



# AMERICAN METEOROLOGICAL SOCIETY

*Bulletin of the American Meteorological Society*

## **EARLY ONLINE RELEASE**

This is a preliminary PDF of the author-produced manuscript that has been peer-reviewed and accepted for publication. Since it is being posted so soon after acceptance, it has not yet been copyedited, formatted, or processed by AMS Publications. This preliminary version of the manuscript may be downloaded, distributed, and cited, but please be aware that there will be visual differences and possibly some content differences between this version and the final published version.

The DOI for this manuscript is doi: 10.1175/BAMS-D-17-0138.1

The final published version of this manuscript will replace the preliminary version at the above DOI once it is available.

If you would like to cite this EOR in a separate work, please use the following full citation:

Beck, H., E. Wood, M. Pan, C. Fisher, D. Miralles, A. van Dijk, T. McVicar, and R. Adler, 2018: MSWEP V2 global 3-hourly 0.1° precipitation: methodology and quantitative assessment. Bull. Amer. Meteor. Soc. doi:10.1175/BAMS-D-17-0138.1, in press.



# MSWEP V2 global 3-hourly $0.1^\circ$ precipitation: methodology and quantitative assessment

Hylke E. Beck\*, Eric F. Wood, Ming Pan, and Colby K. Fisher

*Princeton University, Department of Civil and Environmental Engineering, Princeton, NJ, USA*

Diego G. Miralles

*Ghent University, Laboratory of Hydrology and Water Management, Ghent, Belgium*

Albert I.J.M. van Dijk

*Australian National University, Fenner School of Environment & Society, Canberra, ACT,  
Australia*

Tim R. McVicar

*CSIRO Land and Water, Canberra, ACT, Australia*

*Australian Research Council Centre of Excellence for Climate System Science, Sydney, Australia*

Robert F. Adler

*University of Maryland, Earth System Science Interdisciplinary Center, College Park, MD, USA*

\*Corresponding author address: Hylke Beck, Princeton University, Department of Civil and Environmental Engineering, 59 Olden Street, Princeton, NJ 08544, USA

E-mail: hylke.beck@gmail.com



## ABSTRACT

We present Multi-Source Weighted-Ensemble Precipitation (MSWEP) Version 2 (V2), a gridded precipitation ( $P$ ) dataset spanning 1979–2017. MSWEP V2 is unique in several aspects: (i) full global coverage (all land and oceans); (ii) high spatial ( $0.1^\circ$ ) and temporal (3 hourly) resolution; (iii) optimal merging of  $P$  estimates based on gauges (WorldClim, GHCN-D, GSOD, GPCC, and others), satellites (CMORPH, GridSat, GSMaP, and TMPA 3B42RT), and reanalyses (ERA-Interim and JRA-55); (iv) distributional bias corrections, mainly to improve the  $P$  frequency; (v) correction of systematic terrestrial  $P$  biases using river discharge ( $Q$ ) observations from 13 762 stations across the globe; (vi) incorporation of daily observations from 76 747 gauges worldwide; and (vii) correction for regional differences in gauge reporting times. MSWEP V2 compares substantially better with Stage-IV gauge-radar  $P$  data than other state-of-the-art  $P$  datasets for the US, demonstrating the effectiveness of the MSWEP V2 methodology. Global comparisons suggest that MSWEP V2 exhibits more realistic spatial patterns in mean, magnitude, and frequency. Long-term mean  $P$  estimates for the global, land, and ocean domains based on MSWEP V2 are 955, 781, and 1025 mm  $y^{-1}$ , respectively. Other  $P$  datasets consistently underestimate  $P$  amounts in mountainous regions. Using MSWEP V2,  $P$  was estimated to occur 15.5 %, 12.3 %, and 16.9 % of the time on average for the global, land, and ocean domains, respectively. MSWEP V2 provides unique opportunities to explore spatio-temporal variations in  $P$ , improve our understanding of hydrological processes and their parameterization, and enhance hydrological model performance. The dataset is available via [www.gloh2o.org](http://www.gloh2o.org).

## 42 Capsule summary

43 MSWEP V2 is the first fully global precipitation dataset with a  $0.1^\circ$  resolution derived by optimally  
44 merging a range of gauge, satellite, and reanalysis estimates.

## 45 1. Introduction

46 Precipitation ( $P$ ) drives the terrestrial hydrological cycle (Oki and Kanae 2006; Trenberth et al.  
47 2007). It is also among the most difficult meteorological variables to estimate due to its high  
48 spatio-temporal heterogeneity (Daly et al. 1994; Adler et al. 2001; Roe 2005; Stephens et al. 2010;  
49 Herold et al. 2016; Prein and Gobiet 2017). A plethora of regional, quasi-global, and fully global  
50 gridded  $P$  datasets have been developed over the past decades (for an overview see Maggioni et al.  
51 2016; Beck et al. 2017c; Levizzani et al. 2018; Sun et al. 2018; <http://ipwg.isac.cnr.it>; and  
52 <http://reanalyses.org>). These datasets differ in terms of design objective (instantaneous ac-  
53 curacy, temporal homogeneity, record length, or combinations thereof), data source (gauge, ground  
54 radar, satellite, analysis, reanalysis, or combinations thereof), spatial resolution (from  $0.05^\circ$  to  
55  $2.5^\circ$ ), and temporal resolution (30 minutes to monthly).

