Land Surface Precipitation in MERRA-2

ROLF H. REICHLE
Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland

Q. LIU

RANDAL D. KOSTER
Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland

CLARA S. DRAPER
Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, and Sciences Technology and Research, Universities Space Research Association, Columbia, Maryland

SARITH P. P. MAHANAMA AND GARY S. PARTYKA

(Manuscript received 1 August 2016, in final form 7 November 2016)

ABSTRACT

The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2), features several major advances from the original MERRA reanalysis, including the use, outside of high latitudes, of observations-based precipitation data products to correct the precipitation falling on the land surface in the MERRA-2 system. The method for merging the observed precipitation into MERRA-2 has been refined from that of the (land-only) MERRA-Land reanalysis. This paper describes the method and evaluates the MERRA-2 land surface precipitation. Compared to monthly GPCPv2.2 observations, the corrected MERRA-2 precipitation (M2CORR) is better than the precipitation generated by the atmospheric models within the cycling MERRA-2 and MERRA systems. M2CORR is also better than MERRA-Land precipitation over Africa because in MERRA-2 a merged satellite–gauge precipitation product is used instead of the gauge-only data used for MERRA-Land. Compared to 3-hourly TRMM observations, the M2CORR diurnal cycle has better amplitude but less realistic phasing than MERRA-2 model-generated precipitation. Because correcting the precipitation within the coupled atmosphere–land modeling system allows the MERRA-2 near-surface air temperature and humidity to respond to the improved precipitation forcing, MERRA-2 provides more self-consistent surface meteorological data than were available from MERRA-Land, which is important for applications such as land-only modeling studies. Where precipitation observations of sufficient quality are available for use in the reanalysis, the corrections facilitate the seamless spinup of the land surface initial conditions across the MERRA-2 production streams. At high latitudes, however, the lack of reliable precipitation observations results in undesirable land spinup effects that impact mostly the first published year of each MERRA-2 stream (1980, 1992, 2001, and 2011).

1. Introduction

Retrospective analysis (reanalysis) data products provide global, subdaily estimates of atmospheric and land surface conditions across several decades. Such products are based on the assimilation of a large amount of in situ and remote sensing observations into an atmospheric general circulation model (AGCM) and are among the most widely used datasets in Earth science. The recent Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2; Gelaro et al. 2016, manuscript submitted...
to J. Climate) provides data beginning in 1980. MERRA-2 was introduced to replace and extend the original MERRA dataset (Rienecker et al. 2011; Luchesu 2012), which ended in February 2016. The MERRA-2 spatial resolution of about 50 km and the publication latency of a few weeks remain similar to MERRA. MERRA-2, however, includes recent upgrades to the atmospheric assimilation that enable the use of newer satellite observations that could not be assimilated in the original MERRA system. Additionally, MERRA-2 benefits from advances in the Goddard Earth Observing System, version 5 (GEOS-5) AGCM.

MERRA-2 also uses observations-based precipitation data products to drive the land surface water budget. In most reanalysis systems, including the original MERRA, the precipitation seen by the land surface is generated by the system’s AGCM following the assimilation of atmospheric temperature, humidity, and wind observations, among others. The MERRA-2 model-generated precipitation, however, is corrected with precipitation observations before reaching the land surface. Observation-corrected precipitation was also used in the Climate Forecast System Reanalysis (CFSR; Saha et al. 2010; Meng et al. 2012) and in MERRA-Land (Reichle et al. 2011; Reichle 2012). The latter is an offline, land-only reanalysis product. It provides significantly better land surface moisture storage dynamics than the original MERRA product because the MERRA-Land precipitation forcing was corrected using a global, gauge-based precipitation product, and because important changes were made to the rainfall interception parameterization in the land surface model (Reichle et al. 2011).

The land surface model and parameters in MERRA-2 closely resemble those of MERRA-Land, thus carrying the land model improvements into the coupled atmosphere–land MERRA-2 reanalysis. The precipitation corrections in MERRA-2 are refined from those used in MERRA-Land, with three main differences. First, and perhaps most importantly, the precipitation corrections in MERRA-2 were implemented in the coupled atmosphere–land reanalysis system. This allows the observed precipitation to impact, via evapotranspiration, the near-surface air temperature and humidity, thereby yielding a near-surface meteorological dataset that is more self-consistent than that of MERRA-Land. Second, the gauge-based precipitation corrections in MERRA-2 do not extend to the high latitudes. Third, over Africa MERRA-2 uses a (temporally and spatially) coarser precipitation data product that is based on satellite as well as gauge observations. The latter two changes were made because the sparse coverage and poor quality of the gauge-only precipitation product in Africa and at high latitudes had a detrimental impact on the MERRA-Land product (Mudryk et al. 2015, their Fig. 2; Reichle et al. 2017).

Reanalysis products are used extensively for land surface research and applications. In particular, many studies rely on near-surface meteorological data from reanalysis products to force land surface models (Nijssen et al. 2001; Dirmeyer et al. 2006; Qian et al. 2006; Sheffield et al. 2006; Koster et al. 2011; to name a few). Since precipitation is the dominant driver of the land surface water cycle, the objectives of this paper are as follows:

1) to document the land surface precipitation correction approach used in MERRA-2,
2) to provide an initial evaluation of the resulting MERRA-2 precipitation estimates, and
3) to demonstrate the self-consistency between precipitation and other near-surface meteorological data from MERRA-2.

A related study by Reichle et al. (2017) evaluates the MERRA-2 land surface hydrology against independent observations and compares the results with other reanalysis datasets, including MERRA, MERRA-Land, ERA-Interim, and ERA-Interim/Land. Furthermore, Bosilovich et al. (2017) and Robertson et al. (2016) investigate the global (land and atmospheric) water cycle in MERRA-2 and other major reanalysis data products. Finally, Gehne et al. (2016) conducted a comprehensive comparison of global, daily precipitation products that includes MERRA-2 along with other common reanalysis products. They find that no one precipitation product is better than all the others, and that the most suitable product changes with the intended application, location, and season.

The present study is organized as follows. After a brief description of the MERRA-2 system (section 2a) and the precipitation data (section 2b), the MERRA-2 precipitation correction method is summarized (section 2c), followed by an illustration of the climatological differences in land surface precipitation data between MERRA-2, MERRA-Land, and MERRA (section 2d). Our results compare the reanalysis precipitation to GPCPv2.2 observations (section 3a) and then investigate the diurnal cycle of the MERRA-2 precipitation (section 3b), the consistency of the corrected precipitation with near-surface air temperatures (section 3c), the MERRA-2 land surface spinup (section 3d), and, briefly, the global terrestrial water budget (section 3e). Finally, a summary and conclusions are given (section 4).
2. Data and methods

a. The MERRA-2 data product and system

MERRA-2 is a reanalysis product generated by the NASA Global Modeling and Assimilation Office (GMAO) using the GEOS-5.12.4 system (Bosilovich et al. 2015, 2016; Gelaro et al. 2016, manuscript submitted to J. Climate; http://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/). MERRA-2 replaces the original MERRA reanalysis (Rienecker et al. 2011) and includes updates to the AGCM (Molod et al. 2012, 2015) and to the Global Statistical Interpolation atmospheric analysis scheme of Wu et al. (2002). In addition to the atmospheric in situ and remotely sensed observations used in MERRA, the MERRA-2 system also ingests observations from newer microwave sounders and hyperspectral infrared radiance instruments, and other new data types. Moreover, MERRA-2 preserves the global water balance during the analysis, thereby mitigating the water balance discontinuities associated with observing system changes that occurred in MERRA and other reanalysis products (Takacs et al. 2016). Other developments in MERRA-2 include a comprehensive aerosol analysis (Randles et al. 2016, manuscript submitted to J. Climate; Buchard et al. 2016, manuscript submitted to J. Climate) and a mass balance over glaciated land surfaces (Cullather et al. 2014).

