

Assessment of MERRA-2 Land Surface Energy Flux Estimates

CLARA S. DRAPER^a

GESTAR, Universities Space Research Association, and NASA Global Modeling and Assimilation Office, Greenbelt, Maryland

ROLF H. REICHLE AND RANDAL D. KOSTER

NASA Global Modeling and Assimilation Office, Greenbelt, Maryland

(Manuscript received 24 February 2017, in final form 11 October 2017)

ABSTRACT

In the Modern-Era Retrospective Analysis for Research and Applications version 2 (MERRA-2) system the land is forced by replacing the model-generated precipitation with observed precipitation before it reaches the surface. This approach is motivated by the expectation that the resultant improvements in soil moisture will lead to improved land surface latent heating (LH). Here aspects of the MERRA-2 land surface energy budget and 2-m air temperatures T^{2m} are assessed. For global land annual averages, MERRA-2 appears to overestimate the LH (by 5 W m^{-2}), the sensible heating (by 6 W m^{-2}), and the downwelling shortwave radiation (by 14 W m⁻²) while underestimating the downwelling and upwelling (absolute) longwave radiation (by $10-15 \,\mathrm{Wm^{-2}}$ each). These results differ only slightly from those for NASA's previous reanalysis, MERRA. Comparison to various gridded reference datasets over boreal summer (June-August) suggests that MERRA-2 has particularly large positive biases ($>20 \text{ Wm}^{-2}$) where LH is energy limited and that these biases are associated with evaporative fraction biases rather than radiation biases. For time series of monthly means during boreal summer, the globally averaged anomaly correlations Ranom with reference data were improved from MERRA to MERRA-2, for LH (from 0.39 to 0.48 vs Global Land Evaporation Amsterdam Model data) and the daily maximum T^{2m} (from 0.69 to 0.75 vs Climatic Research Unit data). In regions where T^{2m} is particularly sensitive to the precipitation corrections (including the central United States, the Sahel, and parts of South Asia), the changes in the $T^{2m} R_{anom}$ are relatively large, suggesting that the observed precipitation influenced the T^{2m} performance.

1. Introduction

The NASA Global Modeling and Assimilation Office recently released the Modern-Era Retrospective Analysis for Research and Applications version 2 (MERRA-2; Gelaro et al. 2017). This new global reanalysis product replaces and extends the original MERRA atmospheric reanalysis (Rienecker et al. 2011), as well as the MERRA-Land reanalysis (Reichle et al. 2011). In addition to several other major advances, MERRA-2 uses observed precipitation in place of model-generated precipitation at the land surface during the atmospheric model integration. The use of observed precipitation in MERRA-2 was refined from the approach used for MERRA-Land (Reichle et al. 2017b), which was an offline (land only) replay of MERRA forced by atmospheric fields from MERRA but with the precipitation forcing corrected using gauge-based observations.

The motivation for using observed precipitation in reanalyses is that precipitation is the main driver of soil moisture, which in turn controls the partitioning of incident surface radiation between latent heat (LH) and sensible heat (SH) fluxes back to the atmosphere. Reichle et al. (2017a) show that both MERRA-2 and

DOI: 10.1175/JCLI-D-17-0121.1

© 2018 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-17-0121.s1.

^a Current affiliation: Physical Sciences Division, NOAA/ESRL, and CIRES, Boulder, Colorado.

Corresponding author: Clara Draper, clara.draper@noaa.gov

MERRA-Land have improved upon the land surface hydrology of MERRA, showing better agreement with independent observational time series of soil moisture, terrestrial water storage, streamflow, and snow amount. Here, we extend this work, by evaluating the MERRA-2 surface energy budget and 2-m temperatures T^{2m} over land. In particular, we focus on whether the improved hydrology in both the (offline) MERRA-Land and the (coupled landatmosphere) MERRA-2 datasets translates into the expected improvements to the monthly mean LH and SH. We also expand previous work by evaluating the reanalyses' land surface output globally, rather than focusing on locations with high-quality ground-based observations.

We start by comparing the long-term annual global energy budget over land from MERRA-2, MERRA-Land, and MERRA to state-of-the-art estimates from the literature. These literature estimates, from Trenberth et al. (2009), Wild et al. (2015), and the NASA Energy and Water Cycle Studies (NEWS) program (NEWS Science Integration Team 2007; L'Ecuyer et al. 2015), were each produced by carefully combining multiple input datasets with global energy balance constraints. Taken together they represent our best understanding of the long-term annual mean energy budget over land.

Next, we consider global maps of the performance of the land surface turbulent heat fluxes from each reanalysis, as a step toward linking differences in performance to the dominant local physical processes and to the potential improvements obtained from the use of the observed precipitation in MERRA-2. We focus on the boreal summer [June–August (JJA)], since land–atmosphere coupling is strongest and surface turbulent heat fluxes are most active in the summer.

Unfortunately, there are no standard global gridded reference datasets against which the reanalysis LH and SH can be evaluated. Several recent efforts have compared global LH estimates from different combinations of reanalyses, offline land surface models, and diagnostic methods. Most estimates generally agree on the regional patterns and local seasonal cycle of LH, although there is considerable disagreement in the absolute values and temporal behavior across different flux estimates (Jiménez et al. 2011; Mueller et al. 2011; Miralles et al. 2011). Additionally, uncertainty in the basic model structure is the largest source of disagreement (Schlosser and Gao 2010; Mueller et al. 2013). While ground-based observations are available from tower-mounted eddy covariance sensors (e.g., Baldocchi et al. 2001), the number of towers (in the hundreds) is well below the sampling needed for global estimation (and their locations are not designed to sample globally representative land cover types). Additionally, the measurements themselves have considerable uncertainty and limited spatial representativeness (up to 1 km).

In the absence of a standard reference, we compare the JJA reanalysis turbulent heat flux estimates to two different gridded reference datasets: the Global Land Evaporation Amsterdam Model (GLEAM) (Miralles et al. 2011; Martens et al. 2017) for LH and FLUXNET-Model Tree Ensembles (MTE) (Jung et al. 2010) for LH and SH. These datasets were selected for several reasons: (i) they are among the state of the art, (ii) they are available globally for multidecadal time periods, (iii) they are independent of each other, and (iv) they rely on very different estimation methodologies (water balance modeling for GLEAM and upscaling of tower measurements for MTE). Since neither GLEAM nor MTE represents direct observations of the turbulent heat fluxes, we also compare each reanalysis to tower-based eddy covariance observations from the FLUXNET2015 dataset (FLUXNET 2015). To determine the potential contribution of radiation biases to regional LH and SH biases, we also compare the reanalyses' surface radiation fields for JJA against gridded observations from the Clouds and the Earth's Radiant Energy System (CERES) and Energy Balanced and Filled (EBAF) dataset (Kato et al. 2013).

Finally, to test whether the changes in the surface energy budget from MERRA to MERRA-2 have affected the atmospheric boundary layer, we also evaluate the JJA monthly mean daily minimum and maximum T^{2m} against observations from the Climatic Research Unit (CRU) at the University of East Anglia (Harris et al. 2014). Improvements in MERRA-2 due to the use of observed precipitation cannot be isolated from the many other advances distinguishing MERRA-2 from MERRA. Consequently, we establish whether the improvements in the surface turbulent fluxes and T^{2m} are at least consistent with the expected improvements from the use of observed precipitation, by cross-referencing the evaluation results against the regional sensitivity to precipitation and/or soil moisture.

This paper is organized as follows. Section 2 summarizes the reanalysis and reference datasets used, and section 3 presents the results, including evaluation of the (i) reanalyses' annual global land energy budget averages, (ii) the spatially distributed mean JJA energy budget and T^{2m} , and (iii) the temporal behavior of the JJA turbulent heat fluxes and T^{2m} . We also identify regions of sensitivity to the observed precipitation forcing in MERRA-2, for cross-reference against the evaluation results. Our findings are summarized in section 4.

2. Methodology and data

a. The reanalyses

The coverage and resolution of each reanalysis is summarized in Table 1, with further details below. MERRA

Dataset	Variables used	Output coverage and resolution (variable dataset citation, where available)
MERRA-2		1980–ongoing, hourly, $5/8^{\circ} \times 0.5^{\circ}$
	LH, SH, LW _{net} , SW _{net} ,	Global land (Global Modeling and Assimilation Office 2015b)
	LW_d	Global surface (Global Modeling and Assimilation Office 2015a)
	$T_{\rm max}^{2\rm m}, T_{\rm min}^{2\rm m}$	Global surface (Global Modeling and Assimilation Office 2015c)
MERRA-Land		1980–February 2016, hourly, $2/3^{\circ} \times 0.5^{\circ}$
	LH, SH, LW _{net}	Global land (Global Modeling and Assimilation Office 2008c)
MERRA		1979–February 2015, hourly, $2/3^{\circ} \times 0.5^{\circ}$
	LH, SH, LW _{net} , SW _{net} ,	Global land (Global Modeling and Assimilation Office 2008b)
	LW_d	Global surface
	$T_{\rm max}^{2\rm m}, T_{\rm min}^{2\rm m}$	Global surface (Global Modeling and Assimilation Office 2008a)
ERA-Interim		1979–ongoing, monthly mean, 79-km
	LH, SH	Global surface

TABLE 1. The reanalyses.

(Rienecker et al. 2011) and MERRA-2 (Gelaro et al. 2017) are atmospheric reanalyses produced with the NASA Goddard Earth Observing System Model, version 5 (GEOS-5), modeling and data assimilation system and were designed to provide historical analyses of the hydrological cycle across a broad range of climate time scales. To address shortcomings in the land surface hydrology from MERRA, MERRA-Land (Reichle et al. 2011) was released as an offline (land only) replay of MERRA, with the model-generated precipitation corrected using rain gauge observations and with minor, but important, model parameter changes. MERRA-2 features several major advances from MERRA, including an updated atmospheric general circulation model, an updated atmospheric assimilation system, an interactive aerosol scheme, and the use of observed precipitation at the land surface (and to compute wet aerosol deposition). In addition to the land model updates from MERRA-Land, MERRA-2 includes several more updates relevant to the land, as outlined in Reichle et al. (2017a). Most notably, the surface turbulence scheme was revised, generally resulting in enhanced SH over land (Molod et al. 2015).

The method used to apply the observed precipitation at the land surface in MERRA-2 was refined from that used in MERRA-Land (Reichle and Liu 2014; Reichle et al. 2017b). In MERRA-Land the precipitation was corrected with daily Climate Prediction Center (CPC) Unified (CPCU; Chen et al. 2008) precipitation observations everywhere. For MERRA-2 the input precipitation differs in two ways: (i) in the high latitudes the MERRA-2 modelgenerated precipitation is retained, and (ii) over Africa the MERRA-2 precipitation is corrected with pentad-scale blended satellite and gauge-based observations from the CPC Merged Analysis of Precipitation (CMAP; Xie and Arkin 1997) and the Global Precipitation Climatology Project (GPCP; Huffman et al. 2009), version 2.1.