56 Multi-Source Weighted-Ensemble Precipitation (MSWEP) is a recently released global  $P$  dataset  
57 with a 3-hourly temporal resolution, covering the period 1979 to the near-present (Beck et al.  
58 2017b). The dataset is unique in that it takes advantage of the complementary strengths of gauge-,  
59 satellite-, and reanalysis-based data to provide reliable  $P$  estimates over the entire globe. Since  
60 the release of V1 ( $0.25^\circ$  spatial resolution) in May 2016, MSWEP has been successfully ap-  
61 plied at global scales for a variety of purposes, such as modeling soil moisture and evaporation  
62 (Martens et al. 2017), estimating plant rooting depth (Yang et al. 2016), water resources reanalysis  
63 (Schellekens et al. 2017), and evaluating climatic controls on vegetation (Papagiannopoulou et al.  
64 2017a,b). MSWEP has also been successfully used for several purposes regionally, for example, to

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

71 Since the release of MSWEP V1, considerable improvements were implemented, resulting in  
72 MSWEP V2, the focus of the present study. Improvements include: (i) the introduction of cu-  
73 mulative distribution function (CDF) and  $P$  frequency corrections, to account for spurious drizzle,  
74 attenuated peaks, and temporal discontinuities evident in V1 (Nair and Indu 2017; Zhang et al.  
75 2017); (ii) increasing spatial resolution from  $0.25^\circ$  to  $0.1^\circ$  to increase the local relevance of the  
76  $P$  estimates (especially important for high water-yield mountainous regions); (iii) the inclusion  
77 of ocean areas to enable oceanic studies and terrestrial hydrology studies for coastal areas and  
78 small islands; (iv) the addition of  $P$  estimates derived from Gridded Satellite (GridSat) thermal  
79 infrared (IR) imagery (Knapp et al. 2011) for the pre-TRMM era to supplement the reanalysis and  
80 gauge data; (v) the use of a daily (rather than monthly) gauge correction scheme that accounts for  
81 regional differences in reporting times, to minimize timing mismatches when applying the daily  
82 gauge corrections; (vi) the use of a large database of daily gauge observations compiled from  
83 several sources to replace the  $0.5^\circ$  CPC Unified dataset (Xie et al. 2007; Chen et al. 2008); and  
84 (vii) extension of the data record to 2017 (MSWEP V1 finished in 2015).

85 MSWEP V2 is the first fully global  $P$  dataset with a spatial resolution of  $0.1^\circ$  (11 km at the  
86 equator), supporting global-scale land surface modeling at hyper-resolution (Wood et al. 2011;  
87 Bierkens et al. 2015). Other  $P$  datasets with a high spatial resolution ( $\leq 0.1^\circ$ ) include CHIRPS  
88 ( $0.05^\circ$ ; Funk et al. 2015b), CMORPH ( $0.07^\circ$ ; Joyce et al. 2004), GSMaP ( $0.1^\circ$ ; Ushio et al. 2009;

89 Mega et al. 2014), IMERG (0.1°; Huffman et al. 2014), and PERSIANN-CCS (0.04°; Hong et al.  
90 2004). However, these datasets are limited to latitudes  $\leq 60^\circ\text{N/S}$  ( $\leq 50^\circ\text{N/S}$  for CHIRPS), do  
91 not take advantage of river discharge ( $Q$ ) observations for bias correction, and do not incorporate  
92 reanalysis-based  $P$  estimates (with the arguable exception of CHIRPS, which uses them to tem-  
93 porally disaggregate from 5-day to daily estimates). Additionally, none of these datasets apply  $P$   
94 gauge corrections at the daily time scale, with the exception of GSMaP, although it fails to account  
95 for differences in gauge reporting times. Moreover, CHIRPS and PERSIANN-CCS do not inte-  
96 grate passive microwave-based  $P$  retrievals, and the daily temporal resolution of CHIRPS renders  
97 it less suitable in highly dynamic  $P$  environments. Finally, with the exception of CHIRPS, these  
98 datasets span  $\leq 20$  years, which is less optimal to assess long-term hydrological changes/trends  
99 (Weatherhead et al. 1998).

100 Here, we describe the data and methodology underlying MSWEP V2, evaluate the performance  
101 of the dataset for the conterminous US (CONUS), and assess spatio-temporal  $P$  patterns globally.

## 102 **2. Data and methods**

### 103 *a. MSWEP V2 methodology*

104 Figure 1a flowcharts the main processing steps implemented to produce MSWEP V2. The com-  
105 plete methodology is provided in the Appendix. The main steps can be summarized as follows:

- 106 1. Daily  $P$  gauge observations were used for three purposes: (i) to determine the merg-  
107 ing weights for the six 3-hourly non-gauge-based  $P$  datasets incorporated in MSWEP V2  
108 (CMORPH, ERA-Interim, GridSat, GSMaP, JRA-55, and TMPA 3B42RT; see Table 1 for  
109 details on the datasets); (ii) to calculate the wet-day biases for the reanalyses (ERA-Interim  
110 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>; Figure 2a). After quality control, a final gauge dataset comprising 76 747 gauges remained (Figure 2b). See Appendix a for details.

2. 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 Appendix b for details.
3. 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 ( $0.07^\circ$  resolution; Knapp et al. 2011), to improve the  $P$  estimates in convection-dominated regions. Rainfall was estimated using a parsimonious CDF-matching approach. See Appendix c for details.
4. 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 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  $\beta_{WD}$  (unitless) is the bias in number of wet days, and  $WD_{gridded}$  and  $WD_{gauge}$  represent the mean annual number of wet days in the reanalysis and the gauge observations, respectively.  $WD_{gridded}$  was computed from daily accumulations to be consistent with the gauge observations. Wet days were identified using a  $0.5 \text{ mm d}^{-1}$  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 Appendix d for details.

5. 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  $\beta_{WD}$  were produced for the reanalyses, to correct the  $P$  frequency prior to the data merging. See Appendix e for details.
6. MSWEP V1 used CHPclim ( $0.05^\circ$  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), due to the better  $P$  gauge coverage. Systematic  $P$  underestimation over land due to gauge under-catch and orographic effects was corrected similarly to MSWEP V1, by inferring the “true”  $P$  using river discharge ( $Q$ ) observations. See Appendix f 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 (Figure 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 Appendix g 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 Appendix h 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 (Figures 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 Appendix i 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 (Figures 1e and 1f). We subsequently applied monthly gauge corrections using the GPCC FDR V7 dataset ( $0.5^\circ$  resolution; Schneider et al. 2014b), 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 Appendix j for details.

