Evaluating Precipitation Datasets Using Surface Water and Energy Budget Closure[✉]

SANAA HOBEICHI AND GAB ABRAMOWITZ

Climate Change Research Centre, and ARC Centre of Excellence for Climate Extremes, University of New South Wales, Sydney, New South Wales, Australia

STEEFAN CONTRACTOR

School of Mathematics and Statistics, University of New South Wales, Sydney, New South Wales, Australia

JASON EVANS

Climate Change Research Centre, and ARC Centre of Excellence for Climate Extremes, University of New South Wales, Sydney, New South Wales, Australia

(Manuscript received 28 October 2019, in final form 14 February 2020)

ABSTRACT

Evaluation of global gridded precipitation datasets typically entails using the in situ or satellite-based data used to derive them, so that out-of-sample testing is usually not possible. Here we detail a methodology that incorporates the physical balance constraints of the surface water and energy budgets to evaluate gridded precipitation estimates, providing the capacity for out-of-sample testing. Performance conclusions are determined by the ability of precipitation products to achieve closure of the linked budgets using adjustments that are within their prescribed uncertainty bounds. We evaluate and compare five global gridded precipitation datasets: IMERG, GPCP, GPCC, REGEN, and MERRA-2. At the spatial level, we show that precipitation is best estimated by GPCC over the high latitudes, by GPCP over the tropics, and by REGEN over North Africa and the Middle East. IMERG and REGEN appear best over Australia and South Asia. Furthermore, our results give insight into the adequacy of prescribed uncertainties of these products and shows that MERRA-2, while being less competent than the other four products in estimating precipitation, has the best representation of uncertainties in its precipitation estimates. The spatial extent of our results is not only limited to grid cells with in situ observations. Therefore, the approach enables a robust evaluation of precipitation estimates and goes some way to addressing the challenge of validation over observation scarce regions.

1. Introduction

Precipitation P is the main driver of the hydrological cycle and plays a key role in the surface energy cycle by influencing the partitioning of the outgoing energy from land between latent and sensible heat flux. It is measured in situ using rain gauges and ground-radar detectors (see review by Sun et al. 2018); however, the density of in situ observations varies widely. For example, gauges tend to be dense over Europe, eastern Australia,

India, and the United States but sparse over much of the rest of the land, particularly over Greenland, the Middle East and North African (MENA) region, the high latitudes, inner Asia, and western South America (Fig. 1). This makes understanding water resource availability difficult in many regions on land and hinders the improvement of water resource management, weather forecasting, and the detection of natural hazards such as floods and droughts (Hossain and Anagnostou 2004; Wu et al. 2012; Zhan et al. 2016; Golian et al. 2019).

The advancement of satellite technology and remote sensing retrieval algorithms together with computational capabilities have led to the development of a suite of satellite-driven estimates of precipitation at the global gridded scale. These include upscaled ground P observations (hereafter ground-based) that

[✉] Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JHM-D-19-0255.s1>.

Corresponding author: Sanaa Hobeichi, s.hobeichi@unsw.edu.au

FIG. 1. In situ observations that have been used in to constrain the budgets components. This includes streamflow measurements from gauged basins (blue), energy flux measurements for net radiation R_n , sensible heat flux H , latent heat flux (ET/LH), and ground precipitation measurements. Sources of these datasets are shown in bold text in Table 2.

use sophisticated interpolation techniques that, in some cases, incorporate other remote-sensed variables (e.g., elevation) (Legates and Willmott 1990; Chen et al. 2008; Becker et al. 2013; Harris et al. 2014; Contractor et al. 2020), entirely satellite-driven datasets (Joyce et al. 2004; Hong et al. 2004; Huffman et al. 2007; Ushio et al. 2009; Ashouri et al. 2015; Brocca et al. 2014; Huffman et al. 2015), re-analysis products (Saha et al. 2010; Dee et al. 2011; Kobayashi et al. 2015; Derber et al. 1991; Suarez et al. 2005), and any combination of these (Huffman et al. 1997; Joyce et al. 2004; Ashouri et al. 2015; Weedon et al. 2014; Funk et al. 2015; Bosilovich et al. 2015; Beck et al. 2017a; Reichle et al. 2017; Beck et al. 2019). These global P estimates have been able to fill the spatial gaps in in situ measurements. However, their performance relies heavily on the density of available in situ observations, which are used in the development of these global products directly or indirectly (through calibration, parameterization, or correction). Therefore, there is a lack of confidence in the ability of global P datasets to provide accurate representation of precipitation over regions with sparse or no in situ observations (Hughes 2006; Ward et al. 2011; Satgé et al. 2016; El Kenawy et al. 2019; Zhang and Anagnostou 2019).

Several studies employed available observational networks to characterize the quality of global precipitation products at regional (e.g., Gottschalck et al. 2005; Brown 2006; Herold et al. 2016; Wang et al. 2018) and global scales (Adler et al. 2001; Maggioni et al. 2016; Beck et al. 2017b; Dirmeyer et al. 2018; Adler et al. 2018). However, conclusions obtained from these

studies are mainly describing performance of global precipitation datasets over regions with dense observational P networks and cannot generalize to regions with scattered in situ observation (Beck et al. 2017b).

To enable the evaluation of precipitation datasets over data-sparse regions, several studies have used hydrological modeling constrained with measurements from climatic variables that are directly dictated by precipitation such as streamflow (Grimes and Diop 2003; Bitew et al. 2012; Thiemig et al. 2012; Tong et al. 2014; Sirisena et al. 2018) and soil moisture (Pan et al. 2010; Azarderakhsh et al. 2011; Martens et al. 2017). This has enabled the assessment of global precipitation datasets over regions with sparse P observations with more confidence. However, the success of these studies in expanding the spatial domain over which P datasets can be evaluated is limited because most regions also lack information about the other hydrological variables (Hughes 2006). Recently, Munier and Aires (2018) have characterized the performance of multiple precipitation datasets through an approach that combines ground-based observation of multiple hydrologic variables along with performance metrics based on surface water budget closure. Their approach offers advancement over previous studies by allowing a more reliable evaluation of precipitation datasets over regions with sparse observations by relying on the physical conservation constraints offered by the water balance. In the same context, Hobeichi et al. (2020) implement an approach that enforces the simultaneous closure of the surface

water and energy budgets at 0.5° and monthly space and time scales. They find that the physical constraints of the water and energy budgets can complement in situ observations by allowing observation-based evaluation of global estimates of any components of these budgets at the grid scale including in observationally poor regions.

A common finding from the available evaluation studies is that all existing precipitation datasets suffer from biases that vary in time and space. Details about the type and source of errors in precipitation datasets are provided in Bytheway and Kummerow (2013), Dunn et al. (2014), Bytheway and Kummerow (2013), Avila et al. (2015), Awange et al. (2016), Sun et al. (2018), and others. Unfortunately, few datasets provide estimates for uncertainties in precipitation (e.g., Hou et al. 2014; Reichle et al. 2017; Schneider et al. 2017; Adler et al. 2018; Contractor et al. 2020). Examining the adequacy of the uncertainty representation in precipitation is necessary for better characterization of the sources of errors (Steiner et al. 2003) and for use of precipitation products in many applications such as water resources management and other hydrological applications (Maggioni et al. 2016). Despite this, current evaluation studies of precipitation datasets do not include the anticipated uncertainties when evaluating precipitation fields, and do not use performance metrics suitable for the assessment of the range of precipitation values that include uncertainty.

The aim of the present study is to evaluate and compare precipitation and uncertainty estimates of five global precipitation datasets at the monthly 0.5° grid scales over 2003–09. To achieve this, we employ the same approach as in Hobeichi et al. (2020) by examining the adjustments that precipitation datasets must undergo to enable the simultaneous closure of the water and energy budgets and comparing these with uncertainty estimates. We make use of in situ measurements of several hydrological and energy fluxes along with the physical constraints of the water and energy budgets to aid evaluation at larger spatial domains than those covered by in situ P observations only. We address four objectives:

- evaluating and comparing the performance of monthly estimates of P datasets,
- assessing the consistency of the P datasets with the other budget terms,
- examining the adequacy of the uncertainty estimates in P fields, and
- highlighting regions on land where the precipitation range of datasets is overestimated or underestimated.

The paper is organized as follows: section 2 describes the participating precipitation datasets, the methods, and the metric of performance. Section 3 presents and discusses our findings before we conclude in section 4.

2. Datasets and method

The evaluation and comparison approach adopted in this study is based on the physical conservation constraints provided by the water and energy budgets. The broad methodology is very similar to the one established in Hobeichi et al. (2020) where a data assimilation technique (DAT) is implemented to enforce the balance of the linked surface terrestrial water and energy budgets. This utilizes existing estimates of all other terms in the surface energy and water budgets, including uncertainty estimates, as described in Hobeichi et al. (2018, 2019, 2020). It involves simultaneous adjustments to all the individual components of the budgets based on their relative uncertainties while minimizing deviation from their initial estimates. In this work, we implement the DAT five times, each time with a different precipitation dataset while maintaining the same estimates of the other components of the water and energy budgets. Since uncertainties associated with the budget fluxes constitute an essential element of the DAT, only estimates of budget terms with prescribed uncertainties are incorporated. We apply the DAT on every grid cell in the spatial domain, for every month. We infer performance conclusions for the precipitation datasets as well as their associated uncertainties by 1) examining and analyzing the sign and magnitude of the adjustments applied to the precipitation datasets, 2) comparing those adjustments to the uncertainty estimates, and 3) examining the goodness of the fit that we establish using a χ^2 test detailed later in this section. While this approach could equally be used to give insight into the skills of all other budget terms, their estimates are already hybrid products (as described in Hobeichi et al. 2020).

This section presents the employed precipitation datasets, the DAT, and the performance metrics.

a. Precipitation datasets

Despite the availability of a wide range of global precipitation datasets [a database of 30 precipitation datasets detailed in Roca et al. (2019)], very few provide an estimate of errors. As explained earlier, only precipitation datasets that have prescribed uncertainties can be incorporated in our evaluation study. These are the Global Precipitation Climatology Centre (GPCC; Schneider et al. 2017), the Global Precipitation Climatology Project (GPCP) monthly analysis version 2.3 (Adler et al. 2018), Integrated Multisatellite Retrievals for GPM (IMERG; Hou et al. 2014) monthly product version 06 final, the Modern-Era Retrospective Analysis for Research and Applications version 2 (MERRA-2; Reichle et al. 2017), and Rainfall Estimates on a Gridded Network (REGEN; Contractor et al. 2020).