The land surface turbulent fluxes from the NASA reanalyses (MERRA-2, MERRA-Land, and MERRA)

have not been explicitly evaluated globally. However, Jiménez et al. (2011) and Mueller et al. (2011) both included MERRA LH when merging multiple LH global land datasets into a single enhanced estimate (see section 2b), and in both studies MERRA was among the highest of the input LH estimates used. Additionally, Jiménez et al. (2011) noted a sharp gradient in the MERRA LH around 10°S in the tropics that was not present in other LH estimates. This bias gradient was traced to MERRA's excessive rainfall canopy interception and precipitation errors (Reichle et al. 2011). Consequently, the interception reservoir parameters were revised for MERRA-Land (and MERRA-2) to eliminate this feature (the interception reservoir update was the most significant modeling change from MERRA to MERRA-Land).

An additional reanalysis, ERA-Interim, from the European Centre for Medium-Range Weather Forecasts (Dee et al. 2011), is included in the evaluation of the temporal behavior of the turbulent fluxes. In contrast to the NASA reanalyses, ERA-Interim includes a land surface updating scheme (de Rosnay et al. 2014). Specifically, the soil moisture, soil temperature, and snow temperatures are updated to minimize errors in the forecast screen-level relative humidity and temperature, while the snow depths are updated using satellite- and ground-based snow-cover and snow-depth observations.

b. Annual global land energy budget estimates

We compare the reanalyses' annual global land energy budgets to three state-of-the-art estimates, from Trenberth et al. (2009), Wild et al. (2015), and the NEWS program estimates of L'Ecuyer et al. (2015). Each of these is based on a weighted merger of multiple modeled and observed datasets, and each applies to the energy budget at the start of the twenty-first century. For Trenberth et al. (2009) we have used their estimates for the CERES period of 2000–04; Wild et al. (2015) nominally refers to the same period, while L'Ecuyer et al.

Dataset	Variables used	Output coverage and resolution	Dependencies, error estimates where available			
GLEAM v3.1a	LH	1980–2016 daily mean, 0.25° global land Uses a precipitation dataset that includes CPCU (use in MERRA-2, MERRA-Land) and ERA-Interim precipitation, uses T^{2m} and radiation from ERA-I Compare tower obs, average ubRMSE: 20 W m ⁻² ; average R_{mode} : 0.42, Full details in section 2c(1).				
MTE	LH, SH	1982–2011 monthly mean, 0.5° global land, excluding nonvegetated regions	 Trained on an earlier generation of the FLUXNET2015 dataset. Uses a CRU-based T^{2m} dataset and CPCU precipitation (neither strongly influences temporal behavior). Compare withheld tower obs, average RMSE: 15 W m⁻² (LH and SH); average Ragon; 0.57 (LH); 0.60 (SH). Full details in section 2c(2). 			
CRU v4.00	$T_{\min}^{2m}, T_{\max}^{2m}$	1901–2015 monthly mean, 0.5° global land (data not informed by station obs have been removed)	 Input station obs will overlap with T^{2m} assimilated into ERA-Interim. Locally, it will be more uncertain where input station obs are sparse. Full details in section 2c(3). 			
CERES-EBAF, v4.0	SW_d, SW_u, LW_d, LW_u	March 2000–February 2016 monthly mean, 1° global surface	Uses atmospheric profile and T_{skin} from same system as used in the NASA reanalyses (results in strong dependence for LW _u , LW _d) Compare ground obs average RMSE: 12 (SW _d), 10 W m ⁻² (LW _d). Full details in section 2c(4).			

TABLE 2. The gridded reference datasets.

(2015) nominally refers to 2000–09. Note that the MERRA LH and SH over land were used as one of the inputs in NEWS.

These three global energy budget studies all provide continental and oceanic energy estimates, where "continental" is defined as nonocean and so includes land, land ice, and lakes but excludes inland seas. By contrast, the land estimates from MERRA-2, MERRA-Land, and MERRA apply to the area modeled by the land surface model, excluding land ice, lakes, and inland seas. The discrepancy due to the inclusion or exclusion of land ice is significant: land ice accounts for 10% of the continental area, with Antarctica making up 95% of this. NEWS provides energy budgets for each continent separately (L'Ecuyer et al. 2015), and we use their (balance constrained) energy budget estimates to approximate the land-only energy budget terms by subtracting the area-weighted Antarctica estimates from the global continental estimates. We then use our land-only NEWS estimates to approximate the continental to land ratio for each NEWS energy budget term. By assuming that the same ratios apply to Trenberth et al. (2009) and Wild et al. (2015) we then approximate land-only estimates for the latter two studies. L'Ecuyer et al. (2015) and Wild et al. (2015) both provide uncertainty ranges for their globally averaged continental estimates, which we have applied unchanged to our approximated land-only estimates.

For LH, we have also used three additional global land annual average estimates from the hydrology community, from Jiménez et al. (2011), Mueller et al. (2011), and Mueller et al. (2013). These estimates are also based on merging modeled and observed estimates. Jiménez et al. (2011) applies to global land (using a similar land definition to the NASA reanalyses) for 1994, while Mueller et al. (2011) applies to the global land area, excluding the Sahara, from 1989 to 1995, and Mueller et al. (2013) applies to the global land plus Greenland for 1989–2005. As previously noted, MERRA LH was one of the inputs used in the multiproduct mergers of Jiménez et al. (2011) and Mueller et al. (2011).

c. Gridded reference datasets

The coverage and resolution of each gridded reference dataset, together with a brief summary of important interdependencies with other datasets or reanalyses used in the study and uncertainty estimates (where available), are summarized in Table 2, with further details provided below.

1) GLEAM

GLEAM (version 3.1a) provides daily estimates of terrestrial evapotranspiration, estimated from satellite and reanalysis forcing using a Priestley and Taylor– based model (Miralles et al. 2011; Martens et al. 2017). The precipitation is from the Multi-Source Weighted-Ensemble Precipitation, which is a multimodel merger of established precipitation datasets, including the same CPCU dataset used in MERRA-Land and MERRA-2, as well as ERA-Interim precipitation [the latter is used predominantly in the high latitudes, where observed precipitation datasets are more uncertain (Beck et al. 2017)]. The net surface radiation and T^{2m} are from ERA-Interim. Compared to independent observations from 91 flux towers, GLEAM has an average unbiased root-mean-square error (ubRMSE; or error standard deviation) of 20 Wm^{-2} and an average anomaly correlation of 0.42 (Martens et al. 2017).

2) MTE

MTE provides global estimates of carbon dioxide, energy, and water fluxes at the land surface, calculated using a machine learning technique to upscale halfhourly energy-balance-corrected eddy covariance observations from 253 FLUXNET tower observations (Jung et al. 2011). The input FLUXNET observations are from the La Thuile data release, an earlier generation of the FLUXNET2015 dataset used here (to be introduced in section 2d). CPCU precipitation (again, used directly in MERRA-Land and MERRA-2) and a T^{2m} dataset based on CRU data (Jung et al. 2011) are used as predictive (regression) variables in the MTE. However, these meteorological data have little impact on the MTE monthly anomalies, which are instead driven by the vegetation variability as observed by the fraction of absorbed photosynthetically active radiation (fPAR; Jung et al. 2010). When 20% of the FLUXNET training data was withheld from the algorithm, the average root-mean-square error (RMSE) with the withheld data was 15 Wm^{-2} , for both LH and SH, and the average anomaly correlation was 0.57 for LH and 0.60 for SH (Jung et al. 2011). In general, the MTE method is better suited to estimating spatial variability and the seasonal cycle than it is to capturing interannual anomaly patterns (Jung et al. 2009).

3) CRU TEMPERATURE DATA

CRU time series version 4.00 (TS v4.00) provides gridded monthly means of the daily mean, minimum, and maximum temperature over land (Harris et al. 2014a,b). The temperatures are calculated from qualitycontrolled climate station data, which are interpolated onto the grid according to an assumed correlation decay distance (set to 1200 km for temperature variables). In instances where no station data are available within the assumed decay distance, the published data value defaults to the climatology. Here, such climatological values have been screened out. Also, we require at least 10 data points to estimate each statistic for a given grid cell. Even with this screening, the gridded output will be much less certain when/where station coverage is less dense, which occurs over Africa, South America, central Australia, and the high latitudes.

4) CERES-EBAF RADIATION DATA

CERES-EBAF version 4.00 surface radiances are produced with a radiative transfer model after adjusting modeled and observed input data for consistency with topof-atmosphere (TOA) CERES-EBAF radiation (Kato et al. 2013). The input data (surface, cloud, and atmospheric properties) are adjusted according to their observation-based estimated uncertainties. The input temperature and humidity profiles and land surface skin temperature T_{skin} are from NASA's GEOS-5.4.1 modeling and assimilation system, the same system (although a different version) used in MERRA and MERRA-2.

The CERES output shortwave irradiances are primarily determined by (observation based) TOA radiation and clouds; hence, they are reasonably independent of the MERRA and MERRA-2 reanalyses (Kato et al. 2013). On the other hand, the CERES output longwave irradiances, and particularly the upwelling longwave LW_{u} , are strongly dependent on the GEOS-5 T_{skin} input. However, the CERES algorithm does adjust its input GEOS-5 T_{skin} with observation-based cloud information, so comparison between the CERES-EBAF and GEOS-5 LW_u partly reflects these observation-based adjustments, even though the two fields are not independent. Compared to independent ground-based observations from 24 sites over land, the RMSE of the CERES-EBAF radiation is 12 Wm^{-2} for downwelling shortwave SW_d and 10 Wm^{-2} for downwelling longwave LW_d (Kato et al. 2017). For the regional estimates over land, Kato et al. (2017) estimated the uncertainty to be $12 \text{ W} \text{ m}^{-2}$ for SW_d , $4 \text{ W} \text{ m}^{-2}$ for upwelling shortwave SW_u , 10 W m⁻² for LW_d, and 18 W m⁻² for LW_u.