174 *b. Evaluation using Stage-IV gauge-radar data for the CONUS*

175 We evaluated the performance of MSWEP V2, and for the sake of comparison, MSWEP V1,  
 176 a widely used satellite-based dataset (CMORPH), a widely used reanalysis (ERA-Interim), and  
 177 a state-of-the-art reanalysis corrected using daily gauge observations (MERRA-2; Table 1). The  
 178 evaluation was performed at a 3-hourly temporal and  $0.1^\circ$  spatial resolution for 2002–2015. As  
 179 reference, we used the National Centers for Environmental Prediction (NCEP) Stage-IV dataset  
 180 (Lin and Mitchell 2005), which has a  $0.04^\circ$  spatial and hourly temporal resolution, and merges data  
 181 from 140 radars and  $\sim 5500$  gauges for the CONUS. Stage-IV provides high-quality  $P$  estimates  
 182 and has therefore been widely used as reference for the evaluation of  $P$  datasets (e.g., Hong et al.  
 183 2006; Habib et al. 2009; AghaKouchak et al. 2011, 2012; Zhang et al. 2018). To reduce systematic  
 184 biases, the Stage-IV dataset was rescaled such that its long-term mean matches that of the PRISM  
 185 dataset (Daly et al. 2008) for the evaluation period (2002–2015).

186 As performance metric, we used the Kling-Gupta Efficiency (KGE; Gupta et al. 2009; Kling  
 187 et al. 2012), an objective performance metric combining correlation, bias, and variability, intro-  
 188 duced in Gupta et al. (2009) and modified in Kling et al. (2012). The KGE is calculated as follows:

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

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

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

192 where  $\mu$  and  $\sigma$  are the distribution mean and standard deviation, respectively, and the subscripts  
 193  $s$  and  $o$  indicate estimate and reference, respectively. Three-hourly accumulations were calculated  
 194 for the  $P$  datasets with a temporal resolution  $< 3$  h (CMORPH, MERRA-2, and Stage-IV). The

195  $P$  datasets with a spatial resolution  $> 0.1^\circ$  (MSWEP V1, ERA-Interim, and MERRA-2) were  
196 downscaled to  $0.1^\circ$  using nearest neighbour, while the dataset with a spatial resolution  $< 0.1^\circ$   
197 (CMORPH) was upsampled to  $0.1^\circ$  using bilinear interpolation.

### 198 **3. Results and discussion**

#### 199 *a. Gauge reporting times*

200 For the GHCN-D database, we found marked differences in reporting times between neighbor-  
201 ing countries (e.g., between Canada and USA, and Portugal and Spain) and sometimes within  
202 countries (e.g., Mexico, Namibia, and South Africa; Figure 2c), reflecting differences in report-  
203 ing practices among hydrological and meteorological agencies. Our reporting times correspond  
204 well with published times available for Australia (Viney and Bates 2004), Brazil (Liebmann and  
205 Allured 2005), the eastern CONUS (DeGaetano 2000), India (Yatagai et al. 2012), the Nether-  
206 lands (Holleman 2006), and Japan (Yatagai et al. 2012). Although the GSOD gauges represent  
207 automated gauges with reporting times officially at around midnight UTC (Menne et al. 2012),  
208 our analysis yielded considerably earlier reporting times averaging at around  $-12$  h UTC (except  
209 for eastern Australia; Figure 2d). A potential explanation for this discrepancy could be that satel-  
210 lites represent radiation from an atmospheric column rather than  $P$  that has reached the surface.  
211 However, Villarini and Krajewski (2007) obtained timing differences ranging from 30 to 90 min  
212 for TMPA 3B42 using 5-min rain gauge data for a single  $0.25^\circ$  grid-cell in Oklahoma, suggesting  
213 that this explanation is insufficient to account for the full 12-h difference. Additionally, the differ-  
214 ences are also found in high-latitudes ( $> 60^\circ\text{N/S}$ ), where the reporting times were inferred using  
215 reanalysis data. An alternative, more likely explanation is that the daily GSOD values incorporate

216 a significant portion of  $P$  from the previous day. Overall, these results highlight the importance of  
217 accounting for reporting times in time-critical applications relying on daily gauge observations.

218 *b. Gauge-based assessment of satellite and reanalysis datasets*

219 Figures 3a and 3b present  $r_{3\text{ day}}$  (temporal correlation) values obtained for CMORPH and ERA-  
220 Interim, respectively. Since the results were very similar for all satellite datasets (with the excep-  
221 tion of GridSat) and for all reanalysis datasets, we only present results for one dataset of each  
222 kind. ERA-Interim is most skillful in mid- and high-latitude coastal regions in the path of the  
223 prevailing westerlies (notably along the Pacific coast of North America, in southern Chile, and in  
224 western Europe; Figure 3a), whereas CMORPH performs best in moist mid-latitude regions with  
225 mild winters (e.g., the southeastern US, eastern South America, and eastern China; Figure 3b).  
226 When we calculate the difference in  $r_{3\text{ day}}$  values between the datasets, a clear picture emerges:  
227 CMORPH consistently performs better at low-latitudes and ERA-Interim at high-latitudes (Fig-  
228 ure 3d). These results underscore the long-recognized but sometimes overlooked complementary  
229  $P$  estimation performance of satellites and weather models (e.g., Janowiak 1992; Huffman et al.  
230 1995; Xie and Arkin 1997; Adler et al. 2001; Ebert et al. 2007; Massari et al. 2017). MSWEP is the  
231 only  $P$  dataset besides CMAP (Xie and Arkin 1997) to exploit this complementary relationship.