The predecessors of these datasets also provide uncertainty estimates; these are not included here as we only incorporate the latest release in each product. The datasets are all available over the period of study 2003–09 and remapped to a 0.5° grid using nearest neighborhood. REGEN daily estimates are aggregated to monthly averages. Only a brief description of these datasets is provided here and in Table 1. Readers are encouraged to refer to the associated publications for more details about each dataset.

1) REGEN ALL STATIONS v1–2019

REGEN is a gauge-based, global gridded dataset of daily precipitation with a 1° latitude/longitude resolution. For this study, the daily estimates were aggregated to produce monthly totals and remapped to a 0.5° resolution using the nearest neighborhood method. REGEN's temporal coverage spans from 1950 to 2016. This dataset was developed in collaboration with researchers from the Global Precipitation Climatology Centre (GPCC), Deutscher Wetterdienst, Germany, and the National Centers for Climatic Information (NCEI), National Oceanic and Atmospheric Administration (NOAA), United States. These organizations maintain the largest independent archives of in situ observations of daily precipitation, which were combined by the developers of REGEN to create a database of more than 135 000 stations. The dataset employs ordinary block kriging to interpolate the combined database to create a gridded product. REGEN is published and freely available via <https://dx.doi.org/10.25914/5ca4c380b0d44>.

2) GPCC FULL DATA MONTHLY PRODUCT VERSION 2018

This is a gridded gauge-analysis dataset derived by merging station data from more than 85 000 stations over land. GPCC has one of the two largest databases of in situ observations from a variety of sources. It mostly comprises monthly totals provided by global networks, including those distributed by the Global Telecommunication System, datasets from 190 countries obtained from the national meteorological and/or hydrological services (NMHSs), research projects, regional networks such as the former Soviet Union and the African rainfall archive from Nicholson and other sources (Schneider et al. 2014). Precipitation datasets from all different sources undergo thorough checks and corrections through a sophisticated system that includes an automated module in addition to visual examination. Corrections include accounting for gauge undercatch by applying 85% of the correction factors computed for each calendar month by Legates and Willmott (1990). The dataset employs a modified SPHEREMAP interpolation scheme (Becker et al. 2013).

TABLE 1. Name, resolution, sources and data access details of the employed precipitation datasets.

Name and abbreviation used in this study	Temporal and spatial resolution	Data source(s)	Data download link and reference
GPM IMERG Final Precipitation L3 V06; GPM_3IMERGM (IMERGE)	Monthly, 0.1°	Satellite-based	https://gpm1.gesdisc.eosdis.nasa.gov/data/GPM_L3/GPM_3IMERGM.06/ (Huffman et al. 2019)
GPCC Full Data Monthly Product Version 2018 (GPCC)	Monthly, 0.5°	Ground-based	https://opendata.dwd.de/climate_environment/GPCC/html/fulldata-monthly_v2018_doi_download.html (Schneider et al. 2018)
NASA GPCP V2p3 CDR (GPCP)	Monthly, 2.5°	Satellite-based	https://iridl.ldeo.columbia.edu/SOURCES/.NASA/.GPCP/.V2p3/.CDR/index.html?Set-Language=en (Adler et al. 2003)
MERRA-2 Surface flux diagnostics (M2TMNXFLX) V5.12.4 (MERRA-2)	Monthly, 0.5° × 0.625°	Reanalysis	https://disc.gsfc.nasa.gov/datasets/M2TMNXFLX_V5.12.4/summary (MERRA-2 tavgm_2d_flux_Nx; GMAO 2015)
REGEN All Stations v1–2019 (REGEN)	Daily, 1°	Ground-based	http://dapds00.ncl.org.au/thredds/catalogs/ks32/CLEX/REGEN/REGEN_AllStns.html (Contractor 2018)

3) GPCP MONTHLY ANALYSIS V2.3

GPCP is a multisatellite-based product produced under NOAA's Reference Environmental Data Record (REDR) Program. It combines infrared estimates from geostationary satellites and polar orbiting satellites, multiple microwave estimates, and outgoing longwave radiation estimates. GPCP fields are corrected to match in situ analysis fields from the GPCC V7 Full analysis (over 1979–2013) which applies the full monthly correction factors by Legates and Willmott (1990). Datasets contributing to the composite precipitation fields in GPCP are different at latitudes below and above 40°N/S. More details about the employed version of GPCP can be found in Adler et al. (2018).

4) IMERG FINAL PRECIPITATION L3 1 MONTH V06

IMERG is a multisatellite product developed by NASA and Japan Aerospace and Exploration Agency (JAXA). It incorporates measurements from the Global Precipitation Measurement (GPM) *Core Observatory* satellite after 2014 and TRMM instrument prior to 2014 along with measurements from partner microwave sensors and infrared based observations from geosynchronous satellites (Hou et al. 2014). IMERG is derived by the Goddard profiling algorithm (GPROF) that enables blending and intercalibrating heterogeneous datasets in a consistent way. Some of the satellite data used to derive IMERG are also used to derive GPCP. The monthly dataset used here is produced by the “final” run and is bias corrected using GPCC networks.

5) MERRA-2 SURFACE FLUX DIAGNOSTICS

MERRA-2 is a global atmospheric reanalysis developed by the NASA Global Modeling and Assimilation Office (GMAO). The modeling system used in the development of MERRA-2 is the Goddard Earth Observing System Model, version 5 (GEOS-5) with the Atmospheric General Circulation Model (AGCM) configuration. The AGSM incorporates an Atmospheric Data Assimilation System (ADAS) that enables the integration of observations into the model. The derived precipitation from the atmospheric model is then corrected before being used to force the land surface scheme in MERRA-2. For instance, over the low and midlatitudes (between 42.5°S and 42.5°N) excluding Africa, MERRA-2 model generated precipitation is fully corrected by daily P fields from the gauge-based gridded product of the NOAA Climate Prediction Center (CPC) Unified Gauge-Based Analysis of Global Daily Precipitation (CPCU; Chen et al. 2008) along with five daily P fields from the CPC Merged Analysis of Precipitation (CMAP; Xie and Arkin 1997) product. The correction with CPCU and CMAP decreases gradually toward 62.5°N/S and does not expand poleward from

62.5°N/S due to the sparsity of gauge observations. Over Africa, MERRA-2 model generated precipitation is entirely substituted by CMAP and rescaled to match the seasonal climatology of GPCP v2.1 pentad product. For more details, we refer the reader to Reichle et al. (2017). In this study we use MERRA-2 corrected precipitation and we refer to it as MERRA-2.

We refer to IMERG and GPCP as satellite products, and to GPCC and REGEN as ground-based products. While GPCP, GPCC, REGEN, and IMERG share observational data, these products have substantial differences in their types (satellite/ground-based), retrieval algorithms (for satellite datasets) and interpolation methods (for ground-based datasets), and the temporal resolution of the employed observational data (daily for ground-based datasets and monthly for satellite datasets). The average precipitation of each of the five P datasets is calculated over 2003–09 and presented in Fig. 2. The global mean climatology shows differences across the three types of datasets that have been described in Herold et al. (2016). These plots highlight similarities and differences in the climatology of these datasets. As expected, datasets of the same type (satellite and ground-based) show very similar spatial patterns. Also, satellite datasets and ground-based datasets show agreement over most of the land except the high latitudes, the Amazon, the upper Andes foothills, and the tropical wet coastlines of South Asia. MERRA-2 is the only reanalysis dataset; it exhibits a similar global average as the other four datasets, however the spatial pattern of P fields is different over many regions on land, particularly the high latitudes, the tropics, and South Asia.

b. Water and energy estimates (other than P)

For the components of the surface water and energy budgets other than P and the change in water storage ΔS , we use hybrid estimates derived in previous work (Hobeichi et al. 2018, 2019, 2020). These studies use an optimal weighting technique developed by Bishop and Abramowitz (2013) that can combine several estimates of the same variable in such a way to maximize the performance of the merged product with respect to in situ observations. Also, it accounts for the error dependency between the constituent estimates and computes uncertainties in the derived hybrid product based on its discrepancy with available in situ observation (Abramowitz and Bishop 2015). The variables for which hybrid estimates and uncertainties have been computed are net radiation (R_n), sensible heat flux (H), latent heat flux (LH) [or alternatively evapotranspiration (ET)], ground heat flux (G), and runoff (Q). A range of datasets has been incorporated to derive each hybrid estimate. These datasets are listed in Table 2. Parameter

FIG. 2. Annual mean precipitation (mm day^{-1}) calculated for 2003–09 computed for each of IMERG, GPCP, GPCC, REGEN, and MERRA-2.

ΔS is derived from GRACE Mascons water storage anomalies (Watkins et al. 2015). Except for ΔS , all the components of the budgets are observationally constrained hybrid estimates. We limit our analysis to the short period 2003–09 due to the time span of the hybrid LH that spans over 2000–09 and because GRACE inferred ΔS is not available prior to 2003. We show in Fig. 1 all the in situ observations that have been used to constrain the budgets components. All these in situ observations in addition to the physical relationship offered by the water and energy cycles are incorporated in this study to carry out an evaluation and comparison of five precipitation products including their uncertainty estimates.

c. The data assimilation technique

The DAT employed in this study is based on a theory detailed in L'Ecuyer and Stephens (2002). The same method is applied and detailed in Hobeichi et al. (2020) and at different temporal and spatial scales in L'Ecuyer et al. (2015) and Rodell et al. (2015). Therefore, we

briefly recount the method here, and we refer the reader to Hobeichi et al. (2020) for further details.