5) GRIDDED DATASET PROCESSING

As noted in Tables 1 and 2 some of the reference datasets and reanalyses used here publish output that applies only to the land fraction within each grid cell, while others publish a single estimate that applies to all surface types (land, permanent land ice, lakes, and ocean) within each grid cell. All of the gridded datasets and reanalyses were screened by removing all grid cells where the MERRA-2 land fraction was less than 50% (after interpolation to the relevant resolution) and then aggregated up to monthly means and 1° spatial resolution. All maps of global statistics are based on the boreal summer months of JJA only, and each comparison is made over the maximum available coincident time period, with the time periods noted in the relevant figure captions. The anomaly correlations R_{anom} are evaluated based on anomalies from the mean seasonal cycle (calculated by subtracting the time period mean separately for each calendar month). The gridded reference datasets were also used to estimate the annual global land average values, for which the (interpolated) MERRA-2 land area in each grid cell was used.

d. FLUXNET2015 tower observations

The FLUXNET2015 (FLUXNET 2015) sites were selected by downloading all Tier 1 observations at nonirrigated sites within grid cells classified as land at 1° resolution [as derived previously in section 2c(5)] and for which at least a 10-yr data record is available. Eddy covariance sensors underestimate turbulent heat fluxes and do not generally close the energy balance (Wilson et al. 2002); hence, we used the FLUXNET2015 energy balance closure-corrected LH and SH [see FLUXNET (2015) for details of the correction method]. While these corrections are rather uncertain, the corrected LH and SH showed better agreement with all of the reanalyses in Table 1 in terms of the means across all sites and the correlation of the means between the sites (while having negligible impact on the mean time series anomaly correlations). The balance-corrected FLUXNET data were screened to retain only days with less than 10% gap-filled data and only sites with data for at least 2550 days (\sim 70% of 10 years). The monthly means were then calculated for months with at least 15 days of observations after the above screening, and the corresponding reanalysis monthly means were estimated using the same days. The resulting FLUXNET monthly time series were visually inspected, and obviously unrealistic features were removed. Four sites with unrealistic time series were removed. Of the remaining 21 stations, just one was in the Southern Hemisphere. Since our evaluation focuses on the boreal summertime, this site was excluded. The remaining 20 sites that have been used in this study are listed in Table 1 of the supplemental material.

3. Results

a. Annual global land energy budgets

The globally averaged annual land energy budget estimates for MERRA-2, MERRA-Land, and MERRA are illustrated in Fig. 1, with numerical values given in Table 3. For each term, the estimates for MERRA-2 and MERRA are similar (within $2-3 \text{ Wm}^{-2}$), while the partitioning of R_{net} into LH and SH differs for MERRA-Land, which is shifted toward greater SH. Compared to MERRA, MERRA-Land has 11 Wm^{-2} more SH, and 8 Wm^{-2} less LH, with the difference in R_{net} due to decreased LW_u (recall that in the offline MERRA-Land SW_{net} and LW_d are taken directly from MERRA).

Figure 1 also includes the energy budget estimates from the literature (see section 2b), as well as the annual

global land averages for each of the gridded reference datasets in Table 2. In Fig. 1a, the MERRA-2 and MERRA global land LH are higher than all of the other estimates (although MERRA-2 is within the Jiménez et al. (2011) and Wild et al. (2015) confidence intervals). The three (land adjusted) LH estimates from the global energy budget studies (Trenberth et al. 2009; Wild et al. 2015; NEWS) are very similar to each other and to MTE, GLEAM, Mueller et al. (2011), and MERRA-Land (all are within 1 Wm^{-2}). While the other two LH estimates from the hydrology community (Jiménez et al. 2011; Mueller et al. 2013) are higher, they are not as high as MERRA-2 and MERRA. Compared to the average of the three global land energy budget estimates, the MERRA-2 LH is biased high by 6 W m^{-2} (15%), while MERRA is biased high by 9 Wm^{-2} (21%), and MERRA-Land is much closer, being biased high by just $1 \,\mathrm{W}\,\mathrm{m}^{-2}$ (2%).

For the global land SH in Fig. 1b, MERRA-2 and MERRA are both higher than Trenberth et al. (2009) and Wild et al. (2015), although lower than NEWS (but within the NEWS confidence interval) and very close (within 1 Wm^{-2}) to MTE. Compared to the average of the three global land energy budget estimates, MERRA-2 is biased high by 5 Wm^{-2} (15%) and MERRA by 4 Wm^{-2} (12%), while MERRA-Land is much higher, with a bias of 15 Wm^{-2} (42%).

The positive biases in both LH and SH from the reanalyses indicate a positive bias in the incident energy at the land surface. Indeed, Fig. 1g shows that $R_{\rm net}$ from the reanalyses exceeds the three global energy budget estimates, although MERRA-2 (the lowest of the reanalyses) is only slightly higher (2 W m^{-2}) than the CERES-EBAF value. Compared to the average of the three global energy budget estimates, the R_{net} biases are 12 Wm^{-2} (15%) for MERRA-2, 13 Wm^{-2} (17%) for MERRA, and 16 Wm^{-2} (21%) for MERRA-Land. Figures 1c-f show that the positive R_{net} bias in MERRA-2 and MERRA is made up of a large positive bias in SW_d combined with insufficient LW_u , both partly offset by underestimated LW_d . For SW_d (Fig. 1c) MERRA-2 and MERRA are higher than all three global land energy budget estimates and CERES-EBAF, with a bias compared to the three-product average of 14 Wm^{-2} (7%) for MERRA-2 and 16 Wm^{-2} (8%) for MERRA. For SW_u (Fig. 1d), MERRA-2 and MERRA are both above NEWS, Trenberth et al. (2009), and CERES-EBAF, but below Wild et al. (2015) (although within the confidence interval). Both are biased high by 3 W m^{-2} (8%), compared to the threeproduct average. For LW_d (Fig. 1e), MERRA-2 and MERRA are lower than the other estimates, with biases of -11 Wm^{-2} (-3%) for MERRA-2 and -10 Wm^{-2}





FIG. 1. The global annual mean energy budget over land from the reanalyses [MERRA-2 (M-2); MERRA-Land (M-L); MERRA (M)], the literature [NEWS (NEW), Trenberth et al. (2009) (Tre), Wild et al. (2015) (Wil), Jiménez et al. (2011) (Jim), Mueller et al. (2011) (Mu1), and Mueller et al. (2013) (Mu3)], and the gridded reference datasets [MTE, GLEAM (GLM), and CERES (CER)], for (a) LH, (b) SH, (c) SW_d, (d) SW_u, (e) LW_d, (f) LW_u, and (g) R_{net} . For NEW, Tre, and Wil, the land mean has been approximated from published continental means as described in section 2b. Error bars are included where provided; for NEW and Wil these span the possible range described by multiple products, and for Jim and Mu1 these represent one standard deviation across multiple products (see citations for full details).

(-3%) for MERRA against the three-product average. For LW_u (Fig. 1f) MERRA-2, MERRA-Land, and MERRA are again lower than the other plotted estimates, with biases of -11 Wm^{-2} (-3%) for MERRA-2, -13 Wm^{-2} (-3%) for MERRA-Land, and -10 Wm^{-2} (-3%) for MERRA.

The literature estimates in Fig. 1 are presented as long-term means, and each represents different temporal and spatial coverage. Likewise, the annual global land averages for the gridded reference datasets in Fig. 1 are based on the full available (spatial and temporal) coverage for each. However, the gridded reference datasets and reanalyses can be cross-screened to ensure that they are compared with consistent coverage. With this cross-screening, the MERRA-2 LH bias estimate is 7 Wm^{-2} versus GLEAM, or 9 Wm^{-2} versus MTE, while the SH bias is 1 Wm^{-2} versus MTE, and the

TABLE 3. Global annual land average energy budget from the NASA reanalyses (W m⁻²), estimated over an area of 130.2×10^6 km².

	SW_d	SW_u	LW_d	LW_u	R _{net}	LH	SH
MERRA-2	204.6	40.7	312.6	385.5	91.0	47.8	42.2
MERRA-Land	As for MERRA			384.1	95.1	42.5	52.1
MERRA	206.5	40.9	313.7	386.7	92.6	50.4	41.2

radiation biases versus CERES-EBAF are 10 Wm^{-2} for SW_u, 2 Wm^{-2} for SW_d, -18 Wm^{-2} for LW_d, -11 Wm^{-2} for LW_u, and $< 0.5 \text{ Wm}^{-2}$ for R_{net}. In general, the above-quoted biases (calculated after cross-screening) are all close (within 1 Wm^{-2}) to the values estimated from the data plotted in Fig. 1 (which does not include cross-screening), with the exception of the LH bias versus MTE, which is 6 Wm^{-2} without cross-screening (compared to 9 Wm^{-2}). This discrepancy is due to the MTE global mean being lower than it otherwise would be, due to the lack of coverage over the Sahara (which has near-zero annual mean LH).

b. Land–atmosphere coupling and the MERRA-2 precipitation corrections

Here, we identify regions where, in MERRA-2, (i) LH is sensitive to precipitation (or soil moisture), and (ii) the daily maximum T^{2m} (T^{2m}_{max}) is sensitive to the applied precipitation corrections. These regions can then be used to determine where the change in performance from MERRA to MERRA-2 is most likely associated with the precipitation corrections. Note that for part ii above, the diurnal temperature range could be expected to have a stronger signal of the daytime turbulent heat fluxes (Betts et al. 2017); however, a preliminary comparison (not shown) revealed similar results for the diurnal temperature range (DTR) and T^{2m}_{max} , and we have presented the results for T^{2m}_{max} since this variable is included in the published MERRA-2 datasets.

1) SOIL MOISTURE AND LATENT HEATING

To first order, LH (or evapotranspiration) from soil and vegetation surfaces can be conceptualized as either a moisture- or energy-limited process. In drier conditions (i.e., for soil moisture below some critical point), LH is moisture-limited in that it is restricted by the amount of soil moisture available for evapotranspiration. Temporal variations in LH will then be correlated with the plant available soil moisture (principally, the soil moisture in the root zone). In contrast, in more humid conditions LH is energy limited; there is sufficient soil moisture available for evapotranspiration, so LH proceeds at the maximum rate determined by atmospheric water demand, and temporal variations in LH are accordingly correlated with temporal variations in atmospheric demand (net radiation, atmospheric humidity deficit, and wind) rather than soil moisture.

Figure 2 shows the squared correlation between the JJA monthly anomaly MERRA-2 LH and root-zone soil moisture (SM) R_{anom}^2 (LH, SM). Lower R_{anom}^2 (LH, SM) indicates a tendency toward energy-limited LH, which for the boreal summer occurs in the high latitudes, central and eastern Europe, the eastern United States, southern



FIG. 2. The R_{anom}^2 between monthly anomalies of LH and rootzone SM in MERRA-2 for JJA. No value is plotted where the correlation is negative.

China, and much of the tropics (the Amazon, equatorial Africa, and Southeast Asia). On the other hand, higher R_{anom}^2 (LH, SM) indicates a tendency toward moisturelimited LH and occurs across the remainder of the low and midlatitudes. While we have plotted JJA to focus on the boreal summer, there are still regions of moisturelimited LH in the Southern Hemisphere during austral winter, specifically in arid regions (southern Africa, much of Australia, and the desert and steppe regions of South America).