232 Figure 3c presents  $r_{3\text{ day}}$  values for the GridSat IR-based rainfall dataset, which has been pro-  
233 duced to complement the gauge and reanalysis data during the pre-TRMM era (Appendix c). The  
234  $r_{3\text{ day}}$  values for GridSat are consistently lower than those obtained for CMORPH (Figure 3a),  
235 which was expected since cloud-top IR brightness temperatures are only indirectly related to sur-  
236 face rainfall (Adler and Negri 1988; Vicente et al. 1998; Scofield and Kuligowski 2003). Com-  
237 pared to  $r_{3\text{ day}}$  values obtained using the IR-based PERSIANN dataset (Sorooshian et al. 2000)  
238 presented in Beck et al. (2017b, their Figure 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  $\beta_{WD}$  (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 (Figure 3e). Conversely, ERA-Interim strongly overestimates the  $P$  frequency across the entire globe (Figure 3f), due to 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 (Appendix g). When interpreting these results, it must be kept in mind that point observations from gauges tend to underestimate 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).

### *c. 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) 1979–1982; (ii) 1983–1999; and (iii) 2000–2017. The gauge weights were calculated as a function of distance to surrounding gauges (Appendix j), whereas the satellite and reanalysis weights were calculated based on the performance of the respective satellite and reanalysis datasets at surrounding gauges (Appendix e). Gauge-based  $P$  estimates provide the main contribution over the terrestrial surface for all periods (Figure 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 striping pattern can be observed in some regions caused by gaps in the GridSat data (Figure 4a). In regions without rain gauges, reanalyses provide the dominant contribution over most of the globe until 1999, while from 2000 onwards the dominant contribution comes from satellite data at low and mid latitudes and reanalysis data at high-latitudes (Figure 4).

#### *d. 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 9053 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 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 (Figure 5a). The second and third best performing  $P$  datasets were MSWEP V1 (Beck et al. 2017b; Figure 5e) and MERRA-2 (Reichle et al. 2017; Figure 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 (Section 3a), which has resulted in temporal mismatches when applying the corrections (Figures 1e and 1f). CMORPH (Joyce et al. 2004; Figure 5i) and ERA-Interim (Dee et al. 2011; Figure 5m) obtained

lower median KGE scores of 0.36 and 0.35, respectively. Performance was markedly worse for all datasets in the western CONUS, due to the more complex topography and greater spatio-temporal heterogeneity of  $P$  (Daly et al. 2008).

*e. Global patterns in long-term mean  $P$*

Figure 6a presents a global map of long-term mean  $P$  from MSWEP V2 (Appendix f). Figure 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;  $0.25^\circ$ ; Beck et al. 2017b); (ii) GPCC V2015 (1951–2000; monthly,  $0.5^\circ$ ; Schneider et al. 2014b, 2017); (iii) GPCP V2.3 (1979–2013; monthly,  $2.5^\circ$ ; Adler et al. 2003, 2017, 2018); (iv) HOAPS V3.2 (1987–2008;  $0.5^\circ$ , 6 hourly; Schlosser and Houser 2007; Andersson et al. 2010); and (v) MERRA-2 (1980–2017;  $\sim 50$  km, hourly; Reichle et al. 2017). The differences between MSWEP V1 and V2 (Figure 6b) primarily reflect the change from CHPclim to WorldClim in V2. Compared to MSWEP V2, the fully gauge-based GPCC V2015 dataset shows consistently lower mean  $P$  at high northern latitudes (Figure 6c), whereas the gauge- and satellite-based GPCP V2.3 dataset exhibits lower mean  $P$  only in northern North America and northeastern Asia, but generally higher mean  $P$  in Europe and northwestern Asia (Figure 6d). These differences probably reflect the use of different gauge under-catch correction schemes; GPCC V2015 (Legates and Willmott 1990) and GPCP V2.3 (Legates 1988) employ more conventional approaches using WMO gauge under-catch correction equations in combination with daily observations of  $P$ ,  $T_a$ , and wind speed from a relatively sparse station network. Conversely, MSWEP V2 infers the “true”  $P$  using  $Q$  observations and  $P_e$  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 Africa; Figure 6f). Compared to MSWEP V2,

the other  $P$  datasets (GPCC V2015, GPCP V2.3, and MERRA-2) exhibit substantially less  $P$  at high elevations (e.g., in the Rocky Mountains, the southern Andes, and most Asian mountainous regions; Figures 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 2000–2017, during which the meridional location of the maximum intertropical convergence zone (ITCZ) convection was more northerly (Schneider et al. 2014a).

Using MSWEP V2, we obtained a long-term mean global  $P$  estimate of  $955 \text{ mm y}^{-1}$  (Table 2) or  $488 \text{ 100 km}^3 \text{ yr}^{-1}$ . 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  $839 \text{ mm y}^{-1}$ , corresponding to  $113 \text{ 100 km}^3 \text{ y}^{-1}$ . The same estimate for MSWEP V1 is  $858 \text{ mm y}^{-1}$ , slightly (2.3 %) higher due to the switch from CHPClim to WorldClim and the reduction of the Chilean and Iranian bias correction factors in V2 (Appendix f). The estimate for GPCP V2.3 is  $853 \text{ mm y}^{-1}$ , also slightly (1.7 %) higher than the MSWEP V2 estimate. For GPCC V2015, the corresponding estimate is  $793 \text{ mm y}^{-1}$ , which is considerably (5.5 %) lower for the reasons previously explained. The estimate for MERRA-2 is  $785 \text{ mm y}^{-1}$ , also considerably (6.4 %) lower than the MSWEP V2 estimate, mainly due to the aforementioned differences in tropical and mountainous regions. The long-term mean  $P$  for ocean areas based on MSWEP V2 amounted to  $1025 \text{ mm y}^{-1}$  (Table 2), corresponding to  $373 \text{ 200 km}^3 \text{ y}^{-1}$ . Arguably the most comprehensive  $P$  datasets with ocean coverage currently available are the satellite-based GPCP V2.3 and HOAPS V3.2 datasets. Compared to our estimate, GPCP V2.3 yields a 3.1 % higher estimate of  $1057 \text{ mm y}^{-1}$  (Figure 6d). Over the area for which

HOAPS V3.2 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 (1037 versus 1068 mm y<sup>-1</sup>; Figure 6e). Another estimate of 1074 mm y<sup>-1</sup> for the entire ocean area was derived from satellite radar reflectivities (2007–2009) 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).

