

# MSWEP V2 GLOBAL 3-HOURLY 0.1° PRECIPITATION

## Methodology and Quantitative Assessment

HYLKE E. BECK, ERIC F. WOOD, MING PAN, COLBY K. FISHER, DIEGO G. MIRALLES,  
ALBERT I. J. M. VAN DIJK, TIM R. McVICAR, AND ROBERT F. ADLER

MSWEP V2 is the first fully global precipitation dataset with a 0.1° resolution derived by optimally merging a range of gauge, satellite, and reanalysis estimates.

Precipitation  $P$  drives the terrestrial hydrological cycle (Oki and Kanae 2006; Trenberth et al. 2007). It is also among the most difficult meteorological variables to estimate because of its high spatiotemporal heterogeneity (Daly et al. 1994; Adler et al. 2001; Roe 2005; Stephens et al. 2010; Herold

et al. 2016; Prein and Gobiet 2017). A plethora of regional, quasi-global, and fully global gridded  $P$  datasets have been developed over the past decades (for an overview, see Maggioni et al. 2016; Beck et al. 2017c; Levizzani et al. 2018; Sun et al. 2018; <http://ipwg.isac.cnr.it>; <http://reanalyses.org>). These datasets differ in terms of design objective (instantaneous accuracy, temporal homogeneity, record length, or combinations thereof), data source (gauge, ground radar, satellite, analysis, reanalysis, or combinations thereof), spatial resolution (from 0.05° to 2.5°), and temporal resolution (30 min to monthly).

Multi-Source Weighted-Ensemble Precipitation (MSWEP) is a recently released global  $P$  dataset with a 3-hourly temporal resolution, covering the period 1979 to the near present (Beck et al. 2017b). The dataset is unique in that it takes advantage of the complementary strengths of gauge-, satellite-, and reanalysis-based data to provide reliable  $P$  estimates over the entire globe. Since the release of version 1 (0.25° spatial resolution) in May 2016, MSWEP has been successfully applied at global scales for a variety of purposes, such as modeling soil moisture and evaporation (Martens et al. 2017), estimating plant rooting depth (Yang et al. 2016), water resources reanalysis (Schellekens et al. 2017), and evaluating climatic controls on vegetation (Papagiannopoulou

**AFFILIATIONS:** BECK, WOOD, PAN, AND FISHER—Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey; MIRALLES—Laboratory of Hydrology and Water Management, Ghent University, Ghent, Belgium; VAN DIJK—Fenner School of Environment and Society, Australian National University, Canberra, Australian Capital Territory, Australia; McVICAR—CSIRO Land and Water, Canberra, Australian Capital Territory, and Australian Research Council Centre of Excellence for Climate System Science, Sydney, New South Wales, Australia; ADLER—Earth System Science Interdisciplinary Center, University of Maryland, College Park, College Park, Maryland

**CORRESPONDING AUTHOR:** Hylke Beck,  
[hylke.beck@gmail.com](mailto:hylke.beck@gmail.com)

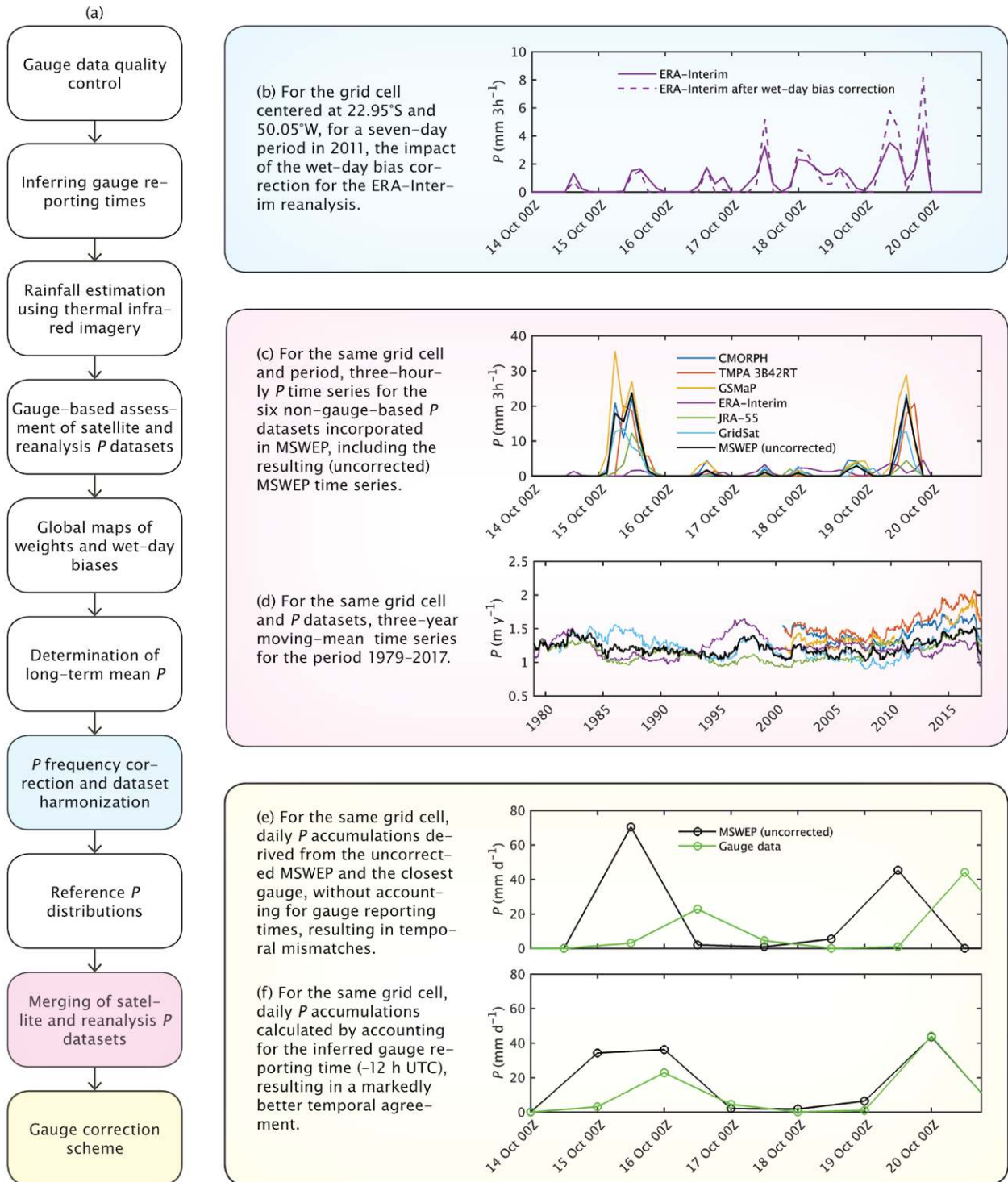
*The abstract for this article can be found in this issue, following the table of contents.*

DOI:10.1175/BAMS-D-17-0138.1

In final form 1 September 2018

©2019 American Meteorological Society

For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#).



**FIG. 1. (a) Flowchart** outlining the main processing steps implemented to produce MSWEP V2. See the appendix for detailed descriptions of each step. **(b) Example of the wet-day bias correction for ERA-Interim.** **(c),(d)** Time series of the satellite and reanalysis  $P$  datasets and MSWEP V2. **(e),(f)** Illustration of the importance of accounting for reporting times when applying gauge corrections.

et al. 2017a,b). MSWEP has also been successfully used for several purposes regionally, for example, to analyze diurnal variations in rainfall (Chen and Dirmeyer 2017; Chen et al. 2017), investigate lake dynamics (Satgé et al. 2017; Wang et al. 2017), evaluate root-zone soil moisture patterns (Zohaib et al. 2017), and drive a dynamic ecohydrological model (Liu et al. 2016). In addition, MSWEP has been included in at least four regional  $P$  dataset evaluation studies focusing, respectively, on the Amazon (Correa et al. 2017), Chile (Zambrano-Bigiarini et al. 2017), India (Nair and Indu 2017), and the Sahel (Zhang et al. 2017).

Since the release of MSWEP version 1 (MSWEP V1), considerable improvements were implemented, resulting in MSWEP version 2 (MSWEP V2), the focus of the present study. Improvements include i) the introduction of cumulative distribution function (CDF) and  $P$  frequency corrections, to account for spurious drizzle, attenuated peaks, and temporal discontinuities evident in version 1 (Nair and Indu 2017; Zhang et al. 2017); ii) increasing the spatial resolution from  $0.25^\circ$  to  $0.1^\circ$  to increase the local relevance of the  $P$  estimates (especially important for high-water-yield mountainous regions); iii) the inclusion of ocean areas to enable oceanic studies and terrestrial hydrology studies for coastal areas and small islands; iv) the addition of  $P$  estimates derived from Gridded Satellite (GridSat) thermal infrared (IR) imagery (Knapp et al. 2011) for the pre-TRMM era to supplement the reanalysis and gauge data; v) the use of a daily gauge correction scheme that accounts for regional differences in reporting times, to minimize timing mismatches when applying the daily gauge corrections; vi) the use of a large database of daily gauge observations compiled from several sources to replace the  $0.5^\circ$  Climate Prediction Center (CPC) unified dataset (Xie et al. 2007; Chen et al. 2008); and vii) extension of the data record to 2017 (MSWEP V1 finished in 2015).

MSWEP V2 is the first fully global  $P$  dataset with a spatial resolution of  $0.1^\circ$  (11 km at the equator), supporting global-scale land-surface modeling at hyperresolution (Wood et al. 2011; Bierkens et al. 2015). Other  $P$  datasets with a high spatial resolution ( $\leq 0.1^\circ$ ) include Climate Hazards Group Infrared Precipitation with Stations (CHIRPS;  $0.05^\circ$ ; Funk et al. 2015b), CPC morphing technique (CMORPH;  $0.07^\circ$ ; Joyce et al. 2004), Global Satellite Mapping of Precipitation (GSMaP;  $0.1^\circ$ ; Ushio et al. 2009; Mega et al. 2014), Integrated Multisatellite Retrievals for Global Precipitation Measurement (IMERG;  $0.1^\circ$ ; Huffman et al. 2014), and Precipitation Estimation from Remotely Sensed Information Using Artificial

Neural Networks–Cloud Classification System (PERSIANN-CCS;  $0.04^\circ$ ; Hong et al. 2004). However, these datasets are limited to latitudes  $\leq 60^\circ\text{N/S}$  ( $\leq 50^\circ\text{N/S}$  for CHIRPS), do not take advantage of river discharge  $Q$  observations for bias correction, and do not incorporate reanalysis-based  $P$  estimates (with the arguable exception of CHIRPS, which uses them to temporally disaggregate from 5-day to daily estimates). Additionally, none of these datasets apply  $P$  gauge corrections at the daily time scale, with the exception of GSMaP, although it fails to account for differences in gauge reporting times. Moreover, CHIRPS and PERSIANN-CCS do not integrate passive microwave-based  $P$  retrievals, and the daily temporal resolution of CHIRPS renders it less suitable in highly dynamic  $P$  environments. Finally, with the exception of CHIRPS, these datasets span  $\leq 20$  years, which is less optimal to assess long-term hydrological changes/trends (Weatherhead et al. 1998).

Here, we describe the data and methodology underlying MSWEP V2, evaluate the performance of the dataset for the conterminous United States (CONUS), and assess spatiotemporal  $P$  patterns globally.

**DATA AND METHODS. MSWEP V2 methodology.** Figure 1a flowcharts the main processing steps implemented to produce MSWEP V2. The complete methodology is provided in the appendix. The main steps can be summarized as follows:

- 1) Daily  $P$  gauge observations were used for three purposes: i) to determine the merging weights for the six 3-hourly non-gauge-based  $P$  datasets incorporated in MSWEP V2 [CMORPH, European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis (ERA-Interim), GridSat, GSMaP, Japanese 55-year Reanalysis (JRA-55), and Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) 3B42RT; see Table 1 for details on the datasets], ii) to calculate the wet-day biases for the reanalyses (ERA-Interim and JRA-55), and iii) to correct the  $P$  estimates near gauge stations. Initially, 117,759  $P$  gauges were compiled from various global and national databases. Extensive quality control was applied, for example, to remove erroneous zeros frequently present in records from the Global Summary of the Day (GSOD) database (<https://data.noaa.gov>; Fig. 2a). After quality control, a final gauge dataset composed of 76,747 gauges remained (Fig. 2b). See “Gauge data quality control” appendix section for details.

**TABLE 1. Overview of the gridded *P* datasets incorporated in MSWEP V2 and used for comparison. Abbreviations: gauge (G); satellite (S); reanalysis (R); radar (N); near-real time (NRT). In the spatial coverage column, “global” indicates fully global coverage including ocean areas, whereas “land” indicates that the coverage is restricted to the land surface. MSWEP V2 has been added for the sake of completeness. Symbols specify dataset availability: \* indicates it is available until the present with a delay of several hours; \*\* indicates it is available until the present with a delay of several days; \*\*\* indicates it is available until the present with a delay of several months.**

Name	Details	Data source(s)	Spatial resolution	Spatial coverage	Temporal resolution	Temporal coverage	Reference(s)
<b>Datasets incorporated in MSWEP V2</b>							
CMORPH	CPC morphing technique (CMORPH) V1.0 and V0.x	S	0.07°	≤60°N/S	30 min	1998–NRT*	Joyce et al. (2004)
Daily gauge data	Compiled from GHCN-D, GSOD, and other sources	G	—	Land	Daily	1979–2017	This study (“Gauge correction scheme” in appendix)
ERA-Interim	European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis	R	~80 km	Global	3 hourly	1979–NRT***	Dee et al. (2011)
GPCC FDR	Global Precipitation Climatology Centre (GPCC) Full Data Reanalysis (FDR) V7 extended using First Guess	G	0.5°/1°	Land	Monthly	1951–NRT**	U. Schneider et al. (2014, 2017)
GridSat	Derived from the Gridded Satellite (GridSat) B1 infrared archive V02R01 using CDF matching	S	0.1°	<~70°N/S	3 hourly	1980–2016	Knapp et al. (2011); this study (“Rainfall estimation using thermal infrared imagery” in appendix)
GSMaP	Global Satellite Mapping of Precipitation (GSMaP) Moving Vector with Kalman (MVK) standard V5 supplemented with V6	S	0.1°	≤60°N/S	Hourly	2000–NRT**	Ushio et al. (2009)
JRA-55	Japanese 55-year Reanalysis (JRA-55)	R	~60 km	Global	3 hourly	1959–NRT**	Kobayashi et al. (2015)
TPMA 3B42RT	TRMM Multi-satellite Precipitation Analysis (TPMA) 3B42RT V7	S	0.25°	≤50°N/S	3 hourly	2000–NRT*	Huffman et al. (2007)
WorldClim	WorldClim V2.0 monthly climatic dataset, corrected for gauge-undercatch and orographic effects	G	~1 km	Global	Monthly	Climatic	Fick and Hijmans (2017); this study (“Determination of long-term mean <i>P</i> ” in appendix)

- Information about gauge reporting times is crucial to avoid timing mismatches when applying daily gauge corrections but is generally not provided. We developed a procedure to infer reporting times for all gauges based on correlations with four non-gauge-based *P* datasets (CMORPH, ERA-Interim, GSMaP, and JRA-55). See “Inferring gauge reporting times” appendix section for details.
- MSWEP V1 relied entirely on reanalysis and gauge data during the pre-TRMM era (prior to 2000; Beck et al. 2017b). For MSWEP V2, we supplemented the

reanalysis and gauge data during the pre-TRMM era with rainfall estimates based on IR data from the GridSat B1 archive (0.07° resolution; Knapp et al. 2011) to improve the *P* estimates in convection-dominated regions. Rainfall was estimated using a parsimonious CDF-matching approach. See “Rainfall estimation using thermal infrared imagery” in the appendix for details.