2) PRECIPITATION FEEDBACK ON AIR TEMPERATURE

Figure 3 shows maps of the squared anomaly correlation R_{anom}^2 between anomaly time series of JJA MERRA-2 monthly T_{max}^{2m} and anomaly time series of 2-month (current + previous month) averaged MERRA-2 precipitation. For example, the June T_{\max}^{2m} is compared to the (May + June) precipitation, while the July $T_{\text{max}}^{2\text{m}}$ is compared to the (June + July) precipitation, and so on. The precipitation is lagged like this to allow the precipitation signal to accumulate in the soil and influence the subsequent $T_{\text{max}}^{2\text{m}}$. In Fig. 3a the MERRA-2 model-generated precipitation (PRECTOT) is used, while in Fig. 3b the MERRA-2 observation-corrected precipitation (PRECTOTCORR) is used. The R_{anom}^2 values are plotted only for negative R values, since the dominant local relationship between precipitation and daytime temperature is negative (i.e., under moisture-limited conditions, reduced precipitation leads to reduced soil moisture, which limits LH and increases SH and T^{2m}). Figure 3b reflects the modeled relationship in MERRA-2 between precipitation falling on the surface and $T_{\text{max}}^{2\text{m}}$. Even with the difference in time periods, the patterns are similar to those found across the contiguous United States from observations by Koster et al. (2015).

Figure 3c then shows the difference between $R_{\text{anom}}^2(T_{\text{max}}^{2m}, \text{PRECTOTCORR})$ and $R_{\text{anom}}^2(T_{\text{max}}^{2m}, \text{PRECTOT})$. This



FIG. 3. JJA sensitivity of the monthly mean $T_{\text{max}}^{2\text{m}}$ to precipitation in MERRA-2: R_{anom}^2 between the monthly mean $T_{\text{max}}^{2\text{m}}$ anomalies and the two-monthly (current + previous months) precipitation anomalies for (a) PRECTOT and (b) PRECTOTCORR, together with their difference (c) $\Delta R_{\text{anom}}^2 = R_{\text{anom}}^2$ (PRECTOTCORR, $T_{\text{max}}^{2\text{m}}$) – R_{anom}^2 (PRECTOT, $T_{\text{max}}^{2\text{m}}$). Values are plotted only where the correlation between $T_{\text{max}}^{2\text{m}}$ and precipitation is negative.

difference ΔR^2_{anom} is the increase in the fraction of variance in T^{2m}_{max} explained by the (observed) precipitation seen by the land (PRECTOTCORR) over that explained by the model-generated precipitation (PRECTOT). It thus provides a measure of the local impact of the observed precipitation on the MERRA-2 T^{2m}_{max} . This measure is sensitive to both the magnitude of the precipitation corrections and the local response of the atmospheric model to those corrections. Note that the lack of sensitivity in the high latitudes was inevitable for this metric, since the model-generated precipitation is used there.

For the boreal summer, the strongest impact of the observed precipitation, which can explain more than 25% of the $T_{\text{max}}^{2\text{m}}$ variance, is indicated in the central United States, Central America, the northern tip of

South America, a broad swath along the Sahel, and parts of South Asia. Note that these regions do not directly correspond to the regions of strongest moisture-limited LH in Fig. 2, for at least two reasons. First, a strong sensitivity of evapotranspiration to soil moisture (Fig. 2) does not imply that the soil moisture variations are locally strong enough to induce large evapotranspiration variations and thus large impacts on air temperature (Fig. 3c). Second, as noted previously, the plotted sensitivity also includes a signal of the size of the precipitation corrections, and so will be enhanced where the differences between the model-generated and observation-corrected precipitation are larger.

Figure 3c is consistent with previous studies identifying hot spots of strong coupling between the land and T^{2m} . In particular Koster et al. (2006) and Miralles et al. (2012) both identify similar regions of strong coupling centered on the central United States/Central America and the Sahel, although they do not agree as well over South Asia. Over South Asia Koster et al. (2006) does not locate a hot spot, while Miralles et al. (2012) identifies India as having the strongest coupling, and Fig. 3c suggests patchy regions of coverage spanning from Southeast Asia through the north of India.

For reference, the corresponding maps for the austral summer (December-February) are shown in Fig. 1 of the supplemental material for R_{anom}^2 (LH, SM) and Fig. 2 in the supplemental material for the sensitivity to the precipitation corrections. In Fig. 1 of the supplemental material, the R^2_{anom} (LH, SM) over austral summer again shows the expected pattern of moisture-limited LH in drier areas of the summer hemisphere (almost everywhere, outside of the tropics). As with the boreal summer, regions of moisture-limitation LH extend into the winter hemisphere. However, the effect of reduced radiation close to the poles is now evident in the switch to energy-limited LH, even in arid regions that are poleward of around 50° (such as central Asia). Figure 2 in the supplemental material shows strong sensitivity of $T_{\text{max}}^{2\text{m}}$ to the precipitation corrections across nearly all of the Southern Hemisphere, including the Amazon and tropical Africa. Since these latter two areas typically have saturated soils, this strong signal is unlikely due to the precipitation-soil moisture pathway and is perhaps due to sensitivity of evaporative cooling from the canopy interception to changes in the precipitation supply to the interception reservoir.

c. Biases over boreal summer

In section 3a, the biases in the reanalyses' global land energy budgets were provided as annual means. The seasonal cycle of the monthly mean global land biases (not shown) reveal that the largest global land biases for all budget terms occur in the boreal summer (JJA).



FIG. 4. The mean JJA turbulent fluxes, with (a) GLEAM LH, (b) MTE LH, and (c) MTE SH reference data, and the difference from the reference data for (d)–(f) MERRA-2, (g)–(i) MERRA-Land, and (j)–(l) MERRA. The statistics span 1980–2015 for GLEAM and 1982–2011 for MTE.

Below, maps of these JJA biases are presented and discussed, together with the corresponding biases in 2-m air temperatures.

1) ENERGY BUDGET TERMS

Figure 4 shows maps of the reanalyses' JJA biases in LH and SH compared to each of GLEAM and MTE. For LH, the regions of positive and negative biases relative to GLEAM or MTE are similar (cf. Figs. 4a,d,g,j and Figs. 4b,e,h,k). For both, the LH biases depend on the local LH regime, with energy-limited regions [low R_{anom}^2 (LH, SM) in Fig. 2] generally having larger positive LH biases (>20 Wm⁻²; e.g., for MERRA-2 in Figs. 4d,e across the tropics, South Asia, and the northern high latitudes), while moisture-limited regions [high R_{anom}^2 (LH, SM) in Fig. 2] tend to have smaller biases (magnitude < 10 Wm⁻²). Consequently, the spatial correlation between R_{anom}^2 (LH, SM) (as plotted in Fig. 2) and the MERRA-2 LH biases is -0.65 for GLEAM and -0.73 for MTE.

The MERRA LH biases (Figs. 4j,k) show some of the same features as for MERRA-2, again with a tendency

for large positive biases in energy-limited LH regimes. The most prominent difference is the sharp bias gradient in MERRA around 10°S (most notable in South America). As discussed in section 2b, this is associated with the unrealistically large rainfall interception reservoir in MERRA, combined with the MERRA precipitation errors; these problems have been alleviated in MERRA-2 (and MERRA-Land). Additionally, there are some isolated regions of large positive biases in moisture-limited regimes in MERRA that are removed in MERRA-2 (and MERRA-Land), such as in Mexico and southern India.

Overall, in energy-limited regions $[R_{anom}^2(LH, SM) < 0.5$ in Fig. 2] the area-averaged LH bias in MERRA-2 (25.5 W m⁻² compared to GLEAM, and 29.9 W m⁻² compared to MTE) was slightly higher than for MERRA (24.1 W m⁻² compared to GLEAM, 27.6 W m⁻² compared to MTE), both of which are much higher than for MERRA-Land (11.3 W m⁻² compared to GLEAM, and 7.6 W m⁻² compared to MTE). In contrast, in moisture-limited LH regions $[R_{anom}^2(LH, SM) > 0.5$ in Fig. 2], the area-averaged LH bias is highest in MERRA (7.0 W m⁻²)



FIG. 5. Separation of mean JJA turbulent flux into EF and incoming radiation biases, with the (a) MTE EF, (b) MTE LH + SH, and (c) CERES-EBAF reference data and the difference from the reference data for (d)–(f) MERRA-2, (g)–(i) MERRA-Land, and (j)–(l) MERRA. The statistics span 1982–2011 for MTE and 2000–15 for CERES-EBAF.

compared to GLEAM, 5.2 W m^{-2} compared to MTE), and reduced in MERRA-2 (3.8 W m^{-2} compared to GLEAM, 1.5 W m^{-2} compared to MTE), and even further reduced in MERRA-Land (0.3 W m^{-2} compared to GLEAM, -2.9 W m^{-2} compared to MTE).

Figures 4c,f,i,l show the reanalyses' biases in SH compared to MTE. In general, the SH biases for each reanalyses have an inverse relationship with the LH biases in Figs. 4b,c,e,f,h,i,k,l (for MERRA-2, the spatial correlation between the SH biases and the LH biases is -0.68 for GLEAM LH and -0.78 for MTE LH). Consequently, the evaporative fraction [EF = LH/(LH + SH)] biases compared to MTE in Figs. 5a,d,g,j show a spatial pattern very similar to that of the LH biases (for MERRA-2, the spatial correlation between MTE LH and EF biases is 0.83).

The sum of LH and SH approximates the net incoming radiation (after neglecting the ground heat flux and temporal change in T_{skin}). Figures 5b,e,h,k and 5c,f,i,l show, respectively, the biases in the reanalyses LH + SH sum compared to MTE and the biases in their R_{net} compared to CERES-EBAF. There is a weak

agreement between the R_{net} biases suggested by MTE and CERES-EBAF (for MERRA-2, the spatial correlation is 0.46). Comparison to MTE (Figs. 5b,e,h,k) suggests that the reanalyses net surface radiation tends to be overestimated, with the largest biases $(>30 \text{ W m}^{-2})$ occurring over the Amazon, the Horn of Africa, and the Tibetan Plateau. While comparison to CERES-EBAF (Figs. 5c,f,i,l,) also suggests relatively large positive biases over the Tibetan Plateau and the Horn of Africa, these positive biases are smaller in both magnitude and regional extent than was suggested by MTE. Additionally, CERES-EBAF also indicates strong negative biases $(<-30 \,\mathrm{W \,m^{-2}})$ over the Sahel and the southeastern United States, particularly in MERRA-Land (Fig. 5i) and MERRA (Fig. 5l). Finally, intercomparing the R_{net} biases for each reanalysis shows qualitatively that the broad patterns are similar in MERRA-2 and MERRA (also MERRA-Land), although MERRA has a tendency toward larger (positive and negative) biases.

There is no obvious correspondence between the regional biases in the LH (compared to GLEAM or



FIG. 6. The mean JJA radiation terms, from (a)–(c) the CERES-EBAF reference data, and (d)–(f) difference from the reference data for MERRA-2, for (left) SW_{net} , (center) LW_d , and (right) LW_u . The statistics span 2000–15.