#### *f. 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 year) 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 mid-latitude 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) also found CMORPH to overestimate summer  $P$  extremes strongly in the CONUS. As expected, ERA-Interim fails to resolve small-scale orographic features due to its coarse ( $\sim 0.7^\circ$ ) resolution and consistently estimates lower 99.99th percentile  $P$  amounts due to 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 and Indices (ETCCDI)  $P$  dataset (Dietzsch et al. 2017, their Figure 5d), our 99.99th percentile 3-hourly  $P$  map (Figure 7a) exhibits more plausible patterns. Most importantly, Dietzsch et al.’s (2017) map fails to represent small-scale  $P$  variations, due mainly to 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 pres-

ence of slight discontinuities in the MSWEP V2 map at approximately 50°S (Figure 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 Figure 7a) reflect the loss of variance between  $P$  gauges due to interpolation (Hutchinson 1998; Haberlandt 2007).

#### *g. 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 (Figure 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 (Figure 3f), ERA-Interim severely overestimates the  $P$  frequency across the entire globe. Our  $P$  frequency map (Figure 8a) visually compares well with an equivalent map for the land surface derived from gauge observations from 1840–2001 produced by Sun et al. (2006, their Figure 1). Additionally, our map agrees closely with ocean maps based on CloudSat data from 2006–2007 (Ellis et al. 2009, their Figure 3a) and CloudSat, TRMM, and AMSR-E data from 2007–2009 (Behrangi et al. 2014, their Figure 1a). We did, however, obtain a somewhat higher  $P$  frequency over the Southern Ocean, possibly due to uncertainties in the  $P$  frequency corrections caused by the near-complete absence of gauges south of 60°S (Figure 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,  $0.25^\circ$  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 3h<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  $\sim 25$  % higher than the ones calculated using hourly data. Based on MSWEP V2 (3-hourly, 0.1° resolution), using the same 0.06 mm 3h<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 ( $\sim 9$  %) 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 Figure 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 due to the detection limits of satellite sensors ( $\sim 0.8$  mm h<sup>-1</sup> over land and  $\sim 0.02$  mm h<sup>-1</sup> over ocean; Wolff and Fisher 2008) and rain gauges ( $\sim 0.25$  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).

#### h. Trends in mean annual $P$

Figure 9 presents global maps of the linear trend in mean annual  $P$  for MSWEP V2 and V1, CHIRPS V2.0, CMAP V1707, GPCC FDR V7, GPCP V2.3, and HOAPS V3.2 (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 CMAP V1707 and MERRA-2), which was expected since all datasets use similar gauge data sources. MERRA-2 exhibits suspect trend patterns over tropical land areas (Figure 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 V2 (e.g., over the Amazon and the southwest Indian Ocean islands; Figures 9a and 9b) are attributable to the new daily gauge data (Appendix j). The correspondence in trends is considerably less over the oceans, presumably due to the lack of gauge observations to constrain uncertainty (Figure 2b). CMAP V1707 (and MERRA-2, which has been bias adjusted using CMAP over the oceans) generally tends more toward negative trends (Figure 9d), which Yin et al. (2004) attributed to discontinuities caused by changes in gauge coverage and satellite input data. HOAPS V3.2 exhibits a substantially noisier trend pattern and more pronounced trends overall (Figure 9g), which are both likely attributable to its shorter data record. In the Southern Ocean HOAPS V3.2 not only shows  $P$  underestimation (Figure 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 due to 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; Schneider et al. 2014b); 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, due to 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 2000–2016

425 and found that MSWEP V2 exhibits more reliable trends overall than MSWEP V1 as well as other  
426 *P* datasets.

#### 427 **4. Conclusion**

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

- 438 1. There are marked differences in reporting times between neighboring countries and some-  
439 times within countries. Contrary to expectations, the automated GSOD gauges exhibited  
440 reporting times averaging at around 1200 h UTC rather than midnight (i.e., 2400 h) UTC.  
441 These findings underscore the importance of accounting for reporting times when applying  
442 daily gauge corrections.
- 443 2. The gauge-based assessment of the satellite and reanalysis *P* datasets revealed that the reanal-  
444 yses strongly overestimate the *P* frequency across the globe. Confirming previous studies,  
445 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 1025 mm  $y^{-1}$ , respectively. This is in close agreement with the published estimates, yet importantly for hydrological applications other datasets appear to consistently underestimate  $P$  amounts in mountainous regions due to 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 are generally agree over land, at least partly due to 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 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 & Dataset (LACA&D) project: IDEAM (Colombia), INAMEH (Venezuela), INAMHI (Ecuador), SENAMHI (Peru), SENAMHI (Bolivia), and DMC (Chile). We further wish to thank Ali Alijanian, 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 no. 715254 (DRY-2-DRY).

## APPENDIX

Here, we describe in detail the different processing steps involved in the production of MSWEP V2 (Figure 1a).

### *a. Gauge data quality control*

Daily gauge observations were used to determine the merging weights and wet-day biases for the individual  $P$  datasets (Appendix e) and to improve  $P$  estimates near gauge stations (Appendix j). Our initial database comprises 117 759 gauges worldwide compiled from the Global Historical Climatology Network-Daily (GHCN-D) database (Menne et al. 2012), the Global Summary Of the Day (GSOD) database (<https://data.noaa.gov>), the Latin American Climate Assessment & Dataset (LACA&D) database (<http://lacad.ciifen.org/>), the Chile Climate Data Library

490 (<http://www.climatedatalibrary.cl>), and national databases for Mexico, Brazil, Peru, and  
491 Iran.