- To assess the individual performance of the six non-gauge-based *P* datasets incorporated in MSWEP V2, we calculated, for each of the 76,747

TABLE 1. Continued.							
Name	Details	Data source(s)	Spatial resolution	Spatial coverage	Temporal resolution	Temporal coverage	Reference(s)
<b>Datasets used for comparison</b>							
CHIRPS2.0	Climate Hazards group Infrared Precipitation with Stations (CHIRPS) V2.0	G, S, R	0.05°	Land, ≤50°N/S	Daily	1981–NRT**	Funk et al. (2015b)
CMAPI707	CPC Merged Analysis of Precipitation (CMAP) V1707	G, S, R	2.5°	Global	5 days	1979–2017	Xie and Arkin (1997)
GPCC2015	Global Precipitation Climatology Centre (GPCC) Climatology V2015	G	0.25°	Land	Monthly	1951–2000	U. Schneider et al. (2014, 2017)
GPCP2.3	Global Precipitation Climatology Project (GPCP) monthly analysis product V2.3	G, S	2.5°	Global	Monthly	1996–NRT**	Adler et al. (2003, 2018)
HOAPS3.2	Hamburg Ocean Atmosphere Parameters and fluxes from Satellite data (HOAPS) V3.2	S	0.5°	Ocean	6 hourly	1987–2008	Andersson et al. (2010)
MERRA-2	Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2)	G, R	~50 km	Global	Hourly	1980–NRT***	Reichle et al. (2017)
MSWEP V1	Multi-Source Weighted-Ensemble Precipitation (MSWEP) V1	G, S, R	0.25°	Land	3 hourly	1979–2015	Beck et al. (2017b)
MSWEP V2	Multi-Source Weighted-Ensemble Precipitation (MSWEP) V2	G, S, R	0.1°	Global	3 hourly	1979–2017	This study
Stage IV	Stage-IV gauge-adjusted, radar-based dataset	G, N	~5 km	CONUS	Hourly	2002–NRT*	Lin and Mitchell (2005)

gauges, Pearson correlation coefficients between 3-day mean gauge and gridded  $P$  time series ( $r_{3\text{day}}$ ). In addition, since reanalyses tend to consistently overestimate the  $P$  frequency and underestimate the intensity (Zolina et al. 2004; Sun et al. 2006; Lopez 2007; Stephens et al. 2010; Skok et al. 2015; Herold et al. 2016), for ERA-Interim and JRA-55, we calculated the bias in the number of wet days per year, using the gauge observations as reference, according to

$$\beta_{\text{WD}} = \frac{\text{WD}_{\text{gridded}}}{\text{WD}_{\text{gauge}}}, \quad (1)$$

where  $\beta_{\text{WD}}$  (unitless) is the bias in number of wet days and  $\text{WD}_{\text{gridded}}$  and  $\text{WD}_{\text{gauge}}$  represent the mean annual number of wet days in the reanalysis and the gauge observations, respectively.  $\text{WD}_{\text{gridded}}$  was computed from daily accumulations to be consistent with the gauge observations. Wet days were identified using a 0.5 mm day<sup>-1</sup> threshold, similar to several previous studies (e.g., Akinremi

et al. 1999; Haylock et al. 2008; Driouech et al. 2009; Trenberth and Zhang 2018). See “Gauge-based assessment of satellite and reanalysis  $P$  datasets” appendix section for details.

- Global weight maps were derived for each of the six non-gauge-based  $P$  datasets incorporated in MSWEP V2 based on the  $r_{3\text{day}}$  values calculated in the preceding step. The  $r_{3\text{day}}$  values were squared to yield the coefficient of determination and subsequently interpolated to yield gap-free global weight maps. Similarly, gap-free global maps of  $\beta_{\text{WD}}$  were produced for the reanalyses to correct the  $P$  frequency prior to the data merging. See “Global maps of weights and wet-day biases” in the appendix for details.
- MSWEP V1 used Climate Hazards Group’s precipitation climatology (CHPclim; 0.05° resolution; Funk et al. 2015a) to determine the long-term mean over the land surface. For MSWEP V2, we used WorldClim (1-km resolution; Fick and Hijmans 2017) because of the better  $P$  gauge



coverage. Systematic  $P$  underestimation over land due to gauge undercatch and orographic effects was corrected similarly to MSWEP V1, by inferring the “true”  $P$  using river discharge  $Q$  observations. See “Determination of long-term mean  $P$ ” appendix section for details.

- 7) To correct the  $P$  frequency of the reanalyses, we subtracted, for each grid cell, a small amount of  $P$  calculated using the interpolated  $\beta_{WD}$  values from step 5 (Fig. 1b). In addition, the six non-gauge-based  $P$  datasets incorporated in MSWEP V2 were resampled to  $0.1^\circ$  and rescaled to minimize the presence of temporal discontinuities after merging. See “ $P$  frequency correction and dataset harmonization” appendix section for details.
- 8) Three-hourly reference  $P$  distributions were calculated by weighted averaging of the distributions of five non-gauge-based  $P$  datasets (CMORPH, ERA-Interim, GSMaP, JRA-55, and TMPA 3B42RT) using the interpolated weight maps from step 5. See “Reference  $P$  distributions” in the appendix for details.
- 9) The six non-gauge-based  $P$  datasets incorporated in MSWEP V2 were merged for every possible  $P$  dataset combination by weighted averaging using the interpolated weight maps from step 5. The merged  $P$  estimates of each dataset combination were subsequently CDF matched to the reference  $P$  distributions derived in step 8, after which we selected, for each 3-hourly time step and  $0.1^\circ$  grid cell, the merged and CDF-corrected  $p$  value from the dataset combination with the highest cumulative weight (Figs. 1c and 1d). The CDF matching corrects the spurious drizzle and attenuated peaks and ensures that temporal transitions from one dataset combination to another are largely unnoticeable. See “Merging of satellite and reanalysis  $P$  datasets” in the appendix for details.
- 10) The 3-hourly merged  $P$  estimates were corrected using daily and monthly  $P$  gauge observations through a multiplicative approach. For each grid cell, we looped over the five closest gauges and corrected the 3-hourly merged  $P$  data at the daily time scale. When applying the daily corrections, we accounted for the gauge reporting times derived in step 2 to reduce temporal mismatches (Figs. 1e and 1f). We subsequently applied monthly gauge corrections using the Global Precipitation Climatology Centre (GPCC) Full Data Reanalysis, version 7 (FDR7), dataset ( $0.5^\circ$  resolution; U. Schneider et al. 2014), which incorporates a more extensive collection of gauges, following the same procedure but without accounting for

gauge reporting times, to yield the final gauge-corrected MSWEP V2. See “Gauge correction scheme” in the appendix for details.

**Evaluation using Stage IV gauge–radar data for the CONUS.** We evaluated the performance of MSWEP V2 and, for the sake of comparison, MSWEP V1, a widely used satellite-based dataset (CMORPH), a widely used reanalysis (ERA-Interim), and a state-of-the-art reanalysis corrected using daily gauge observations [Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2); Table 1]. The evaluation was performed at a 3-hourly temporal and  $0.1^\circ$  spatial resolution for 2002–15. As reference, we used the National Centers for Environmental Prediction (NCEP) Stage IV dataset (Lin and Mitchell 2005), which has a  $0.04^\circ$  spatial and hourly temporal resolution and merges data from 140 radars and ~5,500 gauges for the CONUS. Stage IV provides high-quality  $P$  estimates and has therefore been widely used as a reference for the evaluation of  $P$  datasets (e.g., Hong et al. 2006; Habib et al. 2009; AghaKouchak et al. 2011, 2012; Zhang et al. 2018). To reduce systematic biases, the Stage IV dataset was rescaled such that its long-term mean matches that of the Parameter-Elevation Regressions on Independent Slopes Model (PRISM) dataset (Daly et al. 2008) for the evaluation period (2002–15).

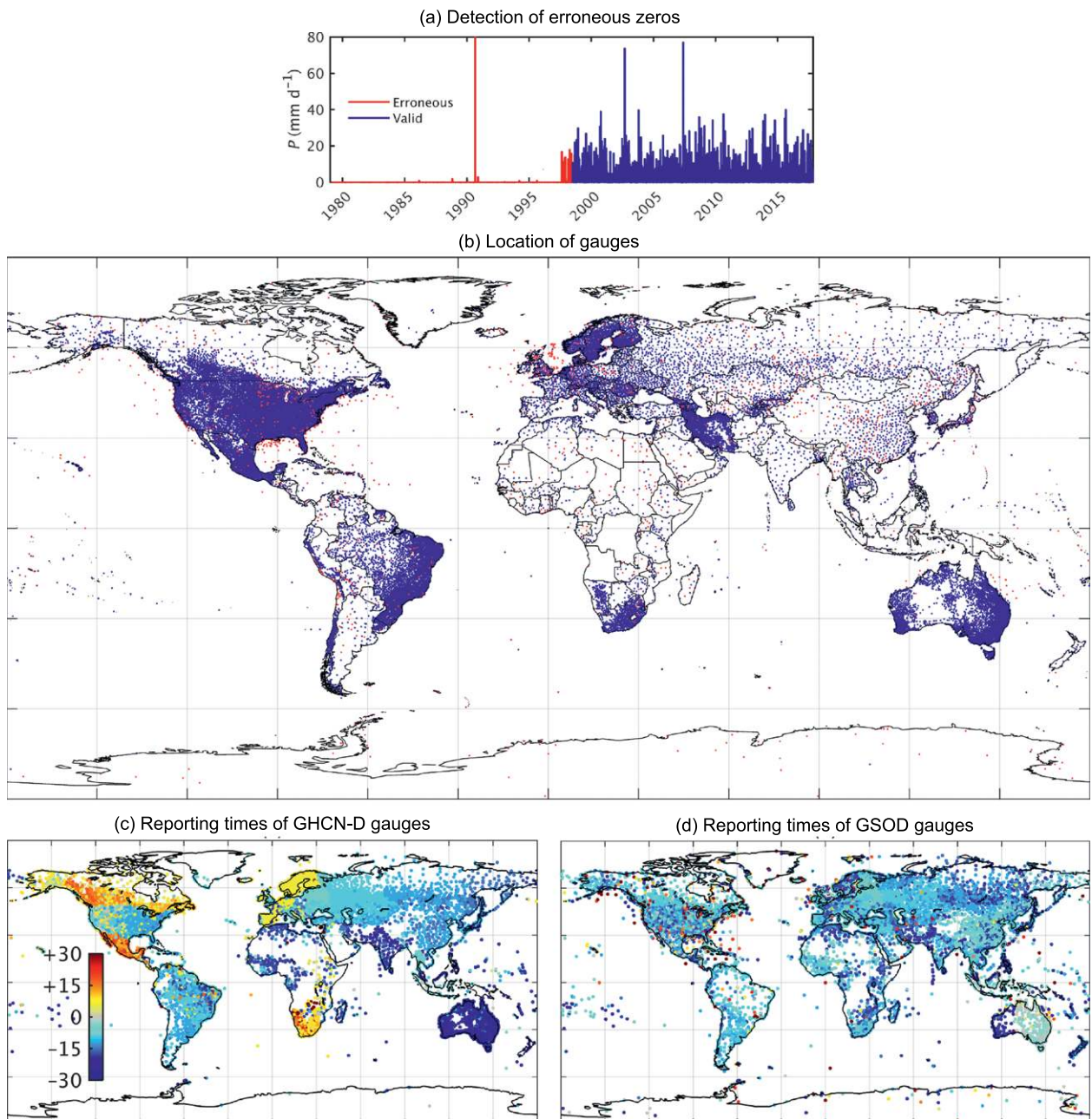
As a performance metric, we used the Kling–Gupta efficiency (KGE; Gupta et al. 2009; Kling et al. 2012), an objective performance metric combining correlation, bias, and variability, introduced in Gupta et al. (2009) and modified in Kling et al. (2012). The KGE is calculated as follows:

$$\text{KGE} = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2}, \quad (2)$$

where the correlation component  $r$  is represented by the (Pearson’s) coefficient of correlation, the bias component  $\beta$  by the ratio of estimated and observed means, and the variability component  $\gamma$  by the ratio of the estimated and observed coefficients of variation:

$$\beta = \frac{\mu_s}{\mu_o} \text{ and } \gamma = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o}, \quad (3)$$

where  $\mu$  and  $\sigma$  are the distribution mean and standard deviation, respectively, and the subscripts  $s$  and  $o$  indicate estimate and reference, respectively. Three-hourly accumulations were calculated for the  $P$  datasets with a temporal resolution  $<3$  h (CMORPH, MERRA-2, and Stage IV). The  $P$  datasets with a spatial resolution  $>0.1^\circ$  (MSWEP V1, ERA-Interim, and MERRA-2) were downscaled to  $0.1^\circ$  using nearest neighbor, while the dataset with a spatial resolution

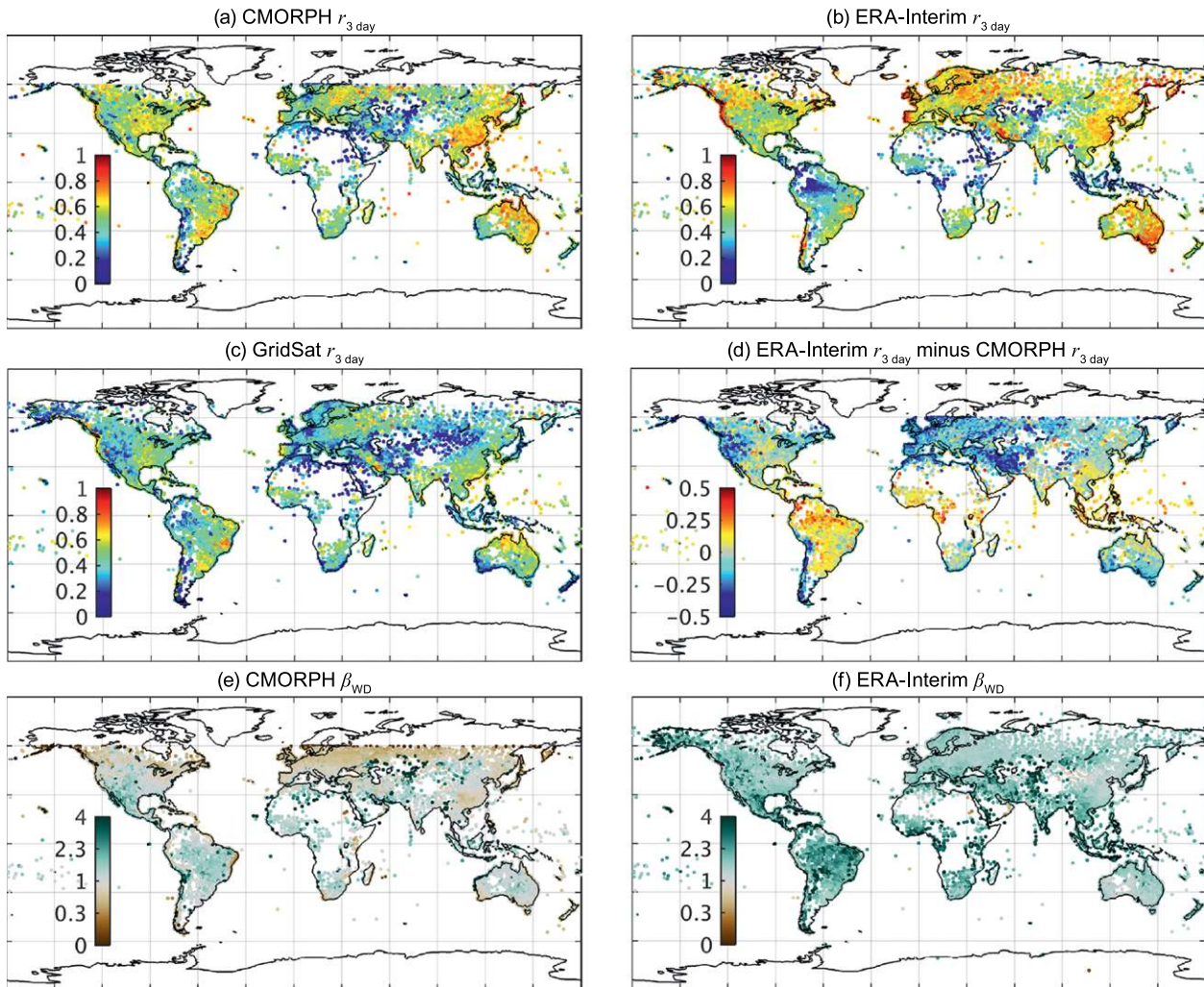


**FIG. 2.** (a) Daily  $P$  measured at GSOD station 038660 ( $50.58^{\circ}\text{N}$ ,  $1.30^{\circ}\text{W}$ ) with the automatically detected erroneous zeros indicated in red. (b) The gauges used to produce MSWEP V2 in blue ( $n = 76,747$ ) and the gauges that did not pass the quality control in red ( $n = 4,300$ ). Also shown are the inferred reporting times (expressed in h UTC) for gauges from the (c) GHCN-D and (d) GSOD databases. A reporting time of +6 h UTC, for example, means that the daily gauge accumulations represent the 24-h period starting at 0600 UTC of the current day and ending at 0600 UTC of the next day.

$<0.1^{\circ}$  (CMORPH) was upscaled to  $0.1^{\circ}$  using bilinear interpolation.