MERRA-2, MERRA-Land, and MERRA all use the Catchment land surface model, which explicitly addresses subgrid-scale soil moisture variability and its effect on runoff and evaporation (Koster et al. 2000; Stieglitz et al. 2001; Mahanama et al. 2015). For MERRA-2, the Catchment model of the original MERRA system has been updated with the interception and snow model parameters of the MERRA-Land version (Table 2 of Reichle et al. 2011) and the “BLM2” soil parameters of De Lannoy et al. (2014; see their Table 3). Besides the use of observed precipitation over land, which can be thought of as a poor-man’s soil moisture and snow analysis, MERRA-2 does not assimilate land surface observations.

MERRA-2 covers the period from 1980 to the present and continues to be updated with a latency on the order of weeks. The AGCM is run on a cube-sphere grid with an approximate resolution of 50 km × 50 km, the atmospheric analysis operates on a Gaussian grid of the same resolution, and the output fields are interpolated to a 0.5° × 0.625° regular latitude–longitude grid prior to publication. Note that the MERRA and MERRA-Land products have the same latitudinal resolution of 0.5° but a slightly different longitudinal resolution of 0.67°.

There are two kinds of precipitation estimates in the MERRA-2 system: 1) the precipitation generated by the AGCM within the cycling MERRA-2 system, hereinafter referred to as M2AGCM, and 2) the corrected precipitation that is seen by the MERRA-2 land surface and that modulates aerosol wet deposition over land and ocean, hereinafter referred to as M2CORR (Table 1). Both are available in the published MERRA-2 data product (Bosilovich et al. 2016) under the variable

<table>
<thead>
<tr>
<th>Reanalysis product</th>
<th>M2CORR</th>
<th>M2AGCM</th>
<th>MERRA-Land</th>
<th>MERRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precipitation dataset label used here</td>
<td>Corrected precipitation seen by the land surface and aerosol wet deposition over land and ocean (see Table 2, Figs. 1 and 2)</td>
<td>Precipitation generated by the AGCM within the cycling MERRA-2 data assimilation system (see Fig. 1)</td>
<td>Corrected precipitation seen by the land surface (see Table 2, Figs. 1 and 2)</td>
<td>Precipitation generated by the AGCM within the cycling MERRA data assimilation system</td>
</tr>
<tr>
<td>Variable name in data product file</td>
<td>PRECOTCORR</td>
<td>PRECOT</td>
<td>PRECOT</td>
<td>PRECOT</td>
</tr>
<tr>
<td>Data product “Collection” name</td>
<td>tavg1_2d_flx_Nx</td>
<td>tavg1_2d_mld_Nx</td>
<td>tavg1_2d_flx_Nx</td>
<td></td>
</tr>
<tr>
<td>Data product ECS short name</td>
<td>M2T1NXFLX</td>
<td>MSTINXMFLD</td>
<td>AT1NXFLX</td>
<td></td>
</tr>
<tr>
<td>Output grid spacing (latitude × longitude)</td>
<td>0.5° × 0.625°</td>
<td>0.5° × 0.667°</td>
<td>0.5° × 0.667°</td>
<td></td>
</tr>
</tbody>
</table>
names PRECOTT and PRECOTCORR, respectively (Table 1). The remainder of this section is devoted to a detailed explanation of how M2CORR is computed from M2AGCM and from ancillary data, including observational precipitation products and MERRA data. The specific MERRA-2 data used here include model-generated and corrected hourly and monthly precipitation (GMAO 2015a,b), monthly soil moisture, terrestrial water storage, evapotranspiration, and runoff (GMAO 2015c), hourly surface meteorological forcing data (GMAO 2015d,e), and time-invariant land fractions (GMAO 2015f).

b. Precipitation observations and selection criteria

Precipitation data products suitable for use in the MERRA-2 precipitation correction algorithm must 1) provide global or near-global coverage, 2) be available from 1979 to the present, 3) be updated operationally with a latency less than about one week, and 4) include gauge measurements where available. The first three requirements flow directly from the spatiotemporal coverage and latency objectives of MERRA-2. The fourth requirement to include gauge measurements is important for the quality of the resulting land surface estimates, in particular in the earlier years of the reanalysis, for which satellite precipitation estimates are less reliable. Furthermore, a ~0.5° spatial and a daily temporal resolution are highly desirable.

For MERRA-2, two precipitation products from the NOAA Climate Prediction Center (CPC) offered the best compromise regarding the above requirements: the CPC Unified Gauge-Based Analysis of Global Daily Precipitation (CPCU) product and the CPC Merged Analysis of Precipitation (CMAP) product. These two products are the same as those used in CFSR (Saha et al. 2010). The CPCU product provides gauge-based estimates of total precipitation over global land areas (Xie et al. 2007; Chen et al. 2008; ftp://ftp.cpc.ncep.noaa.gov/precip/CPC_UCI_PRCP/GAUGE_GLB). The CPCU data are posted on a 0.5° grid and are available for the period from 1979 to the present. The CMAP product provides global estimates of total precipitation for each pentad (5-day period) at 2.5° spatial resolution based on gauge observations and satellite estimates for the period from 1979 to present (Xie and Arkin 1997; ftp://ftp.cpc.ncep.noaa.gov/precip/cmap). MERRA-2 uses CPCU version “V1.0” through 2005, CPCU real-time “V1.0 RT” version thereafter, CMAP version “v0905” through 2006, and CMAP real-time “rt_v0011” thereafter. See Reichle and Liu (2014) for more details on the CPC products used here.

Here, we compare the reanalysis precipitation against the monthly Global Precipitation Climatology Project product, version 2.2 (GPCPv2.2; Adler et al. 2003; Huffman et al. 2009; http://precip.gsfc.nasa.gov/gpcp_v2.2_data.html).

The GPCPv2.2 data are carefully scrutinized by the data providers and many users, and are widely considered to be the best available long-term global precipitation product. Monthly GPCPv2.2 data are available on a 2.5° × 2.5° global grid from 1979 to present. The product is updated to the present within two to three months (and thus would not meet the MERRA-2 latency requirement). The GPCPv2.2 estimates, like those of CMAP, are based on merging (infrared and microwave) satellite measurements with global rain gauge observations, but the GPCPv2.2 and CMAP products differ in the input and processing of the satellite observations and in the approach for combining the satellite and gauge inputs. Nevertheless, GPCPv2.2, CMAP, and, to a lesser extent, CPCU data share a substantial portion of their raw gauge and/or satellite radiance inputs. As our results will illustrate (section 3), the GPCPv2.2 data are a useful but by no means independent reference for the evaluation of the corrected reanalysis precipitation.

Prior to their use in the MERRA-2 precipitation correction algorithm, the CMAP data are rescaled to match the (seasonally varying) climatology of the GPCPv2.1 pentad product (Reichle and Liu 2014). This rescaling of the CMAP data simply merges two observationally based products and should not be confused with the correction algorithm that merges the observations with the model-based background (section 2c). GPCPv2.1 was the most current version when the preparation of MERRA-2 inputs began; it differs over land from GPCPv2.2 mainly in the version of the input gauge data from the Global Precipitation Climatology Centre. Rescaling the CMAP data with the GPCPv2.1 pentad product is straightforward because both products are on the same 2.5° spatial grid, are provided at the same (pentad) time resolution, and cover land and ocean. In contrast, the CPCU product is a 0.5°, daily, land-only dataset. Rescaling the CPCU data using the coarser GPCPv2.1 product would thus be more difficult, especially along the coasts, and was therefore not considered for MERRA-2.

