers (PSPs) and Eppley Precision Infrared Radiometers (PIRs) at the ARM SGP site. (e)–(h) As in (a)–(d), but for all-sky conditions.
exist in the boundary layer and in the upper troposphere near and above the tropopause. In the boundary layer, the reanalyses and ARM forcing have dry and moist biases, respectively, and these biases are greatest during the spring and summer seasons. Near and above the tropopause, the ARM forcing has a large positive bias for temperature, a moist bias for humidity, and a negative bias for zonal winds. These biases are larger than the reanalyses except for NARR humidity, which also has significant positive bias as compared to the sounding profiles. These humidity issues may cause errors in cloud simulations for SCMs/CRMs that use these datasets as forcing.

2) Significant differences exist for the omega field. The largest differences occur for upwelling during the spring months and the magnitude of downwelling during the summer. Although NARR and MERRA share many resemblances to each other, ARM slightly outperforms these reanalyses in terms of correlation with CF. Given that the ARM forcing is constrained by precipitation to give the adequate mass, momentum, heat, and moisture budgets, this indicates that some of the precipitation (especially during the late spring and early summer) is caused by smaller-scale forcing that is not captured by the reanalyses. This also suggests that SCMs based on the forcing derived from reanalyses...
would not be able to model precipitation adequately during this period. Combined with known issues such as SGSP in NARR documented by West et al. (2007) and within this study, vertical velocity values in re-analyses should be used with caution.

3) Given the findings 1 and 2, this study suggests that reanalysis-based forcing is a worthy endeavor. Efforts are already underway to produce forcing from the NARR at the ARM SGP site. It is believed that this forcing will provide a good resource when used either in tandem with the ARM continuous forcing or at locations that do not frequently experience convection.

4) ARM and NARR have excellent agreement for monthly precipitation amounts, which are a testament to the improved precipitation assimilation into NARR. NARR has a slight (−3 mm) bias for monthly precipitation but with more variability for daily precipitation, suggesting that the assimilation of precipitation may sometimes be mistimed or misplaced. Despite this, both monthly and daily correlations are still high. Conversely, MERRA only captures the monthly variation of precipitation well and contains considerable negative biases at monthly (−22.2 mm) and daily (−0.7 mm) intervals.

5) As found in Kennedy et al. (2010) and Xi et al. (2010), total CF at the ARM SGP site has good agreement between ARM and GOES satellite observations. From 1999 to 2001, CF peaked during the months of March and June before reaching a minimum during the summer months. Both NARR and MERRA capture this change, as evidenced by high correlations (0.92–0.78), although they have negative biases (14% and 3%, respectively). MERRA correlations for CF are highest with satellite observations, while NARR correlations are highest with the ARM surface observations. This is not surprising, given the amount of satellite information being assimilated into MERRA.

6) Surface radiative fluxes within this study agree well with those from Dong et al. (2006). Of the two reanalyses, MERRA shows better agreement with ARM observations for all fluxes except for LW-down. NARR has significant positive biases for SW-down, SW-up, and LW-up, and these are attributed to a combination of too few clouds and a lack of sufficient extinction by aerosols and water vapor in the atmospheric column. These results are consistent with previous studies that have investigated NARR elsewhere in the United States and the Eta Model at the ARM SGP site. MERRA biases for LW-down are attributed to the negative bias of water vapor within the atmospheric column.

The results presented here are not advertised as a general assessment of the reanalyses and only represent one location within the well-constrained continental midlatitudes with a limited period. However, in a companion study over the Arctic region (Zib et al. 2011, manuscript submitted to J. Climate), similar results were found—albeit with smaller biases. This study and Zib et al. (2011, manuscript submitted to J. Climate) have indicated that MERRA generally agrees better than the NARR/NCEP reanalyses, with ARM in both the midlatitudes and Arctic regions for CF and radiative fluxes. A potential avenue of research is expanding this analysis for a longer period using the newly developed CMBE cloud-radiation dataset by ARM (Xie et al. 2010). There are also plans to expand the ARM continuous forcing from 2001 to the present time over the ARM SGP site, as well as other surface sites.

Acknowledgments. The authors kindly acknowledge the anonymous reviewers, who provided helpful suggestions for this paper. NARR data were provided by NOAA/OAR/ESRL PSD, Boulder, Colorado, from their Web site (http://www.esrl.noaa.gov/psd). MERRA was obtained from the Goddard Earth Sciences Data and Information Services Center, Greenbelt, Maryland, from their Web site (http://disc.sci.gsfc.nasa.gov/mdisc). The University of North Dakota authors were supported by the NASA NEWS project under Grant NNX07AW05G, the DOE ARM under Grant DE-AC52-07NA27344/B589973, and the NASA CERES project under Grant NNL04AA11G.

REFERENCES


<table>
<thead>
<tr>
<th>TABLE 3. Annual mean TOA radiative fluxes and their biases compared to ARM continuous forcing.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clear sky (W m⁻²)</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>SWUP</td>
</tr>
<tr>
<td>ARM</td>
</tr>
<tr>
<td>NARR</td>
</tr>
<tr>
<td>MERRA</td>
</tr>
<tr>
<td>Bias</td>
</tr>
<tr>
<td>NARR</td>
</tr>
<tr>
<td>MERRA</td>
</tr>
</tbody>
</table>