We first consider the water and energy budgets written in the form $\Delta S = P - ET - Q$ and $R_n = H + LH + G$. When all the estimates are converted to the same units (e.g., W m^{-2}) we can use ET and LH interchangeably. In this case, we can write: $\mathbf{R} = \mathbf{A} \cdot \mathbf{F}$, where

$$\mathbf{R} = \begin{pmatrix} \Delta S \\ R_n \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 1 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix}, \quad \text{and} \quad \mathbf{F} = \begin{pmatrix} P \\ LH \\ Q \\ H \\ G \end{pmatrix}.$$

Since the initial budget terms do not precisely balance, we denote our initial estimates of $\mathbf{F}$ and $\mathbf{R}$ by $\mathbf{F}_i$ and $\mathbf{R}_i$ (i refers to imbalance) respectively. We aim to find

new vectors, $\mathbf{F}_b$ and $\mathbf{R}_b$ that conserve the budget closure by adjusting each term in $\mathbf{F}_i$ based on its relative uncertainty in such a way that the deviation from $\mathbf{R}_i$ and $\mathbf{F}_i$ is minimized, when weighted by their uncertainty distributions. This can be achieved using a Bayesian framework by finding $\mathbf{F}_b$ and $\mathbf{R}_b$ that maximize the joint probability:

$$P(\mathbf{F}_b | \mathbf{F}_i, \mathbf{R}_i) = \exp[-(\mathbf{F}_b - \mathbf{F}_i)^T \mathbf{S}_{\mathbf{F}_i}^{-1} (\mathbf{F}_b - \mathbf{F}_i)] \times \exp[-(\mathbf{R}_b - \mathbf{R}_i)^T \mathbf{S}_{\mathbf{R}_i}^{-1} (\mathbf{R}_b - \mathbf{R}_i)]. \quad (1)$$

The terms $\mathbf{S}_{\mathbf{F}_i}$ and $\mathbf{S}_{\mathbf{R}_i}$ represent the error covariance of $\mathbf{F}_i$ and $\mathbf{R}_i$ derived from the uncertainty of each element in $\mathbf{F}_i$ and $\mathbf{R}_i$ and subscript b refers to “balance.” Maximizing (1) requires maximizing the exponent term, or alternatively minimizing J_a , where

$$J_a = (\mathbf{F}_b - \mathbf{F}_i)^T \mathbf{S}_{\mathbf{F}_i}^{-1} (\mathbf{F}_b - \mathbf{F}_i) + (\mathbf{R}_b - \mathbf{R}_i)^T \mathbf{S}_{\mathbf{R}_i}^{-1} (\mathbf{R}_b - \mathbf{R}_i). \quad (2)$$

The solution to minimizing (2) is a set of coherent fluxes (i.e., $\mathbf{F}_b$, $\mathbf{R}_b$) that can close the water and energy budgets simultaneously, such that

$$\mathbf{F}_b = \mathbf{F}_i + \mathbf{S}_{\mathbf{F}_i} \mathbf{K}^T \mathbf{S}_{\mathbf{R}_i}^{-1} (\mathbf{R}_i - \mathbf{K} \mathbf{F}_i), \quad (3)$$

where $\mathbf{K}$ represents the Jacobian of $\mathbf{R}$ at $\mathbf{F}_i$ and $\mathbf{S}_{\mathbf{F}_b} = (\mathbf{K}^T \mathbf{S}_{\mathbf{R}_i}^{-1} \mathbf{K} + \mathbf{S}_{\mathbf{F}_i}^{-1})^{-1}$ is the error covariance of $\mathbf{F}_b$, while $\mathbf{S}_{\mathbf{R}_b} = \mathbf{S}_{\mathbf{R}_i}$.

Under the assumption that the uncertainty estimates are uncorrelated, unbiased and are Gaussian, the estimated $\mathbf{F}_b$ and $\mathbf{R}_b$ are the best possible fit to $\mathbf{F}_i$ and $\mathbf{R}_i$ respectively, given the uncertainty estimates $\mathbf{S}_{\mathbf{F}_i}$ and $\mathbf{S}_{\mathbf{R}_i}$.

The estimates used for $\mathbf{F}_i$ and $\mathbf{R}_i$ are those referred to in sections 2a and 2b. These mostly consist of observationally constrained (hybrid) estimates. Parameter P_b is one of IMERG, GPCC, GPCP, MERRA-2, and REGEN, which are all observationally constrained. The new balanced estimates $\mathbf{F}_b$ and $\mathbf{R}_b$ give insight into the quality of $\mathbf{F}_i$ and $\mathbf{R}_i$ as well as their uncertainties by looking at the adjustments that these estimates (i.e., $\mathbf{F}_i$ and $\mathbf{R}_i$) have undergone to achieve the balance.

The data assimilation technique employed in this study has been previously implemented by Rodell et al. (2015) and L’Ecuyer et al. (2015) primarily to enforce the closure of multiple linked budgets and also to infer conclusion on the performance of the original imbalanced estimates. Other techniques that apply maximum likelihood estimation are available (e.g., Shumway and Stoffer 2017; Zia et al. 2008) and could be explored and adapted to enforce the simultaneous closure of the budgets.

d. Performance metrics

The quality of the final fit is tested using a χ^2 test to the resultant balanced estimates (i.e., $\mathbf{F}_b$ and $\mathbf{R}_b$). As established in L’Ecuyer et al. (2015), we compute the χ^2 scores from the ratios of adjustments applied to the imbalanced estimates (numerator) to the uncertainties of the original imbalanced estimates (denominator) instead of the imbalanced estimates (denominator) as in the Pearson’s χ^2 test. Consequently, the χ^2 scores are calculated by substituting the original and the balanced estimates of the budgets’ components in the right side of (2) such as

$$\chi^2 \text{ score} = (\mathbf{F}_b - \mathbf{F}_i)^T \mathbf{S}_{\mathbf{F}_i}^{-1} (\mathbf{F}_b - \mathbf{F}_i) + (\mathbf{R}_b - \mathbf{R}_i)^T \mathbf{S}_{\mathbf{R}_i}^{-1} (\mathbf{R}_b - \mathbf{R}_i). \quad (4)$$

The χ^2 scores are calculated at every grid for every month during 2003–09 and for each of the five employed datasets. This assumes that the uncertainties of the original estimates are uncorrelated, unbiased and follow a Gaussian distribution. As in the Pearson’s χ^2 test, the computed χ^2 scores are compared to a critical value, which in our case is the number of fluxes being optimized, that is, seven. A score larger than seven indicates that on average, adjustments larger than one standard deviation of quoted uncertainties are applied to one or more original estimates of the budgets components as a result of inconsistencies between them. Therefore, comparing the χ^2 scores achieved by each of the five employed precipitation datasets can reveal regions where a particular precipitation dataset contributes to a better/worse fit than the others.

In addition to the χ^2 test, the adjustments applied to the precipitation datasets are directly compared to their associated uncertainties. This helps pointing out grid cells where precipitation adjustments violate the uncertainty assumptions, and hence allows us to evaluate the adequacy of uncertainty estimates to reflect likely errors in the precipitation fields at the grid level. Furthermore, for each employed precipitation dataset, we examine the sign of adjustments applied to its original estimates to 1) identify areas where the original estimates are underestimated or overestimated, and 2) identify areas where the precipitation range (i.e., [precipitation – uncertainty, precipitation + uncertainty]) underestimates the balanced precipitation estimates. In the first, the modified precipitation estimates do not violate uncertainties assumptions. In contrast, the second is linked to grid cells of large χ^2 scores (i.e., greater than seven).

3. Results and discussions

In this section, we address the research objectives highlighted in section 1.

TABLE 2. Source, resolution, and data access details of the employed datasets for the components of the surface water and energy budgets. These include products used to derive hybrid estimates for R_n , H , and G (CLASS; Hobeichi 2018a), for ET/LH (DOLCE; Hobeichi 2018b), and for Q (LORA; Hobeichi 2018b), along with the employed dataset for ΔS . Bold text refers to in situ observations.

Budget component	Source	Resolution	Data download link/reference
R_n hybrid	CERES-EBAF	Monthly 1°	https://eosweb.larc.nasa.gov/project/ceres/ebaf-surface_ed4.0_table ; Loeb (2017)
	ERA-Interim	Monthly 0.5°	http://apps.ecmwf.int/datasets/data/interim-mdfa/levtype=sfc/ ; Dee et al. (2011)
	GLDAS-NOAH	Monthly 0.5°	https://hydro1.gesdisc.eosdis.nasa.gov/data/GLDAS/GLDAS_NOAH025_M2.1/ ; Beaudoin and Rodell (2016)
	MERRA-2 Land surface diagnostics	Monthly 0.5° × 0.625°	https://goldsmr4.gesdisc.eosdis.nasa.gov/data/MERRA2_MONTHLY/M2TMNXLND.5.12.4/ ; Gelaro et al. (2017)
	NCEP-DOE	Monthly T62 Gaussian grid with 192 × 94 points	https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis2.surface.html ; Kanamitsu et al. (2002)
	BSRN	Minutely in situ	https://dataportal.pangaea.de/bsrn/
	FLUXNET 2015-tier1	Daily in situ	https://fluxnet.ornl.gov/
	LaThuile Free Fair Use 2007	Daily in situ	http://www.fluxdata.org/DataDownload/default.aspx
	GLDAS-NOAH	Monthly 0.5°	https://hydro1.gesdisc.eosdis.nasa.gov/data/GLDAS/GLDAS_NOAH025_M2.1/ ; Beaudoin and Rodell (2016)
	MERRA-2 land surface diagnostics	Monthly 0.5° × 0.625°	https://goldsmr4.gesdisc.eosdis.nasa.gov/data/MERRA2_MONTHLY/M2TMNXLND.5.12.4/ ; Gelaro et al. (2017)
H hybrid	MERRA-2 surface flux diagnostics	Monthly 0.5° × 0.625°	https://goldsmr4.gesdisc.eosdis.nasa.gov/data/MERRA2_MONTHLY/M2TMNXFLX.5.12.4/ ; Gelaro et al. (2017)
	MPiBGC	Monthly 0.5°	https://vesgint-data.ipsl.upmc.fr/thredds/dodsC/IPSLFS/igcmg/IGCM/ReferenceDatasets/ls/land/mo/fluxsens/fluxsens_Lmon_EnsembleHcor_observations_198201-201112.nc.html ; Jung et al. (2011)
	NCEP-DOE	Monthly T62 Gaussian grid with 192 × 94 points	https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis2.surface.html ; Kanamitsu et al. (2002)
	Princeton- H	Daily 0.5°	Princeton University, New Jersey; Siemann et al. (2018)
	FLUXNET 2015-tier1	Daily in situ	https://fluxnet.ornl.gov/
	LaThuile Free Fair Use 2007	Daily in situ	http://www.fluxdata.org/DataDownload/default.aspx
	DOLCE	Monthly 0.5°	https://dx.doi.org/10.4225/41/58980b55b0495
	GLEAM-V2A	Daily 0.25°	Miralles et al. (2011)
	GLEAM-V2B	Daily 0.25°	Miralles et al. (2011)
	GLEAM-V3A	Daily 0.25°	Martens et al. (2016)
DOLCE hybrid ET/LH	MOD16 MODIS global ET products	Monthly 0.5°	Mu et al. (2011)
	MPiBGC	Monthly 0.5°	Jung et al. (2011)
	PML PM-Leuning model	Monthly 0.5°	Zhang et al. (2016)
	FLUXNET 2015-tier1	Daily in situ	https://fluxnet.ornl.gov/
	LaThuile Free Fair Use 2007	Daily in situ	http://www.fluxdata.org/DataDownload/default.aspx