We further use the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Algorithm (TMPA) 3B42 data (TRMM 2011; Huffman et al. 2007) to evaluate the diurnal cycle of the reanalysis precipitation (section 3b). TMPA data are available from late 1998 to present on a 0.25° × 0.25° grid within 50° latitude from the equator. Precipitation is estimated using TRMM radar and radiometer observations (along with other satellite inputs) until about the end of 2014 and observations from the Global Precipitation Measurement (GPM) mission since its launch in 2014. The 3-hourly observations match the satellite–gauge TMPA 3B43 product at the monthly scale.
To estimate the diurnal cycle phase we followed the approach that Dirmeyer et al. (2012) used for their 3-hourly data. Phase estimates are unreliable and thus masked out for areas with mean precipitation rate below 0.5 mm day$^{-1}$ or relative amplitude (maximum minus minimum normalized by the mean) less than 0.2. MERRA-2 estimates were aggregated to 3-hourly data prior to estimating the diurnal cycle amplitude and phase.

c. Precipitation correction algorithm

It is important to emphasize that the MERRA-2 (and MERRA-Land) precipitation corrections are very different from a simple climatological rescaling. In a nutshell, the MERRA-2 precipitation corrections involve the merger and disaggregation of the observational products (section 2b) with precipitation estimates from MERRA and MERRA-2. In this section, we briefly review the method. See Reichle and Liu (2014) for further details.

The algorithm consists of two main steps (Fig. 1). Because the observations-based, global, long-term precipitation datasets that satisfy the requirements for MERRA-2 have coarser temporal resolution and also, in the case of CMAP, coarser spatial resolution than the reanalysis product, step 1 of the algorithm consists of downscaling the observed precipitation data to hourly 0.5° resolution in order to provide the subdaily variations and spatial detail required to force the land surface in the reanalysis. This step uses MERRA data as the model background through 2014 and the GEOS-5 FP-IT product version 5.9.1 (Lucchesi 2015) thereafter. (The current plan is to use FP-IT version 5.12.4 starting in 2017.) For a given day (or pentad) and 0.5° (or 2.5°) grid cell, the correction factor used in the downscaling is the ratio of the observed to the (aggregated) background precipitation. This factor is then applied uniformly to all background (hourly) time steps and grid cells that make up the given day (or pentad) and 0.5° (or 2.5°) grid cell. When a correction factor is undefined because the background precipitation is zero but the observation is nonzero, the observed precipitation is added evenly to each of the hourly corrected estimates between midnight and 3 a.m. local time. Throughout the process, regridding is mass conservative and carried out via fine-resolution exchange grids.

Files with downscaled, corrected precipitation data are first generated separately for the CPCU and CMAP observations (Fig. 1). By construction, this precipitation matches, after suitable aggregation, the values of the CPCU (or CMAP) observations for each day (or pentad) and for each 0.5° (or 2.5°) grid cell. Note that the resulting CPCU-corrected precipitation is the same as that used for

FIG. 1. Schematic of the precipitation correction algorithm. See text (section 2c), Fig. 2, and Table 2 for more details.
MERRA-Land. To complete step 1 of the algorithm, composite files are generated from the CMAP-corrected data over Africa (and the oceans) and the CPCU-corrected data over all other land. Finally, in step 2 of the algorithm the composited precipitation data are applied—separately for each hour and at each grid cell—such that the precipitation experienced by the MERRA-2 land surface (M2CORR) is the weighted average of the composited precipitation from step 1 and the precipitation generated by the AGCM within the cycling MERRA-2 system (M2AGCM), with weights depending on latitude.

The resulting four distinct precipitation correction regimes in MERRA-2 are illustrated in Fig. 2 and summarized in Table 2. Over Africa, the satellite- and gauge-based CMAP product (after climatological rescaling to the GPCPv2.1 pentad product; section 2b) is applied in full. Over low- to midlatitude land (equatorward of 42.5° latitude) except Africa, CPCU-based corrections are applied in full in MERRA-2 (as in MERRA-Land). In high-latitude land areas (poleward of 62.5° latitude), corrections are not applied because of the sparse coverage and poor quality of the gauge observations (section 2b). In this region, the MERRA-2 land surface experiences the M2AGCM precipitation (i.e., \( w = 0 \) in Figs. 1 and 2). In the Northern and Southern Hemisphere latitude bands between 42.5° and 62.5°, the precipitation corrections are tapered linearly between full corrections based on CPCU (at 42.5°) and no corrections (at 62.5°). Over the oceans (not shown in Fig. 2), the M2CORR precipitation is based on (rescaled) CMAP observations and uses the same latitude-dependent tapering as over land (not shown for clarity).

MERRA-2 precipitation correction algorithm shares key elements with that used in CFSR (Meng et al. 2012), including the choice of the CPCU and CMAP input precipitation products and the latitude-dependent merger of observed and modeled precipitation. But there are also key differences. Perhaps most importantly, in MERRA-2 the observed precipitation is applied directly within the coupled atmosphere–land system at hourly intervals. By contrast, CFSR employs a separate land modeling subsystem that is forced with the observation-corrected precipitation. The soil moisture and temperature states in the CFSR atmosphere–ocean model are then replaced once per day at 0000 UTC with those from the land-only subsystem. Moreover, the latitude- and region-dependent weights for the CPCU, CMAP, and model-generated precipitation differ between MERRA-2 and CFSR. MERRA-2 uses CMAP only in and across all of Africa (and after rescaling to the GPCPv2.1 climatology), whereas CFSR uses primarily CMAP throughout the tropics. Note also that CFSR, unlike MERRA-2, includes a snow analysis.

d. Comparison of MERRA-2, MERRA-Land, and MERRA precipitation

In this section, we further illustrate the precipitation correction algorithm (section 2c) by highlighting the climatological differences between the corrected MERRA-2 precipitation (M2CORR), the
precipitation generated by the AGCM within the MERRA-2 system (M2AGCM), the corrected precipitation used in MERRA-Land, and the MERRA precipitation (Table 1). Throughout the manuscript, all comparisons and spatial averages are over global land excluding permanently ice-covered surfaces, that is, excluding Antarctica and most of Greenland.

Figure 3a shows the long-term mean (1980–2015) M2CORR precipitation. Averaged over global land, the mean annual precipitation is 2.16 mm day\(^{-1}\). Figure 3b shows the mean difference between M2AGCM and M2CORR, reflecting the net adjustment made to the model-generated precipitation from MERRA-2 before it reaches the land surface. The greatest mean adjustments are made in the tropics and subtropics, where precipitation is greatest. In most regions, the M2AGCM precipitation is wetter than M2CORR, and the correction procedure decreases the mean precipitation experienced by the land surface. In parts of central South America, central Africa and the Sahel, however, the M2AGCM precipitation is too low, and the correction procedure increases the mean precipitation. Averaged over global land, the correction procedure reduces the M2AGCM precipitation by 0.8 mm day\(^{-1}\). Note that Fig. 3b also illustrates the (intentional) absence of precipitation corrections at high latitudes.

Next, Fig. 3c shows the difference between MERRA-Land and M2CORR precipitation. By construction, MERRA-Land closely matches the M2CORR precipitation in low- and midlatitude regions except Africa. In most of tropical Africa and in the high latitudes, the MERRA-Land precipitation is much drier than M2CORR. At high latitudes, this is a reflection of the strong wet bias in the model-generated precipitation used in M2CORR (see below), combined with the fact that the CPCU (and thus MERRA-Land) precipitation tends to be lower than that of other products (e.g., Robertson et al. 2014). Over Africa, the lower MERRA-Land precipitation is a reflection of the dry bias in the CPCU product with respect to the GPCPv2.1 climatology that was used in M2CORR in this region (section 2b).