**RESULTS AND DISCUSSION.** *Gauge reporting times.* For the Global Historical Climatology Network-Daily (GHCN-D) database, we found marked differences in reporting times between neighboring

countries (e.g., between Canada and the United States and between Portugal and Spain) and sometimes within countries (e.g., Mexico, Namibia, and South Africa; Fig. 2c), reflecting differences in reporting practices among hydrological and meteorological agencies. Our reporting times correspond well with published times available for Australia (Viney and Bates 2004),



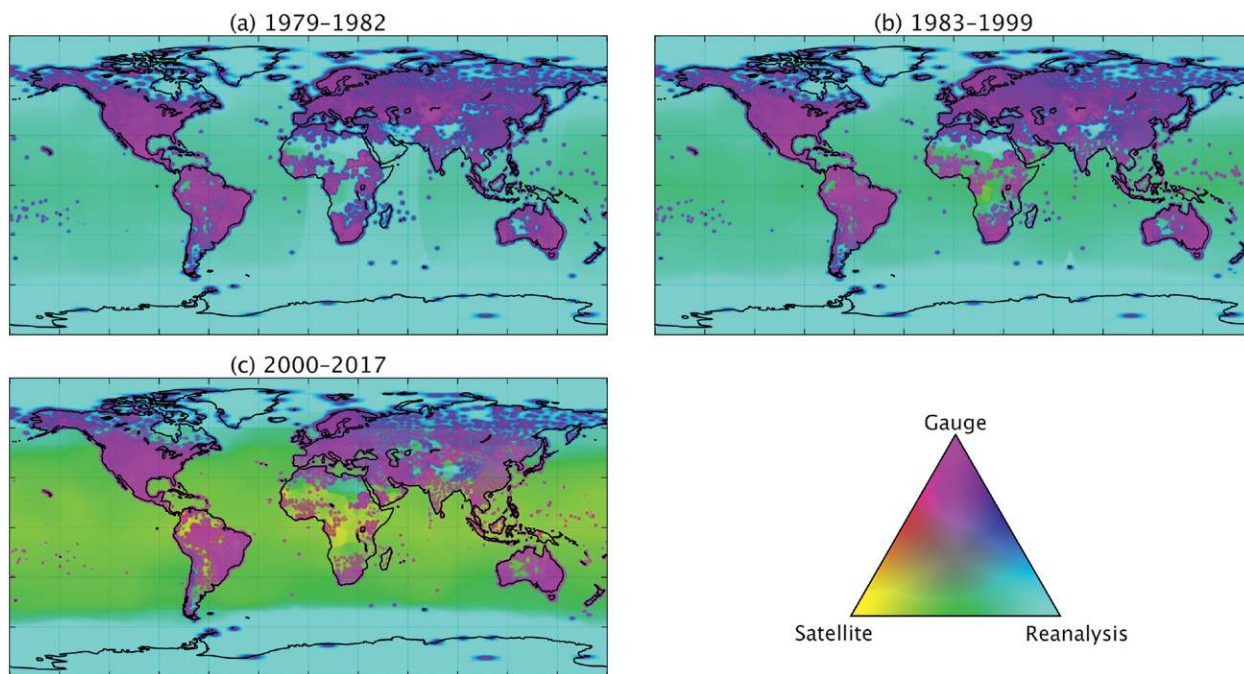
**FIG. 3.** Temporal correlations  $r_{3\text{day}}$  for (a) CMORPH, (b) ERA-Interim, and (c) GridSat. (d) The difference in  $r_{3\text{day}}$  values between CMORPH and ERA-Interim. Also shown is the bias in the number of wet days  $\beta_{\text{WD}}$  for (e) CMORPH and (f) ERA-Interim (note the nonlinear color scale). The results for CMORPH and ERA-Interim are representative of the other satellite and reanalysis datasets, respectively. CMORPH is limited to latitudes  $\leq 60^\circ$ . Each data point represents a gauge.

Brazil (Liebmann and Allured 2005), the eastern CONUS (DeGaetano 2000), India (Yatagai et al. 2012), the Netherlands (Holleman 2006), and Japan (Yatagai et al. 2012). Although the GSOD gauges represent automated gauges with reporting times officially at around 2400 UTC (Menne et al. 2012), our analysis yielded considerably earlier reporting times averaging at around  $-12$  h (UTC) (except for eastern Australia; Fig. 2d). A potential explanation for this discrepancy could be that satellites represent radiation from an atmospheric column rather than  $P$  that has reached the surface. However, Villarini and Krajewski (2007) obtained timing differences ranging from 30 to 90 min for TMPA 3B42 using 5-min rain gauge data for a single  $0.25^\circ$  grid cell in Oklahoma, suggesting that this explanation is insufficient to account for the full 12-h

difference. Additionally, the differences are also found in high latitudes ( $>60^\circ\text{N/S}$ ), where the reporting times were inferred using reanalysis data. An alternative, more likely explanation is that the daily GSOD values incorporate a significant portion of  $P$  from the previous day. Overall, these results highlight the importance of accounting for reporting times in time-critical applications relying on daily gauge observations.

**Gauge-based assessment of satellite and reanalysis datasets.** Figures 3a and 3b present  $r_{3\text{day}}$  (temporal correlation) values obtained for CMORPH and ERA-Interim, respectively. Since the results were very similar for all satellite datasets (with the exception of GridSat) and for all reanalysis datasets, we only present results for one dataset of each kind. ERA-Interim is most





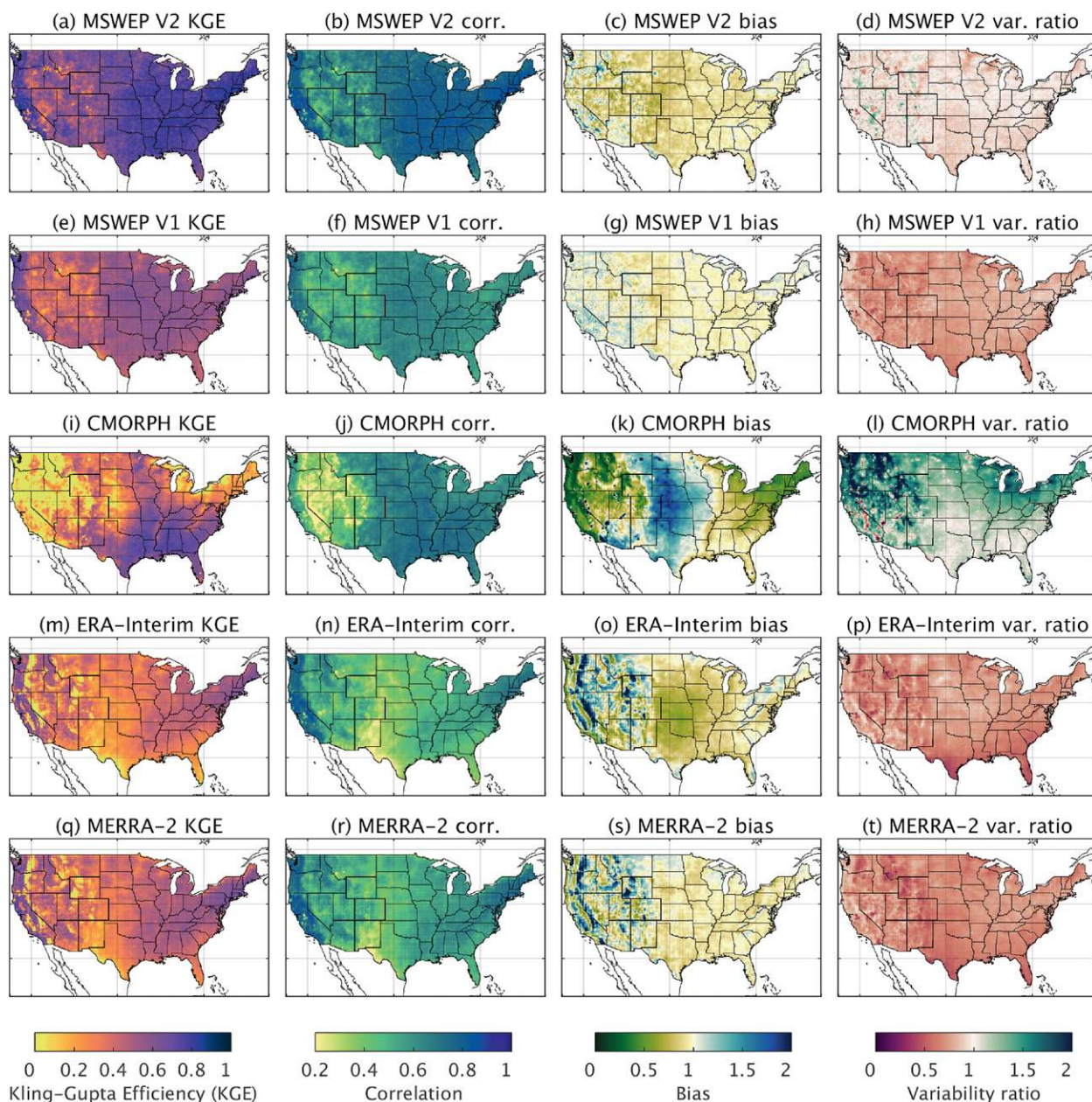
**FIG. 4.** Relative weights assigned to the gauge-, satellite-, and reanalysis-based  $P$  estimates shown using a barycentric color map for the periods (a) 1979–82, (b) 1983–99, and (c) 2000–17. The weights represent averages over the respective periods. The satellite and reanalysis weights represent cumulative weights assigned to the respective satellite and reanalysis  $P$  datasets.

skillful in mid- and high-latitude coastal regions in the path of the prevailing westerlies (notably along the Pacific coast of North America, in southern Chile, and in western Europe; Fig. 3a), whereas CMORPH performs best in moist midlatitude regions with mild winters (e.g., the southeastern United States, eastern South America, and eastern China; Fig. 3b). When we calculate the difference in  $r_{3\text{day}}$  values between the datasets, a clear picture emerges: CMORPH consistently performs better at low latitudes and ERA-Interim at high latitudes (Fig. 3d). These results underscore the long-recognized but sometimes overlooked complementary  $P$  estimation performance of satellites and weather models (e.g., Janowiak 1992; Huffman et al. 1995; Xie and Arkin 1997; Adler et al. 2001; Ebert et al. 2007; Massari et al. 2017). MSWEP is the only  $P$  dataset besides CMAP (Xie and Arkin 1997) to exploit this complementary relationship.

Figure 3c presents  $r_{3\text{day}}$  values for the GridSat IR-based rainfall dataset, which has been produced to complement the gauge and reanalysis data during the pre-TRMM era (“Rainfall estimation using thermal infrared imagery” appendix section). The  $r_{3\text{day}}$  values for GridSat are consistently lower than those obtained for CMORPH (Fig. 3a), which was expected since cloud-top IR brightness temperatures are only indirectly related to surface rainfall (Adler and Negri 1988; Vicente et al. 1998; Scofield and Kuligowski

2003). Compared to  $r_{3\text{day}}$  values obtained using the IR-based PERSIANN dataset (Sorooshian et al. 2000) presented in Beck et al. (2017b, their Fig. 3c), the Gridsat-based  $r_{3\text{day}}$  values are slightly lower in some regions, suggesting there may still be some opportunity for improving the GridSat-based rainfall estimates. We refer to Beck et al. (2017c) for a more comprehensive evaluation of the GridSat rainfall.

Figures 3e and 3f present  $\beta_{\text{WD}}$  (bias in the number of wet days per year) values for CMORPH and ERA-Interim, respectively. The results were again similar among satellite datasets and among reanalysis datasets, and therefore, we again present results for only one of each. Globally, CMORPH represents the  $P$  frequency substantially better than ERA-Interim. CMORPH slightly overestimates (underestimates) the  $P$  frequency at low (high) latitudes (Fig. 3e). Conversely, ERA-Interim strongly overestimates the  $P$  frequency across the entire globe (Fig. 3f), because of deficiencies in the parameterization of the processes controlling  $P$  generation (Zolina et al. 2004; Sun et al. 2006; Lopez 2007; Stephens et al. 2010; Skok et al. 2015; Herold et al. 2016). These findings highlight the importance of the  $P$  frequency corrections implemented in MSWEP V2 (“ $P$  frequency correction and dataset harmonization” appendix section). When interpreting these results, it must be kept in mind that point observations from gauges tend to underestimate



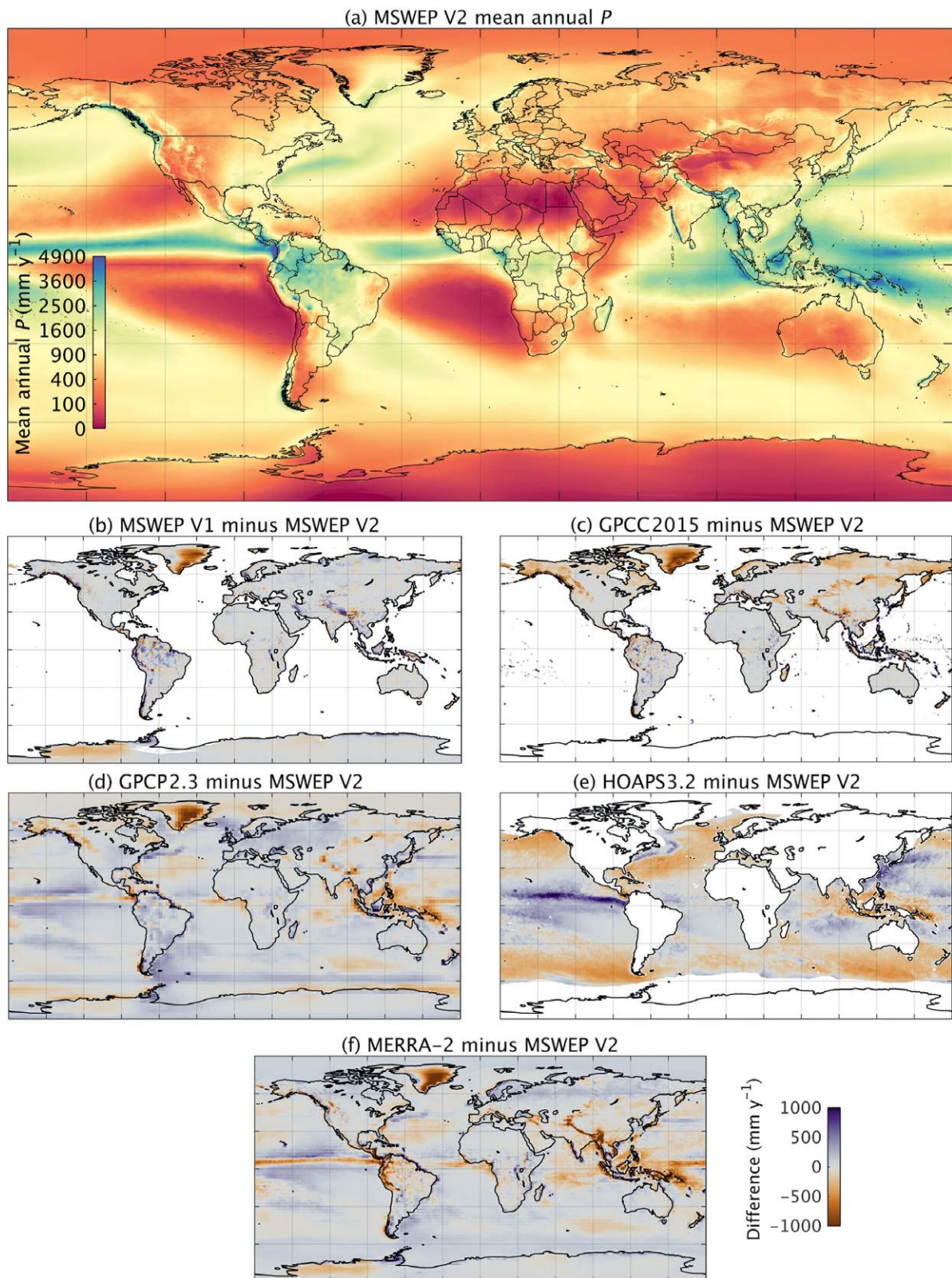
**FIG. 5. KGE, correlation, bias, and variability ratio scores for the CONUS calculated from 3-hourly  $P$  time series using the Stage IV gauge–radar dataset as reference. Regions without Stage IV coverage are shown in white.**

the number of wet days compared to similar estimates from gridded data from satellites and reanalyses (as the former samples a much smaller area; Osborn and Hulme 1997; Ensor and Robeson 2008).

**Global patterns in weights.** Figure 4 shows global maps of the relative weights assigned to the gauge-, satellite-, and reanalysis-based  $P$  estimates for three periods: i) 1979–82, ii) 1983–99, and iii) 2000–17. The gauge weights were calculated as a function of distance to surrounding gauges (“Gauge correction scheme” appendix section), whereas the satellite and

reanalysis weights were calculated based on the performance of the respective satellite and reanalysis datasets at surrounding gauges (“Global maps of weights and wet-day biases” appendix section). Gauge-based  $P$  estimates provide the main contribution over the terrestrial surface for all periods (Fig. 4). GridSat data are introduced in 1980 and represent the only satellite-based source of  $P$  estimates until 2000, when passive microwave-based estimates are introduced (CMORPH, GSMaP, and TMPA 3B42RT). Prior to 1982, however, GridSat provides limited coverage over South Asia and particularly Africa, and a horizontal





**FIG. 6.** (a) The long-term mean  $P$  ( $\text{mm yr}^{-1}$ ) for MSWEP V2. Also shown are the differences in long-term mean  $P$  between MSWEP V2 and (b) MSWEP VI, (c) GPCC2015, (d) GPCP2.3, (e) HOAPS3.2, and (f) MERRA-2. The values represent 1979–2015 for MSWEP VI, 1987–2008 for HOAPS3.2, 1980–2017 for MERRA-2, and 1979–2017 for the other datasets. Areas with no data are shown in white. For HOAPS2.3, only grid cells with continuous data are displayed.

striping pattern can be observed in some regions caused by gaps in the GridSat data (Fig. 4a). In regions without rain gauges, reanalyses provide the dominant contribution over most of the globe until 1999, while from 2000 onward, the dominant contribution comes from satellite data at low and midlatitudes and reanalysis data at high latitudes (Fig. 4).