TABLE 2. (Continued)

Budget component	Source	Resolution	Data download link/reference
<i>G</i> hybrid	MERRA-2 Land surface diagnostics	Monthly $0.5^{\circ} \times 0.625^{\circ}$	https://goldsmr4.gesdisc.eosdis.nasa.gov/data/MERRA2_MONTHLY/M2TMNXLND.5.12.4/ ; Gelaro et al. (2017)
	MERRA-2 Surface flux diagnostics	Monthly $0.5^{\circ} \times 0.625^{\circ}$	https://goldsmr4.gesdisc.eosdis.nasa.gov/data/MERRA2_MONTHLY/M2TMNXFLX.5.12.4/ ; Gelaro et al. (2017)
	GLDAS-NOAH	Monthly 0.5°	https://hydro1.gesdisc.eosdis.nasa.gov/data/GLDAS/GLDAS_NOAH025_M.2.1/ ; Beaudoin and Rodell (2016)
	NCEP-NCAR	Monthly T62 Gaussian grid with 192×94 points	https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.surfaceflux.html ; Kalnay et al. (1996)
	NCEP-DOE		https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis2.surface.html ; Kanamitsu et al. (2002)
	BSRN	Minutely in situ	https://dataportal.pangaea.de/bsrn/
	FLUXNET 2015-tier1	Daily in situ	https://fluxnet.ornl.gov/
	LaThuile Free Fair Use 2007	Daily in situ	http://www.fluxdata.org/DataDownload/default.aspx
	LORA	Monthly 0.5°	http://dx.doi.org/10.25914/5b612e993d8ea
	HBV-SIMREG	Monthly 0.5°	Beck et al. (2016); ftp://wci.earth2observe.eu/
LORA <i>Q</i> hybrid	HTESSEL	Monthly 0.5°	Balsamo et al. (2009, 2011); ftp://wci.earth2observe.eu/
	HTESSEL	Monthly 0.25°	Balsamo et al. (2009, 2011); ftp://wci.earth2observe.eu/
	JULES	Monthly 0.5°	Best et al. (2011) and Clark et al. (2011); ftp://wci.earth2observe.eu/
	JULES	Monthly 0.25°	Best et al. (2011) and Clark et al. (2011); ftp://wci.earth2observe.eu/
	LISFLOOD	Monthly 0.5°	Burek et al. (2013) and van der Knijff et al. (2010); ftp://wci.earth2observe.eu/
	PCR-GLOBWB	Monthly 0.5°	van Beek and Bierkens (2008); ftp://wci.earth2observe.eu/
	SURFEX	Monthly 0.5°	Decharme et al. (2011, 2013); ftp://wci.earth2observe.eu/
	SURFEX	Monthly 0.25°	Decharme et al. (2011, 2013); ftp://wci.earth2observe.eu/
	W3RA	Monthly 0.5°	van Dijk and Warren (2010) and van Dijk et al. (2014); ftp://wci.earth2observe.eu/
	WaterGAP3	Monthly 0.5°	Flörke et al. (2013); ftp://wci.earth2observe.eu/
	USGS		http://water.usgs.gov
	GAGES-II		Falcone et al. (2010)
	Dai (2016)		Dai (2016)
	GRACE Mascons-JPL	Monthly 0.5°	ftp://podaac-ftp.jpl.nasa.gov/allData/tellus/L3/mascon/RL06/JPL/ ; Watkins et al. (2015)
	RL06M1MSCNV01		
ΔS

FIG. 3. Change induced by the data assimilation technique in IMERG as a percentage of the original estimate, averaged over 2003–09 for each calendar month.

a. Evaluating the performance of monthly P datasets

The DAT induced changes to each precipitation product and to the other components of the water and energy budgets in order to establish the closure of these budgets. We illustrate in Figs. 3–7 the changes applied to each of IMERG, GPCP, GPCC, REGEN, and MERRA-2, respectively, averaged for each calendar month over 2003–09 as a percentage of their original mean estimates. The plots can be interpreted as follows: (i) the larger the modifications a dataset undergoes, the less consistent it is with other estimated budget terms, and (ii) a relative increase (decrease) indicates that P has originally been underestimated (overestimated). These plots enable us to compare the five products and identify areas where a particular product outperforms or is less reliable than the others. The plots in Figs. 3–7 show that MERRA-2 is the most modified dataset by the DAT and exhibits large differences in modification

patterns with the other datasets over all land and throughout the year.

Figures 3 and 4 show that over the Siberian high latitudes, P estimates in the two satellite products are increased in fall, particularly during the first snow (i.e., in October), while spring P estimates undergo a decrease in May over the Siberian plateau and northeastern Siberia, and in June along the coasts of the Laptev Sea and the East Siberian Sea. Figures 5 and 6 show minimal adjustments in ground-based P over the Siberian high latitudes in fall, and similar adjustments to the satellite-based P in spring. On the other hand, the two ground-based datasets exhibit more modifications over South America, particularly GPCC over the Brazilian highlands, the Southern Andes, and the Atacama Desert. Both satellite and ground-based datasets undergo the least relative changes over Europe and North America, which are the most observationally dense regions. Over the tropics the adjustments in the satellite and ground-based

FIG. 4. Change induced by the data assimilation technique in GPCP as a percentage of the original estimate, averaged over 2003–09 for each calendar month.

datasets mostly occur in the African horn, the Ethiopian highlands, and some parts of the Amazon. Both satellite and ground-based P exhibit a decrease in the Amazon during the wet season and an increase in September and October. In general, the least relative change in P over the Amazon is found in GPCP (Fig. 4). Furthermore, both satellite and ground-based P exhibit large relative change in the MENA region throughout the year. In India, the DAT does not change the monsoonal precipitation (June–September) in the satellite and ground-based estimates, except for GPCC, which exhibits a high relative decrease along the wet coastlines of southwest India. On the other hand, the DAT applies a decrease to all these products in northern India and an increase in southern India during the early monsoon (April and May). In northern Australia, GPCP, IMERG, and REGEN exhibit minimal changes during the wet season. It is noticeable from Figs. 3–6 that IMERG and REGEN are the least changed products over Australia.

In contrast, MERRA-2 undergoes consistent high-latitude increases during July–October and increases in the Northern Hemisphere spring. Changes are overall notably larger than for ground or satellite based products.

Our results suggest that GPCC appears most skillful over the high latitudes followed by REGEN, while GPCP and IMERG tend to underestimate fall and winter snow. None of the products perform well over the MENA region, particularly GPCC. This is not surprising given that the MENA region is the most observationally scarce region. Despite this, REGEN is able to estimate precipitation in many parts of this region, which has been shown in Contractor et al. (2020). In general, the best performance obtained over the tropics is by GPCP. This is consistent with Azarderakhsh et al. (2011) who assessed the coherence of multiple global gridded datasets with previous versions of GPCC and GPCP with other components of the water budgets and

FIG. 5. Change induced by the data assimilation technique in GPCC as a percentage of the original estimate, averaged over 2003–09 for each calendar month.

found that GPCP outperforms the other products over the Amazon.

The lower performance exhibited by GPCC and REGEN in the Amazon are explained by the sparse observational data over this complex ecosystem. On the other hand, IMERG performs the best in Australia particularly during the wet season (November–March) followed by REGEN. The latter performs the best in India, particularly during the monsoon season. Satellite and ground-based products underestimate P in the Western Ghats and overestimate it toward the north of India during the early and late monsoon season. [Sunilkumar et al. \(2015\)](#) report similar results but during the whole monsoon season. Despite the large similarity between REGEN and GPCC in terms of using the same observational dataset over most of the land, the improvement of GPCC over REGEN at the high latitudes is probably attributed to the application of wind-induced undercatch which has not been applied in REGEN. On

the other hand, the better skill shown by REGEN relative to GPCC over most of the land is attributed to the inclusion of additional in situ observations. This is particularly evident in North Africa, South America, and India. Interestingly, most adjustments in the satellite-based estimates occur during fall and winter except over the Sahara where adjustments throughout along the year.

It is noticeable that the changes applied to the two satellite products are clustered by latitude and exhibit a less stochastic pattern than those applied to the two ground-based products. In the satellite-based products this is partially attributable to the differences in the source of raw satellite imagery that contributed to precipitation fields over different latitudes. The scattered pattern of performance features in ground-based datasets is a result of the sensitivity of the interpolation method to the density of in situ observation and to the spatial coherence of grid cells, which are widely variable

FIG. 6. Change induced by the data assimilation technique in REGEN as a percentage of the original estimate, averaged over 2003–09 for each calendar month.

across land. Most differences in the patterns of adjustment applied by the DAT to the five products occur over the land when light or no precipitation occur such as in the MENA region, inner Australia, eastern Siberia, the Tibetan Plateau, South America, and northern North America. This is consistent with results from [Bytheway and Kummerow \(2013\)](#) who report that most differences between precipitation datasets occur when little or no rain accumulates, and findings from [Donat et al. \(2019\)](#) who report large differences in precipitation estimates over grid cells located in arid land. Also, [Tian and Peters-Lidard \(2010\)](#) report large relative spread between global satellite precipitation datasets over highlands including the Tibetan Plateau and the Andes in addition to the coastlines and the complex terrains receiving light precipitation. Furthermore, [Harrison et al. \(2019\)](#) show that satellite-based products and ground-based products do not agree over sub-Saharan Africa.