Finally, Fig. 3d shows the difference between MERRA and M2CORR precipitation. Outside of the high latitudes, the differences between M2AGCM and M2CORR (Fig. 3b) resemble those between MERRA and M2CORR (Fig. 3d), suggesting that overall the model-generated precipitation estimates from MERRA-2 and MERRA are closer to each other than to the observations. Note, however, that the dry bias in the M2AGCM precipitation in South America and central Africa is diminished in comparison with that of MERRA. This is probably a reflection of the improvements in the MERRA-2 AGCM and atmospheric assimilation, but it may also be facilitated by the MERRA-2 precipitation corrections maintaining a wetter land surface and thus increased evapotranspiration and local precipitation recycling. Further investigation of this feedback is left for future studies.

3. Results

a. Comparison against GPCPv2.2 precipitation

It cannot be overstated that the quality of land surface estimates from reanalysis products depends critically on the accuracy of the precipitation forcing. In this section, we examine the MERRA-2, MERRA-Land, and MERRA precipitation (Table 1) through comparisons

<table>
<thead>
<tr>
<th>Reanalysis product</th>
<th>Land region (temporal/spatial resolution)</th>
<th>Observations (temporal/spatial resolution)</th>
<th>Correction method</th>
</tr>
</thead>
<tbody>
<tr>
<td>MERRA-2</td>
<td>Africa</td>
<td>CMAP/GPCPv2.1 satellite and gauge data (pentad / 2.5°)</td>
<td>Full correction to observations.</td>
</tr>
<tr>
<td></td>
<td>Low- and midlatitude land (lat &lt; 42.5), except Africa</td>
<td>CPCU gauge data (daily / 0.5°)</td>
<td>Full correction to observations.</td>
</tr>
<tr>
<td></td>
<td>Mid- to high-latitude land (42.5° &lt; lat &lt; 62.5°)</td>
<td>CPCU gauge data (daily / 0.5°)</td>
<td>Linear tapering between full observation correction for lat = 42.5°, and no correction (full use of MERRA-2 model-generated precipitation) for lat = 62.5°.</td>
</tr>
<tr>
<td></td>
<td>High-latitude land (lat &gt; 62.5°)</td>
<td>None</td>
<td>No correction.</td>
</tr>
<tr>
<td>MERRA-Land</td>
<td>Global land</td>
<td>CPCU gauge data (daily / 0.5°)</td>
<td>Full correction to observations.</td>
</tr>
</tbody>
</table>
against monthly GPCPv2.2 observations (section 2b). For MERRA-2, we evaluate the corrected precipitation that is seen by the MERRA-2 land surface (M2CORR) as well as the precipitation generated by the AGCM within the cycling MERRA-2 system (M2AGCM). We focus first on global maps of seasonal biases and monthly time series correlations and RMSE values. Thereafter, we illustrate some artifacts resulting from changes in the gauge network by investigating time series of regionally averaged precipitation.

First, Fig. 4 shows seasonal mean differences between each of the reanalysis products and the GPCPv2.2 precipitation. Differences are shown for (Northern Hemisphere) winter (DJF) and summer (JJA). Generally, the M2CORR precipitation is closer to the GPCPv2.2 product than that of M2AGCM, MERRA-Land, or MERRA (where closeness is measured in the global average sense, that is, the mean across the difference map, and in terms of an average magnitude of the bias, that is, the standard deviation across the difference map). For winter (DJF; Fig. 4a) and summer (JJA; Fig. 4b), the seasonal mean M2CORR precipitation is typically within 0.7 and 1.1 mm day\(^{-1}\), respectively, of the GPCPv2.2 precipitation in a global RMS sense, and within 0.2 mm day\(^{-1}\) in the global average. The larger discrepancies in JJA are partly driven by the large seasonal bias at high latitudes. In this region, precipitation is not corrected to observations and M2CORR is equal to the M2AGCM precipitation (Fig. 2). But note that the assumed observational truth (GPCPv2.2) is quite uncertain here as well. Very large differences in JJA are also seen in southern Asia. The minor differences in Africa reflect the differences between the GPCPv2.1 climatology (which was imposed on the CMAP observations in the corrections process; section 2b) and that of GPCPv2.2.

Unsurprisingly, the bias in the M2AGCM precipitation exceeds that of M2CORR, with global RMS values of 2.9 mm day\(^{-1}\) for both DJF and JJA (Figs. 4c,d), and global average differences of 0.6 and 0.7 mm day\(^{-1}\), respectively. Generally, M2AGCM suffers from excessive precipitation over topography (Bosilovich et al. 2015). These errors are present in a relatively small surface area, but are sufficiently large to adversely impact the global RMS bias statistics. In winter, the M2AGCM precipitation (Fig. 4c) is too high over most of North America and Australia and much too high over large portions of South America, Africa, and the Maritime Continent, but much too low over central South America and central Africa, and too low over Europe and western Eurasia. In summer, the bias of the M2AGCM precipitation (Fig. 4d) is similar in absolute terms in many tropical and subtropical areas. But in contrast to winter, the M2AGCM summer precipitation shows a dry bias over southeastern North America and a marked wet bias over all of the high latitudes. The latter has important implications for the land surface spinup in MERRA-2 (section 3d).

By construction, the bias of MERRA-Land precipitation against GPCPv2.2 (Figs. 4e,f) closely matches that of M2CORR in low and middle latitudes (except Africa). In the high latitudes and Africa, MERRA-Land precipitation is generally too low in both summer and winter, which is simply a reflection of the mean differences between the CPCU and GPCPv2.2 products.
These differences partly motivated the changes in the MERRA-2 precipitation corrections over Africa and the high latitudes. Finally, Figs. 4g and 4h show the seasonal bias for MERRA precipitation, which is less biased than M2AGCM in winter and in summer, both in the global average sense and in terms of typical absolute bias values, because of the large errors over topography in M2AGCM. Generally, the seasonal biases for spring (March–May) resemble those for summer, and the seasonal biases for fall (September–November) resemble those for winter (not shown).

Next, Fig. 5a shows a map of the monthly time series correlation coefficients ($R$) versus GPCPv2.2 observations for M2CORR, with a global average $R$ of 0.82. Low values are found only in the Sahara and Arabian deserts, where precipitation amounts are small anyway, and in a few small mountainous, tropical, or high-latitude regions. The generally high $R$ values of M2CORR in most of Africa are not surprising because during the corrections process the seasonal cycle of the CMAP precipitation was rescaled to match the climatology of the GPCPv2.1 data (section 2b), which is similar to that of
FIG. 5. (a) Time series correlation coefficient ($R$) for M2CORR precipitation vs monthly GPCPv2.2 observations. (b)–(d) Difference between $R$ value for M2CORR and that for M2AGCM, MERRA-Land, and MERRA, respectively. (e)–(h) As in (a)–(d), but for anomaly $R$. (i)–(l) As in (a)–(d), but for RMSE. In all skill difference plots, yellow and red colors indicate higher skill for M2CORR than the other product, and blue colors indicate lower skill for M2CORR.
GPCPv2.2. Estimates of 95% confidence intervals for the \( R \) values shown in Fig. 5a depend on the \( R \) values themselves and range from \( \pm 0.16 \) for \( R = 0.2 \) to \( \pm 0.06 \) for \( R = 0.8 \). These estimates are based on a Fisher Z transform and assume \( N = 144 \) degrees of freedom (out of the 432 monthly data points in the 36-yr time series, with the ad hoc reduction meant to provide a rough correction for autocorrelation in the time series). This implies that \( R \) values greater than \( \sim 0.2 \) are statistically different from zero at the 5% significance level, which includes nearly all of the domain.