**Evaluation using Stage IV gauge–radar data for the CONUS.** Beck et al. (2017c) evaluated MSWEP V2 and 20 other *P* datasets globally using observations from 76,086 gauges and hydrological modeling for 9,053 catchments at daily and monthly time steps. However, evaluation at the 3-hourly time step was lacking. We therefore evaluated MSWEP V2 and, for comparison purposes, MSWEP V1, CMORPH, ERA-Interim, and MERRA-2 (details provided in Table 1) at the 3-hourly time step for the CONUS using the Stage IV gauge–radar *P* dataset (Lin and Mitchell 2005) as a reference. Consistent with the global evaluation by Beck et al. (2017c), MSWEP V2 was found to perform best overall, yielding a median KGE score of 0.70 (Fig. 5a). The second- and third-best performing *P* datasets were MSWEP V1 (Beck et al. 2017b; Fig. 5e) and MERRA-2 (Reichle et al. 2017; Fig. 5q), exhibiting median KGE scores of 0.53 and of 0.41, respectively. Similar to MSWEP V2, MSWEP V1 and MERRA-2 include daily gauge corrections (based on the CPC unified dataset; Xie et al. 2007; Chen et al. 2008). However, in contrast to MSWEP V2, they did not account for gauge reporting times (“Gauge reporting times” section), which has resulted in temporal mismatches when applying the corrections (Figs. 1e and 1f). CMORPH (Joyce et al. 2004; Fig. 5i) and ERA-Interim (Dee et al. 2011; Fig. 5m) obtained lower median KGE scores of 0.36 and 0.35, respectively. Performance was markedly worse for all datasets in the western CONUS, because of the more complex topography and greater spatiotemporal heterogeneity of *P* (Daly et al. 2008).

**Global patterns in long-term mean *P*.** Figure 6a presents a global map of long-term mean *P* from MSWEP V2 (“Determination of long-term mean *P*” appendix section). Figures 6b–f, respectively, present the difference in long-term mean *P* between MSWEP V2 and five other *P* datasets (Table 1): i) MSWEP V1 (1979–2015; 3 hourly; 0.25°; Beck et al. 2017b), ii) GPCC, version 2015 (GPCC2015; 1951–2000; monthly; 0.5°; U. Schneider et al. 2014, 2017), iii) Global Precipitation Climatology Project, version 2.3 (GPCP2.3; 1979–2013; monthly; 2.5°; Adler et al. 2003, 2017, 2018), iv) Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data, version 3.2 (HOAPS3.2; 1987–2008; 0.5°; 6 hourly; Schlosser and Houser 2007; Andersson et al. 2010), and v) MERRA-2 (1980–2017; ~50 km; hourly; Reichle et al. 2017). The differences between MSWEP V1 and MSWEP V2 (Fig. 6b) primarily reflect the change from CHPClim to WorldClim in version 2. Compared to MSWEP V2, the fully gauge-based GPCC2015 dataset shows consistently lower mean *P* at high northern latitudes (Fig. 6c), whereas the gauge- and satellite-based GPCP2.3 dataset exhibits lower mean *P* only in northern North America and northeastern Asia but generally higher mean *P* in Europe and northwestern Asia (Fig. 6d). These differences probably reflect the use of different gauge undercatch correction schemes; GPCC2015 (Legates and Willmott 1990) and GPCP2.3 (Legates 1988) employ more conventional approaches using World Meteorological Organization (WMO) gauge undercatch correction equations in combination with daily observations of *P*, *T<sub>a</sub>*, and wind speed from a relatively sparse station network. Conversely, MSWEP V2 infers the “true” *P* using *Q* observations and *P<sub>e</sub>* estimates from 13,762 catchments globally (Beck et al. 2017b). The gauge- and reanalysis-based MERRA-2 dataset exhibits good agreement with MSWEP V2 at high latitudes but shows substantially lower *P* over tropical regions (except in

**TABLE 2. Long-term mean annual *P* estimates (mm yr<sup>−1</sup>) for global, land, and ocean domains from various sources.**

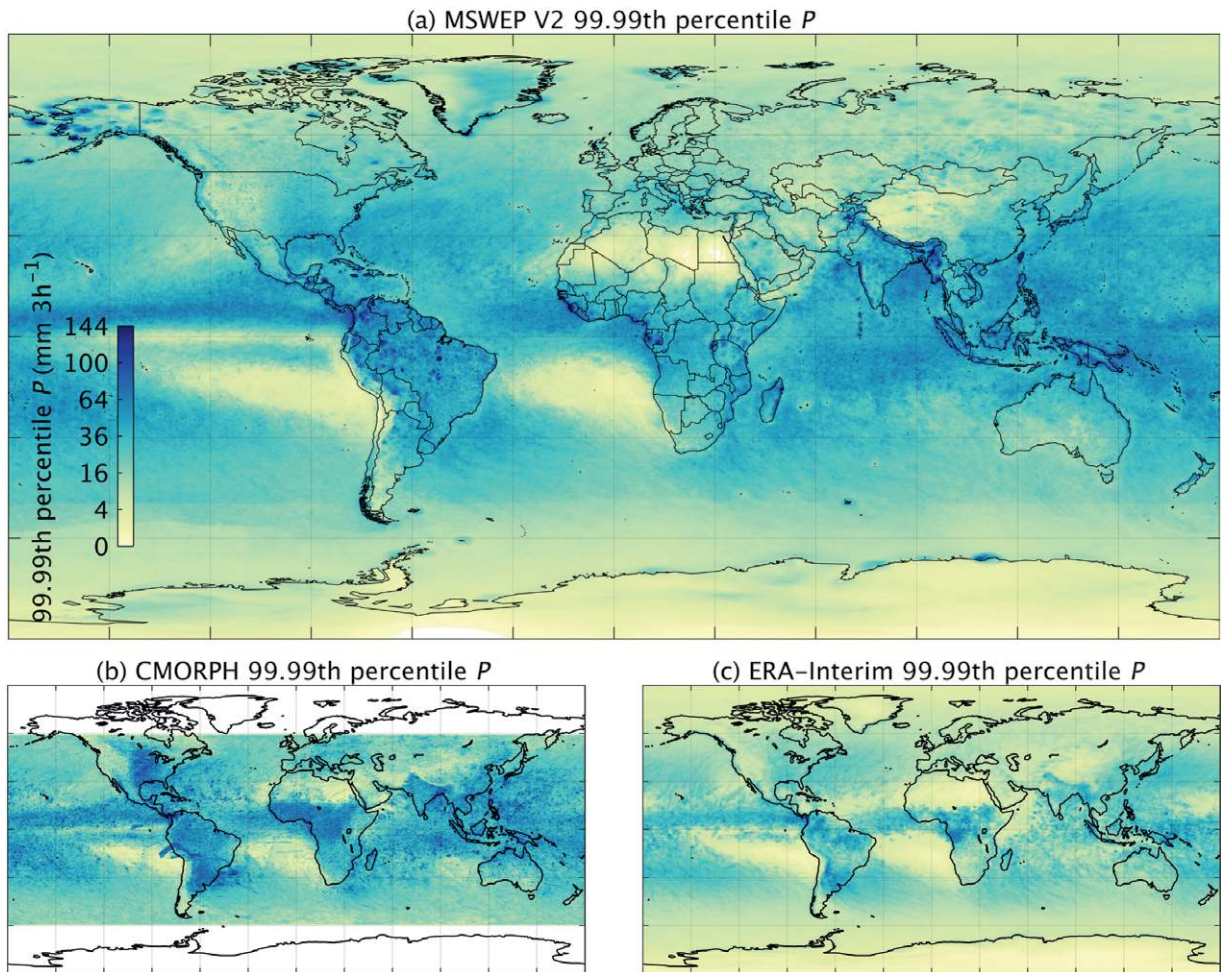
Domain	MSWEP V2	MSWEP V1	GPCC2015	GPCP2.3	HOAPS3.2	MERRA-2	Behrangi et al. (2014)
Global	955	—	—	982	—	946	—
Land (excluding Antarctica)	839	858	793	853	—	785	—
Land (including Antarctica)	781	798	—	798	—	735	—
Ocean <sup>a</sup>	1,025	—	—	1,057	—	1,031	1,074 <sup>b</sup>
Ocean (HOAPS mask) <sup>c</sup>	1,068	—	—	1,101	1,037	1,066	—

<sup>a</sup> Includes the Hudson Bay and the Baltic, Mediterranean, Black, Caspian, and Red Seas.

<sup>b</sup> Based on the 2007–09 period.

<sup>c</sup> Smaller mask that excludes coastal and high-latitude regions for which HOAPS does not provide continuous data.





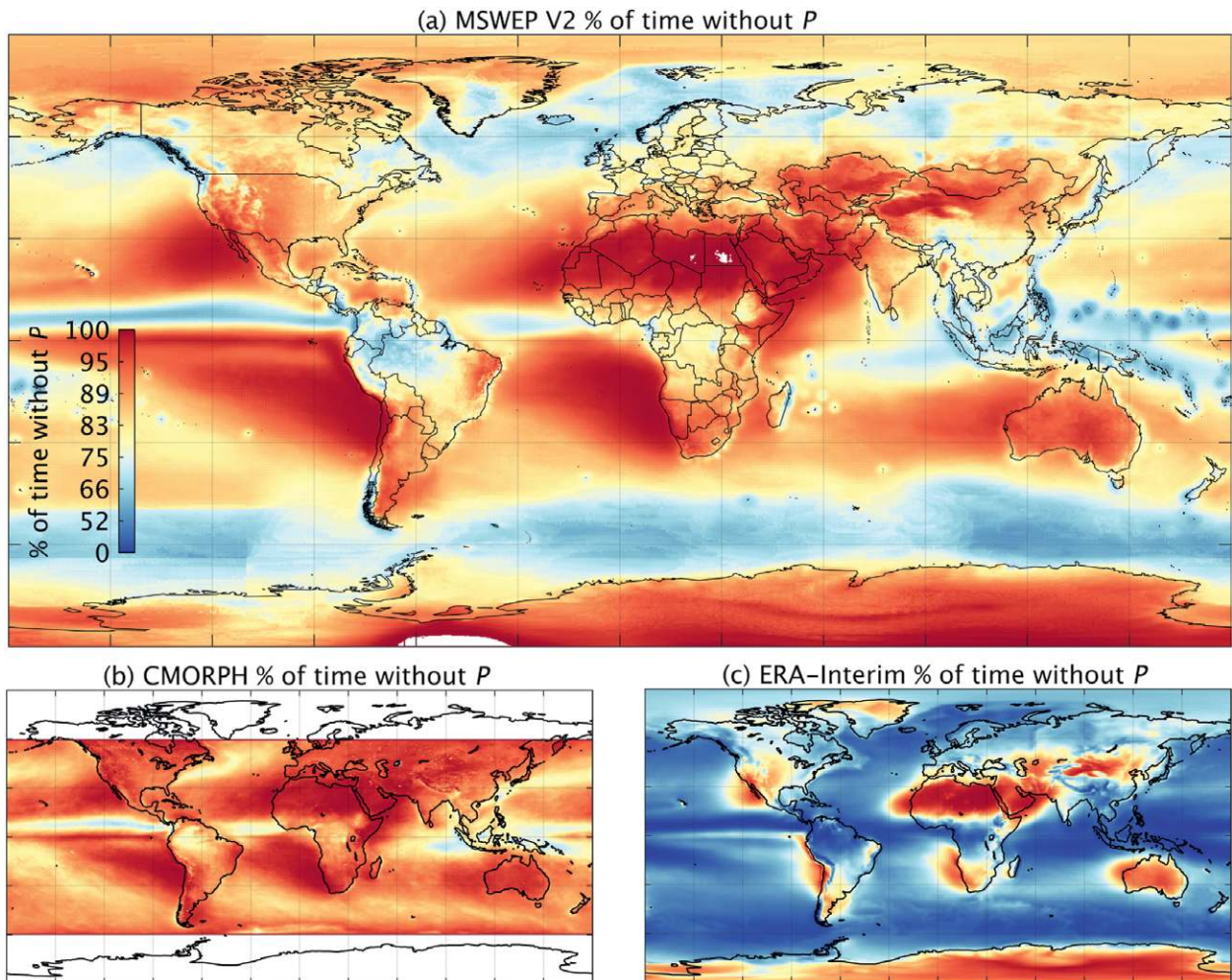
**FIG. 7. Global maps of 99.99th-percentile 3-hourly  $P$  amounts [ $\text{mm (3 h)}^{-1}$ ] for (a) MSWEP V2, (b) CMORPH, and (c) ERA-Interim for 2000–17. CMORPH is limited to latitudes  $\leq 60^\circ$ . Note the nonlinear color scale.**

Africa; Fig. 6f). Compared to MSWEP V2, the other  $P$  datasets (GPCC2015, GPCP2.3, and MERRA-2) exhibit substantially less  $P$  at high elevations (e.g., in the Rocky Mountains, the southern Andes, and most Asian mountainous regions; Figs. 6c, 6d, and 6f, respectively). This is attributable to their coarser resolutions ( $0.5^\circ$ ,  $2.5^\circ$ , and  $0.5^\circ$ , respectively) and lack of explicit orographic corrections. The differences between MSWEP V2 and the other  $P$  datasets over the equatorial oceans are probably at least partly because MSWEP V2 computes the long-term mean using satellite data from 2000 to 2017, during which the meridional location of the maximum intertropical convergence zone (ITCZ) convection was more northerly (T. Schneider et al. 2014).

Using MSWEP V2, we obtained a long-term mean global  $P$  estimate of  $955 \text{ mm yr}^{-1}$  (Table 2) or  $488,100 \text{ km}^3 \text{ yr}^{-1}$ . This estimate is based on terrestrial  $P$  data representative of 1970–2000 (i.e., the range of the WorldClim gauges; Fick and Hijmans 2017) and

oceanic  $P$  data representative of 1979–2017 (i.e., the range of the satellite and reanalysis datasets). The long-term mean  $P$  of MSWEP V2 over land (excluding Antarctica) is  $839 \text{ mm yr}^{-1}$ , corresponding to  $113,100 \text{ km}^3 \text{ yr}^{-1}$ . The same estimate for MSWEP V1 is  $858 \text{ mm yr}^{-1}$ , slightly (2.3%) higher because of the switch from CHPClim to WorldClim and the reduction of the Chilean and Iranian bias correction factors in version 2 (“Determination of long-term mean  $P$ ” in the appendix). The estimate for GPCP2.3 is  $853 \text{ mm yr}^{-1}$ , also slightly (1.7%) higher than the MSWEP V2 estimate. For GPCC2015, the corresponding estimate is  $793 \text{ mm yr}^{-1}$ , which is considerably (5.5%) lower for the reasons previously explained. The estimate for MERRA-2 is  $785 \text{ mm yr}^{-1}$ , also considerably (6.4%) lower than the MSWEP V2 estimate, mainly because of the aforementioned differences in tropical and mountainous regions. The long-term mean  $P$  for ocean areas based on MSWEP V2 amounted to  $1,025 \text{ mm yr}^{-1}$  (Table 2), corresponding to  $373,200 \text{ km}^3 \text{ yr}^{-1}$ . Arguably,





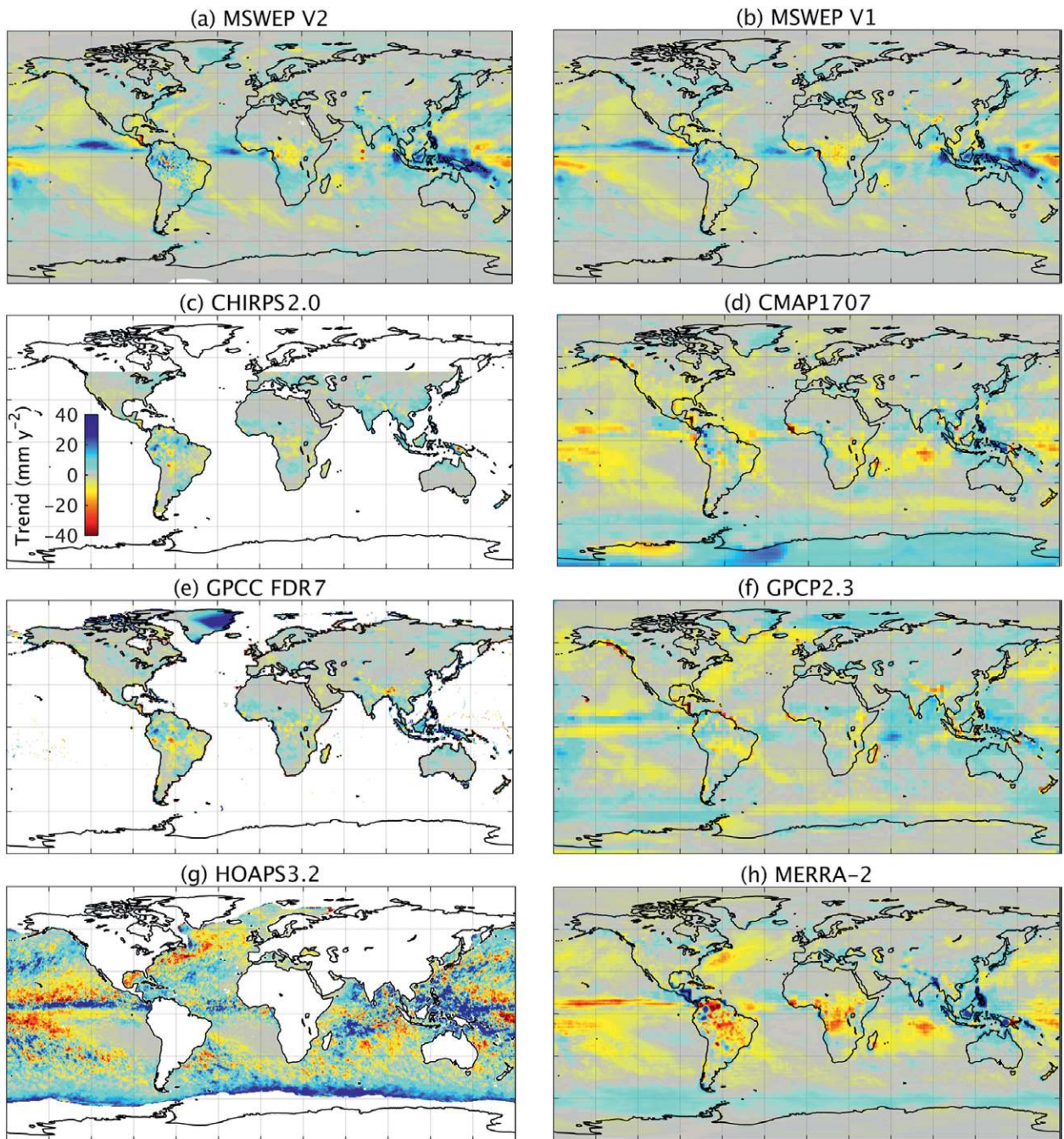
**FIG. 8.** The percentage of time without  $P$  for (a) MSWEP V2, (b) CMORPH, and (c) ERA-Interim for 2000–17. A  $0.06 \text{ mm } (3 \text{ h})^{-1}$  threshold was used to identify 3-hourly intervals with  $P$ . Areas with no data are shown in white. Note the highly nonlinear color scale.

the most comprehensive  $P$  datasets with ocean coverage currently available are the satellite-based GPCP2.3 and HOAPS3.2 datasets. Compared to our estimate, GPCP2.3 yields a 3.1% higher estimate of  $1,057 \text{ mm yr}^{-1}$  (Fig. 6d). Over the area for which HOAPS3.2 has continuous data (coastal areas are missing, and there are seasonal gaps at high latitudes), the dataset yields a 2.9% lower long-term mean  $P$  than MSWEP V2 ( $1,037$  vs  $1,068 \text{ mm yr}^{-1}$ ; Fig. 6e). Another estimate of  $1,074 \text{ mm yr}^{-1}$  for the entire ocean area that was derived from satellite radar reflectivities (2007–09) by the TRMM and *CloudSat* instruments (Behrangi et al. 2014) is 4.8% higher than our estimate. In summary, our  $P$  estimate is close to the average of previous estimates (Table 2).