492 Gauge data can have considerable measurement errors and therefore quality control is important  
493 (Goodison et al. 1998; Viney and Bates 2004; Sevruk et al. 2009; Schneider et al. 2014b). For  
494 example, GSOD records frequently contain long series of erroneous zero rainfall (Durre et al.  
495 2010; Funk et al. 2015b). To identify and discard these periods, we developed an automated  
496 procedure entailing the following steps: (i) for each month, we computed the fraction of days  
497 without  $P$  ( $fD$ ); (ii) we excluded months without any  $P$  ( $fD = 1$ ) and computed the distribution  
498 mean ( $\mu$ ) and standard deviation ( $\sigma$ ); (iii) if the CDF of the normal distribution with  $\mu$  and  $\sigma$   
499 evaluated at  $fD = 0.9$  exceeds 0.85, the gauge was considered to be sufficiently ‘wet’ for detecting  
500 the erroneous zeros and we proceeded to the next step; (iv) a year was marked as erroneous if  
501 the median of the 12 monthly  $fD$  values exceeded 0.9; and (v) the six months preceding and  
502 following each erroneous year were also marked as erroneous. Figure 2a illustrates the procedure  
503 for an arbitrarily selected GSOD gauge with the described issue.

504 Additionally, we eliminated all days with  $P > 2000$  mm (approximately the maximum recorded  
505 24-h rainfall; Cervený et al. 2007), and discarded gauges with record length  $< 4$  years during  
506 1979–2017. From the remaining set of 81 047 gauges we also discarded those matching one  
507 or more of the following criteria (% of remaining gauges satisfying the criteria reported between  
508 parentheses): (i) 3-day Pearson correlation coefficient ( $r_{3\text{ day}}$ ) with five non-gauge-based  $P$  datasets  
509 (CMORPH, ERA-Interim, GSMaP, JRA-55, and TMPA 3B42RT; Table 1)  $< 0.4$ , and  $r_{3\text{ day}}$  with  
510 the nearest gauge also  $< 0.4$  (1.01 %); (ii) more than half of the 3-day intervals contain missing  
511 values (1.62 %); (iii) less than 15 unique values in the entire record (1.02 %); (iv) the highest  
512 and/or second highest values were present  $> 3$  times in the record, indicative of truncated peaks  
513 (0.60 %); and (v)  $> 99.5$  % of the record is dry ( $< 0.5$  mm d<sup>-1</sup>; 3.05 %). In total, 4300 (5.31 %) of

the remaining gauges fulfilled one or more of these criteria and hence were discarded; the resultant dataset comprised 76 747 gauges (Figure 2b).

### *b. 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  $+36$  hours, 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-hour 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 due to 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).

### *c. Rainfall estimation using thermal infrared imagery*

MSWEP V1 relied exclusively on reanalysis and gauge data during the pre-TRMM era ( $< 1998$ ; 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 accu-

536 racy in convection-dominated regions. Although several IR-based rainfall datasets already ex-  
537 ist (e.g., CHIRP, Hydro-Estimator, PERSIANN, PERSIANN-CCS, PERSIANN-CDR, and TAM-  
538 SAT), none of these meet all of our requirements: (i) quasi-global coverage over land and ocean;  
539 (ii) temporal coverage from the 1980s to the near present; (iii) spatial resolution  $\leq 0.1^\circ$ ; (iv) tem-  
540 poral resolution  $\leq 3$  hours; and (v) no gauge corrections. We therefore produced a new 3-hourly  
541  $0.1^\circ$  rainfall dataset based on the GridSat B1 IR archive (V02R01; 3-hourly,  $0.07^\circ$  resolution;  
542 1980 to the near present) containing IR imagery from various intercalibrated geostationary satel-  
543 lites (Knapp et al. 2011).

544 Although the GridSat archive has already had some quality control applied, we still observed  
545 numerous navigation, calibration, and masking errors (particularly prior to 1983). To ensure that  
546 the data were robust, several additional quality control steps were applied. First, all grid-cells with  
547 values  $< 173\text{K}$  (the record minimum, Ebert and Holland 1992) were assumed to be erroneous  
548 and discarded. Additionally, if the percentage of grid-cells with temperature  $< 173\text{K}$  exceeded  
549 1 %, the entire image was discarded. Furthermore, if the spatial (Pearson) correlation between  
550 the current image and the previous image (both resampled to  $1^\circ$  using bilinear interpolation) was  
551  $< 0.75$ , both images were discarded. Finally, assuming that sudden isolated changes in the record  
552 are indicative of errors, images were discarded if the global mean deviated  $> 3\text{K}$  from the 24-hour  
553 running global mean. Note that prior to 1998 there are extensive periods of missing data due to a  
554 poorer spatial coverage.

555 IR data can be used to estimate rainfall in several ways (Scofield and Kuligowski 2003; Stephens  
556 and Kummerow 2007; Michaelides et al. 2009; Kidd and Levizzani 2011). Hydro-Estimator, for  
557 example, employs an empirical equation calibrated using ground radar data to obtain an initial  
558 rain rate estimate which is subsequently corrected using precipitable water and relative humid-  
559 ity 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 235K 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 it would require IR data with a temporal resolution  $< 3$  hours 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  $< 5^\circ\text{C}$ , given 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 Appendix h. 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 (Appendix e).

#### *d. 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 2000–2017 (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 (Appendix b). 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-hour 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  $1^\circ\text{C}$  threshold, which we increased in V2 to further reduce the likelihood of incorporating potentially unreliable satellite data. For  $T_a$ , we used ERA-Interim (Dee et al. 2011) downscaled to  $0.1^\circ$  using nearest neighbour resampling and offset to match the long-term mean of the high-resolution, station-based WorldClim V2.0 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 due to 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 Equation 1. Wet days were identified using a  $0.5\text{ mm d}^{-1}$  threshold, similar to several previous studies (e.g., Akinremi et al. 1999; Haylock

et al. 2008; Driouech et al. 2009; Trenberth and Zhang 2018).  $\beta_{WD}$  values range from 0 to  $\infty$ , with values closer to unity corresponding to better performance.