Difference maps between the time series correlation (\( R \) value) of M2CORR precipitation and that of M2AGCM, MERRA-Land, and MERRA precipitation, respectively, are shown in Figs. 5b–d. For these skill differences, yellow and red colors indicate higher skill for M2CORR compared to the other MERRA products, and blue colors indicate lower skill for M2CORR. From the figure it is immediately obvious that the \( R \) values of M2CORR are higher than those of the model-generated precipitation, by 0.09 on average for M2AGCM (Fig. 5b) and by 0.07 for MERRA (Fig. 5d). The greatest improvements in the \( R \) values of M2CORR over the model-generated (M2AGCM and MERRA) precipitation can be found in South America and Africa, but there are modest skill improvements almost everywhere. The \( R \) values of M2CORR are worse than those of the model-generated precipitation only in the very dry Arabian Peninsula and Iran, along coastal mountains in Central and South America, in Myanmar (see below), and, in the case of MERRA, in a small Eurasian high-latitude region. By construction, the \( R \) values of M2CORR and MERRA-Land precipitation match in the low and middle latitudes except Africa (Fig. 5c). Compared to MERRA-Land, M2CORR precipitation is slightly better in Africa, where the precipitation corrections differ between the two products. In the high latitudes, M2CORR is better than MERRA-Land in North America but worse in Eurasia.

The skill differences shown in Figs. 5b–d can be considered statistically significant if the 95% confidence intervals of the two \( R \) values included in the difference do not overlap. Based on the 95% confidence interval estimates above, skill differences for low skill (\( R \sim 0.2 \)) are significant only if the (absolute) difference in the \( R \) values exceeds 0.32. For high skill (\( R \sim 0.8 \)) the (absolute) difference in the \( R \) values needs to exceed 0.12 for the skill differences to be significant. This means that skill differences shown in orange-to-red colors and medium-to-dark blue colors in Figs. 5b–d are typically significant at the 5% level.

The results are similar in terms of the anomaly \( R \) values, which are computed after removing the mean seasonal cycle from the data and therefore measure primarily the skill in each product’s interannual variability. M2CORR again shows the greatest skill, with a global average anomaly \( R \) value of 0.70 (Fig. 5d). The low M2CORR skill values in much of Africa reflect the skill of CMAP (and not GPCPv2.1) because the anomaly correlation metric is largely insensitive to the climatological rescaling that was applied to CMAP (section 2b). Based on the estimated 95% confidence intervals above, most of the anomaly \( R \) values shown in Fig. 5e are statistically different from zero at the 5% significance level. Skill improvements in the M2CORR precipitation over the model-generated precipitation (M2AGCM and MERRA) are also seen in terms of the anomaly \( R \) values (Figs. 5f,h), and the improvements are again greatest in South America and Africa, with areas of slightly worse skill in M2CORR being mostly limited to dry areas. The anomaly \( R \) skill is the same for M2CORR and MERRA-Land in low and mid latitudes (except Africa) by construction, but M2CORR again shows mostly improvements over MERRA-Land in Africa and in the Canadian high latitudes (Fig. 5g). Following the reasoning above, skill differences shown in orange-to-red colors and medium-to-dark blue colors in Figs. 5f–h are typically significant at the 5% level.

Finally, the bottom row of Fig. 5 investigates the skill of the reanalysis products in terms of the RMSE against the monthly GPCPv2.2 time series. For M2CORR, the highest RMSE values (exceeding 2 mm day\(^{-1}\)) occur in the tropics (Fig. 5i), where precipitation is largest (see, e.g., Fig. 3a). Errors are lowest in dry regions, including southwestern North America, northern Africa, central Eurasia, and the interior of Australia. Figures 5j–l show the differences in the RMSE between the M2CORR precipitation and that of M2AGCM, MERRA-Land, and MERRA, respectively. As with the correlation difference maps, yellow-to-red colors indicate that M2CORR is better (lower RMSE) than the other product. M2CORR shows considerable improvements in terms of RMSE when compared to the M2AGCM precipitation, particularly in parts of tropical South America, Africa, and Southeast Asia (Fig. 5j). M2CORR also shows large improvements against M2AGCM along steep mountain ranges such as the Andes and the Himalayas (Fig. 5j). As shown above (Figs. 4c,d), M2AGCM has much larger biases against GPCPv2.2 than M2CORR in these regions, which contributes to the larger RMSE values for M2AGCM.

The RMSE values for M2CORR are also lower than those of MERRA-Land in much of tropical Africa (Fig. 5k) because M2CORR relies on the CMAP satellite–gauge product (rescaled to the GPCPv2.1 climatology), whereas MERRA-Land uses the CPCU gauge-based product, which is unreliable in this region. At high latitudes, however, M2CORR has higher RMSE...
values than MERRA-Land (Fig. 5k), primarily because of the strong wet bias in spring and summer of the M2AGCM precipitation (Fig. 4d), which is identical to M2CORR in this region (Fig. 2), although the assumed observational truth (GPCPv2.2) is quite uncertain here. Last, M2CORR has much lower RMSE values than MERRA in South America, along with modest improvements everywhere else except in the high latitudes, where MERRA has lower RMSE values (Fig. 5l).

Particularly high RMSE values and low $R$ and anomaly $R$ values in the corrected precipitation are found in Myanmar (Fig. 5). Here, the CPCU product determines the M2CORR and MERRA-Land precipitation. Figure 6 illustrates that the annual mean precipitation anomalies from CPCU have virtually no skill when compared to GPCPv2.2. Perhaps the most notable feature in the CPCU precipitation anomaly time series is the sharp discontinuity in 2008. The poor skill in the reanalysis products in Myanmar can be traced back directly to errors in the CPCU precipitation, which in turn stem from persistent errors in the input gauge measurements from Myanmar prior to May 2008 (P. Xie 2015, personal communication). Note, though, that the inaccurate inputs also impact the quality of the GPCPv2.2 estimates (because gauge observations were not available to anchor the product), and the discrepancies shown in Fig. 6 thus reflect the combined impact of the erroneous gauge data in the CPCU and GPCPv2.2 estimates. Note also that in the larger region including and surrounding Myanmar, the M2CORR and MERRA-Land precipitation (both based on CPCU observations) also exhibit a considerable dry bias during summer (Figs. 4b,f).

While the problem with the precipitation gauge measurements in Myanmar is fairly easy to detect from the global time series correlation plots (Fig. 5), other artifacts in the observed precipitation are more difficult to identify. One approach to revealing potential issues is to investigate the precipitation data where there are changes in the gauge network. Figure 3 of Lorenz and Kunstmann (2012) forcefully demonstrates the dramatic changes in the number of gauges used in the CPCU product. Overall, the network density decreases considerably over the course of the reanalysis period, with South America experiencing the most pronounced decline.