**Global patterns in  $P$  extremes.** Figure 7 presents global maps of 99.99th-percentile 3-hourly  $P$  amounts (equivalent to a return period of 3.42 years) for MSWEP

V2 and, for illustrative purposes, CMORPH and ERA-Interim. CMORPH agrees well with MSWEP V2 in the tropics but appears to overestimate the 99.99th-percentile  $P$  with respect to MSWEP V2 in some midlatitude regions (e.g., in the central CONUS and in Argentina). Indeed, Beck et al. (2017c) recently found CMORPH to overestimate the 99th-percentile daily  $P$  magnitude in precisely these regions, and Tian et al. (2009) found CMORPH to overestimate summer  $P$  extremes strongly in the CONUS. As expected, ERA-Interim fails to resolve small-scale orographic features because of its coarse ( $\sim 0.7^\circ$ ) resolution and consistently estimates lower 99.99th-percentile  $P$  amounts because of the model parameterization challenges mentioned. Compared to a global map ( $1^\circ$ ) of 99th-percentile daily  $P$  amounts (equivalent to a return period of 100 days) derived from the Expert Team on Climate Change Detection and Indices (ETCCDI)  $P$  dataset (Dietzsch et al. 2017, their Fig. 5d), our 99.99th-percentile 3-hourly  $P$





**FIG. 9.** Linear trends in mean annual  $P$  ( $\text{mm yr}^{-2}$ ) for (a) MSWEP V2, (b) MSWEP V1, (c) CHIRPS2.0, (d) CMAP1707, (e) GPCC FDR7, (f) GPCP2.3, (g) HOAPS3.2, and (h) MERRA-2. The trends represent 1979–2015 for MSWEP V1, 1981–2017 for CHIRPS2.0, 1987–2008 for HOAPS3.2, 1980–2017 for MERRA-2, and 1979–2017 for the other datasets. Areas with no data are shown in white.

map (Fig. 7a) exhibits more plausible patterns. Most importantly, Dietzsch et al.'s (2017) map fails to represent small-scale  $P$  variations, mainly because of its coarse resolution, and shows unrealistically low values over land compared to the oceans, reflecting the use of different  $P$  data sources for land and ocean areas (the gauge-based GPCC and satellite-based HOAPS

datasets, respectively). The presence of slight discontinuities in the MSWEP V2 map at approximately  $50^{\circ}\text{S}$  (Fig. 7a) suggests that there are still inhomogeneities among the incorporated datasets, despite the frequency correction and harmonization applied. The higher 99.99th-percentile amounts near gauge locations (most noticeable in the Amazon in Fig. 7a) reflect the

loss of variance between  $P$  gauges due to interpolation (Hutchinson 1998; Haberlandt 2007).

**Global patterns in  $P$  occurrence.** Figure 8 presents global maps of the percentage of time without  $P$  for MSWEP V2, CMORPH, and ERA-Interim. CMORPH agrees fairly well overall with MSWEP V2 in the tropics, although it exhibits less frequent  $P$  at mid- and high latitudes (notably in southern Chile and along the Pacific coast of North America), in agreement with our  $P$  gauge-based assessment (Fig. 3e). This reflects the inability of current-generation satellites to detect  $P$  signals at high latitudes (Ebert et al. 2007; Tian et al. 2009; Tian and Peters-Lidard 2010; Behrangi et al. 2012; Massari et al. 2017; Beck et al. 2017c). Also in agreement with our  $P$  gauge-based assessment (Fig. 3f), ERA-Interim severely overestimates the  $P$  frequency across the entire globe. Our  $P$  frequency map (Fig. 8a) visually compares well with an equivalent map for the land surface derived from gauge observations during the period 1840 to 2001 produced by Sun et al. (2006, their Fig. 1). Additionally, our map agrees closely with ocean maps based on *CloudSat* data from 2006 to 2007 (Ellis et al. 2009, their Fig. 3a) and *CloudSat*, TRMM, and Advanced Microwave Scanning Radiometer for Earth Observing System (AMS-E) data from 2007 to 2009 (Behrangi et al. 2014, their Fig. 1a). We did, however, obtain a somewhat higher  $P$  frequency over the Southern Ocean, possibly because of uncertainties in the  $P$  frequency corrections caused by the near-complete absence of gauges south of 60°S (Fig. 2b).

Trenberth and Zhang (2018) examined how often it rains (or snows) worldwide for latitudes  $\leq 60^\circ\text{N/S}$ , using a gauge-corrected version of CMORPH (hourly, 0.25° resolution) and a 0.02 mm h<sup>-1</sup> threshold, and found that  $P$  occurs 11.0% of the time on average (8.2% over land and 12.1% over oceans). Using 3-hourly accumulations and a 0.06 mm (3 h)<sup>-1</sup> threshold (triple the hourly threshold), they found that  $P$  occurs 13.8% of the time on average (10.7% over land and 15.0% over oceans). The averages calculated using 3-hourly data are thus ~25% higher than the ones calculated using hourly data. Based on MSWEP V2 (3-hourly, 0.1° resolution), using the same 0.06 mm (3 h)<sup>-1</sup> threshold, we found that  $P$  occurs 15.0% of the time on average (11.5% over land and 16.2% over oceans) for the same region ( $\leq 60^\circ$  latitude). Therefore, our estimates are similar to but slightly (~9%) higher than those of Trenberth and Zhang (2018) but possibly more accurate given that the corrected CMORPH exhibits difficulties in detecting northern  $P$  (Beck et al. 2017c, their Fig. 2b). For the entire globe, based on MSWEP V2,  $P$  occurs

15.5% of the time on average, while  $P$  occurs 12.3% of the time over the land surface (excluding Antarctica) and 16.9% of the time over ocean areas. All estimates should, however, be interpreted with some caution because of the detection limits of satellite sensors (~0.8 mm h<sup>-1</sup> over land and ~0.02 mm h<sup>-1</sup> over ocean; Wolff and Fisher 2008) and rain gauges (~0.25 mm; Kuligowski 1997), as well as the scale discrepancy between point observations from rain gauges and gridded data from satellites and reanalyses (Osborn and Hulme 1997; Ensor and Robeson 2008).

**Trends in mean annual  $P$ .** Figure 9 presents global maps of the linear trend in mean annual  $P$  for MSWEP V2 and MSWEP V1; CHIRPS, version 2.0 (CHIRPS2.0); Climate Prediction Center (CPC) Merged Analysis of Precipitation, version 1707 (CMAP1707); GPCC FDR7; GPCP2.3; and HOAPS3.2 (details in Table 1). The trends were estimated at each grid cell using simple linear regression (Kenney and Keeping 1962). Over land, the datasets exhibit good agreement overall (with the exception of CMAP1707 and MERRA-2), which was expected since all datasets use similar gauge data sources. MERRA-2 exhibits suspect trend patterns over tropical land areas (Fig. 9h), which could be related to the bias adjustment using CMAP and CPC unified (Reichle et al. 2017). The small differences in trends between MSWEP V1 and MSWEP V2 (e.g., over the Amazon and the southwest Indian Ocean islands; Figs. 9a and 9b) are attributable to the new daily gauge data (“Gauge correction scheme” in the appendix). The correspondence in trends is considerably less over the oceans, presumably because of the lack of gauge observations to constrain uncertainty (Fig. 2b). CMAP1707 (and MERRA-2, which has been bias adjusted using CMAP over the oceans) generally tends more toward negative trends (Fig. 9d), which Yin et al. (2004) attributed to discontinuities caused by changes in gauge coverage and satellite input data. HOAPS3.2 exhibits a substantially noisier trend pattern and more pronounced trends overall (Fig. 9g), which are both likely attributable to its shorter data record. In the Southern Ocean, HOAPS3.2 not only shows  $P$  underestimation (Fig. 6e) but also a spurious upward trend, as reported in previous studies (Romanova et al. 2010; Liu et al. 2011). We refer to Adler et al. (2017) and Schneider et al. (2017) for a more comprehensive overview of the current state of knowledge with respect to trends in  $P$  worldwide.

Any  $P$  trend estimates should, however, be interpreted with caution because of the potential presence of temporal inhomogeneities. For gauge data, inhomogeneities tend to be caused by measurement errors



and changes in station coverage (Sevruk et al. 2009; U. Schneider et al. 2014); for satellite estimates, by instrument changes, sensor degradation, and algorithm changes (Kummerow et al. 1998; Biswas et al. 2013); and for reanalyses, by production stream transitions and changes in the observing systems (Dee et al. 2011; Trenberth et al. 2011; Kang and Ahn 2015; Kobayashi et al. 2015). Additionally, agreement in trends among different *P* datasets does not necessarily imply less uncertainty because the input data may be the same. MSWEP V2 trends are likely subject to much less uncertainty after the year 2000, because of the relative stability of the observing systems and the addition of multiple passive microwave-based *P* datasets. Beck et al. (2017c) recently evaluated 22 *P* datasets using observations from 76,086 gauges worldwide covering 2000–16 and found that MSWEP V2 exhibits more reliable trends overall than MSWEP V1 as well as other *P* datasets.

**CONCLUSIONS.** We presented MSWEP V2, a gridded *P* dataset spanning 1979–2017, which has several unique aspects: i) fully global coverage including all land and oceans (most satellite-based datasets are limited to 50° or 60° latitude); ii) high spatial (0.1°) and temporal (3 hourly) resolution, increasing the local relevance of the *P* estimates; iii) optimal merging of a wide range of gauge, satellite, and reanalysis *P* datasets, to obtain the best possible *P* estimates at any location; iv) correction for distributional biases, to eliminate spurious drizzle and restore attenuated peaks; v) correction of systematic terrestrial *P* biases due to gauge undercatch using observed *Q* from 13,762 catchments worldwide; vi) corrections using daily observations from 76,747 gauges across the globe; and vii) a gauge correction scheme that accounts for gauge reporting times. The main findings are as follows:

- 1) There are marked differences in reporting times between neighboring countries and sometimes within countries. Contrary to expectations, the automated GSOD gauges exhibited reporting times averaging at around 1200 UTC rather than at 2400 UTC. These findings underscore the importance of accounting for reporting times when applying daily gauge corrections.
- 2) The gauge-based assessment of the satellite and reanalysis *P* datasets revealed that the reanalyses strongly overestimate the *P* frequency across the globe. Confirming previous studies, we found that reanalyses exhibit lower skill than the satellite estimates in the (sub)tropics, whereas the

opposite was the case at high latitudes. MSWEP is the only high-resolution *P* dataset to date that exploits this complementary relationship.

- 3) For the CONUS, we evaluated MSWEP V2 and four other *P* datasets at a 3-hourly time scale using Stage IV gauge–radar *P* data as reference. MSWEP V2 provided the best overall performance, followed in order by MSWEP V1, MERRA-2, CMORPH, and ERA-Interim. These results confirm the effectiveness of the MSWEP V2 methodology.
- 4) Long-term mean *P* estimates for the global, land, and ocean domains based on MSWEP V2 are 955, 781, and 1,025 mm yr<sup>−1</sup>, respectively. This is in close agreement with previously published estimates, yet importantly for hydrological applications, other datasets appear to consistently underestimate *P* amounts in mountainous regions because of a lack of orographic corrections and coarser spatial resolutions.
- 5) Compared to other state-of-the-art *P* datasets, MSWEP V2 shows more plausible spatial patterns in mean, magnitude, and frequency. Using MSWEP V2, we estimated that *P* occurs 15.5%, 12.3%, and 16.9% of the time on average for the global, land, and ocean domains, respectively, slightly more frequent than previous estimates based on CMORPH.
- 6) Trends in 1979–2017 mean annual *P* among state-of-the-art *P* datasets generally agree over land, at least partly because of the use of common input datasets. Over oceans, the agreement is considerable less, possibly reflecting the lack of marine gauge observations.

**ACKNOWLEDGMENTS.** MSWEP V2 is available online (via [www.gloh2o.org](http://www.gloh2o.org)). We thank the four anonymous *BAMS* reviewers and the *BAMS* editorial team for constructively critical comments. We gratefully acknowledge the *P* dataset developers for producing and making available their datasets. The Water Center for Arid and Semi-Arid Zones in Latin America and the Caribbean (CAZALAC) and the Centro de Ciencia del Clima y la Resiliencia (CR) 2 (FONDAP 15110009) are thanked for sharing the Mexican and Chilean gauge data, respectively. We also acknowledge the gauge data providers in the Latin American Climate Assessment and Dataset (LACA&D) project: IDEAM (Colombia), INAMEH (Venezuela), INAMHI (Ecuador), SENAMHI (Peru), SENAMHI (Bolivia), and DMC (Chile). We further wish to thank Ali Aljani, Koen Verbist, and Piyush Jain for providing additional gauge data and Francesco Lin, Kevin Trenberth, Mauricio Zambrano Bigiarini, Noemi Vergopolan, Vincenzo Levizzani, and Yuting Yang for comments and suggestions on earlier versions of the

manuscript. Hylke E. Beck was supported by the U.S. Army Corps of Engineers' International Center for Integrated Water Resources Management (ICIWaRM), under the auspices of UNESCO. Albert I. J. M. van Dijk was supported under Australian Research Council's Discovery Projects funding scheme (Project DP140103679). Diego G. Miralles acknowledges support from the European Research Council (ERC) under Grant Agreement 715254 (DRY-2-DRY).

**APPENDIX: DETAILED DESCRIPTION OF THE MSWEP V2 METHODOLOGY.** Here, we describe in detail the different processing steps involved in the production of MSWEP V2 (Fig. 1a).

*Gauge data quality control.* Daily gauge observations were used to determine the merging weights and wet-day biases for the individual  $P$  datasets ("Global maps of weights and wet-day biases" appendix section) and to improve  $P$  estimates near gauge stations ("Gauge correction scheme" appendix section). Our initial database comprises 117,759 gauges worldwide compiled from the GHCN-D database (Menne et al. 2012), the GSOD database (<https://data.noaa.gov>), the Latin American Climate Assessment and Dataset (LACA&D) database (<http://lacad.ciifcn.org/>), the Chile Climate Data Library ([www.climatedatalibrary.cl](http://www.climatedatalibrary.cl)), and national databases for Mexico, Brazil, Peru, and Iran.

Gauge data can have considerable measurement errors, and therefore, quality control is important (Goodison et al. 1998; Viney and Bates 2004; Sevruk et al. 2009; U. Schneider et al. 2014). For example, GSOD records frequently contain long series of erroneous zero rainfall (Durre et al. 2010; Funk et al. 2015b). To identify and discard these periods, we developed an automated procedure entailing the following steps: i) for each month, we computed the fraction of days without  $P$  ( $fD$ ); ii) we excluded months without any  $P$  ( $fD = 1$ ) and computed the distribution mean  $\mu$  and standard deviation  $\sigma$ ; iii) if the CDF of the normal distribution with  $\mu$  and  $\sigma$  evaluated at  $fD = 0.9$  exceeds 0.85, the gauge was considered to be sufficiently "wet" for detecting the erroneous zeros, and we proceeded to the next step; iv) a year was marked as erroneous if the median of the 12 monthly  $fD$  values exceeded 0.9; and v) the 6 months preceding and following each erroneous year were also marked as erroneous. Figure 2a illustrates the procedure for an arbitrarily selected GSOD gauge with the described issue.