MTE) and the regional biases in R_{net} (compared to either MTE LH + SH or CERES-EBAF). For example, the spatial correlations are less than 0.1 between the MERRA-2 LH bias (implied by comparison to GLEAM or MTE) and the MERRA-2 LH + SH bias (implied by MTE). Likewise, the spatial correlations are again less than 0.1 between the MERRA-2 LH bias (implied by GLEAM of MTE) and the MERRA-2 R_{net} bias (implied by CERES-EBAF). This suggests then that the pattern of regional biases in the reanalyses LH for JJA (compared to either GLEAM or MTE) are associated with differences in the partitioning of incoming radiation into LH and SH, rather than with differences in the surface radiation (compared to MTE or CERES-EBAF) itself.

While radiation biases do not appear to be the main predictor of LH biases, biased radiation will result in biased LH and/or SH. Hence, we have partitioned the JJA R_{net} bias between MERRA-2 and CERES-EBAF into the individual contributions from each radiation term. Figure 6 shows the JJA biases between MERRA-2 and CERES-EBAF for SW_{net} , LW_d , and LW_u . In terms of the direction of the biases, the broad patterns of regional biases in the radiation terms are unchanged from MERRA (not shown). The direction of the regional R_{net} biases for MERRA-2 in Fig. 5f largely mirror the regional SW_{net} biases in Fig. 6d (spatial correlation: 0.75), the main exception being over the southeastern United States. The LW biases are somewhat balanced, in that both are negative across most of the domain, with the LW_d bias in Fig. 6e typically being slightly more negative than the LW_u bias in Fig. 6f. Both have relatively large negative biases (magnitude $> 30 \,\mathrm{W \,m^{-2}}$) in Northern Hemisphere desert regions and smaller (magnitude: $10-20 \text{ W m}^{-2}$) negative biases elsewhere. The spatial distribution of the SW_{net} biases mirrors that of the downwelling shortwave SW_d (not shown), indicating that the SW_{net} biases are primarily driven by SW_d differences rather than differences in the surface albedo used in CERES-EBAF and GEOS-5. The above patterns of overestimated SW_{net} (or SW_d) and underestimated LW_d across much of the globe are consistent with a known tendency for the GEOS-5 systems to underestimate midlatitude continental cloud cover (Molod et al. 2012; Wang and Dickinson 2013; Gelaro et al. 2015).

The LW_{*u*} is calculated from T_{skin} , and the negative biases in MERRA-2 (and also MERRA and MERRA-Land) indicate a cool bias in the model T_{skin} . At 285 K, an LW_u bias of 10 W m⁻² is roughly equivalent to a T_{skin} bias of 2 K. Recall that the CERES-EBAF LW_u is not independent of the MERRA suite of reanalyses, due to its use of GEOS-5 T_{skin} . However, the input GEOS-5 $T_{\rm skin}$ is adjusted within the CERES-EBAF algorithm to constrain the TOA irradiance, so comparison of GEOS-5 and CERES LW_u indicates the adjustment required to the GEOS-5 T_{skin} to balance the TOA fluxes. Previous work has also suggested that the GEOS-5 T_{skin} is underestimated, particularly in dry regions. For example, in agreement with our Fig. 6f, Draper et al. (2015) found large cool biases in the GEOS-5 T_{skin} over desert regions in summer (their Fig. 5), compared to remotely sensed observations. As argued in Draper et al. (2015), this GEOS-5 T_{skin} cool bias is, at least in part, caused by the model's T_{skin} definition differing from that of a true skin layer from which LW_u is emitted (or as is observed in the thermal infrared).

In summary, the pattern of regional LH biases in the reanalyses suggested by GLEAM and MTE are very similar. This result adds confidence to the use of GLEAM and MTE for estimating regional biases in the



FIG. 7. The mean JJA T^{2m} from (a)–(c) CRU reference data and the difference from the reference data for (d)–(f) MERRA-2 and (g)–(i) MERRA, for (left) T^{2m}_{min} , (center) T^{2m}_{max} , and (right) DTR. The statistics span 1980–2015, and white plotted over land indicates insufficient CRU data.

reanalyses. As with the annual global land averages in Fig. 1, the maps presented here suggest that MERRA-2 and MERRA (but not MERRA-Land) have a general tendency to overestimate LH. If the GLEAM, MTE, and CERES-EBAF regional means are assumed to be more accurate than the reanalyses, the above comparisons suggest that in energy-limited regions, MERRA-2 (and MERRA) overestimates LH as a result of an overestimated evaporative fraction (i.e., too much incoming radiation is converted to LH rather than SH). There is little change in the global average biases from MERRA to MERRA-2. However, there are some isolated regions in Mexico and South Asia that are typified by moisture-limited LH, where MERRA has positive LH biases associated with overestimated EF, while MERRA-2 and MERRA-Land have much smaller biases. The precipitation corrections in MERRA-2 (and MERRA-Land) removed a relatively large amount of precipitation across these locations (Reichle et al. 2017b, their Fig. 3b), strongly suggesting that the use of precipitation observations in these products reduced the LH biases.

2) AIR TEMPERATURE

The biases in the MERRA-2 and MERRA JJA monthly mean daily minimum, daily maximum, and diurnal range in T^{2m} , relative to the CRU dataset, are shown in Fig. 7 (T^{2m} is not calculated by the land-only

MERRA-Land system). For the daily minimum T^{2m} (T_{\min}^{2m}) in Figs. 7a,d,g, both reanalyses tend toward positive (warm) biases, particularly MERRA. For $T_{\text{max}}^{2\text{m}}$ in Figs. 7b,e,h, MERRA-2 tends toward cool biases, with patches of warm biases across central Asia and the Arabian Peninsula (investigation of the large positive bias in the Arabian Peninsula suggests it is associated with an error in the CRU reference data, rather than the reanalyses). For MERRA, these patches of positive bias are expanded to cover most of the desert region in the Northern Hemisphere and also much of the Southern Hemisphere. For the DTR in Figs. 7c,f,i, the MERRA-2 biases inherit the broad spatial pattern of the $T_{\rm max}^{\rm 2m}$ biases, while for MERRA some of the large positive $T_{\text{max}}^{2\text{m}}$ biases are offset in the DTR by collocated positive T_{\min}^{2m} .

The LH and SH biases in Fig. 4 and the DTR biases in Fig. 7 show some of the expected regional similarities. In particular, in the high latitudes and the Amazon MERRA-2 has relatively large positive LH biases (and negative SH biases) and relatively large negative DTR biases. MERRA also has overestimated LH and underestimated DTR in the same regions, as well as in Southeast Asia and Central America. This is consistent with an underestimated DTR caused by underestimated SH (and overestimated LH), particularly given that the R_{net} bias is generally neutral in these regions in Fig. 5. It should however be noted that the high latitudes and the



FIG. 8. The R_{anom} between (left) GLEAM LH, (center) MTE LH, and (right) MTE SH for (a)–(c) MERRA-2, (d)–(f) MERRA-Land, (g)–(i) MERRA, and (j)–(l) ERA-Interim for JJA. Statistics span 1980–2015 for GLEAM and 1982–2011 for MTE.

Amazon regions are both data scarce, and both the reanalyses and reference datasets are less well constrained. In other regions there is less correspondence. For example, the western United States also has underestimated DTR for MERRA and MERRA-2, while neither GLEAM nor MTE suggests overestimated LH. Overall, the spatial correlations between the LH biases and DTR biases are rather low (for MERRA-2, they are -0.38 for GLEAM and -0.47 for MTE).

Recall that in section 3c(1) above, the CERES-EBAF comparison suggested that the MERRA-2 (and MERRA) T_{skin} is generally biased cool, with larger cool biases in desert areas. However, a comparison of the LW_u biases in Fig. 6f to the T_{\min}^{2m} and T_{\max}^{2m} biases in Figs. 7d,e shows little correspondence between them, and in particular the regions of relatively large cool $T_{\rm skin}$ biases (underestimated LW_u) in the Northern Hemisphere deserts do not have cool biases in either $T_{\text{max}}^{2\text{m}}$ and T_{\min}^{2m} . This apparent contradiction between the temperature biases suggested by comparison to the CERES-EBAF LW_u ($\sim T_{skin}$) and the CRU T^{2m} does not necessarily imply that one of these datasets is incorrect, given the likelihood mentioned above that the model T_{skin} biases are at least partly associated with the model definition of T_{skin} .

d. Turbulent heat flux anomaly correlations over boreal summer

Here the monthly mean turbulent heat flux time series are evaluated over boreal summer based on their temporal correlations R_{anom} with the reference datasets. Figure 8 shows maps of the JJA Ranom for each of the NASA reanalyses (MERRA-2, MERRA-Land, and MERRA) and ERA-Interim, with the R_{anom} calculated separately versus each of the GLEAM and MTE turbulent heat fluxes. For LH, the regional patterns in R_{anom} versus either GLEAM (Figs. 8a,d,g,j) or MTE (Figs. 8b,e,h,k) show some similar features (for MERRA-2, spatial correlation between Figs. 8a and 8b: 0.69). Comparison to Fig. 2 again suggests some dependence on the LH regime. In the Northern Hemisphere, the LH R_{anom} is generally highest (~0.6) in regions where LH is moisture limited and generally much lower (<0.2) where LH is energy limited. The two exceptions are the high latitudes, which have high LH R_{anom} and energy-limited LH, and the Sahara, which has low LH R_{anom} and is moisture limited (although LH variability in the Sahara is very low, making the signal susceptible to noise).

The R_{anom} patterns for ERA-Interim in Figs. 8j–1 provide some additional context for evaluating the



FIG. 9. Bar plot of the mean annual (a) LH and (b) SH across the 20 FLUXNET site locations from M-2, M-L, M, FLUXNET (FIN), MTE, and GLM (LH only), calculated using each dataset at its native resolution (and screened temporally for FLUXNET availability). For the global datasets, circles are plotted for the global land annual mean (taken from Fig. 1).

NASA reanalyses. The LH R_{anom} values are generally higher for ERA-Interim than for the NASA reanalyses. As for MERRA-2, the ERA-Interim R_{anom} versus MTE is relatively low in many energy-limited LH regimes (including the eastern United States, tropics, and South Asia), while R_{anom} for ERA-Interim versus GLEAM is more spatially consistent, in contrast to R_{anom} for MERRA-2. The relatively high R_{anom} between GLEAM and ERA-Interim LH in energy-limited LH regimes may well be due to GLEAM having used ERA-Interim radiation and temperature, since it is in these regions that these fields will have the strongest influence on the LH. On the other hand, the lower R_{anom} between the NASA reanalyses and the LH reference datasets (and also between ERA-Interim and MTE) could be attributed to errors in both the reference datasets and the reanalyses under energy-limited conditions. For MTE, this result was expected because MTE is thought to be more reliable in estimating temporal variability in moisture-limited areas, since its temporal variability is largely driven by fPAR (Jung et al. 2010).