Based on the Lorenz and Kunstmann (2012) figure, we selected five regions with considerable changes in the gauge network for further analysis. The five regions are outlined in Fig. 7, and the green bars in Fig. 8 illustrate the gauge counts for each region as a function of time. In the eastern U.S. region (Fig. 8a), around 2000 gauges were used between 1980 and 1991, and then again from 2004 to present. Between 1992 and 2003, nearly 4000 gauges were used. In the Amazon region (Fig. 8c), about 350 gauges were used between 1980 and 2004, before a sharp drop to an average of about 65 gauges since then. In eastern Brazil (Fig. 8b), the number of gauges declined more steadily from more than 2000 in the 1980s to about 1000 in 2004, and then also dropped sharply to around 230 from 2005 to 2015. The gauge count is apparently more stable in the central Africa region (Fig. 8d), but on average only about 20 gauges were used there. In the central Australia region (Fig. 8e), the gauge count was stable at about 500 through about 2006. Thereafter, only 40–60 gauges were used. On average, the CPCU product currently uses one gauge every 30 km in the eastern U.S. region. By contrast, gauges are, on average, spaced 250 km apart in the Amazon and central Australia regions, and 400 km apart in the central Africa region, where precipitation is dominated by convective and thus highly spatially variable precipitation events.
Figure 8 also shows the mean annual precipitation from MERRA-2 and GPCPv2.2. Problems with M2CORR are obvious in the Amazon region (Fig. 8b), where it closely tracks GPCPv2.2 until 2004 and then drops sharply by about 25% as the gauge count diminishes. Similarly, in the eastern Brazil region (Fig. 8c) M2CORR tracks GPCPv2.2 through the early 2000s and is between 10% and 20% less than GPCPv2.2 starting in 2005. In the central Australia region (Fig. 8e), a small positive bias in M2CORR develops starting in 2006 as the gauge count
drops. In the central Africa region (Fig. 8d), M2CORR is based on CMAP data (after rescaling to the GPCPv2.1 climatology). The multiyear variations in M2CORR in this region appear to follow those in the CPCU gauge count. In the eastern United States (Fig. 8a), however, there is no obvious correspondence between the number of gauges and the M2CORR bias despite the sudden changes in the gauge count in 1992 and again in 2004. Overall, Fig. 8 suggests that changes in the gauge network do not necessarily cause discontinuities, provided the reduced network is still dense enough to capture the broad spatial structure of precipitation. Where the gauge density is generally low, however, changes in the gauge network can have a detrimental impact on the long-term homogeneity of the dataset.

For reference, Fig. 8 also includes the M2AGCM precipitation (Table 1). In the eastern Brazil and central Australia regions (Figs. 8c,e), M2AGCM is biased high but otherwise shows good agreement with the observed interannual variations. In the Amazon region (Fig. 8b), the difference between M2AGCM and GPCPv2.2 increases suddenly in 2010, presumably in response to a change in the atmospheric observing system feeding into the MERRA-2 analysis. A similar (albeit smaller) increase in M2AGCM compared to GPCPv2.2 can be seen in the eastern U.S. region in 2003 (Fig. 8a). In the central Africa region (Fig. 8d), the M2AGCM precipitation is even less stable than that of M2CORR. Nevertheless, Fig. 8 might suggest that the atmospheric observing system changes that adversely impact M2AGCM are no more problematic than the deficiencies in M2CORR, particularly in South America. It is important to keep in mind, however, that Fig. 8 highlights the regions with the most prominent changes in the gauge network and that over land the M2AGCM precipitation is much more biased and generally of lower quality than M2CORR (Figs. 4 and 5), despite the obvious shortcomings of the latter.

b. Diurnal cycle of precipitation and shortwave radiation

Accurately representing the diurnal cycle of precipitation in global modeling systems remains a challenge (Dai 2006; Dirmeyer et al. 2012; Chao 2013; Kim and Alexander 2013; Chen et al. 2014). To illustrate this challenge in the context of the MERRA-2 precipitation corrections, Fig. 9 shows hourly values of precipitation and shortwave radiation for one week in June 2003 for a grid cell near Gainesville, Florida. During the first 5 days of this week, the area experienced typical summer convective conditions, with locally intense and short-lived precipitation events and the corresponding rapid changes in downward shortwave forcing [see Fig. 3 of Reichle et al. (2011)]. Since the estimates from MERRA-2 (Fig. 9a) and MERRA (Fig. 9b) represent ~0.5° grid cell–scale averages, they are not expected to reflect the intensity or rapid changes in local conditions associated with small-scale storms. However, the diurnal cycle in the MERRA precipitation estimates is dominated by an unrealistic midday maximum, with almost no nighttime precipitation and very little day-to-day variability (Fig. 9b). The corrected MERRA-2 precipitation (M2CORR) relies on the MERRA diurnal cycle and therefore has a similar diurnal cycle by construction, but with some modulation in the day-to-day variations that is informed by the precipitation product used in the corrections (Fig. 9a). Note that in this location the (daily) M2CORR precipitation matches the CPCU observations by design.

The MERRA-2 model-generated precipitation (M2AGCM) is less dominated by regular midday rainfall and shows more nighttime precipitation and more day-to-day variability than MERRA (Fig. 9a). Moreover, the MERRA-2 surface shortwave radiation is more consistent with the M2AGCM precipitation than is the MERRA shortwave radiation with its precipitation because MERRA typically underestimates the attenuation of the radiation during precipitation events. Compared to MERRA, the MERRA-2 modeling system thus exhibits...
some improvements, but by construction the M2CORR precipitation falling on the MERRA-2 land surface still inherits the errors in MERRA’s precipitation diurnal cycle through 2014 and those of GEOS-5 FP-IT thereafter (section 2c).

Next, we compare the diurnal amplitude (maximum minus minimum) of the reanalysis precipitation against that of the TMPA 3B42 observations (section 2b), which is shown in Fig. 10a for Northern Hemisphere summer (JJA). Strong diurnal cycles with amplitudes greater than 6 mm day\(^{-1}\) are present in the southeastern United States, Central America, northern South America, the Sahel, portions of tropical Africa, and most of Southeast Asia. As expected for JJA, amplitudes in the Southern Hemisphere are much smaller. The M2CORR precipitation captures the broad spatial patterns of the amplitude reasonably well, albeit with some underestimation of the magnitude (Fig. 10c). By construction, the ratio of the diurnal amplitude to the mean precipitation (i.e., the relative amplitude) of M2CORR (not shown) is determined by that of MERRA through 2014 and that of GEOS-5 FP-IT thereafter (section 2c). But the (absolute) amplitude of M2CORR also depends on the mean precipitation of M2CORR, which by design matches the observations used in the corrections procedure (and generally differs from that of MERRA). For example, the diurnal amplitude of MERRA in central Africa is much lower than that of M2CORR and therefore less consistent with that of the TMPA 3B42 observations (not shown).

The M2AGCM amplitude, shown in Fig. 10e, also captures the broad spatial pattern of the observed amplitude, but M2AGCM underestimates the observed magnitude by much more than M2CORR. Also, the spatial pattern of the M2AGCM amplitude is less consistent with the observations than that of M2CORR. For example, M2AGCM has very low amplitudes in central Africa, which coincides with the dry bias in M2AGCM (Fig. 4d).

Finally, Fig. 10b shows the observed phase of the diurnal cycle (local solar time of maximum) for JJA precipitation from TMPA 3B42. Magenta, red, and yellow colors indicate a nighttime precipitation maximum, such as in the U.S. Great Plains and portions of northern South America, the Sahel, and tropical Africa and Southeast Asia. Green and blue colors denote a daytime maximum, with darker blues indicating a precipitation maximum in the late afternoon or early evening, which is prevalent in most of the rest of the Northern Hemisphere. Figure 10d shows that in M2CORR the JJA precipitation...
precipitation maximum is very close to local noon almost everywhere. This aspect of the M2CORR diurnal cycle is inherited directly from MERRA through 2014 and from GEOS5 FP-IT thereafter (section 2c). The phase of the M2AGCM precipitation (Fig. 10f) is somewhat more realistic, especially the evening and nighttime precipitation maxima in Central America, the U.S. Great Plains, northern South America, pockets of tropical Africa, and Tibet. However, M2AGCM also exhibits unrealistic morning and midday precipitation maxima in many regions, including the southeastern United States and much of tropical Africa and Southeast Asia.

It is important to reiterate that the M2AGCM precipitation was generated within the MERRA-2 cycling system and thus could not be used to create the “step 1 precipitation” of the corrections algorithm (Fig. 1 and section 2c) without additional, complex, and far-reaching modifications of the MERRA-2 system, which were beyond the scope of the development at the time. The poor representation of the phase of the diurnal cycle in M2CORR was therefore unavoidable. Given the otherwise considerable improvements in the M2CORR over the M2AGCM precipitation, which resulted in significantly better land surface estimates (Reichle et al. 2017), we consider the phase errors in M2CORR acceptable. Nevertheless, their precise impact on the simulated land surface states and fluxes in MERRA-2 requires further study.