Additionally, we eliminated all days with  $P > 2,000$  mm (approximately the maximum recorded 24-h rainfall; Cervený et al. 2007) and discarded gauges with record length  $< 4$  years during 1979–2017.

From the remaining set of 81,047 gauges, we also discarded those matching one or more of the following criteria (percentage of remaining gauges satisfying the criteria reported between parentheses): i) 3-day Pearson correlation coefficient  $r_{3\text{day}}$  with all five non-gauge-based  $P$  datasets (CMORPH, ERA-Interim, GSMaP, JRA-55, and TMPA 3B42RT; Table 1)  $< 0.4$  and  $r_{3\text{day}}$  with the nearest gauge also  $< 0.4$  (1.01%); ii) more than half of the 3-day intervals contain missing values (1.62%); iii) fewer than 15 unique values in the entire record (1.02%); iv) the highest and/or second highest values were present  $> 3$  times in the record, indicative of truncated peaks (0.60%); and v)  $> 99.5\%$  of the record is dry ( $< 0.5$  mm day $^{-1}$ ; 3.05%). In total, 4,300 (5.31%) of the remaining gauges fulfilled one or more of these criteria and hence were discarded; the resultant dataset comprised 76,747 gauges (Fig. 2b).

*Inferring gauge reporting times.* Information about gauge reporting times is crucial to avoid timing mismatches when applying daily gauge corrections but is generally not provided. We developed a procedure to infer gauge reporting times using four gridded 3-hourly non-gauge-based  $P$  datasets (CMORPH, ERA-Interim, GSMaP, and JRA-55; Table 1). Specifically, we calculated, for each gauge, Spearman rank correlation coefficients  $\rho$  between daily grid- and gauge-based time series, with the grid-based time series shifted by offsets of  $-36, -33, -30, \dots, +30, +33$ , and  $+36$  h, resulting in  $4 \times 25 = 100$   $\rho$  values for each gauge. The dataset and temporal-offset combination yielding the highest  $\rho$  value was subsequently taken to reflect the UTC boundary of the 24-h accumulation period for the gauge under consideration. It should be kept in mind, however, that the inferred estimates are subject to a rounding error of at most 1.5 h and on average 45 min because of the 3-hourly temporal resolution of the  $P$  datasets. In addition, the estimates are affected by the fact that satellites represent radiation from an atmospheric column, whereas gauges represent  $P$  that has reached the surface (Villarini and Krajewski 2007). Furthermore, the approach relies on the assumption of a temporally constant reporting time, which may not be true for every gauge (Viney and Bates 2004).

*Rainfall estimation using thermal infrared imagery.* MSWEP V1 relied exclusively on reanalysis and gauge data during the pre-TRMM era (before 1998; Beck et al. 2017b). For MSWEP V2, we supplemented the reanalysis and gauge data with rainfall estimates based on cloud-top IR temperatures during the pre-TRMM era to improve the accuracy in convection-dominated regions. Although several IR-based rainfall

datasets already exist [e.g., CHIRP, Hydro-Estimator, PERSIANN, PERSIANN-CCS, PERSIANN–Climate Data Record (PERSIANN-CDR), and Tropical Applications of Meteorology using Satellite Data and Ground-Based Observations (TAMSAT)], none of these meet all of our requirements: i) quasi-global coverage over land and ocean, ii) temporal coverage from the 1980s to the near present, iii) spatial resolution  $\leq 0.1^\circ$ , iv) temporal resolution  $\leq 3$  h, and v) no gauge corrections. We therefore produced a new 3-hourly  $0.1^\circ$  rainfall dataset based on the GridSat B1 IR archive (V02R01; 3-hourly,  $0.07^\circ$  resolution; 1980 to the near present) containing IR imagery from various intercalibrated geostationary satellites (Knapp et al. 2011).

Although the GridSat archive has already had some quality control applied, we still observed numerous navigation, calibration, and masking errors (particularly prior to 1983). To ensure that the data were robust, several additional quality control steps were applied. First, all grid cells with values  $< 173$  K (the record minimum; Ebert and Holland 1992) were assumed to be erroneous and discarded. Additionally, if the percentage of grid cells with temperature  $< 173$  K exceeded 1%, the entire image was discarded. Furthermore, if the spatial (Pearson) correlation between the current image and the previous image (both resampled to  $1^\circ$  using bilinear interpolation) was  $< 0.75$ , both images were discarded. Finally, assuming that sudden isolated changes in the record are indicative of errors, images were discarded if the global mean deviated  $> 3$  K from the 24-h running global mean. Note that prior to 1998, there are extensive periods of missing data because of a poorer spatial coverage.

IR data can be used to estimate rainfall in several ways (Scofield and Kuligowski 2003; Stephens and Kummerow 2007; Michaelides et al. 2009; Kidd and Levizzani 2011). Hydro-Estimator, for example, employs an empirical equation calibrated using ground radar data to obtain an initial rain-rate estimate, which is subsequently corrected using precipitable water and relative humidity outputs from an atmospheric analysis model (Scofield and Kuligowski 2003). Conversely, CHIRP employs cold-cloud duration (CCD) values derived from IR data using a fixed 235-K threshold to estimate 5-day rain rates, where the CCD–rain relationship is established by linear regression against TMPA 3B42 data (Funk et al. 2015b). Similarly, the African TAMSAT dataset uses IR-based CCD values to estimate 10-day rainfall but uses gauge observations to determine the regression parameters and temperature thresholds (Tarnavsky et al. 2014). CCD-based methods are, however, unsuitable for our purposes as they would require IR data with a temporal resolution  $< 3$  h

to derive 3-hourly CCD values. PERSIANN-CCS employs a more elaborate method using artificial neural networks and IR data patterns to distinguish between cloud types, which are subsequently related to specific rainfall intensities (Ashouri et al. 2015).

Here, we used a parsimonious method entailing the following steps: i) resampling the GridSat IR data to  $0.1^\circ$  using bilinear interpolation; ii) rejecting IR data when daily mean  $T_a$  is  $< 5^\circ\text{C}$ , due to the difficulty of detecting  $P$  signals in cold conditions (Kidd and Levizzani 2011; Beck et al. 2017b); iii) reversing the sign of the values, since lower IR radiances correspond to higher rainfall intensities (Adler and Negri 1988); and iv) converting the values to rain rates by CDF matching against the warm-period reference  $P$  distribution produced in the “Reference  $P$  distributions” appendix section. Our approach bears some resemblance to that of Karbalaee et al. (2017), who CDF matched the IR-based PERSIANN-CCS rainfall dataset to a passive microwave-based reference. The method used here may not perform well in regions with a marked temporal variability in storm type and, correspondingly, in the relationship between IR radiance and rainfall. Any such deficiencies would be reflected in low weights in the merging process (“Global maps of weights and wet-day biases” appendix section).

#### *Gauge-based assessment of satellite and reanalysis $P$ datasets.*

MSWEP V2 incorporates six gridded non-gauge-based  $P$  datasets (CMORPH, ERA-Interim, GridSat, GSMaP, JRA-55, and TMPA 3B42RT; Table 1). To assess the individual performance of these datasets, we calculated, for each  $P$  gauge, Pearson correlation coefficients between 3-day mean gauge- and grid-based  $P$  time series ( $r_{3\text{day}}$ ) for 2000–17 (the start date is limited by the GSMaP and TMPA 3B42RT datasets). To minimize timing mismatches between the gauge- and grid-based time series, prior to calculating the  $r_{3\text{day}}$  values, the records of gauges with reporting times  $> +12$  h UTC were shifted backward by  $-1$  day, while the records of gauges with reporting times  $< -12$  h UTC were shifted forward by  $+1$  day (“Inferring gauge reporting times” appendix section). The use of 3-day rather than daily averages has two benefits: first, it minimizes the impact of any remaining temporal mismatches in the 24-h accumulation period between the gridded datasets and the gauges, and second, it reduces the influence of days with potentially erroneous gauge measurements. The  $r_{3\text{day}}$  values were calculated for the full period of contemporaneous gauge- and grid-based data, as well as for “cold” and “warm” conditions, distinguished using a daily mean air temperature  $T_a$  threshold of  $5^\circ\text{C}$ .

MSWEP V1 employed a 1°C threshold, which we increased in version 2 to further reduce the likelihood of incorporating potentially unreliable satellite data. For  $T_a$ , we used ERA-Interim (Dee et al. 2011) downscaled to 0.1° using nearest-neighbor resampling and offset to match the long-term mean of the high-resolution, station-based WorldClim, version 2.0 (WorldClim2.0), dataset (Fick and Hijmans 2017). We only calculated an  $r_{3\text{day}}$  value if >1 year of simultaneous gauge and gridded 3-day means were available. The  $r_{3\text{day}}$  values range from -1 to 1, with higher values corresponding to better performance.

Reanalyses tend to overestimate the  $P$  frequency and underestimate the intensity because of deficiencies in the parameterization of the physical processes controlling  $P$  generation (Zolina et al. 2004; Sun et al. 2006; Lopez 2007; Stephens et al. 2010; Skok et al. 2015; Herold et al. 2016). To quantify and correct for this, we calculated the bias in the number of wet days per year, using the  $P$  gauge observations as reference, according to Eq. (1). Wet days were identified using a 0.5 mm day<sup>-1</sup> threshold, similar to several previous studies (e.g., Akinremi et al. 1999; Haylock et al. 2008; Driouech et al. 2009; Trenberth and Zhang 2018). The  $\beta_{\text{WD}}$  values range from 0 to  $\infty$ , with values closer to unity corresponding to better performance.

**Global maps of weights and wet-day biases.** Global weight maps were derived for the entire period and for warm and cold conditions for each of the non-gauge-based satellite and reanalysis  $P$  datasets (Table 1) from the gauge-based  $r_{3\text{day}}$  values (“Gauge-based assessment of satellite and reanalysis  $P$  datasets” appendix section). The  $r_{3\text{day}}$  values were truncated at zero, squared to yield the coefficient of determination, and subsequently interpolated to yield gap-free global weight maps by calculating, for each 0.1° grid cell, the median of the 10 nearest gauges. The cold-condition weights were set to zero for the satellite datasets. Similarly, gap-free global maps of  $\beta_{\text{WD}}$  were produced for the reanalyses to correct the  $P$  frequency prior to the merging.

Because of a lack of gauges over ocean areas, the use of the 10 nearest gauges in the interpolation frequently resulted in strong discontinuities in the middle of oceans as a result of contrasting values on opposite sides of the oceans. To eliminate these discontinuities, we applied an exponential smoothing kernel with a bandwidth of 1,000 km over the ocean areas of the interpolated weight and  $\beta_{\text{WD}}$  maps.

**Determination of long-term mean  $P$ .** The long-term mean  $P$  over the land surface was determined in

version 2 using the WorldClim dataset (1-km resolution; version 2.0; Fick and Hijmans 2017) rather than the CHPClim dataset (0.05° resolution; Funk et al. 2015a). We switched from CHPClim to WorldClim because of the better gauge coverage in South America, Scandinavia, India, Australia, and New Zealand. Systematic  $P$  underestimation over land due to gauge undercatch and orographic effects (Kauffeldt et al. 2013; Beck et al. 2015, 2017a; Prein and Gobiet 2017) was corrected similarly to MSWEP V1 by inferring catchment-average  $P$  using the Zhang et al. (2001) relationship in combination with river discharge  $Q$  observations and potential evaporation  $E_p$  estimates (Beck et al. 2017b). However, for MSWEP V2, the correction factors inferred for Chilean and Iranian catchments were set to 1 prior to the interpolation, because of suspected issues with the observed  $Q$  data.

The long-term mean  $P$  over the oceans was estimated by weighting the long-term means of five satellite and reanalysis datasets (CMORPH, ERA-Interim, GSMaP, JRA-55, and TMPA 3B42RT; Table 1). The weights for the satellite datasets  $w_s$  were set to 1 for latitudes <20° and 0 for latitudes >40°, decreasing linearly from 1 at 20° to 0 at 40°. The weights for the reanalyses  $w_r$  were set to  $1 - w_s$ . Thus,  $w_r$  was set to 0 at latitudes <20°, due to the tendency of reanalyses to overestimate tropical  $P$  amounts (Trenberth et al. 2011; Kang and Ahn 2015).

**$P$  frequency correction and dataset harmonization.** The following three steps were implemented to reduce the  $P$  frequency of the two reanalyses and harmonize the six non-gauge-based  $P$  datasets incorporated in MSWEP V2 (CMORPH, ERA-Interim, GridSat, GSMaP, JRA-55, and TMPA 3B42RT; Table 1):

- 1) The datasets with spatial resolutions higher or lower than 0.1° (CMORPH, ERA-Interim, JRA-55, and TMPA 3B42RT) were resampled to 0.1 using nearest-neighbor resampling, and 3-hourly means were calculated for the datasets with temporal resolutions <3 h (CMORPH and GSMaP).
- 2) The Water and Global Change (WATCH; Weedon et al. 2011) and WATCH Forcing Data ERA-Interim (WFDEI; Weedon et al. 2014) datasets [derived respectively from the 40-yr ECMWF Re-Analysis (ERA-40) and ERA-Interim] were corrected for overestimations in  $P$  frequency by progressively removing the smallest events until the  $P$  frequency matched that of the gauge-based CRU dataset. However, this approach results in  $P$  distributions with a lack of light  $P$  events. We therefore employed an alternative approach to



correct the  $P$  frequency of the reanalyses (ERA-Interim and JRA-55). First, for grid cells with interpolated  $\beta_{WD}$  values  $>1$ , we calculated the “correct” annual number of wet days ( $WD_{objective}$ ) according to  $WD_{objective} = WD_{gridded} / \beta_{WD}$ , where  $WD_{gridded}$  was calculated from daily accumulations and  $\beta_{WD}$  represents the interpolated value (“Global maps of weights and wet-day biases” appendix section). Next, we iteratively carried out the following steps: i) subtract  $d$  mm  $(3\text{ h})^{-1}$  from the original 3-hourly time series, starting with  $d = 0.01$  mm  $(3\text{ h})^{-1}$ ; ii) truncate the resulting values to zero and rescale them to restore the original long-term mean; iii) calculate the annual number of wet days from daily accumulations ( $WD_{new}$ ); and iv) return to step i, increasing  $d$  in  $0.01$  mm  $(3\text{ h})^{-1}$  increments, until  $WD_{new} \leq WD_{objective}$ . Figure 1b illustrates the procedure for ERA-Interim.

- 3) The reanalysis datasets, which are valid for the entire period, and the satellite datasets, which are only valid for warm conditions, were rescaled to minimize the presence of spurious temporal discontinuities after merging. For this purpose, we first rescaled the reanalyses to match the long-term  $P$  estimates derived in the “Determination of long-term mean  $P$ ” appendix section. Next, means were calculated for the entire period and for warm and cold conditions based on the rescaled reanalyses, using the full-period weight maps derived in the “Global maps of weights and wet-day biases” appendix section. Finally, the satellite datasets were rescaled to match the rescaled warm-condition reanalysis mean.

**Reference  $P$  distributions.** In MSWEP V2, the 3-hourly merged satellite and reanalysis  $P$  estimates were CDF matched to reference  $P$  distributions (Fig. 1) to correct the spurious drizzle and attenuated peaks evident in version 1 (Nair and Indu 2017; Zhang et al. 2017). Two separate 3-hourly reference distributions ( $0.1^\circ$  resolution) were calculated: one representing warm conditions and one representing cold conditions (as before distinguished using a daily mean  $T_a$  threshold of  $5^\circ\text{C}$ ). The reference distribution for warm conditions was calculated by weighted-median averaging of the distributions of five satellite and reanalysis  $P$  datasets (CMORPH, ERA-Interim, GSMaP, JRA-55, and TMPA 3B42RT; Table 1). The GridSat dataset was excluded because it does not represent an independent estimate, being derived using the reference distributions (“Rainfall estimation using thermal infrared imagery” appendix section).

For cold conditions, the reference distribution was calculated by weighted-mean averaging of only the two reanalysis  $P$  datasets (ERA-Interim and JRA-55). Prior to the averaging, the  $P$  frequency of the reanalyses was corrected and the datasets were homogenized as described in the previous section. We only used data observed since 2000 to derive the reference distributions for two reasons: i) to avoid inconsistencies between the warm- and cold-condition reference distributions due to the much longer temporal coverage of the reanalyses and ii) because satellite data prior to 2000 are subject to more uncertainty (Xie et al. 2017).