Moving on to SH, Figs. 8c,f,i,l show R_{anom} versus MTE for each reanalysis. The regional patterns are similar to those for LH, with higher R_{anom} (>0.5) in moisturelimited LH regions and lower (<0.2) values elsewhere. ERA-Interim R_{anom} versus MTE is generally higher than the NASA reanalyses, with values greater than 0.5 across most of the globe (and particularly in the Northern Hemisphere). Despite the improved LH from MERRA-Land, the SH R_{anom} versus MTE is lower than for MERRA (or MERRA-2).

Globally averaged, the rank order of the mean LH R_{anom} , while rather low, is the same versus either

GLEAM or MTE and follows the expected progression of improvement from MERRA, to MERRA-Land, and then to MERRA-2. GLEAM suggests a larger improvement, from a globally averaged R_{anom} of 0.39 for MERRA to 0.48 for MERRA-2, with MERRA-Land falling in between (0.45). MTE suggests an improvement from 0.29 for MERRA to 0.34 for MERRA-2, with MERRA-Land again falling in between (0.32). For SH, the globally averaged R_{anom} versus MTE is similar for MERRA (0.36) and MERRA-2 (0.37), but is much lower for MERRA-Land (0.28). For ERA-Interim, the global mean R_{anom} for LH is ~0.1 higher than for MERRA-2 (0.60 vs GLEAM and 0.44 vs MTE) and ~ 0.2 higher for SH (0.46 vs MTE). The better agreement between ERA-Interim and the reference datasets could be a consequence of the land surface updates applied in ERA-Interim, which indirectly targets the turbulent heat fluxes. [Although recall that the relatively strong agreement between the GLEAM and ERA-Interim LH will partly reflect their dependence; see section 2c(2).]

e. Comparison to FLUXNET tower data

Since the reference datasets used above do not represent direct observations, we now compare the globally averaged LH and SH statistics from section 3a (for the annual mean turbulent heat fluxes over land) and section 3d (for the mean JJA R_{anom}) to statistics calculated against FLUXNET2015 (eddy covariance) tower observations. Figure 9 shows the annual mean of the turbulent fluxes for the FLUXNET (eddy covariance) measurements themselves and for each reanalysis and reference dataset averaged across the 20 FLUXNET locations. For the global data sets, the global land annual



FIG. 10. Bar plot of R_{anom} over JJA averaged across the 20 FLUXNET site locations for (a) LH, (b) SH, and (c) LH + SH between each pair of the reanalyses [M-2, M-L, M, and ERA-Interim (E-I)] and the reference data (FIN, MTE, and GLM). The R_{anom} vs the FLUXNET reference data use the reanalysis output at their reported spatial resolution (and screened temporally for FLUXNET availability), while the R_{anom} vs GLEAM and MTE use reanalyses and reference data regridded to 1°. For GLEAM and MTE, circles are plotted for the global mean JJA R_{anom} (averaged over the panels of Fig. 8).

means (from Fig. 1) are also included for reference. For LH, comparison to the FLUXNET observations agrees with the results from the global land comparison in section 3a, again suggesting that the MERRA-2 LH is biased high, although the FLUXNET observations suggest a larger bias (of 12 W m^{-2} , or 30%) than was suggested by the global comparison (estimated as 6 W m^{-2} in section 3a). Averaged across the 20 FLUXNET sites, the MTE LH is very close to the FLUXNET data (within 0.5 W m^{-2}), while GLEAM is slightly higher. For the interested reader, Fig. 3 in the supplemental material shows scatterplots comparing the MERRA-2 and reference dataset LH annual means at the 20 individual sites.

For SH, the FLUXNET observations agree less well with the global land comparison. First, the annual mean of the FLUXNET data is about 10 Wm^{-2} below the global mean estimates from the other reference datasets. For each of the global reference datasets and reanalyses, the annual average over the 20 FLUXNET sites is also $15-20 \text{ Wm}^{-2}$ lower than the global average, suggesting that the relatively low FLUXNET annual mean is associated with the spatial sampling of the FLUXNET sites. Second, averaged across the FLUXNET sites, the FLUXNET mean SH is close to that of MERRA-Land, and above that of MERRA-2 (by 6 W m⁻²; 18%). In contrast, for the global averages in section 3a the reference datasets were all close to MERRA-2 (and MERRA), with MERRA-Land standing out as being biased high.

Figure 10 shows the JJA R_{anom} averaged over the 20 FLUXNET sites for each reanalysis versus each of FLUXNET, GLEAM, and MTE, with the global average JJA R_{anom} from section 3d also included for GLEAM and MTE. The R_{anom} values for the FLUXNET data are quite low, which is somewhat expected due to the mismatch in spatial representation between the tower-based observations and the reanalysis. Nonetheless, the

FLUXNET R_{anom} (as well as the GLEAM and MTE R_{anom} at the same locations) indicates similar relative reanalysis performance as the global mean R_{anom} . In particular, for LH MERRA-2 and MERRA-Land outperform MERRA, as also indicated by the global means. However, the one discrepancy is that the R_{anom} versus the FLUXNET data is similar for ERA-Interim and MERRA-2, while the global comparisons (and also the GLEAM and MTE data averaged across the FLUXNET sites) all suggest that ERA-Interim outperforms MERRA-2 (giving mean R_{anom} around 0.1 higher). For SH, the rank order between the average JJA R_{anom} is the same from the FLUXNET data than from the global reference datasets, with the MERRA-Land R_{anom} again being lower than that for MERRA (and MERRA-2) and the ERA-Interim average R_{anom} being higher than that for MERRA-2.

It is notable that over the FLUXNET tower sites, both GLEAM and MTE have higher average R_{anom} with the reanalyses than the FLUXNET observations do. In particular, MTE was trained on an earlier generation of the FLUXNET data, and the higher mean R_{anom} versus MTE than versus FLUXNET suggests that the MTE algorithm has added coarse-scale information (similar quality control was applied here as was applied to the tower observations used in MTE). For the interested reader, Fig. 4 in the supplemental material shows scatterplots of the MERRA-2 LH R_{anom} versus each reference dataset at the 20 individual sites.

Note that for FLUXNET, R_{anom} for LH + SH, plotted in Fig. 10c, is consistently about 0.1 higher than R_{anom} for either LH or SH separately. Decker et al. (2012) obtained a similar result for the correlation between reanalyses and tower observations. This indicates that the eddy covariance measurements and the reanalyses have a stronger agreement in the implied incoming radiation than in the partitioning of that radiation into LH and SH (this result is unchanged if R_{anom} values are calculated from the FLUXNET data that have not been energy balance corrected). This could be a signal of errors in the partitioning within the reanalyses, or perhaps just as likely, this difference is associated with the spatial representation of the tower observations, since the incoming radiation is more spatially homogeneous than either LH or SH on its own.

f. Precipitation corrections and air temperature performance

Finally, we seek to establish whether the precipitation corrections in MERRA-2 influenced the local $T_{\text{max}}^{2\text{m}}$. We do this by comparing the performance of the MERRA-2 and MERRA $T_{\text{max}}^{2\text{m}}$ to Fig. 3c, which shows the MERRA-2 sensitivity to observed precipitation. Figure 11 shows the T_{\max}^{2m} R_{anom} versus CRU observations over JJA for MERRA-2 and MERRA. In general, the MERRA-2 R_{anom} is high (>0.7) across most of the domain, particularly in the high latitudes, with much lower (<0.4) values across much of the tropics and parts of South America, Africa, and South Asia. Note that the latter regions all have relatively sparsely distributed CRU station data, which is likely contributing to the lower agreement with the reanalyses. Compared to MERRA, the greatest improvements in the MERRA-2 $T_{\text{max}}^{2m} R_{\text{anom}}$ occurred in the eastern United States, much of tropical South America and Africa, the Sahel, and parts of South Asia and China. There are also several regions where the T_{\max}^{2m} R_{anom} is reduced, including northern South America and much of Southeast Asia. Overall, the global averaged $T_{\max}^{2m} R_{\text{anom}}$ versus CRU was increased from 0.69 for MERRA to 0.75 for MERRA-2.

Comparing Fig. 11c to Fig. 3c, the regions with the strongest sensitivity of $T_{\text{max}}^{2\text{m}}$ to the precipitation corrections generally have relatively large changes in the $T_{\rm max}^{2m}$ R_{anom} (including the Sahel, parts of South Asia, and Central America). Consequently, where the metric in Fig. 3c is above 0.25 (i.e., the observation-corrected precipitation explains at least 25% more of the MERRA-2 $T_{\text{max}}^{2\text{m}}$ variance than the model-generated precipitation does), the area-averaged absolute change in R_{anom} is 0.15, compared to an area-average absolute change of 0.07 elsewhere. This tendency toward a relatively large change in the $T_{\max}^{2m} R_{\text{anom}}$ where T_{\max}^{2m} is sensitive to the precipitation corrections suggests that the observed precipitation in MERRA-2 contributed to the change in $T_{\rm max}^{\rm 2m}$ performance. Additionally, the change in $T_{\max}^{2m} R_{\text{anom}}$ in these regions is generally, although not always, positive (giving an area averaged change in R_{anom} of 0.06 where the metric in Fig. 3c is greater than 0.25). In some of the instances where the $T_{\text{max}}^{2\text{m}} R_{\text{anom}}$ is



FIG. 11. The (a) MERRA-2 R_{anom} vs CRU monthly mean T_{max}^{2m} and (b) the improvement in the $T_{\text{max}}^{2m} R_{\text{anom}}$ from MERRA to MERRA-2, both over JJA. Statistics span 1980–2015, and white plotted over land indicates insufficient CRU data.

degraded, this can be traced back to errors in the precipitation observation datasets input into MERRA-2. For example, over Myanmar, the $T_{max}^{2m} R_{anom}$ is decreased by more than 0.15, likely due to persistent local errors in the precipitation observations input into MERRA-2 (Reichle et al. 2017b). Finally, there are also regions with large changes in the $T_{max}^{2m} R_{anom}$ outside of the regions of T_{max}^{2m} sensitivity to precipitation (the eastern United States, tropical Africa and South America, and central China). The $T_{max}^{2m} R_{anom}$ is increased in MERRA-2 across most of these regions, likely due to other advances (beyond the use of observed precipitation) in the MERRA-2 modeling and assimilation system.