Our results suggest that there is no single precipitation dataset that can outperform all the other ones, and that the choice of the best performing dataset depends on the application such as the region and month of interest and the metric assessed. In this work we have evaluated monthly precipitation rates and their associated uncertainties in five datasets. Assessing these datasets in different temporal scales (e.g., hourly, daily), over particular time periods (e.g., extreme events), or for other skills (e.g., phasing, amplitude, or variability) can lead to different performance results. We therefore call for a careful selection of the best precipitation product depending on the scientific application.

Our results should not be interpreted in a way that undermines MERRA-2. There are several studies that showed high skills exhibited in MERRA-2 with respect to evaluation metrics different to those used here. For instance, [Dirmeyer et al. \(2018\)](#) have shown that MERRA-2 performs as well as other precipitation

FIG. 7. Change induced by the data assimilation technique in MERRA-2 as a percentage of the original estimate, averaged over 2003–09 for each calendar month.

datasets in terms of agreement with in situ observations from FLUXNET2015.

The aim of this study is to carry out an evaluation and comparison of different precipitation datasets. In Hobeichi et al. (2020) a similar analysis has been implemented involving all the other budget variables and REGEN precipitation dataset. It showed that in many regions on land, budget components other than precipitation undergo an adjustment by the DAT that is relatively larger than that applied to precipitation. For instance, R_n is largely modified over the mid- and high latitudes of the Northern Hemisphere in March and April. Sensible heat flux H appears underestimated over most of the Southern Hemisphere during March–May. In October, LH exhibits a significant increase over Europe and the Siberian plains. In many areas over the Amazon, both LH and P exhibit significant changes and contribute to high χ^2 values. We refer the reader to Hobeichi et al. (2020) to gain more insights on the

adjustments made to the other components of the water and energy budgets when REGEN is used.

b. Consistency of the uncertainty estimates of P datasets with other budget terms

In an ideal case, each component of the water and energy budgets receives adjustments that are on average smaller than its quoted uncertainty standard deviation, achieving a χ^2 score smaller than 7. However, Fig. 8 shows that at many regions on land, high χ^2 scores (i.e., greater than 7) are achieved indicating that there have been inconsistencies between the budget terms, which have led at least one budget component to undergo an adjustment greater than its associated uncertainty standard deviation. Figures S3–S7 in the online supplemental material show the χ^2 scores averaged for each calendar month that are achieved when IMERG, GPCP, GPCC, and REGEN has been employed, respectively. Regions where incoherence exists between the budget

FIG. 8. The χ^2 scores averaged for all months over 2003–09. Values exceeding 7 correspond to areas where inconsistencies are found in the original components of the water and energy budgets (pixels in dark blue and orange). Dark blue and orange are used to highlight regions where a P dataset does and does not contribute to these inconsistencies, respectively.

terms are shown in dark blue if P is a contributor to the incoherence, and in orange otherwise [i.e., where precipitation adjustments are greater than one standard deviation (SD) of quoted uncertainty]. We can infer from these plots that, in general, precipitation is the major contributor to the incoherence between the components of the water and energy budgets. This has been previously reported in [Sahoo et al. \(2011\)](#) and [Zhang et al. \(2018\)](#) over South America and [Pan et al. \(2012\)](#) over the Amazon. It is important to highlight that incoherence attributed to P does not necessarily occur over regions where P undergoes large relative adjustments. It is only over regions where the adjustments made to P exceed one standard deviation of quoted uncertainty that P is considered incoherent with the other terms. For instance, despite the fact that MERRA-2 was the most changed dataset globally, it does not exhibit high χ^2 scores indicating that the associated uncertainties are realistic (or perhaps even underconfident) in representing errors in P fields. The plots of IMERG and GPCP in Figs. S3 and S4 respectively show high χ^2 scores in northern Australia during the wet season from December to March, indicating that they

were originally inconsistent with the other components of the water and energy budgets. Despite the large relative changes that both products have undergone over the Siberian high latitudes, no high χ^2 scores are achieved in these regions (except in May and June), indicating that the errors have been well characterized in the associated uncertainty estimates. Interestingly, all the products have reported reasonably adequate uncertainties over North Africa and the Middle East. In contrast, in Europe, despite the small relative changes exhibited by the satellite and ground-based datasets, high χ^2 scores have been achieved, indicating that the uncertainties attributed to these products have likely been too optimistic. With the exception of MERRA-2, all the products are inconsistent with the other budget terms over some parts of the Amazon throughout the year. In India, inconsistencies between the satellite P products and the other fluxes are mainly in June and October, that is, the beginning and the end of the monsoon season.

It seems reasonable to conclude that, with respect to the adequacy of uncertainty estimates to encompass likely bias in the precipitation fields, aside from MERRA-2,

FIG. 9. Regions where a P dataset had to undergo adjustments beyond its uncertainty bounds indicating that it has been originally underestimated or overestimated. Green grid cells are locations where P was found underestimated in at least one calendar month but never overestimated. Orange grid cells refer to locations where P was found overestimated in at least one calendar month but never underestimated. Orchid grid cells show regions where P exhibited different behaviors (underestimated/overestimated) in different calendar months. Beige grid cells are regions where changes applied to P do not exceed its uncertainty.

GPCC performs the best over Northern Australia, REGEN is the best over the tropics and Africa, while GPCP is the best over Siberia except in winter. All the products fail to represent biases in P estimates over the Amazon, the Brazilian highlands and India, particularly the southwestern part.

c. Highlighting regions where uncertainty ranges appear inadequate

Section 3a presents regions where P fields had to undergo an increase (or decrease) for being underestimated (overestimated). Section 3b highlights regions where the adjusted P values are large relative to the quoted uncertainty estimates. In this section, we expand on our results from section 3b by simply highlighting regions where the precipitation adjustments are greater than 1 SD of quoted uncertainty and $\chi^2 > 7$. This is

displayed in Figs. S8–S12, for IMERG, GPCP, GPCC, REGEN, and MERRA-2, respectively, where orange grid cells indicate that the quoted $P \pm 1$ SD range has been underestimated and green grid cells indicate that the quoted $P \pm 1$ SD range has been overestimated. Beige grid cells—labeled “adequate”—represent regions where the quoted $P \pm 1$ SD range encompasses the adjusted P estimate. Figures S8–S12 are summarized in Fig. 9, where we highlight (i) locations where the quoted $P \pm 1$ SD is underestimated in at least one calendar month but never overestimated (green); (ii) locations where this range is overestimated in at least one calendar month but never underestimated (orange); (iii) locations where P exhibits different behaviors (underestimated/overestimated) in different calendar months. The plots in Figs. S8–S12 reveal that much of the underestimations and overestimations in the P fields

shown in Figs. 3–7 are within our ± 1 SD bounds (in beige), and so in some sense consistent with quoted uncertainty. Cases where bias is outside these bounds are similar in the ground-based and satellite datasets and are mainly found over Europe in October where uncertainty bounds appear overly confident, and over the Siberian high latitudes in May and June, where the quoted $P \pm 1$ SD is overestimated. Also, $P \pm 1$ SD in ground-based and satellite datasets mostly overestimate precipitation at the end of the wet season (i.e., May and June) in many regions of the Amazon. As discussed in section 3b, all the biases in MERRA-2 are well within its quoted uncertainty bounds (which may be overly generous). It is therefore expected to see that among all the datasets, MERRA-2 has the largest mean coefficient of variation that we compute from the ratio of uncertainty and precipitation averaged over 2003–09. We show the mean coefficient of variation computed for each dataset in Fig. S1. The results achieved by MERRA-2 in Figs. 7–9, highlight the importance of judging a precipitation dataset by examining not only how well it represents precipitation but also how realistic its uncertainties are.

d. The adjusted precipitation estimates

All the precipitation dataset employed in this study have been corrected using in situ observations. Enforcing the closure of the coupled water and energy budgets has offered additional, out-of-sample constraint to these datasets, as a result they have all been adjusted in such a way to ensure they are coherent with the other surface water and energy budget variables and that simultaneous closure of the water and energy balance is achieved. Figure 2 and Fig. S2 show the annual mean precipitation before and after the adjustment respectively, the mean global adjustment is relatively small and led to decreasing the global annual averages of each of IMERG, GPCP, and REGEN by 0.1 mm day^{-1} . The adjustments to both GPCC and MERRA-2 have not changed the global mean P in the original estimates (i.e., 2.1 mm day^{-1}). The global average of the adjusted estimates ranges from 2 mm day^{-1} achieved by REGEN and 2.2 mm day^{-1} achieved by IMERG and GPCP.

4. Conclusions

In this work, we shed light on performance differences in IMERG, GPCP, GPCC, REGEN, and MERRA-2. We incorporate available observations of surface energy and hydrological fluxes along with the physical constraints provided by the water and energy budgets, which arguably provide one of the most robust references ever used for evaluating and comparing precipitation datasets. The performance maps that we produce throughout this

work give insights into the performance of precipitation estimates and the adequacy of their anticipated uncertainties to reflect likely bias in precipitation. At the spatial level, this has been achieved for every grid in our spatial domain including observation-limited regions.

We show that precipitation estimates present in IMERG, GPCP, GPCC, and REGEN outperform that in MERRA-2 across land. However, the uncertainty in MERRA-2 is the best in encompassing precipitation errors. It is therefore reasonable to conclude that the choice of the best precipitation dataset should be carefully considered and depends on the application, location, and the month of the year.