*e. 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 (Appendix d). 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  $0.1^\circ$  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  $\beta_{WD}$  were produced for the reanalyses, to correct the  $P$  frequency prior to the merging.

Due to 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 due to contrasting values on opposite sides of the oceans. To eliminate these discontinuities, we applied an exponential smoothing kernel with a bandwidth of 1000 km over the ocean areas of the interpolated weight and  $\beta_{WD}$  maps.

*f. Determination of long-term mean  $P$*

The long-term mean  $P$  over the land surface was determined in V2 using the WorldClim dataset (1-km resolution; V2.0; Fick and Hijmans 2017) rather than the CHPclim dataset ( $0.05^\circ$  resolution; Funk et al. 2015a). We switched from CHPclim to WorldClim due to the better gauge coverage in South America, Scandinavia, India, Australia, and New Zealand. Systematic  $P$  underestimation over land due to gauge under-catch and orographic effects (Kauffeldt et al. 2013; Beck et al. 2015, 2017a; Prein and Gobiet 2017) was corrected similarly to MSWEP V1, by in-

ferring 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, due to 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  $< 20^\circ$  and 0 for latitudes  $> 40^\circ$ , decreasing linearly from 1 at  $20^\circ$  to 0 at  $40^\circ$ . The weights for the reanalyses ( $w_r$ ) were set to  $1 - w_s$ . Thus,  $w_r$  was set to 0 at latitudes  $< 20^\circ$ , due to the tendency of reanalyses to overestimate tropical  $P$  amounts (Trenberth et al. 2011; Kang and Ahn 2015).

#### *g. 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  $0.1^\circ$  (CMORPH, ERA-Interim, JRA-55, and TMPA 3B42RT) were resampled to  $0.1^\circ$  using nearest neighbor resampling, and 3-hourly means were calculated for the datasets with temporal resolutions  $< 3$  hours (CMORPH and GSMaP).
2. The WATCH (Weedon et al. 2011) and WFDEI (Weedon et al. 2014) datasets (derived respectively from the ERA-40 and ERA-Interim reanalyses) 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 (Appendix e). Next, we iteratively carried out the following steps: (i) subtract  $d \text{ mm } 3\text{h}^{-1}$  from the original 3-hourly time series, starting with  $d = 0.01 \text{ mm } 3\text{h}^{-1}$ ; (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 ( $WD_{new}$ ); (iv) return to step (i), increasing  $d$  in  $0.01 \text{ mm } 3\text{h}^{-1}$  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 Appendix f. 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 Appendix e. Finally, the satellite datasets were rescaled to match the rescaled warm-condition reanalysis mean.

#### *h. Reference $P$ distributions*

In MSWEP V2, the 3-hourly merged satellite and reanalysis  $P$  estimates were CDF matched to reference  $P$  distributions (Figure 1), to correct the spurious drizzle and attenuated peaks evident in V1 (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 dis-

tribution 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 (Appendix c). 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).

#### *i. 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 (Appendix e). 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 (Appendix g). 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 2000–2017 of each dataset combination, for cold and warm conditions separately, to the reference  $P$  distributions (which represent 2000–2017; see Appendix h). Similar CDF-matching approaches have been used to correct other  $P$  datasets, including CMORPH (Xie et al. 2017), 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 1979–1999.
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.

#### j. Gauge correction scheme

The merged 3-hourly satellite- and reanalysis-based  $P$  data (referred to hereafter as  $p_{\text{merge}}$ ; Appendix i) were corrected using gauge  $P$  observations through an iterative, multiplicative approach that accounts for variability in the reporting times of gauges (Appendix b). We used a multiplicative rather than an additive correction method (Vila et al. 2009) to preserve the sub-daily 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 (Appendix f), 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 due to 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 hereafter as  $p_{\text{drizzly}}$ .
2. The five nearest (as the crow flies) gauges were selected (Appendix a), 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-hour 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. Figures 1e and 1f

illustrate the importance of accounting for reporting times. 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  $w_g = 4 \exp\left(\frac{-d_i}{d_0}\right)$ , 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 (Appendix i step 3) and the gauge weights from the previous loop iterations.

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 FDR V7 dataset (Schneider et al. 2014b, 2017) following the same procedure but without accounting for gauge reporting times to yield the final gauge-corrected MSWEP V2.

## References

- Adler, R. F., G. Gu, M. Sapiiano, J.-J. Wang, and G. J. Huffman, 2017: Global precipitation: means, variations and trends during the satellite era (1979–2014). *Surveys in Geophysics*, **38** (4), 679–699.
- Adler, R. F., C. Kidd, G. Petty, M. Morissey, and H. M. Goodman, 2001: Intercomparison of global precipitation products: The third precipitation intercomparison project (PIP-3). *Bulletin of the American Meteorological Society*, **82** (7), 1377–1396.
- Adler, R. F., and A. J. Negri, 1988: A satellite infrared technique to estimate tropical convective and stratiform rainfall. *Journal of Applied Meteorology*, **27** (1), 30–51.
- Adler, R. F., and Coauthors, 2003: The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *Journal of Hydrometeorology*, **4** (6),



1147–1167.

Adler, R. F., and Coauthors, 2018: The Global Precipitation Climatology Project (GPCP) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere*, **9** (4), 138.

AghaKouchak, A., A. Behrangi, S. Sorooshian, K. Hsu, and E. Amitai, 2011: Evaluation of satellite retrieved extreme precipitation rates across the central United States. *Journal of Geophysical Research: Atmospheres*, **116** (D2), doi:10.1029/2010JD014741.