4. Summary and conclusions

The land surface energy budgets of three reanalyses from NASA (MERRA, MERRA-Land, and MERRA-2) are compared here to the best available estimates from the literature and to (largely) independent global reference datasets. In terms of the global land annual averages, the results suggest that the MERRA-2 LH and SH are biased high by 5 and 6 W m⁻², respectively, while SW_u has a large positive bias of 14 W m⁻², SW_d is biased high by 3 W m⁻², and the upwelling and downwelling LW components are biased low, by 11 and 13 W m⁻², respectively. Compared to MERRA, this is a slight $(\sim 2 \text{ W m}^{-2})$ reduction in the LH and SW_{net} biases, while the difference is even smaller for the LW terms $(\sim 1 \text{ W m}^{-2})$. The radiation biases are associated with known issues in the GEOS-5 models used in the reanalyses, specifically a tendency to underestimate midlatitude continental clouds (Wang and Dickinson 2013) and a cool bias in the model T_{skin} (Draper et al. 2015).

Compared to reference flux estimates from GLEAM and MTE over the boreal summer (when both the fluxes themselves and their biases are greatest), the largest MERRA-2 LH biases (> 20 W m^{-2} , vs either GLEAM or MTE) occur in regions where LH is energy limited, such as in the high latitudes, the tropics, parts of South Asia, and the eastern United States. The MERRA-2 LH biases are typically smaller in regions where LH is moisture limited, which include the drier regions of the mid and low latitudes. In some of these moisture-limited regions (parts of South Asia and Mexico) the high bias in the MERRA LH was largely removed in MERRA-2 (and MERRA-Land), likely because the observed precipitation used in the latter was lower than that produced by the MERRA (or MERRA-2) modeling systems. Finally, comparison to the evaporative fraction from MTE and to R_{net} from CERES-EBAF or as inferred from MTE LH + SH indicates that the regional biases in the reanalyses LH are generally associated with differences in the partitioning of R_{net} into LH and SH rather than with differences in the radiation input.

The temporal agreement between the reanalyses and the reference datasets over boreal summer was measured using the monthly anomaly correlation R_{anom} over JJA. For LH, the R_{anom} between the reanalyses and the reference datasets (GLEAM and MTE) again showed some dependency on the LH regime, with a tendency toward better agreement where LH is moisture limited than where it is energy limited. The lower agreement in energy-limited regions does not necessarily imply poorer performance in the reanalyses, as it may be due to errors in the reference datasets. The globally averaged R_{anom} values show the expected improvement in skill with each new NASA reanalyses. For example, MERRA-2 has slightly better globally averaged LH R_{anom} (0.48 vs GLEAM) than MERRA-Land (0.45), which is substantially better than MERRA (0.39). The R_{anom} value was also calculated for the monthly mean daily $T_{\text{max}}^{2\text{m}}$ versus CRU reference data over JJA. Averaged over global land, the JJA $T_{\text{max}}^{2\text{m}} R_{\text{anom}}$ versus CRU increased from 0.69 for MERRA to 0.75 for MERRA-2. The results presented above for the regional biases and R_{anom} were based on the boreal summer; however, the same analysis has been performed over the austral summer (not shown), yielding qualitatively similar results.

The use of observed precipitation in MERRA-2 was motivated by the hope that the subsequent improvements in simulated soil moisture would lead to the improved partitioning of incoming radiation between latent and sensible heating, ultimately leading to improvements in the diurnal evolution of the boundary layer. It is difficult, however, to unequivocally attribute the improvements in MERRA-2 to the use of observed precipitation because MERRA-2 includes many other modeling and assimilation advances relative to MERRA. Nonetheless, many of the improvements in the MERRA-2 LH and T^{2m} are consistent with the changes expected from the use of observed precipitation. MERRA-2 and MERRA-Land have smaller positive LH biases and higher LH R_{anom} than MERRA in regions where LH is moisture limited and thus sensitive to precipitation (South Asia and the western United States). This is most easily explained by the forcing of the land surface with observed precipitation in MERRA-2. Additionally, regions where the MERRA-2 JJA $T_{\text{max}}^{2\text{m}}$ was most sensitive to the precipitation corrections (the Sahel, central United States, and parts of South Asia) generally experience larger changes in the $T_{\text{max}}^{2\text{m}} R_{\text{anom}}$ from MERRA to MERRA-2. However, the changes in R_{anom} in these areas are not uniformly positive, and in some cases degraded $T_{\text{max}}^{2\text{m}} R_{\text{anom}}$ can be traced back to problems in the input precipitation datasets (e.g., over Myanmar). In the future, the use of precipitation corrections could be enhanced by also implementing a land data assimilation scheme to update the model soil moisture according to observations (e.g., Draper et al. 2011; Dharssi et al. 2011; De Lannoy and Reichle 2016). By making use of remotely sensed observations, the land data assimilation would be particularly valuable in regions where the rain gauge network is sparse or has known problems (e.g., in Africa and parts of Southeast Asia).

However, some of the largest biases and lowest R_{anom} for the MERRA-2 turbulent fluxes occur where the LH is energy limited and thus less sensitive to improvements in the precipitation and soil moisture. Hence, future efforts to improve the MERRA-2 land surface turbulent fluxes would best be focused on other facets of the modeling and assimilation. Specifically, future GEOS-5 development should focus on the overestimated evaporative fraction where LH is energy limited. Additionally, even though the MERRA-2 R_{net} is relatively unbiased (compared to CERES-EBAF), there are large compensating biases in the individual SW and LW radiation fluxes that are 2-3 times the magnitude of the LH biases in terms of the global land annual averages. Reducing the cloud bias in the atmospheric model will help these biases, as will redefining the model T_{skin} to generate a LW_u more consistent with observations.

689

Finally, the SH results for MERRA-Land are troubling. While MERRA-Land did have the desired reduction in the LH biases compared to MERRA (to $1 \,\mathrm{W}\,\mathrm{m}^{-2}$ in the global land annual average), it also had a compensating, and much larger, increase in the SH bias (up to 15 W m^{-2} in the global land average). Additionally, the JJA R_{anom} compared to MTE were reduced from MERRA to MERRA-Land (from a global average of 0.36–0.28), despite the LH R_{anom} being increased. The cause of the degraded SH in MERRA-Land is presently unknown, but given the otherwise similar MERRA and MERRA-Land land surface models and meteorological forcing, an obvious possibility is that the use of observed precipitation in an offline (land only) replay of an analysis, such as MERRA-Land, can lead to inconsistencies in the forcing (e.g., warm and dry air, stemming from dry conditions in MERRA, overlying cold ground induced by high antecedent rainfall from the observations). Such inconsistencies would not appear in MERRA or (as much) in MERRA-2, given the coupling in the reanalyses of the land surface state with the overlying atmosphere.

While this work focused on evaluating surface energy fluxes in MERRA-2, the findings have relevance to anyone interested in designing a methodology to evaluate global estimates of turbulent heat fluxes. The gridded LH reference datasets (GLEAM and MTE) had better agreement with the reanalyses' time series (as measured by R_{anom}) and were more useful for evaluating the reanalysis output than were the tower observations. In particular they offer (near) global coverage across several decades, at similarly coarse resolution to the reanalyses. In the absence of a recognized truth for LH (or other similar terms), the recommended evaluation strategy is to compare the product under evaluation to multiple datasets. However, given the uncertainty in the available reference datasets, extra care is necessary to understand the methodology, input data, assumptions, and potential dependencies and weaknesses of each reference dataset. This process relies on expert judgement and inevitably introduces some subjectivity into the interpretation of the results. Further development of global gridded LH datasets (including the quality and quantity of "ground-truth" observations), to increase their confidence would obviously be of great benefit to this process.

The GLEAM and MTE reference datasets used here are independent of each other and are based on very different methodologies, thus providing complementary information for use in an evaluation. However, given the use of the common precipitation input data in GLEAM as in MERRA-2, and the fact that MTE data are not optimized to estimate interannual variability, LH estimates from a third reference dataset would be useful. Emerging global and multidecadal land surface flux datasets based on an energy balance approach (Anderson et al. 2011), or alternative observational frameworks (Alemohammad et al. 2017) would provide useful complements to GLEAM and MTE for a more comprehensive analysis.

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. The authors acknowledge the teams that produce and publish the GLEAM, MTE, CERES-EBAF, ERA-Interim, MERRA, MERRA-Land, and MERRA-2 products. Additionally, we are grateful to Diego Miralles (VU University Amsterdam/Ghent University), Martin Jung (Max Planck Institute for Biogeochemistry), and Seiji Kato (NASA Langley Research Center) for their thoughtful feedback on this work and detailed advice on the use of GLEAM, MTE, and CERES-EBAF, respectively. The FLUXNET eddy covariance data processing and harmonization was carried out by the European Fluxes Database Cluster, AmeriFlux Management Project, and Fluxdata project of FLUXNET, with the support of CDIAC and ICOS Ecosystem Thematic Center, and the OzFlux, ChinaFlux, and AsiaFlux offices.

REFERENCES

- Alemohammad, S. H., and Coauthors, 2017: Water, Energy, and Carbon with Artificial Neural Networks (WECANN): A statistically based estimate of global surface turbulent fluxes and gross primary productivity using solar-induced fluorescence. *Biogeosciences*, 15, 4101–4124, https://doi.org/10.5194/ bg-14-4101-2017.
- Anderson, M., and Coauthors, 2011: Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery. *Hydrol. Earth Syst. Sci.*, 15, 223–239, https://doi.org/10.5194/hess-15-223-2011.
- Baldocchi, D., and Coauthors, 2001: FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bull. Amer. Meteor. Soc.*, 82, 2415–2434, https://doi.org/10.1175/ 1520-0477(2001)082<2415:FANTTS>2.3.CO;2.
- Beck, H., A. van Dijk, V. Levizzani, J. Schellekens, D. Miralles, B. Martens, and A. de Roo, 2017: MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrol. Earth Syst. Sci.*, 21, 589– 615, https://doi.org/10.5194/hess-21-589-2017.
- Betts, A., A. Tawfik, and R. Desjardins, 2017: Revisiting hydrometeorology using cloud and climate observations. *J. Hydrometeor.*, 18, 939–955, https://doi.org/10.1175/JHM-D-16-0203.1.
- Chen, M., W. Shi, P. Xie, V. B. S. Silva, V. E. Kousky, R. W. Higgins, and J. E. Janowiak, 2008: Assessing objective techniques for gauge-based analyses of global daily precipitation. J. Geophys. Res., 113, D04110, https://doi.org/ 10.1029/2007JD009132.