It is important to mention that this study examines monthly climatology and draws performance conclusions based on a particular metric. However, these results do not necessarily hold if another metric is used or different features of the precipitation datasets are assessed such as variability, intensity, or extreme events.

Acknowledgments. Sanaa Hobeichi, Gab Abramowitz, and Jason Evans acknowledge the support of the Australian Research Council Centre of Excellence for Climate Extremes (CE170100023). The authors thank the NASA MEaSUREs Program for making GRACE data available at <http://grace.jpl.nasa.gov>. They also extend their gratitude to the research communities that contributed to the development of GPCP, GPCC, IMERG, REGEN, and MERRA-2 and in making them freely available. This research was undertaken with the assistance of resources and services from the National Computational Infrastructure (NCI), which is supported by the Australian Government. The authors declare that they have no conflict of interest.

REFERENCES

- Abramowitz, G., and C. H. Bishop, 2015: Climate model dependence and the ensemble dependence transformation of CMIP projections. *J. Climate*, **28**, 2332–2348, <https://doi.org/10.1175/JCLI-D-14-00364.1>.
- Adler, R. F., C. Kidd, G. Petty, M. Morissey, and H. M. Goodman, 2001: Intercomparison of global precipitation products: The third Precipitation Intercomparison Project (PIP-3). *Bull. Amer. Meteor. Soc.*, **82**, 1377–1396, [https://doi.org/10.1175/1520-0477\(2001\)082<1377:IOGPPT>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<1377:IOGPPT>2.3.CO;2).
- , and Coauthors, 2003: The version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–present). *J. Hydrometeorol.*, **4**, 1147–1167, [https://doi.org/10.1175/1525-7541\(2003\)004<1147:TVGPCP>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2).
- , and Coauthors, 2018: The Global Precipitation Climatology Project (GPCP) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere*, **9**, 138, <https://doi.org/10.3390/atmos9040138>.
- Ashouri, H., K.-L. Hsu, S. Sorooshian, D. K. Braithwaite, K. R. Knapp, L. D. Cecil, B. R. Nelson, and O. P. Prat, 2015: PERSIANN-CDR:

- Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bull. Amer. Meteorol. Soc.*, **96**, 69–83, <https://doi.org/10.1175/BAMS-D-13-00068.1>.
- Avila, F. B., S. Dong, K. P. Menang, J. Rajczak, M. Renom, M. G. Donat, and L. V. Alexander, 2015: Systematic investigation of gridding-related scaling effects on annual statistics of daily temperature and precipitation maxima: A case study for south-east Australia. *Wea. Climate Extremes*, **9**, 6–16, <https://doi.org/10.1016/j.wace.2015.06.003>.
- Awange, J. L., V. G. Ferreira, E. Forootan, Khandu, S. A. Andam-Akorful, N. O. Agutu, and X. F. He, 2016: Uncertainties in remotely sensed precipitation data over Africa. *Int. J. Climatol.*, **36**, 303–323, <https://doi.org/10.1002/joc.4346>.
- Azarderakhsh, M., W. B. Rossow, F. Papa, H. Norouzi, and R. Khanbilvardi, 2011: Diagnosing water variations within the Amazon basin using satellite data. *J. Geophys. Res.*, **116**, D24107, <https://doi.org/10.1029/2011JD015997>.
- Balsamo, G., A. Beljaars, K. Scipal, P. Viterbo, B. van den Hurk, M. Hirschi, and A. K. Betts, 2009: A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the integrated forecast system. *J. Hydrometeorol.*, **10**, 623–643, <https://doi.org/10.1175/2008JHM1068.1>.
- , F. Pappenberger, E. Dutra, P. Viterbo, and B. van den Hurk, 2011: A revised land hydrology in the ECMWF model: A step towards daily water flux prediction in a fully-closed water cycle. *Hydrol. Processes*, **25**, 1046–1054, <https://doi.org/10.1002/hyp.7808>.
- Beaudoin, H. K., and M. Rodell, 2016: GLDAS Noah Land Surface Model L4 monthly 0.25×0.25 degree V2.1. Goddard Earth Sciences Data and Information Services Center (GES DISC), accessed 31 May 2017, <http://doi.org/10.5067/SXAVCZFAQLNO>.
- Beck, H. E., A. I. J. M. van Dijk, A. de Roo, D. G. Miralles, T. R. McVicar, J. Schellekens, and L. A. Bruijnzeel, 2016: Global-scale regionalization of hydrologic model parameters. *Water Resour. Res.*, **52**, 3599–3622, <https://doi.org/10.1002/2015WR018247>.
- , —, V. Levizzani, J. Schellekens, D. G. Miralles, B. Martens, and A. De Roo, 2017a: 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>.
- , and Coauthors, 2017b: Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrol. Earth Syst. Sci.*, **21**, 6201–6217, <https://doi.org/10.5194/hess-21-6201-2017>.
- , E. F. Wood, M. Pan, C. K. Fisher, D. G. Miralles, A. I. J. M. van Dijk, T. R. McVicar, and R. F. Adler, 2019: MSWEP V2 global 3-hourly 0.1° precipitation: Methodology and quantitative assessment. *Bull. Amer. Meteor. Soc.*, **100**, 473–500, <https://doi.org/10.1175/BAMS-D-17-0138.1>.
- Becker, A., P. Finger, A. Meyer-Christoffer, B. Rudolf, K. Schamm, U. Schneider, and M. Ziese, 2013: A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present. *Earth Syst. Sci. Data*, **5**, 71–99, <https://doi.org/10.5194/essd-5-71-2013>.
- Best, M. J., and Coauthors, 2011: The Joint UK Land Environment Simulator (JULES), model description – Part I: Energy and water fluxes. *Geosci. Model Dev.*, **4**, 677–699, <https://doi.org/10.5194/gmd-4-677-2011>.
- Bishop, C. H., and G. Abramowitz, 2013: Climate model dependence and the replicate Earth paradigm. *Climate Dyn.*, **41**, 885–900, <https://doi.org/10.1007/s00382-012-1610-y>.
- Bitew, M. M., M. Gebremichael, L. T. Ghebremichael, and Y. A. Bayissa, 2012: Evaluation of high-resolution satellite rainfall products through streamflow simulation in a hydrological modeling of a small mountainous watershed in Ethiopia. *J. Hydrometeorol.*, **13**, 338–350, <https://doi.org/10.1175/2011JHM1292.1>.
- Bosilovich, M., and Coauthors, 2015: MERRA-2: Initial evaluation of the climate. NASA Tech. Memo. NASA/TM-2015-104606, Vol. 43, 145 pp., <https://gmao.gsfc.nasa.gov/pubs/docs/Bosilovich803.pdf>.
- Brocca, L., and Coauthors, 2014: Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data. *J. Geophys. Res. Atmos.*, **119**, 5128–5141, <https://doi.org/10.1002/2014JD021489>.
- Brown, J. E. M., 2006: An analysis of the performance of hybrid infrared and microwave satellite precipitation algorithms over India and adjacent regions. *Remote Sens. Environ.*, **101**, 63–81, <https://doi.org/10.1016/j.rse.2005.12.005>.
- Burek, P., J. van der Knijff, and A. de Roo, 2013: LISFLOOD, distributed water balance and flood simulation model: Revised user manual. JRC Tech. Rep. 78917, 139 pp., <https://doi.org/10.2788/24719>.
- Bytheway, J. L., and C. D. Kummerow, 2013: Inferring the uncertainty of satellite precipitation estimates in data-sparse regions over land. *J. Geophys. Res. Atmos.*, **118**, 9524–9533, <https://doi.org/10.1002/JGRD.50607>.
- Chen, M., W. Shi, P. Xie, V. B. S. Silva, V. E. Kousky, R. Wayne 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>.
- Clark, D. B., and Coauthors, 2011: The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation dynamics. *Geosci. Model Dev.*, **4**, 701–722, <https://doi.org/10.5194/GMD-4-701-2011>.
- Contractor, S., 2018: Rainfall estimates on a gridded network based on long-term station data v1.0. NCI National Research Data Collection, accessed 15 January 2019, <https://doi.org/10.25914/5b9fa55a8298c>.
- , and Coauthors, 2020: Rainfall Estimates on a Gridded Network (REGEN)—A global land-based gridded dataset of daily precipitation from 1950 to 2016. *Hydrol. Earth Syst. Sci.*, **24**, 919–943, <https://doi.org/10.5194/hess-24-919-2020>.
- Dai, A., 2016: Historical and future changes in streamflow and continental runoff: A review. *Terrestrial Water Cycle and Climate Change: Natural and Human-Induced Impacts*, *Geophys. Monogr.*, Vol. 221, Amer. Geophys. Union, 17–37, <https://doi.org/10.1002/9781118971772.ch2>.
- Decharme, B., A. Boone, C. Delire, and J. Noilhan, 2011: Local evaluation of the Interaction between Soil Biosphere Atmosphere soil multilayer diffusion scheme using four pedotransfer functions. *J. Geophys. Res.*, **116**, D20126, <https://doi.org/10.1029/2011JD016002>.
- , E. Martin, and S. Faroux, 2013: Reconciling soil thermal and hydrological lower boundary conditions in land surface models. *J. Geophys. Res. Atmos.*, **118**, 7819–7834, <https://doi.org/10.1002/JGRD.50631>.
- Dee, D. P., 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>.