**Merging of satellite and reanalysis  $P$  datasets.** Six 3-hourly non-gauge-based  $P$  datasets (CMORPH, ERA-Interim, GridSat, GSMaP, JRA-55, and TMPA 3B42RT; Table 1) were merged through the following steps:

- 1) For cold and warm conditions separately, and for every possible  $P$  dataset combination, the 3-hourly estimates were merged by weighted-mean averaging using the interpolated weight maps (“Global maps of weights and wet-day biases” appendix section). The total number of combinations comprising two or more  $P$  datasets equals 57 for warm conditions, while just one combination (containing both reanalyses) is valid for cold conditions (the satellite data were discarded). Prior to the merging, the  $P$  frequency of the reanalyses was corrected, and the datasets were harmonized (“ $P$  frequency correction and dataset harmonization” appendix section). Satellite data were discarded prior to 2000 and for grid cells with daily mean  $T_a \geq 5^\circ\text{C}$  less than 10% of the time.
- 2) Averaging multiple  $P$  datasets tends to result in spurious drizzle and attenuated peaks, as was the case for MSWEP V1 (Nair and Indu 2017; Zhang et al. 2017). To correct for this, we CDF matched the merged  $P$  estimates from 2000 to 2017 of each dataset combination, for cold and warm conditions separately, to the reference  $P$  distributions (which represent 2000–17; see the “Reference  $P$  distributions” appendix section). Similar CDF-matching approaches have been used to correct other  $P$  datasets, including CMORPH (Xie et al. 2017), Global Ensemble Forecast System (GEFS; Zhu and Luo 2015), and PERSIANN-CCS (Karbalaei et al. 2017). To obtain consistent time series for the entire 1979–2017 period, we first calculated the change in the  $P$  estimates due to the CDF corrections for different  $P$  magnitudes, after which we applied

the same magnitude-specific changes to the  $P$  estimates from 1979 to 1999.

- 3) A side effect of the implemented CDF corrections is that they result in regionally amplified trends. These corrections essentially increase (decrease) the magnitude of large (small)  $P$  events, inadvertently causing the trends associated with large events to become not just stronger but also more prominent in the overall record. We therefore rescaled the merged CDF-corrected estimates for cold and warm conditions separately and for each dataset combination such that their trends match those of the merged non-CDF-corrected estimates. Trends were calculated using simple linear regression (Kenney and Keeping 1962).
- 4) For cold and warm conditions separately, and for each possible dataset combination, we subsequently summed the interpolated weights of the incorporated datasets, yielding the cumulative interpolated weight, which roughly reflects the total information content of the dataset combination in question. Next, we selected, for each 3-hourly time step and  $0.1^\circ$  grid cell, the merged and corrected  $P$  value from the dataset combination with the highest cumulative weight. The applied CDF corrections ensure that temporal transitions from one dataset combination to another are largely unnoticeable. Figures 1c and 1d illustrate the merging procedure for a single grid cell.

**Gauge correction scheme.** The merged 3-hourly satellite- and reanalysis-based  $P$  data (referred to as  $p_{\text{merge}}$ ; “Merging of satellite and reanalysis  $P$  datasets” appendix section) were corrected using gauge  $P$  observations through an iterative, multiplicative approach that accounts for variability in the reporting times of gauges (“Inferring gauge reporting times” appendix section). We used a multiplicative rather than an additive correction method (Vila et al. 2009) to preserve the subdaily distribution of  $p_{\text{merge}}$ . The approach assumes that the long-term mean of  $p_{\text{merge}}$ , being based on the gauge-corrected WorldClim dataset (“Determination of long-term mean  $P$ ” appendix section), is already reliable and therefore only adjusts the temporal variability of  $p_{\text{merge}}$  using the gauge data. The approach entails the following steps:

- 1) For each  $0.1^\circ$  grid cell, very small  $P$  amounts were added to  $p_{\text{merge}}$ , to avoid a high gauge estimate from yielding a zero estimate after the correction when  $p_{\text{merge}} = 0$ , which occurs frequently in MSWEP V2 because of the  $P$  frequency and

CDF corrections. Specifically, we added an almost negligible amount (0.1%) of the non-CDF-matched (and thus drizzly) merged satellite- and reanalysis-based  $P$  data. The resulting estimate will be referred to as  $p_{\text{drizzly}}$ .

- 2) The five nearest (as the crow flies) gauges were selected (“Gauge data quality control” appendix section), and each gauge record was rescaled such that its mean equals that of  $p_{\text{merge}}$  for the period of overlap.
- 3)  $p_{\text{drizzly}}$  was corrected at the daily time scale in an iterative manner by looping through the five nearest gauges. During each loop, daily  $P$  accumulations of  $p_{\text{drizzly}}$  were calculated for the 24-h period ending at the reporting time, after which a blended estimate was calculated by weighted-mean averaging of the daily  $p_{\text{drizzly}}$  and gauge accumulations. The 3-hourly  $p_{\text{drizzly}}$  data were subsequently rescaled to match this blended estimate and passed on to the next loop iteration. The gauge weight  $w_g$  (unitless) was calculated according to  $w_g = 4\exp(-d_i/d_0)$ , where  $d_i$  (km) represents the distance from the grid-cell center to the gauge and  $d_0$  (km) represents the range of influence (set to 25 km using trial and error). The  $p_{\text{drizzly}}$  weight was calculated as the sum of the weights assigned to the incorporated gridded  $P$  datasets (step 3 of the “Merging of satellite and reanalysis  $P$  datasets” appendix section) and the gauge weights from the previous loop iterations. Figures 1e and 1f illustrate the importance of accounting for reporting times when applying daily gauge corrections.
- 4) To take advantage of the wider availability of monthly gauge data, we subsequently corrected  $p_{\text{drizzly}}$  using the monthly  $0.5^\circ$  GPCC FDR7 dataset (U. Schneider et al. 2014, 2017) following the same procedure but without accounting for gauge reporting times to yield the final gauge-corrected MSWEP V2.

## REFERENCES

- Adler, R. F., and A. J. Negri, 1988: A satellite infrared technique to estimate tropical convective and stratiform rainfall. *J. Appl. Meteor.*, **27**, 30–51, [https://doi.org/10.1175/1520-0450\(1988\)027<0030:ASITTE>2.0.CO;2](https://doi.org/10.1175/1520-0450(1988)027<0030:ASITTE>2.0.CO;2).
- , 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. Hydrometeor.*, **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).
- , G. Gu, M. Sapiiano, J.-J. Wang, and G. J. Huffman, 2017: Global precipitation: Means, variations and trends during the satellite era (1979–2014). *Surv. Geophys.*, **38**, 679–699, <https://doi.org/10.1007/s10712-017-9416-4>.
- , 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>.
- AghaKouchak, A., A. Behrangi, S. Sorooshian, K. Hsu, and E. Amitai, 2011: Evaluation of satellite retrieved extreme precipitation rates across the central United States. *J. Geophys. Res.*, **116**, D02115, <https://doi.org/10.1029/2010JD014741>.
- , A. Mehran, H. Norouzi, and A. Behrangi, 2012: Systematic and random error components in satellite precipitation data sets. *Geophys. Res. Lett.*, **39**, L09406, <https://doi.org/10.1029/2012GL051592>.
- Akinremi, O. O., S. M. McGinn, and H. W. Cutforth, 1999: Precipitation trends on the Canadian prairies. *J. Climate*, **12**, 2996–3003, [https://doi.org/10.1175/1520-0442\(1999\)012<2996:PTOTCP>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<2996:PTOTCP>2.0.CO;2).
- Andersson, A., K. Fennig, C. Klepp, S. Bakan, H. Graß, and J. Schulz, 2010: The Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data—HOAPS-3. *Earth Syst. Sci. Data*, **2**, 215–234, <https://doi.org/10.5194/essd-2-215-2010>.
- Ashouri, H., K. Hsu, S. Sorooshian, D. K. Braithwaite, K. R. Knapp, L. D. Cecil, B. R. Nelson, and O. P. Pratt, 2015: PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bull. Amer. Meteor. Soc.*, **96**, 69–83, <https://doi.org/10.1175/BAMS-D-13-00068.1>.
- Beck, H. E., A. I. J. M. van Dijk, and A. de Roo, 2015: Global maps of streamflow characteristics based on observations from several thousand catchments. *J. Hydrometeor.*, **16**, 1478–1501, <https://doi.org/10.1175/JHM-D-14-0155.1>.
- , —, —, E. Dutra, G. Fink, R. Orth, and J. Schellekens, 2017a: Global evaluation of runoff from ten state-of-the-art hydrological models. *Hydrol. Earth Syst. Sci.*, **21**, 2881–2903, <https://doi.org/10.5194/hess-21-2881-2017>.
- , —, V. Levizzani, J. Schellekens, D. G. Miralles, B. Martens, and A. de Roo, 2017b: 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, 2017c: 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>.
- Behrangi, A., M. Lebsock, S. Wong, and B. Lambrigtsen, 2012: On the quantification of oceanic rainfall using spaceborne sensors. *J. Geophys. Res.*, **117**, D20105, <https://doi.org/10.1029/2012JD017979>.
- , Y. Tian, B. H. Lambrigtsen, and G. L. Stephens, 2014: What does CloudSat reveal about global land precipitation detection by other spaceborne sensors? *Water Resour. Res.*, **50**, 4893–4905, <https://doi.org/10.1002/2013WR014566>.
- Bierkens, M. F. P., and Coauthors, 2015: Hyper-resolution global hydrological modelling: What is next? *Hydrol. Processes*, **29**, 310–320, <https://doi.org/10.1002/hyp.10391>.
- Biswas, S. K., S. Farrar, K. Gopalan, A. Santos-Garcia, W. L. Jones, and S. Bilanow, 2013: Intercalibration of microwave radiometer brightness temperatures for the Global Precipitation Measurement Mission. *IEEE Trans. Geosci. Remote Sens.*, **51**, 1465–1477, <https://doi.org/10.1109/TGRS.2012.2217148>.
- Cervený, R. S., J. Lawrimore, R. Edwards, and C. Landsea, 2007: Extreme weather records. *Bull. Amer. Meteor. Soc.*, **88**, 853–860, <https://doi.org/10.1175/BAMS-88-6-853>.
- Chen, L., and P. A. Dirmeyer, 2017: Impacts of land-use/land-cover change on afternoon precipitation over North America. *J. Climate*, **30**, 2121–2140, <https://doi.org/10.1175/JCLI-D-16-0589.1>.
- , —, A. Tawfik, and D. M. Lawrence, 2017: Sensitivities of land cover–precipitation feedback to convective triggering. *J. Hydrometeor.*, **18**, 2265–2283, <https://doi.org/10.1175/JHM-D-17-0011.1>.
- 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>.
- Correa, S. W., R. C. D. de Paiva, J. C. Espinoza, and W. Collischonn, 2017: Multi-decadal hydrological retrospective: Case study of Amazon floods and droughts. *J. Hydrol.*, **549**, 667–684, <https://doi.org/10.1016/j.jhydrol.2017.04.019>.
- Daly, C., R. P. Neilson, and D. L. Phillips, 1994: A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteor.*, **33**, 140–158, [https://doi.org/10.1175/1520-0450\(1994\)033<0140:ASTMFM>2.0.CO;2](https://doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2).

- , M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris, 2008: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol.*, **28**, 2031–2064, <https://doi.org/10.1002/joc.1688>.
- 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>.
- DeGaetano, A. T., 2000: A serially complete simulated observation time metadata file for U.S. daily Historical Climatology Network stations. *Bull. Amer. Meteor. Soc.*, **81**, 49–67, [https://doi.org/10.1175/1520-0477\(2000\)081<0049:ASCOT>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<0049:ASCOT>2.3.CO;2).
- Dietzsch, F., A. Andersson, M. Ziese, M. Schröder, K. Raykova, K. Schamm, and A. Becker, 2017: A global ETCCDI-based precipitation climatology from satellite and rain gauge measurements. *Climate*, **5**, 9, <https://doi.org/10.3390/cli5010009>.
- Driouech, F., M. Déqué, and A. Mokssit, 2009: Numerical simulation of the probability distribution function of precipitation over Morocco. *Climate Dyn.*, **32**, 1055–1063, <https://doi.org/10.1007/s00382-008-0430-6>.
- Durre, I., M. J. Menne, B. E. Gleason, T. G. Houston, and R. S. Vose, 2010: Comprehensive automated quality assurance of daily surface observations. *J. Appl. Meteor. Climatol.*, **49**, 1615–1633, <https://doi.org/10.1175/2010JAMC2375.1>.
- Ebert, E. E., and G. J. Holland, 1992: Observations of record cold cloud-top temperatures in Tropical Cyclone Hilda (1990). *Mon. Wea. Rev.*, **120**, 2240–2251, [https://doi.org/10.1175/1520-0493\(1992\)120<2240:OORCCT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1992)120<2240:OORCCT>2.0.CO;2).
- , J. E. Janowiak, and C. Kidd, 2007: Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bull. Amer. Meteor. Soc.*, **88**, 47–64, <https://doi.org/10.1175/BAMS-88-1-47>.
- Ellis, T. D., T. L'Ecuyer, J. M. Haynes, and G. L. Stephens, 2009: How often does it rain over the global oceans? The perspective from CloudSat. *Geophys. Res. Lett.*, **36**, L03815, <https://doi.org/10.1029/2008GL036728>.
- Ensor, L. A., and S. M. Robeson, 2008: Statistical characteristics of daily precipitation: Comparisons of gridded and point datasets. *J. Appl. Meteor. Climatol.*, **47**, 2468–2476, <https://doi.org/10.1175/2008JAMC1757.1>.
- Fick, S. E., and R. J. Hijmans, 2017: WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.*, **37**, 4302–4315, <https://doi.org/10.1002/joc.5086>.
- Funk, C., A. Verdin, J. Michaelsen, P. Peterson, D. Pedreros, and G. Husak, 2015a: A global satellite assisted precipitation climatology. *Earth Syst. Sci. Data*, **7**, 275–287, <https://doi.org/10.5194/essd-7-275-2015>.
- , and Coauthors, 2015b: The Climate Hazards Infrared Precipitation with Stations—A new environmental record for monitoring extremes. *Sci. Data*, **2**, 150066, <https://doi.org/10.1038/sdata.2015.66>.
- Goodison, B. E., P. Y. T. Louie, and D. Yang, 1998: WMO solid precipitation intercomparison. World Meteorological Organization Tech. Rep. WMO/TD-872, 318 pp.
- Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez, 2009: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.*, **377**, 80–91, <https://doi.org/10.1016/j.jhydrol.2009.08.003>.
- Haberlandt, U., 2007: Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. *J. Hydrol.*, **332**, 144–157, <https://doi.org/10.1016/j.jhydrol.2006.06.028>.
- Habib, E., A. Henschke, and R. F. Adler, 2009: Evaluation of TMPA satellite-based research and real-time rainfall estimates during six tropical-related heavy rainfall events over Louisiana, USA. *Atmos. Res.*, **94**, 373–388, <https://doi.org/10.1016/j.atmosres.2009.06.015>.
- Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. Jones, and M. New, 2008: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *J. Geophys. Res.*, **113**, D20119, <https://doi.org/10.1029/2008JD010201>.
- 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>.
- Holleman, I., 2006: Bias adjustment of radar-based 3-hour precipitation accumulations. Royal Netherlands Meteorological Institute Tech. Rep. TR-290, 56 pp.
- 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. Meteor.*, **43**, 1834–1853, <https://doi.org/10.1175/JAM2173.1>.
- , —, H. Moradkhani, and S. Sorooshian, 2006: Uncertainty quantification of satellite precipitation estimation and Monte Carlo assessment of the error propagation into hydrologic response. *Water Resour. Res.*, **42**, W08421, <https://doi.org/10.1029/2005WR004398>.
- Huffman, G. J., R. F. Adler, B. Rudolf, U. Schneider, and P. R. Keehn, 1995: Global precipitation