VOLUME 31

- Decker, M., M. Brunke, Z. Wang, K. Sakaguchi, X. Zeng, and M. Bosilovich, 2012: Evaluation of the reanalysis products from GSFC, NCEP, and ECMWF using flux tower observations. J. Climate, 25, 1916–1944, https://doi.org/10.1175/ JCLI-D-11-00004.1.
- Dee, D., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, https://doi.org/ 10.1002/qj.828.
- De Lannoy, G., and R. Reichle, 2016: Assimilation of SMOS brightness temperatures or soil moisture retrievals into a land surface model. *Hydrol. Earth Syst. Sci.*, 20, 4895–4911, https:// doi.org/10.5194/hess-20-4895-2016.
- de Rosnay, P., G. Balsamo, C. Albergel, J. Muñoz-Sabater, and L. Isaksen, 2014: Initialisation of land surface variables for numerical weather prediction. *Surv. Geophys.*, 35, 607–621, https://doi.org/10.1007/s10712-012-9207-x.
- Dharssi, I., K. Bovis, B. Macpherson, and C. Jones, 2011: Operational assimilation of ASCAT surface soil wetness at the Met Office. *Hydrol. Earth Syst. Sci.*, **15**, 2729–2746, https://doi.org/ 10.5194/hess-15-2729-2011.
- Draper, C., J.-F. Mahfouf, and J. Walker, 2011: Root-zone soil moisture from the assimilation of screen-level variables and remotely sensed soil moisture. J. Geophys. Res., 116, D02127, https://doi.org/10.1029/2010JD013829.
- —, R. Reichle, G. De Lannoy, and B. Scarino, 2015: A dynamic approach to addressing observation-minus-forecast bias in a land surface skin temperature data assimilation system. *J. Hydrometeor.*, 16, 449–464, https://doi.org/10.1175/JHM-D-14-0087.1.
- FLUXNET, 2015: FLUXNET2015 dataset. Fluxdata, accessed 9 August 2016, http://fluxnet.fluxdata.org/data/fluxnet2015-dataset/.
- Gelaro, R., and Coauthors, 2015: Evaluation of the 7-km GEOS-5 nature run. NASA Tech. Memo. NASA/TM-2014-104606, Vol. 36, 285 pp.
- —, and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). J. Climate, **30**, 5419–5454, https://doi.org/ 10.1175/JCLI-D-16-0758.1.
- Global Modeling and Assimilation Office, 2008a: tavg1_2d_slv_ Nx: MERRA 2D IAU diagnostic, single level meteorology, time average 1-hourly V5.2.0. GES DISC, accessed 1 October 2016, https://doi.org/10.5067/B6DQZQLSFDLH.
- —, 2008b: tavgM_2d_lnd_Nx: MERRA 2D IAU diagnostic, land only states and diagnostics, monthly mean V5.2.0. GES DISC, accessed 1 October 2016, https://doi.org/10.5067/ XOHTIIK0W9RK.
- —, 2008c: tavgM_2d_mld_Nx: MERRA simulated 2D incremental analysis update (IAU) MERRA-Land reanalysis, GEOSIdas-MERRALand, time average monthly mean V5.2.0. GES DISC, accessed 1 October 2016, https://doi.org/ 10.5067/K9PCGOMQ1XP1.
- —, 2015a: MERRA-2 tavg1_2d_lfo_Nx: 2D, 1-hourly, timeaveraged, single-level, assimilation, land surface forcings V5.12.4. GES DISC, accessed 1 October 2016, https://doi. org/10.5067/L0T5GEG1NYFA.
- —, 2015b: MERRA-2 tavgM_2d_Ind_Nx: 2D, monthly mean, time-averaged, single-level, assimilation, land surface diagnostics V5.12.4. GES DISC, accessed 1 October 2016, https://doi.org/ 10.5067/8S35XF81C28F.
- —, 2015c: MERRA-2 tavgM_2d_slv_Nx: 2D, monthly mean, time-averaged, single-level, assimilation, single-level diagnostics V5.12.4. GES DISC, accessed 1 October 2016, https://doi.org/10.5067/AP1B0BA5PD2K.

- Harris, I., P. Jones, T. Osborn, and D. Lister, 2014a: Updated highresolution grids of monthly climatic observations—the CRU TS3.10 dataset. *Int. J. Climatol.*, **34**, 623–642, https://doi.org/ 10.1002/joc.3711.
- —, and Coauthors, 2014b: CRU TS3.22: Climatic Research Unit (CRU) time-series (TS) version 3.22 of high resolution gridded data of month-by-month variation in climate (Jan. 1901-Dec. 2013). NCAS British Atmospheric Data Centre, accessed 17 May 2017, https://doi.org/10.5285/18BE23F8-D252-482D-8AF9-5D6A2D40990C.
- Huffman, G., R. Adler, D. Bolvin, and G. Gu, 2009: Improving the global precipitation record: GPCP version 2.1. *Geophys. Res. Lett.*, 36, L17808, https://doi.org/10.1029/2009GL040000.
- Jiménez, C., and Coauthors, 2011: Global intercomparison of 12 land surface heat flux estimates. J. Geophys. Res., 116, D02102, https://doi.org/10.1029/2010JD014545.
- Jung, M., M. Reichstein, and A. Bondeau, 2009: Towards global empirical upscaling of FLUXNET eddy covariance observations: Validation of a model tree ensemble approach using a biosphere model. *Biogeosciences*, 6, 2001–2013, https:// doi.org/10.5194/bg-6-2001-2009.
- —, and Coauthors, 2010: Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature*, **467**, 951–954, https://doi.org/10.1038/nature09396.
- —, and Coauthors, 2011: Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. J. Geophys. Res., **116**, G00J07, https://doi.org/10.1029/ 2010jg001566.
- Kato, S., N. Loeb, F. Rose, D. Doelling, D. Rutan, T. Caldwell, L. Yu, and R. Weller, 2013: Surface irradiances consistent with CERES-derived top-of-atmosphere shortwave and longwave irradiances. J. Climate, 26, 2719–2740, https://doi.org/10.1175/ JCLI-D-12-00436.1.
- , and Coauthors, 2017: CERES_EBAF-Surface_Ed4.0:
 Data quality summary (May 26, 2017). NASA CERES Rep.,
 29 pp., https://ceres.larc.nasa.gov/documents/DQ_summaries/
 CERES_EBAF-Surface_Ed4.0_DQS.pdf.
- Koster, R., and Coauthors, 2006: GLACE: The Global Land– Atmosphere Coupling Experiment. Part I: Overview. J. Hydrometeor., 7, 590–610, https://doi.org/10.1175/JHM510.1.
- —, G. Salvucci, A. Rigden, M. Jung, G. Collatz, and S. Schubert, 2015: The pattern across the continental United States of evapotranspiration variability associated with water availability. *Front. Earth Sci.*, **3**, 35, https://doi.org/10.3389/feart.2015.00035.
- L'Ecuyer, T., and Coauthors, 2015: The observed state of the energy budget in the early twenty-first century. *J. Climate*, **28**, 8319–8346, https://doi.org/10.1175/JCLI-D-14-00556.1.
- Martens, B., and Coauthors, 2017: GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geosci. Model Dev.*, 10, 1903–1925, https://doi.org/10.5194/gmd-10-1903-2017.
- Miralles, D., T. Holmes, R. de Jeu, J. Gash, A. Meesters, and A. Dolman, 2011: Global land-surface evaporation estimated from satellite-based observations. *Hydrol. Earth Syst. Sci.*, 15, 453–469, https://doi.org/10.5194/hess-15-453-2011.
- —, M. van den Berg, A. Teuling, and R. de Jeu, 2012: Soil moisture-temperature coupling: A multiscale observational analysis. *Geophys. Res. Lett.*, **39**, L21707, https://doi.org/ 10.1029/2012GL053703.
- Molod, A., L. Takacs, M. Suarez, J. Bacmeister, I.-S. Song, and A. Eichmann, 2012: The GEOS-5 atmospheric general circulation model: Mean climate and development from MERRA to Fortuna. NASA Tech. Memo. NASA/TM-2014-104606, Vol. 28, 117 pp.

-, —, , and —, 2015: Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA-2. *Geosci. Model Dev.*, **8**, 1339–1356, https://doi.org/10.5194/gmd-8-1339-2015.

- Mueller, B., and Coauthors, 2011: Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations. *Geophys. Res. Lett.*, 38, L06402, https:// doi.org/10.1029/2010GL046230.
- —, and Coauthors, 2013: Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis. *Hydrol. Earth Syst. Sci.*, **17**, 1661–1679, https://doi.org/10.5194/ hess-17-3707-2013.
- NEWS Science Integration Team, 2007: A NASA Earth science implementation plan for energy and water cycle research: Predicting energy and water cycle consequences of Earth system variability and change. NASA Energy and Water Cycle Study Rep., 89 pp., http://news.cisc.gmu.edu/doc/NEWS_ implementation.pdf.
- Reichle, R., and Q. Liu, 2014: Observation-corrected precipitation estimates in GEOS-5. NASA Tech. Memo. NASA/TM-2014-104606, Vol. 35, 18 pp.
- —, R. Koster, G. De Lannoy, B. Forman, Q. Liu, S. Mahanama, and A. Toure, 2011: Assessment and enhancement of MERRA land surface hydrology estimates. *J. Climate*, 24, 6322–6338, https://doi.org/10.1175/JCLI-D-10-05033.1.
- —, C. Draper, Q. Liu, M. Girotto, S. Mahanama, R. Koster, and G. D. Lannoy, 2017a: Assessment of MERRA-2 land surface hydrology estimates. *J. Climate*, **30**, 2937–2960, https://doi.org/ 10.1175/JCLI-D-16-0720.1.

- —, Q. Liu, R. Koster, C. Draper, S. Mahanama, and G. Partyka, 2017b: Land surface precipitation in MERRA-2. J. Climate, 30, 1643–1664, https://doi.org/10.1175/JCLI-D-16-0570.1.
- Rienecker, M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. J. Climate, 24, 3624–3648, https://doi.org/10.1175/JCLI-D-11-00015.1.
- Schlosser, C., and X. Gao, 2010: Assessing evapotranspiration estimates from the Second Global Soil Wetness Project (GSWP-2) simulations. J. Hydrometeor., 11, 880–897, https:// doi.org/10.1175/2010JHM1203.1.
- Trenberth, K., J. Fasullo, and J. Kiehl, 2009: Earth's global energy budget. Bull. Amer. Meteor. Soc., 90, 311–323, https://doi.org/ 10.1175/2008BAMS2634.1.
- Wang, K., and R. Dickinson, 2013: Global atmospheric downward longwave radiation at the surface from ground-based observations, satellite retrievals, and reanalyses. *Rev. Geophys.*, 51, 150–185, https://doi.org/10.1002/rog.20009.
- Wild, M., D. Folini, and M. Hakuba, 2015: The energy balance over land and oceans: An assessment based on direct observations and CMIP5 climate models. *Climate Dyn.*, 44, 3393–3429, https://doi.org/10.1007/s00382-014-2430-z.
- Wilson, K., and Coauthors, 2002: Energy balance closure at FLUXNET sites. Agric. For. Meteor., 113, 223–243, https:// doi.org/10.1016/S0168-1923(02)00109-0.
- Xie, P., and P. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, 78, 2539–2558, https://doi.org/10.1175/1520-0477(1997)078<2539: GPAYMA>2.0.CO;2.