- Derber, J. C., D. F. Parrish, and S. J. Lord, 1991: The new global operational analysis system at the national meteorological center. *Wea. Forecasting*, **6**, 538–547, [https://doi.org/10.1175/1520-0434\(1991\)006<0538:TNGOAS>2.0.CO;2](https://doi.org/10.1175/1520-0434(1991)006<0538:TNGOAS>2.0.CO;2).
- Dirmeyer, P. A., and Coauthors, 2018: Verification of land-atmosphere coupling in forecast models, reanalyses, and land surface models using flux site observations. *J. Hydrometeorol.*, **19**, 375–392, <https://doi.org/10.1175/JHM-D-17-0152.1>.
- Donat, M. G., O. Angélic, and A. M. Ukkola, 2019: Intensification of precipitation extremes in the world's humid and water-limited regions. *Environ. Res. Lett.*, **14**, 065003, <https://doi.org/10.1088/1748-9326/ab1c8e>.
- Dunn, R. J. H., M. G. Donat, and L. V. Alexander, 2014: Investigating uncertainties in global gridded datasets of climate extremes. *Climate Past*, **10**, 2171–2199, <https://doi.org/10.5194/cp-10-2171-2014>.
- El Kenawy, A. M., and Coauthors, 2019: Spatial assessment of the performance of multiple high-resolution satellite-based precipitation data sets over the Middle East. *Int. J. Climatol.*, **39**, 2522–2543, <https://doi.org/10.1002/joc.5968>.
- Falcone, J. A., D. M. Carlisle, D. M. Wolock, and M. R. Meador, 2010: GAGES: A stream gage database for evaluating natural and altered flow conditions in the conterminous United States. *Ecology*, **91**, 621, <https://doi.org/10.1890/09-0889.1>.
- Flörke, M., E. Kynast, I. Bärlund, S. Eisner, F. Wimmer, and J. Alcamo, 2013: Domestic and industrial water uses of the past 60 years as a mirror of socio-economic development: A global simulation study. *Global Environ. Change*, **23**, 144–156, <https://doi.org/10.1016/j.gloenvcha.2012.10.018>.
- Funk, C. C., A. P. Verdin, J. C. Michaelsen, D. Pedreros, G. J. Husak, and P. Peterson, 2015: A global satellite assisted precipitation climatology. *Earth Syst. Sci. Data*, **8**, 401–425, <https://doi.org/10.5194/essdd-8-401-2015>.
- Gelaro, R., 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>.
- GMAO, 2015: MERRA-2 tavgM_2d_flux_Nx: 2d, Monthly mean, Time-Averaged, Single-Level, Assimilation, Surface Flux Diagnostics V5.12.4. Goddard Earth Sciences Data and Information Services Center (GES DISC), accessed 17 June 2019, <https://doi.org/10.5067/JURLVL8YV2Y4>.
- Golian, S., M. Javadian, and A. Behrangi, 2019: On the use of satellite, gauge, and reanalysis precipitation products for drought studies. *Environ. Res. Lett.*, **14**, 075005, <https://doi.org/10.1088/1748-9326/ab2203>.
- Gottschalck, J., J. Meng, M. Rodell, and P. Houser, 2005: Analysis of multiple precipitation products and preliminary assessment of their impact on global land data assimilation system land surface states. *J. Hydrometeorol.*, **6**, 573–598, <https://doi.org/10.1175/JHM437.1>.
- Grimes, D. I. F., and M. Diop, 2003: Satellite-based rainfall estimation for river flow forecasting in Africa. I: Rainfall estimates and hydrological forecasts. *Hydrol. Sci. J.*, **48**, 567–584, <https://doi.org/10.1623/hysj.48.4.567.51410>.
- Harris, I., P. D. Jones, T. J. Osborn, and D. H. Lister, 2014: Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. *Int. J. Climatol.*, **34**, 623–642, <https://doi.org/10.1002/joc.3711>.
- Harrison, L., C. Funk, and P. Peterson, 2019: Identifying changing precipitation extremes in Sub-Saharan Africa with gauge and satellite products. *Environ. Res. Lett.*, **14**, 085007, <https://doi.org/10.1088/1748-9326/ab2cae>.
- Herold, N., L. V. Alexander, M. G. Donat, S. Contractor, and A. Becker, 2016: How much does it rain over land? *Geophys. Res. Lett.*, **43**, 341–348, <https://doi.org/10.1002/2015GL066615>.
- Hobeichi, S., 2017: Derived optimal linear combination evapotranspiration. NCI National Research Data Collection, accessed 15 January 2019, <https://doi.org/10.4225/41/58980b55b0495>.
- , 2018a: Conserving Land-Atmosphere Synthesis Suite (CLASS) v 1.1. NCI National Research Data Collection, accessed 15 January 2019, <https://doi.org/10.25914/5c872258dc183>.
- , 2018b: Linear Optimal Runoff Aggregate v1.0. NCI National Research Data Collection, accessed 15 January 2019, <https://doi.org/10.25914/5b612e993d8ea>.
- , G. Abramowitz, J. Evans, and A. M. Ukkola, 2018: Derived Optimal Linear Combination Evapotranspiration (DOLCE): A global gridded synthesis ET estimate. *Hydrol. Earth Syst. Sci.*, **22**, 1317–1336, <https://doi.org/10.5194/hess-22-1317-2018>.
- , —, —, and H. E. Beck, 2019: Linear Optimal Runoff Aggregate (LORA): A global gridded synthesis runoff product. *Hydrol. Earth Syst. Sci.*, **23**, 851–870, <https://doi.org/10.5194/hess-23-851-2019>.
- , —, and —, 2020: Conserving Land-Atmosphere Synthesis Suite (CLASS). *J. Climate*, **33**, 1821–1844, <https://doi.org/10.1175/JCLI-D-19-0036.1>.
- Hong, Y., K.-L. Hsu, S. Sorooshian, and X. Gao, 2004: Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System. *J. Appl. Meteorol.*, **43**, 1834–1853, <https://doi.org/10.1175/JAM2173.1>.
- Hossain, F., and E. N. Anagnostou, 2004: Assessment of current passive-microwave-and infrared-based satellite rainfall remote sensing for flood prediction. *J. Geophys. Res.*, **109**, D07102, <https://doi.org/10.1029/2003JD003986>.
- Hou, A. Y., and Coauthors, 2014: The Global Precipitation Measurement mission. *Bull. Amer. Meteor. Soc.*, **95**, 701–722, <https://doi.org/10.1175/BAMS-D-13-00164.1>.
- Huffman, G. J., and Coauthors, 1997: The Global Precipitation Climatology Project (GPCP) combined precipitation dataset. *Bull. Amer. Meteor. Soc.*, **78**, 5–20, [https://doi.org/10.1175/1520-0477\(1997\)078<0005:TGPCPG>2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078<0005:TGPCPG>2.0.CO;2).
- , and Coauthors, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.*, **8**, 38–55, <https://doi.org/10.1175/JHM560.1>.
- , D. T. Bolvin, D. Braithwaite, K. Hsu, R. Joyce, C. Kidd, E. J. Nelkin, and P. Xie, 2015: NASA Global Precipitation Measurement (GPM) Integrated Multi-satellite Retrievals for GPM (IMERG). Algorithm Theoretical Basis Doc., version 4.5, 30 pp., http://pmm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V4.5.pdf.
- , E. F. Stocker, D. T. Bolvin, E. J. Nelkin, and J. Tan, 2019: GPM IMERG Final Precipitation L3 1 month 0.1 degree × 0.1 degree V06. Goddard Earth Sciences Data and Information Services Center (GES DISC), accessed 18 July 2019, <https://doi.org/10.5067/GPM/IMERG/3B-MONTH/06>.
- Hughes, D. A., 2006: Comparison of satellite rainfall data with observations from gauging station networks. *J. Hydrol.*, **327**, 399–410, <https://doi.org/10.1016/j.jhydrol.2005.11.041>.
- Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie, 2004: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeorol.*, **5**, 487–503, [https://doi.org/10.1175/1525-7541\(2004\)005<0487:CAMTPG>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2).
- Jung, M., 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>.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–472, [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).
- Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP-DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, **83**, 1631–1644, <https://doi.org/10.1175/BAMS-83-11-1631>.
- Kobayashi, S., and Coauthors, 2015: The JRA-55 reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48, <https://doi.org/10.2151/jmsj.2015-001>.
- L'Ecuyer, T. S., and G. L. Stephens, 2002: An estimation-based precipitation retrieval algorithm for attenuating radars. *J. Appl. Meteor.*, **41**, 272–285, [https://doi.org/10.1175/1520-0450\(2002\)041<0272:AEBPRA>2.0.CO;2](https://doi.org/10.1175/1520-0450(2002)041<0272:AEBPRA>2.0.CO;2).
- , 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>.
- Legates, D. R., and C. J. Willmott, 1990: Mean seasonal and spatial variability in gauge-corrected, global precipitation. *Int. J. Climatol.*, **10**, 111–127, <https://doi.org/10.1002/joc.3370100202>.
- Loeb, N., 2017: CERES Level 3B EBAF-Surface Terra+ Aqua netCDF file - Edition 4.0. NASA Langley Atmospheric Science Data Center DAAC, accessed 31 October 2017, https://doi.org/10.5067/terra+aquaceres/ebaf-surface_13b004.0.
- Maggioni, V., P. C. Meyers, and M. D. Robinson, 2016: A review of merged high-resolution satellite precipitation product accuracy during the Tropical Rainfall Measuring Mission (TRMM) era. *J. Hydrometeorol.*, **17**, 1101–1117, <https://doi.org/10.1175/JHM-D-15-0190.1>.
- Martens, B., D. Miralles, H. Lievens, R. Van Der Schalie, R. De Jeu, D. Fernández-Prieto, and N. Verhoest, 2016: GLEAM v3: Updated land evaporation and root-zone soil moisture datasets. *Geophysical Research Abstracts*, Vol. 18, Abstract EGU2016-4253, <https://meetingorganizer.copernicus.org/EGU2016/EGU2016-4253.pdf>.
- , 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. G., T. R. H. Holmes, R. A. M. De Jeu, J. H. Gash, A. G. C. A. Meesters, and A. J. 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>.
- Mu, Q., M. Zhao, and S. W. Running, 2011: Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sens. Environ.*, **115**, 1781–1800, <https://doi.org/10.1016/j.rse.2011.02.019>.
- Munier, S., and F. Aires, 2018: A new global method of satellite dataset merging and quality characterization constrained by the terrestrial water budget. *Remote Sens. Environ.*, **205**, 119–130, <https://doi.org/10.1016/j.rse.2017.11.008>.
- Pan, M., H. Li, and E. Wood, 2010: Assessing the skill of satellite-based precipitation estimates in hydrologic applications. *Water Resour. Res.*, **46**, W09535, <https://doi.org/10.1029/2009WR008290>.
- , A. K. Sahoo, T. J. Troy, R. K. Vinukollu, J. Sheffield, and A. E. F. Wood, 2012: Multisource estimation of long-term terrestrial water budget for major global river basins. *J. Climate*, **25**, 3191–3206, <https://doi.org/10.1175/JCLI-D-11-00300.1>.
- Reichle, R. H., Q. Liu, R. D. Koster, C. S. Draper, S. P. P. Mahanama, and G. S. Partyka, 2017: Land surface precipitation in MERRA-2. *J. Climate*, **30**, 1643–1664, <https://doi.org/10.1175/JCLI-D-16-0570.1>.
- Roca, R., L. V. Alexander, G. Potter, M. Bador, R. Jucá, S. Contractor, M. G. Bosilovich, and S. Cloché, 2019: FROGS: A daily 1° × 1° gridded precipitation database of rain gauge, satellite and reanalysis products. *Earth Syst. Sci. Data*, **11**, 1017–1035, <https://doi.org/10.5194/essd-11-1017-2019>.
- Rodell, M., and Coauthors, 2015: The observed state of the water cycle in the early twenty-first century. *J. Climate*, **28**, 8289–8318, <https://doi.org/10.1175/JCLI-D-14-00555.1>.
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1058, <https://doi.org/10.1175/2010BAMS3001.1>.
- Sahoo, A. K., M. Pan, T. J. Troy, R. K. Vinukollu, J. Sheffield, and E. F. Wood, 2011: Reconciling the global terrestrial water budget using satellite remote sensing. *Remote Sens. Environ.*, **115**, 1850–1865, <https://doi.org/10.1016/j.rse.2011.03.009>.
- Satgé, F., M. P. Bonnet, M. Gosset, J. Molina, W. Hernan Yunque Lima, R. Pillco Zolá, F. Timouk, and J. Garnier, 2016: Assessment of satellite rainfall products over the Andean Plateau. *Atmos. Res.*, **167**, 1–14, <https://doi.org/10.1016/j.atmosres.2015.07.012>.
- Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, M. Ziese, and B. Rudolf, 2014: GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theor. Appl. Climatol.*, **115**, 15–40, <https://doi.org/10.1007/s00704-013-0860-x>.
- , P. Finger, A. Meyer-Christoffer, E. Rustemeier, M. Ziese, and A. Becker, 2017: Evaluating the hydrological cycle over land using the newly-corrected precipitation climatology from the Global Precipitation Climatology Centre (GPCC). *Atmosphere*, **8**, 52, <https://doi.org/10.3390/ATMOS8030052>.
- , A. Becker, P. Finger, A. Meyer-Christoffer, and M. Ziese, 2018: GPCC full data monthly product version 2018 at 0.5°: Monthly land-surface precipitation from rain-gauges built on GTS-based and Historical Data. Global Precipitation Climatology Centre, accessed 18 June 2019, https://doi.org/10.5676/DWD_GPCC/FD_M_V2018_050.
- Shumway, R. H., and D. S. Stoffer, 2017: *Time Series Analysis and its Applications with R Examples*. Springer, 562 pp.
- Siemann, A. L., N. Chaney, and E. F. Wood, 2018: Development and validation of a long term, global, terrestrial sensible heat flux dataset. *J. Climate*, **31**, 6073–6095, <https://doi.org/10.1175/JCLI-D-17-0732.1>.
- Sirisena, T. A. J. G., S. Maskey, R. Ranasinghe, and M. S. Babel, 2018: Effects of different precipitation inputs on streamflow simulation in the Irrawaddy River Basin, Myanmar. *J. Hydrol.*, **19**, 265–278, <https://doi.org/10.1016/j.ejrh.2018.10.005>.
- Steiner, M., T. L. Bell, Y. Zhang, and E. F. Wood, 2003: Comparison of two methods for estimating the sampling-related uncertainty of satellite rainfall averages based on a large radar dataset. *J. Climate*, **16**, 3759–3778, [https://doi.org/10.1175/1520-0442\(2003\)016<3759:COTMFE>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<3759:COTMFE>2.0.CO;2).
- Suarez, M. J., and Coauthors, 2005: Documentation and Validation of the Goddard Earth Observing System (GEOS) Data Assimilation System, version 4. NASA/TM-2005-104606, Vol. 26, 187 pp., <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20050175690.pdf>.
- Sun, Q., C. Miao, Q. Duan, H. Ashouri, S. Sorooshian, and K. L. Hsu, 2018: A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Rev. Geophys.*, **56**, 79–107, <https://doi.org/10.1002/2017RG000574>.
- Sunilkumar, K., T. Narayana Rao, K. Saikranthi, and M. Purnachandra Rao, 2015: Comprehensive evaluation of multisatellite precipitation