**c. Imprint of corrected precipitation on near-surface meteorology**

Land surface hydrology is, for the most part, driven by the atmosphere. The atmosphere drives the land surface with precipitation, radiation, near-surface air temperature, near-surface humidity, and other forcing, and the land surface responds with fluxes of moisture and energy (e.g., runoff, evapotranspiration, sensible heat flux, outgoing longwave radiation), updating in the process its internal moisture and energy states. Self-consistency is an often underappreciated facet of atmospheric forcing used for land modeling studies. As illustrated in Fig. 9, incoming solar radiation is reduced during precipitation events due to the associated cloudiness. Subsequent evaporative cooling at the land surface can then lead to reduced near-surface air temperature. Because all of the forcing variables work together to drive the land surface, this self-consistency can be expected to have a significant impact on simulated hydrological variables.

Using an idealized experiment, Fig. 11 illustrates the basic nature of such an impact. The figure shows, for a
location in the U.S. southern Great Plains, time series of precipitation, surface downward shortwave radiation, air temperature (at the lowest atmospheric model level), and evapotranspiration over a 2-week period during July 2013 from (land only) simulations with the Catchment model. The blue lines in the top three panels show the forcing data (including the precipitation, radiation, and air temperature) used to drive the Catchment model in a control simulation; these forcing data were taken directly from MERRA-2 output files. The resulting evapotranspiration fluxes produced in the control simulation are shown in the bottom panel. The red lines show data from a second simulation that differed from the control in only one way: the timing of the precipitation was artificially shifted forward by 4 days, as indicated in the top panel, while the other forcing variables were left unchanged. This 4-day shift in precipitation is designed to illustrate, in an idealized way, the impact of inconsistency in the timing of forcing variables. Notice, for example, that in the control the rainy conditions on 16 and 17 July are associated with contemporaneous lower air temperatures and radiation, whereas in the second simulation the 4-day shift of the rain leads to contemporaneous (and thus presumably unrealistic) rainy and warmer/sunnier forcing conditions on 20 and 21 July. The simulated evapotranspiration fluxes are, as a result, markedly higher in the second simulation during the latter period. Also, relative to the control, the evapotranspiration fluxes from the second simulation on days 16 and 17 July are essentially nonexistent.

Note that by construction, full self-consistency is absent in the MERRA-Land forcing data—MERRA-Land is essentially forced with CPCU precipitation data but uses air temperature, radiation, and other meteorological forcing from the MERRA reanalysis. In contrast, because the observed precipitation rates are ingested into the cycling MERRA-2 system, the observed precipitation can feed back on the near-surface meteorology (e.g., by inducing cooler air temperatures in subsequent days), thereby increasing the inherent self-consistency of the MERRA-2 data. This is illustrated in Fig. 12 with a simple metric: the correlation, computed for June through August, between 4-day accumulated observation-corrected precipitation (days $N-3$ through $N$) and air temperature on day $N$. As expected, MERRA-2 shows higher correlations than MERRA-Land between the quantities, a clear indication that the ingestion of observed precipitation in MERRA-2 has imprinted itself on the subsequent air temperature. The correlations for MERRA-2 are particularly strong in the transition zones between wet and dry areas (e.g., the southern Great Plains and the Sahel), locations for which variations in soil moisture, as affected by variations in precipitation, are known to have a strong impact on near-surface temperature (Koster et al. 2006).

We should note that the ability of the imposed precipitation in MERRA-2 to feed back accurately on the atmospheric radiation is presumably more tenuous than for air temperature. Feedbacks on solar radiation are subject to questionable links in the model between the surface energy balance and the simulation of clouds. We note again, though, that MERRA-2, through improvements in its radiation physics, does show increased consistency between its precipitation and solar radiation fields (Fig. 9). The presence of an overall increased consistency among the land surface forcing variables can be seen, in conjunction with the imposed realism of the precipitation forcing, as a unique strength of the MERRA-2 dataset.

d. Land surface spinup

MERRA-2 was produced using four separate streams, initialized in 1979, 1991, 2000, and 2010, with the first year of each stream designated as spinup (Bosilovich et al. 2015). One year, however, is not sufficient for spinning up the land surface states. Therefore, the land surface restart files for each MERRA-2 stream were themselves spun up for at least 20 years, using the offline (land only) version of the MERRA-2 land model forced with MERRA surface meteorological fields and the downscaled, corrected precipitation data from step 1 of the correction algorithm (labeled “step 1 precipitation” in Fig. 1; section 2c).

In regions with full precipitation corrections (i.e., in low to middle latitudes), the offline spinup procedure ensured that the MERRA-2 land initial conditions were
generated with the same precipitation forcing that was later used in the MERRA-2 production, thereby resulting in land initial conditions for the MERRA-2 production streams that provide continuity across stream boundaries. For example, Fig. 13a illustrates this continuity for a region in southern North America (25°–42.5°N, 60°–130°W) and each stream’s 1-yr spinup (thick gray lines). (c),(d) As in (a),(b), respectively, but for terrestrial water storage. See Fig. 7 for a map of the regions.

At high latitudes, however, the precipitation experienced by the land surface is (fully or in part) determined by the MERRA-2 model-generated precipitation (M2AGCM; Table 1 and section 2c). The latter was, of course, not available for the land surface spinup, which in this region essentially used the MERRA-Land precipitation. Unfortunately, the M2AGCM precipitation at high latitudes turned out to be very different from the MERRA-Land precipitation, as can be seen by comparing Figs. 4d and 4f. This difference in high-latitude precipitation between the MERRA-2 land spinup and production systems thus resulted in land surface moisture discontinuities at the introduction of each new stream. Figure 13b demonstrates these high-latitude discontinuities, showing root-zone soil moisture from each of the MERRA-2 streams in a region in northern Siberia (60°–75°N, 40°–180°E; Fig. 7), with the 1-yr spinup periods plotted alongside the officially released product.
The situation is similar for terrestrial water storage, with successful spinup in regions where full precipitation corrections were applied (e.g., for the southern North American region; Fig. 13c). There are, however, very obvious deficiencies in the northern Siberia region (Fig. 13d), where at least the first (published) year of each product stream is clearly still impacted by the insufficient spinup, with obvious TWS discontinuities of a few tens of kg m$^{-2}$ on 1 January 1992, 1 January 2001, and 1 January 2011. While it is difficult to isolate the natural interannual variability from the effect of the comparatively dry land surface restarts used to initialize each stream, the figure suggests that at least 2 years are required for the regional soil moisture and TWS estimates to recover from each low initialization.

Since the land surface turbulent fluxes in high latitudes are mostly energy limited, the consequences of this high-latitude soil moisture spin up effect are less serious than they would be in other regions. For example, during the overlapping periods between streams (years 1991, 2000, and 2010 in Fig. 13), the inconsistency in soil moisture caused differences of only about 10 W m$^{-2}$ (10%) in the daily maximum latent heat flux, and about 5 W m$^{-2}$ (25%) in the daily maximum sensible heat flux, which then resulted in differences of about 0.5 K in the 2-m air temperature (not shown). The impact on the peak summer runoff is (relatively) larger, at ~1–2 mm day$^{-1}$ (>50%). These values are an upper limit for the consequences of the soil moisture discontinuity in the MERRA-2 product, since roughly half of the soil moisture discontinuity and spinup effect shown in Fig. 13 is removed during the one year of spinup allowed for each MERRA-2 stream. This leaves only the first published year of each stream (1980, 1992, 2001, and 2011) with some impact from the land spinup in regions where observed precipitation is not used in MERRA-2. Elsewhere, the use of precipitation corrections in MERRA-2 ensures continuity across the stream boundaries in the land surface fields.

e. Global land surface water budget

Finally, for a brief look at the global terrestrial water budget, Fig. 14 and Table 3 provide global land average precipitation, evapotranspiration, and total runoff for MERRA-2, MERRA-Land, and MERRA. The fluxes given here are averaged for 1980–2015 over land excluding inland water and permanently frozen surfaces, that is, the land surfaces where the Catchment model is applied. This land area encompasses 130.232 million km$^2$ in MERRA-2 and 130.176 million km$^2$ in MERRA and MERRA-Land. [See Fig. 6–14 of Bosilovich et al. (2015) for estimates of the global fluxes for all land surfaces, including inland water and permanent ice. Note, however, that the number for MERRA-2 runoff in their figure should read 27 200 km$^3$ yr$^{-1}$.] To facilitate comparison with literature estimates, fluxes are given in units of mm day$^{-1}$ and in units of 10$^3$ km$^3$ yr$^{-1}$.