- estimates based on a technique for combining satellite-based estimates, rain gauge analysis, and NWP model precipitation information. *J. Climate*, **8**, 1284–1295, [https://doi.org/10.1175/1520-0442\(1995\)008<1284:GPEBOA>2.0.CO;2](https://doi.org/10.1175/1520-0442(1995)008<1284:GPEBOA>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, 2014: NASA Global Precipitation Measurement (GPM) Integrated Multi-Satellite Retrievals for GPM (IMERG). NASA/GSFC Algorithm Theoretical Basis Doc., 30 pp.
- Hutchinson, M. F., 1998: Interpolation of rainfall data with thin plate smoothing splines—Part I: Two dimensional smoothing of data with short range correlation. *J. Geogr. Inf. Decis. Anal.*, **2**, 168–185.
- Janowiak, J. E., 1992: Tropical rainfall: A comparison of satellite-derived rainfall estimates with model precipitation forecasts, climatologies, and observations. *Mon. Wea. Rev.*, **120**, 448–462, [https://doi.org/10.1175/1520-0493\(1992\)120<0448:TRACOS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1992)120<0448:TRACOS>2.0.CO;2).
- Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xi, 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).
- Kang, S., and J.-B. Ahn, 2015: Global energy and water balances in the latest reanalyses. *Asia-Pac. J. Atmos. Sci.*, **51**, 293–302, <https://doi.org/10.1007/s13143-015-0079-0>.
- Karbalae, N., K. Hsu, S. Sorooshian, and D. Braithwaite, 2017: Bias adjustment of infrared based rainfall estimation using passive microwave satellite rainfall data. *J. Geophys. Res. Atmos.*, **122**, 3859–3876, <https://doi.org/10.1002/2016JD026037>.
- Kauffeldt, A., S. Halldin, A. Rodhe, C.-Y. Xu, and I. K. Westerberg, 2013: Disinformative data in large-scale hydrological modelling. *Hydrol. Earth Syst. Sci.*, **17**, 2845–2013, <https://doi.org/10.5194/hess-17-2845-2013>.
- Kenney, J. F., and E. S. Keeping, 1962: Linear regression and correlation. *Mathematics of Statistics: Part 1*, J. F. Kenney, Ed., Van Nostrand, 252–285.
- Kidd, C., and V. Levizzani, 2011: Status of satellite precipitation retrievals. *Hydrol. Earth Syst. Sci.*, **15**, 1109–1116, <https://doi.org/10.5194/hess-15-1109-2011>.
- Kling, H., M. Fuchs, and M. Paulin, 2012: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J. Hydrol.*, **424–425**, 264–277, <https://doi.org/10.1016/j.jhydrol.2012.01.011>.
- Knapp, K. R., and Coauthors, 2011: Globally gridded satellite observations for climate studies. *Bull. Amer. Meteor. Soc.*, **92**, 893–907, <https://doi.org/10.1175/2011BAMS3039.1>.
- 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>.
- Kuligowski, R. J., 1997: An overview of National Weather Service quantitative precipitation estimates. National Weather Service Techniques Development Laboratory Office Note 97-4, 30 pp.
- Kummerow, C., W. Barnes, T. Kozu, J. Shiue, and J. Simpson, 1998: The Tropical Rainfall Measuring Mission (TRMM) sensor package. *J. Atmos. Oceanic Technol.*, **15**, 809–817, [https://doi.org/10.1175/1520-0426\(1998\)015<0809:TTRMMT>2.0.CO;2](https://doi.org/10.1175/1520-0426(1998)015<0809:TTRMMT>2.0.CO;2).
- Legates, D. R., 1988: A climatology of global precipitation. Ph.D. thesis, University of Delaware, 91 pp.
- , 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>.
- Levizzani, V., and Coauthors, 2018: The activities of the International Precipitation Working Group. *Quart. J. Roy. Meteor. Soc.*, **144**, 3–15, <https://doi.org/10.1002/qj.3214>.
- Liebmman, B., and D. Allured, 2005: Daily precipitation grids for South America. *Bull. Amer. Meteor. Soc.*, **86**, 1567–1570, <https://doi.org/10.1175/BAMS-86-11-1567>.
- Lin, Y., and K. E. Mitchell, 2005: The NCEP stage II/IV hourly precipitation analyses: Development and applications. *19th Conf. Hydrology*, San Diego, CA, Amer. Meteor. Soc., 1.2, <https://ams.confex.com/ams/pdfpapers/83847.pdf>.
- Liu, J., T. Xiao, and L. Chen, 2011: Intercomparisons of air–sea heat fluxes over the Southern Ocean. *J. Climate*, **24**, 1198–1211, <https://doi.org/10.1175/2010JCLI3699.1>.
- Liu, Q., T. R. McVicar, Z. Yang, R. J. Donohue, L. Liang, and Y. Yang, 2016: The hydrological effects of varying vegetation characteristics in a temperate water-limited basin: Development of the dynamic Budyko-Choudhury-Porporato (dBCP) model. *J. Hydrol.*, **543**, 595–611, <https://doi.org/10.1016/j.jhydrol.2016.10.035>.
- Lopez, P., 2007: Cloud and precipitation parameterizations in modeling and variational data assimilation: A review. *J. Atmos. Sci.*, **64**, 3766–3784, <https://doi.org/10.1175/2006JAS2030.1>.

- 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., 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>.
- Massari, C., W. Crow, and L. Brocca, 2017: An assessment of the accuracy of global rainfall estimates without ground-based observations. *Hydrol. Earth Syst. Sci.*, **21**, 4347–4361, <https://doi.org/10.5194/hess-21-4347-2017>.
- Mega, T., T. Ushio, T. Kubota, M. Kachi, K. Aonashi, and S. Shige, 2014: Gauge adjusted global satellite mapping of precipitation (GSMaP\_Gauge). *31st General Assembly and Scientific Symp.*, Beijing, China, Union Radio-Scientifique Internationale, 14694215, <https://doi.org/10.1109/URSIGASS.2014.6929683>.
- Menne, M. J., I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston, 2012: An overview of the Global Historical Climatology Network-Daily database. *J. Atmos. Oceanic Technol.*, **29**, 897–910, <https://doi.org/10.1175/JTECH-D-11-00103.1>.
- Michaelides, S., V. Levizzani, E. Anagnostou, P. Bauer, T. Kasparis, and J. E. Lane, 2009: Precipitation: Measurement, remote sensing, climatology and modeling. *Atmos. Res.*, **94**, 512–533, <https://doi.org/10.1016/j.atmosres.2009.08.017>.
- Nair, A. S., and J. Indu, 2017: Performance assessment of Multi-Source Weighted-Ensemble Precipitation (MSWEP) product over India. *Climate*, **5**, 2, <https://doi.org/10.3390/cli5010002>.
- Oki, T., and S. Kanae, 2006: Global hydrological cycles and world water resources. *Science*, **313**, 1068–1072, <https://doi.org/10.1126/science.1128845>.
- Osborn, T. J., and M. Hulme, 1997: Development of a relationship between station and grid-box rainfall frequencies for climate model evaluation. *J. Climate*, **10**, 1885–1908, [https://doi.org/10.1175/1520-0442\(1997\)010<1885:DOARBS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010<1885:DOARBS>2.0.CO;2).
- Papagiannopoulou, C., D. G. Miralles, W. A. Dorigo, N. E. C. Verhoest, M. Depoorter, and W. Waegeman, 2017a: Vegetation anomalies caused by antecedent precipitation in most of the world. *Environ. Res. Lett.*, **12**, 074016, <https://doi.org/10.1088/1748-9326/aa7145>.
- , —, N. E. C. Verhoest, W. Dorigo, and W. Waegeman, 2017b: A non-linear Granger causality framework to investigate climate–vegetation dynamics. *Geosci. Model Dev.*, **10**, 1945–1960, <https://doi.org/10.5194/gmd-10-1945-2017>.
- Prein, A. F., and A. Gobiet, 2017: Impacts of uncertainties in European gridded precipitation observations on regional climate analysis. *Int. J. Climatol.*, **37**, 305–327, <https://doi.org/10.1002/joc.4706>.
- 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>.
- Roe, G. H., 2005: Orographic precipitation. *Annu. Rev. Earth Planet. Sci.*, **33**, 645–671, <https://doi.org/10.1146/annurev.earth.33.092203.122541>.
- Romanova, V., A. Köhl, D. Stammer, C. Klepp, A. Andersson, and S. Bakan, 2010: Sea surface freshwater flux estimates from GECCO, HOAPS and NCEP. *Tellus*, **62A**, 435–452, <https://doi.org/10.1111/j.1600-0870.2010.00447.x>.
- Satgé, F., and Coauthors, 2017: Role of climate variability and human activity on Poopó Lake droughts between 1990 and 2015 assessed using remote sensing data. *Remote Sens.*, **9**, 218, <https://doi.org/10.3390/rs9030218>.
- Schellekens, J., and Coauthors, 2017: A global water resources ensemble of hydrological models: The EarthH2Observe tier-1 dataset. *Earth Syst. Sci. Data*, **9**, 389–413, <https://doi.org/10.5194/essd-9-389-2017>.
- Schlosser, C. A., and P. R. Houser, 2007: Assessing a satellite-era perspective of the global water cycle. *J. Climate*, **20**, 1316–1338, <https://doi.org/10.1175/JCLI4057.1>.
- Schneider, T., T. Bischoff, and G. H. Haug, 2014: Migrations and dynamics of the intertropical convergence zone. *Nature*, **513**, 45–53, <https://doi.org/10.1038/nature13636>.
- 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>.
- Scofield, R. A., and R. J. Kuligowski, 2003: Status and outlook of operational satellite precipitation algorithms for extreme-precipitation events. *Wea. Forecasting*, **18**, 1037–1051, [https://doi.org/10.1175/1520-0434\(2003\)018<1037:SAOOOS>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<1037:SAOOOS>2.0.CO;2).
- Sevruk, B., M. Ondrás, and B. Chvíla, 2009: The WMO precipitation measurement intercomparisons. *Atmos. Res.*, **92**, 376–380, <https://doi.org/10.1016/j.atmosres.2009.01.016>.

- Skok, G., N. Žagar, L. Honzak, R. Žabkar, J. Rakovec, and A. Ceglar, 2015: Precipitation intercomparison of a set of satellite- and raingauge-derived datasets, ERA Interim reanalysis, and a single WRF regional climate simulation over Europe and the North Atlantic. *Theor. Appl. Climatol.*, **123**, 217–232, <https://doi.org/10.1007/s00704-014-1350-5>.
- Sorooshian, S., K.-L. Hsu, X. Gao, H. V. Gupta, B. Imam, and D. Braithwaite, 2000: Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Amer. Meteor. Soc.*, **81**, 2035–2046, [https://doi.org/10.1175/1520-0477\(2000\)081<2035:EOPSSE>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<2035:EOPSSE>2.3.CO;2).
- Stephens, G. L., and C. D. Kummerow, 2007: The remote sensing of clouds and precipitation from space: A review. *J. Atmos. Sci.*, **64**, 3742–3765, <https://doi.org/10.1175/2006JAS2375.1>.
- , and Coauthors, 2010: Dreary state of precipitation in global models. *J. Geophys. Res.*, **115**, D24211, <https://doi.org/10.1029/2010JD014532>.
- Sun, Q., C. Miao, Q. Duan, H. Ashouri, S. Sorooshian, and K.-L. Hsu, 2018: A review of global precipitation datasets: Data sources, estimation, and inter-comparisons. *Rev. Geophys.*, **56**, 79–107, <https://doi.org/10.1002/2017RG000574>.
- Sun, Y., S. Solomon, A. Dai, and R. W. Portmann, 2006: How often does it rain? *J. Climate*, **19**, 916–934, <https://doi.org/10.1175/JCLI3672.1>.
- Tarnavsky, E., D. Grimes, R. Maidment, E. Black, R. P. Allan, M. Stringer, R. Chadwick, and F. Kayitakire, 2014: Extension of the TAMSAT satellite-based rainfall monitoring over Africa and from 1983 to present. *J. Appl. Meteor. Climatol.*, **53**, 2805–2822, <https://doi.org/10.1175/JAMC-D-14-0016.1>.
- 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>.
- , and Coauthors, 2009: Component analysis of errors in satellite-based precipitation estimates. *J. Geophys. Res.*, **114**, D24101, <https://doi.org/10.1029/2009JD011949>.
- Trenberth, K. E., and Y. Zhang, 2018: How often does it really rain? *Bull. Amer. Meteor. Soc.*, **99**, 289–298, <https://doi.org/10.1175/BAMS-D-17-0107.1>.
- , L. Smith, T. Qian, A. Dai, and J. Fasullo, 2007: Estimates of the global water budget and its annual cycle using observational and model data. *J. Hydrometeorol.*, **8**, 758–769, <https://doi.org/10.1175/JHM600.1>.
- , J. T. Fasullo, and J. Mackaro, 2011: Atmospheric moisture transports from ocean to land and global energy flows in reanalyses. *J. Climate*, **24**, 4907–4924, <https://doi.org/10.1175/2011JCLI4171.1>.
- 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>.
- Vicente, G. A., R. A. Scofield, and W. P. Menzel, 1998: The operational GOES infrared rainfall estimation technique. *Bull. Amer. Meteor. Soc.*, **79**, 1883–1898, [https://doi.org/10.1175/1520-0477\(1998\)079<1883:TOGIRE>2.0.CO;2](https://doi.org/10.1175/1520-0477(1998)079<1883:TOGIRE>2.0.CO;2).
- Vila, D. A., L. G. G. de Goncalves, D. L. Toll, and J. R. Rozante, 2009: Statistical evaluation of combined daily gauge observations and rainfall satellite estimates over continental South America. *J. Hydrometeorol.*, **10**, 533–543, <https://doi.org/10.1175/2008JHM1048.1>.
- Villarini, G., and W. F. Krajewski, 2007: Evaluation of the research version TMPA three-hourly  $0.25^\circ \times 0.25^\circ$  rainfall estimates over Oklahoma. *Geophys. Res. Lett.*, **34**, L05402, <https://doi.org/10.1029/2006GL029147>.
- Viney, N. R., and B. C. Bates, 2004: It never rains on Sunday: The prevalence and implications of untagged multi-day rainfall accumulations in the Australian high quality data set. *Int. J. Climatol.*, **24**, 1171–1192, <https://doi.org/10.1002/joc.1053>.
- Wang, Y., Z. Guo, and G. Li, 2017: Precipitation estimation and analysis of the Three Gorges Dam region (1979–2014) by combining gauge measurements and MSWEP with generalized additive model. *Acta Geogr. Sin.*, **72**, 1207–1220.
- Weatherhead, E. C., and Coauthors, 1998: Factors affecting the detection of trends: Statistical considerations and applications to environmental data. *J. Geophys. Res.*, **103**, 17 149–17 161, <https://doi.org/10.1029/98JD00995>.
- Weedon, G. P., and Coauthors, 2011: Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century. *J. Hydrometeorol.*, **12**, 823–848, <https://doi.org/10.1175/2011JHM1369.1>.
- , 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>.
- Wolff, D. B., and B. L. Fisher, 2008: Comparisons of instantaneous TRMM ground validation and satellite rain-rate estimates at different spatial scales. *J. Appl. Meteor. Climatol.*, **47**, 2215–2237, <https://doi.org/10.1175/2008JAMC1875.1>.
- Wood, E. F., and Coauthors, 2011: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial

- water. *Water Resour. Res.*, **47**, W05301, <https://doi.org/10.1029/2010WR010090>.
- 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).
- , M. Chen, S. Yang, A. Yatagai, T. Hayasaka, Y. Fukushima, and C. Liu, 2007: A gauge-based analysis of daily precipitation over East Asia. *J. Hydrometeorol.*, **8**, 607–626, <https://doi.org/10.1175/JHM583.1>.
- , R. Joyce, S. Wu, S.-H. Yoo, Y. Yarosh, F. Sun, and R. Lin, 2017: Reprocessed, bias-corrected CMORPH global high-resolution precipitation estimates from 1998. *J. Hydrometeorol.*, **18**, 1617–1641, <https://doi.org/10.1175/JHM-D-16-0168.1>.
- Yang, Y., R. J. Donohue, and T. R. McVicar, 2016: Global estimation of effective plant rooting depth: Implications for hydrological modeling. *Water Resour. Res.*, **52**, 8260–8276, <https://doi.org/10.1002/2016WR019392>.
- Yatagai, A., K. Kamiguchi, O. Arakawa, A. Hamada, N. Yasutomi, and A. Kitoh, 2012: APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Bull. Amer. Meteor. Soc.*, **93**, 1401–1415, <https://doi.org/10.1175/BAMS-D-11-00122.1>.
- Yin, X., A. Gruber, and P. Arkin, 2004: Comparison of the GPCP and CMAP merged gauge–satellite monthly precipitation products for the period 1979–2001. *J. Hydrometeorol.*, **5**, 1207–1222, <https://doi.org/10.1175/JHM-392.1>.
- Zambrano-Bigiarini, M., A. Nauditt, C. Birkel, K. Verbist, and L. Ribbe, 2017: Temporal and spatial evaluation of satellite-based rainfall estimates across the complex topographical and climatic gradients of Chile. *Hydrol. Earth Syst. Sci.*, **21**, 1295–1320, <https://doi.org/10.5194/hess-21-1295-2017>.
- Zhang, L., W. R. Dawes, and G. R. Walker, 2001: Response of mean annual evapotranspiration to vegetation changes at catchment scale. *Water Resour. Res.*, **37**, 701–708, <https://doi.org/10.1029/2000WR900325>.
- Zhang, W., M. Brandt, F. Guichard, Q. Tian, and R. Fensholt, 2017: Using long-term daily satellite based rainfall data (1983–2015) to analyze spatio-temporal changes in the Sahelian rainfall regime. *J. Hydrol.*, **550**, 427–440, <https://doi.org/10.1016/j.jhydrol.2017.05.033>.
- Zhang, X., E. N. Anagnostou, and C. S. Schwartz, 2018: NWP-based adjustment of IMERG precipitation for flood-inducing complex terrain storms: Evaluation over CONUS. *Remote Sens.*, **10**, 642, <https://doi.org/10.3390/rs10040642>.
- Zhu, Y., and Y. Luo, 2015: Precipitation calibration based on the frequency-matching method. *Wea. Forecasting*, **30**, 1109–1124, <https://doi.org/10.1175/WAF-D-13-00049.1>.
- Zohaib, M., H. Kim, and M. Choi, 2017: Evaluating the patterns of spatiotemporal trends of root zone soil moisture in major climate regions in East Asia. *J. Geophys. Res. Atmos.*, **122**, 7705–7722, <https://doi.org/10.1002/2016JD026379>.
- Zolina, O., A. Kapala, C. Simmer, and S. K. Gulev, 2004: Analysis of extreme precipitation over Europe from different reanalyses: A comparative assessment. *Global Planet. Change*, **44**, 129–161, <https://doi.org/10.1016/j.gloplacha.2004.06.009>.