- estimates over India using gridded rainfall data. *J. Geophys. Res. Atmos.*, **120**, 8987–9005, <https://doi.org/10.1002/2015JD023437>.
- Thiemig, V., R. Rojas, M. Zambrano-Bigiarini, V. Levizzani, and A. De Roo, 2012: Validation of satellite-based precipitation products over sparsely gauged african river basins. *J. Hydrometeor.*, **13**, 1760–1783, <https://doi.org/10.1175/JHM-D-12-032.1>.
- Tian, Y., and C. D. Peters-Lidard, 2010: A global map of uncertainties in satellite-based precipitation measurements. *Geophys. Res. Lett.*, **37**, L24407, <https://doi.org/10.1029/2010GL046008>.
- Tong, K., F. Su, D. Yang, and Z. Hao, 2014: Evaluation of satellite precipitation retrievals and their potential utilities in hydrologic modeling over the Tibetan Plateau. *J. Hydrol.*, **519**, 423–437, <https://doi.org/10.1016/j.jhydrol.2014.07.044>.
- Ushio, T., and Coauthors, 2009: A Kalman filter approach to the Global Satellite Mapping of Precipitation (GSMaP) from combined passive microwave and infrared radiometric data. *J. Meteor. Soc. Japan*, **87A**, 137–151, <https://doi.org/10.2151/jmsj.87a.137>.
- van Beek, L. P. H., and M. F. P. Bierkens, 2008: The Global Hydrological Model PCR-GLOBWB: Conceptualization, Parameterization and Verification. Utrecht University Department of Physical Geography Tech. Rep., 53 pp., <http://vanbeek.geo.uu.nl/supinfo/vanbeekbierkens2009.pdf>.
- van der Knijff, J. M., J. Younis, and A. P. J. De Roo, 2010: LISFLOOD: A GIS-based distributed model for river basin scale water balance and flood simulation. *Int. J. Geogr. Inf. Sci.*, **24**, 189–212, <https://doi.org/10.1080/13658810802549154>.
- van Dijk, A., and G. Warren, 2010: Landscape model (version 0.5) evaluation against observations. AWRA Tech. Rep. 4, WIRADA/CSIRO Water for a Healthy Country Flagship, 100 pp.
- van Dijk, A. I. J. M., L. J. Renzullo, Y. Wada, and P. Tregoning, 2014: A global water cycle reanalysis (2003–2012) merging satellite gravimetry and altimetry observations with a hydrological multi-model ensemble. *Hydrol. Earth Syst. Sci.*, **18**, 2955–2973, <https://doi.org/10.5194/hess-18-2955-2014>.
- Wang, C., G. Tang, Z. Han X. Guo, and Y. Hong, 2018: Global intercomparison and regional evaluation of GPM IMERG Version-03, Version-04 and its latest Version-05 precipitation products: Similarity, difference and improvements. *J. Hydrol.*, **564**, 342–356, <https://doi.org/10.1016/j.jhydrol.2018.06.064>.
- Ward, E., W. Buytaert, L. Peaver, and H. Wheeler, 2011: Evaluation of precipitation products over complex mountainous terrain: A water resources perspective. *Adv. Water Resour.*, **34**, 1222–1231, <https://doi.org/10.1016/j.advwatres.2011.05.007>.
- Watkins, M. M., D. N. Wiese, D. N. Yuan, C. Boening, and F. W. Landerer, 2015: Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. *J. Geophys. Res. Solid Earth*, **120**, 2648–2671, <https://doi.org/10.1002/2014JB011547>.
- Weedon, G. P., G. Balsamo, N. Bellouin, S. Gomes, M. J. Best, and P. Viterbo, 2014: The WFDEI meteorological forcing data set: WATCH forcing data methodology applied to ERA-Interim reanalysis data. *Water Resour. Res.*, **50**, 7505–7514, <https://doi.org/10.1002/2014WR015638>.
- Wu, H., R. F. Adler, Y. Hong, Y. Tian, and F. Policelli, 2012: Evaluation of global flood detection using satellite-based rainfall and a hydrologic model. *J. Hydrometeor.*, **13**, 1268–1284, <https://doi.org/10.1175/JHM-D-11-087.1>.
- Xie, P., and P. A. 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](https://doi.org/10.1175/1520-0477(1997)078<2539:GPAYMA>2.0.CO;2).
- Zhan, W., K. Guan, J. Sheffield, and E. F. Wood, 2016: Depiction of drought over sub-Saharan Africa using reanalyses precipitation data sets. *J. Geophys. Res. Atmos.*, **121**, 10 555–10 574, <https://doi.org/10.1002/2016JD024858>.
- Zhang, X., and E. N. Anagnostou, 2019: Evaluation of numerical weather model-based satellite precipitation adjustment in tropical mountainous regions. *J. Hydrometeor.*, **20**, 431–445, <https://doi.org/10.1175/JHM-D-18-0008.1>.
- Zhang, Y., and Coauthors, 2016: Multi-decadal trends in global terrestrial evapotranspiration and its components. *Sci. Rep.*, **6**, 19124, <https://doi.org/10.1038/srep19124>.
- Zhang, Y., and Coauthors, 2018: A Climate Data Record (CDR) for the global terrestrial water. *Earth Syst. Sci.*, **22**, 241–263, <https://doi.org/10.5194/hess-22-241-2018>.
- Zia, A., T. Kirubarajan, J. P. Reilly, D. Yee, K. Punithakumar, and S. Shirani, 2008: An EM algorithm for nonlinear state estimation with model uncertainties. *IEEE Trans. Signal Process.*, **56**, 921–936, <https://doi.org/10.1109/TSP.2007.907883>.