Figure 14 again illustrates the significant reduction in MERRA-2 of the model-generated precipitation (M2AGCM; 2.96 mm day$^{-1}$) to the precipitation experienced by the land surface (M2CORR; 2.16 mm day$^{-1}$) through the correction procedure (see also Fig. 3). Moreover, Fig. 14 again reflects the fact that the M2CORR precipitation is somewhat higher than that of MERRA-Land (1.92 mm day$^{-1}$) but less than the MERRA precipitation (2.30 mm day$^{-1}$). The figure further demonstrates that the evapotranspiration and runoff estimates

<table>
<thead>
<tr>
<th></th>
<th>Precipitation</th>
<th>ET</th>
<th>Runoff</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10$^3$ km$^3$ yr$^{-1}$</td>
<td>mm day$^{-1}$</td>
<td>10$^3$ km$^3$ yr$^{-1}$</td>
</tr>
<tr>
<td>MERRA-2</td>
<td>102.9</td>
<td>2.16</td>
<td>79.5</td>
</tr>
<tr>
<td>MERRA-Land</td>
<td>91.5</td>
<td>1.92</td>
<td>70.6</td>
</tr>
<tr>
<td>MERRA</td>
<td>109.5</td>
<td>2.30</td>
<td>84.2</td>
</tr>
</tbody>
</table>
from MERRA-2, MERRA-Land, and MERRA scale with the magnitude of the precipitation, with MERRA-2 estimates for evapotranspiration (1.67 mm day\(^{-1}\)) and runoff (0.49 mm day\(^{-1}\)) falling between those of MERRA-Land and MERRA. Put differently, the figure illustrates that MERRA-2, much like MERRA-Land and MERRA, still overestimates evapotranspiration and underestimates runoff compared to other modeling systems and observations (Trenberth et al. 2007; Koster and Mahanama 2012; Rodell et al. 2015). This bias is caused by the imperfect representation in the land surface model of the soil moisture–evapotranspiration and soil moisture–runoff relationships that exist in nature. Koster and Mahanama (2012) developed a strategy that improves the Catchment model’s evaporative fraction (the fraction of net radiative energy released as latent heat) and runoff ratio (the fraction of incident precipitation that becomes runoff), which will facilitate the generation of improved data products in the future.

4. Summary and conclusions

In the MERRA-2 reanalysis, the precipitation experienced by the land surface (and aerosol wet deposition over land and ocean) is derived by merging precipitation observations and the precipitation generated by the MERRA-2 AGCM, with weights depending on latitude. The precipitation correction approach used for MERRA-2 has evolved from that used for the land-only MERRA-Land data product. Key differences are that MERRA-2 applies precipitation corrections within the coupled atmosphere–land reanalysis system, uses more skillful precipitation observations over Africa, and does not apply corrections over the high latitudes. The latter two changes were motivated by the poor quality in these regions of the gauge-based precipitation product used everywhere else in MERRA-2 and everywhere in MERRA-Land. The present paper describes this precipitation correction algorithm in detail (section 2; Figs. 1–3).

Our results demonstrate the improvements in the corrected MERRA-2 precipitation (M2CORR) over that used in MERRA-Land and over that generated by the AGCMs in the MERRA-2 and MERRA systems. Comparisons against the GPCPv2.2 precipitation observations show that the M2CORR precipitation generally has lower bias and RMSE values and higher correlation and anomaly correlation values than both the MERRA-Land precipitation and the model-generated precipitation from MERRA-2 and MERRA (Figs. 4 and 5). While the diurnal cycle of the M2CORR precipitation has reasonable amplitudes compared to TRMM TMPA 3B42 observations, the time of day of maximum precipitation is inherited from MERRA and is thereby unrealistic (Fig. 10).

The application of the precipitation corrections within the MERRA-2 coupled atmosphere–land system allows the evapotranspiration and thus the subsequent near-surface air temperature and humidity in MERRA-2 to respond to the observed precipitation. Consequently, the MERRA-2 near-surface meteorological data are more self-consistent than those of MERRA-Land (Fig. 12). Compared to MERRA-Land, MERRA-2 not only provides an extended record, it also provides an enhanced forcing dataset for land-only modeling and data assimilation applications.

Moreover, the precipitation corrections facilitate the consistent land surface initialization of the four MERRA-2 production streams at low and middle latitudes, thereby ensuring smooth transitions in the land surface estimates across stream boundaries and avoiding adverse land surface spinup effects in the MERRA-2 product. Such adverse effects are unavoidable in some high-latitude regions where the MERRA-2 land surface experiences the model-generated precipitation, which is biased against the observational precipitation product that was used to generate the MERRA-2 initial conditions (Fig. 13). It should be emphasized, however, that at high latitudes precipitation observations from gauges and satellites are of limited quality. Finally, an analysis of the global land water budget (excluding inland water and permanently frozen surfaces) suggests that land surface evapotranspiration in MERRA-2 is still too high and that runoff is too low (Fig. 14), with a bias in evaporative fraction that is similar to that of the MERRA-Land and MERRA products.

As shown elsewhere, the improved precipitation leads to better-quality land surface estimates in MERRA-2 (Reichle et al. 2017) and MERRA-Land (Reichle et al. 2011), with improvements in soil moisture, terrestrial water storage, and runoff skill. The land surface precipitation forcing in reanalysis products, however, still shows much room for improvement. The observations-based precipitation products used for MERRA-2 and MERRA-Land are subject to errors that result from considerable changes in the gauge network over the reanalysis period, sometimes in combination with errors in the input gauge measurements. Examples include the discontinuities in the observed CPCU precipitation in Myanmar (Fig. 6) and South America (Fig. 8), which carry directly into the corrected precipitation experienced by the land surface in MERRA-2 and MERRA-Land. Such adverse effects are particularly pronounced in regions where the density of the gauge network is low.

Climate-quality precipitation products such as the GPCPv2.2 precipitation are more scrutinized and probably less prone to errors than the operational CPC products, but at the time of MERRA-2 production, only
the monthly GPCPv2.1 product was available for the years prior to 1996. Moreover, GPCP products that include gauge measurements are generally not available within the latency requirements of the MERRA products, which are updated to the present within weeks in order to facilitate a wide variety of applications. Therefore, better-quality precipitation products that combine gauge and satellite measurement with a latency of weeks rather than months are needed to improve the land surface precipitation and thus the terrestrial water budget in forthcoming reanalysis datasets.

Acknowledgments. Funding for this work was provided by the NASA Modeling, Analysis, and Prediction program. Computational resources were provided by the NASA High-End Computing program through the NASA Center for Climate Simulation. We are grateful for the datasets and data archiving centers that supported this work and appreciate those who made our study possible, including the MERRA-2 team at the GMAO and staff at GSFC and NOAA CPC. Thanks also to three anonymous reviewers and to M. Bosilovich for valuable comments and discussion.

REFERENCES