AghaKouchak, A., A. Mehran, H. Norouzi, and A. Behrangi, 2012: Systematic and random error components in satellite precipitation data sets. *Geophysical Research Letters*, **39** (9), doi:10.1029/2012GL051592.

Akinremi, O. O., S. M. McGinn, and H. W. Cutforth, 1999: Precipitation trends on the Canadian prairies. *Journal of Climate*, **12** (10), 2996–3003.

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 System Science Data*, **2** (2), 215–234.

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. *Bulletin of the American Meteorological Society*, **96** (1), 69–83.

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. *Journal of Hydrometeorology*, **16** (4), 1478–1501.

784 Beck, H. E., A. I. J. M. van Dijk, A. de Roo, E. Dutra, G. Fink, R. Orth, and J. Schellekens, 2017a:  
785 Global evaluation of runoff from ten state-of-the-art hydrological models. *Hydrology and Earth*  
786 *System Sciences*, **21** (6), 2881–2903.

787 Beck, H. E., A. I. J. M. van Dijk, V. Levizzani, J. Schellekens, D. G. Miralles, B. Martens, and  
788 A. de Roo, 2017b: MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merg-  
789 ing gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences*, **21** (1), 589–615.

790 Beck, H. E., and Coauthors, 2017c: Global-scale evaluation of 22 precipitation datasets using  
791 gauge observations and hydrological modeling. *Hydrology and Earth System Sciences*, **21** (12),  
792 6201–6217.

793 Behrangi, A., M. Lebsock, S. Wong, and B. Lambriksen, 2012: On the quantification of oceanic  
794 rainfall using spaceborne sensors. *Journal of Geophysical Research: Atmospheres*, **117** (D20),  
795 doi:10.1029/2012JD017979.

796 Behrangi, A., Y. Tian, B. H. Lambriksen, and G. L. Stephens, 2014: What does CloudSat re-  
797 veal about global land precipitation detection by other spaceborne sensors? *Water Resources*  
798 *Research*, **50** (6), 4893–4905.

799 Bierkens, M. F. P., and Coauthors, 2015: Hyper-resolution global hydrological modelling: what is  
800 next? *Hydrological Processes*, **29** (2), 310–320.

801 Biswas, S. K., S. Farrar, K. Gopalan, A. Santos-Garcia, W. L. Jones, and S. Bilanow, 2013: Inter-  
802 calibration of microwave radiometer brightness temperatures for the Global Precipitation Mea-  
803 surement Mission. *IEEE Transactions on Geoscience and Remote Sensing*, **51** (3), 1465–1477.

804 Cervený, R. S., J. Lawrimore, R. Edwards, and C. Landsea, 2007: Extreme weather records.  
805 *Bulletin of the American Meteorological Society*, **88** (6), 853–860.

806 Chen, L., and P. A. Dirmeyer, 2017: Impacts of land-use/land-cover change on afternoon precipi-  
807 tation over North America. *Journal of Climate*, **30** (6), 2121–2140.

808 Chen, L., P. A. Dirmeyer, A. Tawfik, and D. M. Lawrence, 2017: Sensitivities of land cover  
809 precipitation feedback to convective triggering. *Journal of Hydrometeorology*, **18** (8), 2265–  
810 2283.

811 Chen, M., W. Shi, P. Xie, V. B. S. Silva, V. E. Kousky, R. W. Higgins, and J. E. Janowiak, 2008:  
812 Assessing objective techniques for gauge-based analyses of global daily precipitation. *Journal*  
813 *of Geophysical Research*, **113**, D04 110, doi:10.1029/2007JD009132.

814 Correa, S. W., R. C. D. de Paiva, J. C. Espinoza, and W. Collischonn, 2017: Multi-decadal hydro-  
815 logical retrospective: case study of Amazon floods and droughts. *Journal of Hydrology*, **549**,  
816 667–684, doi:10.1016/j.jhydrol.2017.04.019.

817 Daly, C., M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P.  
818 Pasteris, 2008: Physiographically sensitive mapping of climatological temperature and precip-  
819 itation across the conterminous United States. *International Journal of Climatology*, **28** (15),  
820 2031–2064.

821 Daly, C., R. P. Neilson, and D. L. Phillips, 1994: A statistical-topographic model for mapping  
822 climatological precipitation over mountainous terrain. *Journal of Applied Meteorology*, **33** (2),  
823 140–158.

824 Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: configuration and performance  
825 of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society Part A*,  
826 **137** (656), 553–597.

827 DeGaetano, A. T., 2000: A serially complete simulated observation time metadata file for U.S.  
828 Daily Historical Climatology Network stations. *Bulletin of the American Meteorological Soci-*  
829 *ety*, **81** (1), 49–67.

830 Dietzsch, F., A. Andersson, M. Ziese, M. Schröder, K. Raykova, K. Schamm, and A. Becker, 2017:  
831 A global ETCCDI-based precipitation climatology from satellite and rain gauge measurements.  
832 *Climate*, **5** (9), doi:10.3390/cli5010009.

833 Driouech, F., M. Déqué, and A. Mokssit, 2009: Numerical simulation of the probability distribu-  
834 tion function of precipitation over Morocco. *Climate Dynamics*, **32** (7), 1055–1063.

835 Durre, I., M. J. Menne, B. E. Gleason, T. G. Houston, and R. S. Vose, 2010: Comprehensive  
836 automated quality assurance of daily surface observations. *Journal of Applied Meteorology and*  
837 *Climatology*, **49** (8), 1615–1633.

838 Ebert, E. E., and G. J. Holland, 1992: Observations of record cold cloud-top temperatures in  
839 tropical cyclone Hilda (1990). *Monthly Weather Review*, **120** (10), 2240–2251.

840 Ebert, E. E., J. E. Janowiak, and C. Kidd, 2007: Comparison of near-real-time precipitation esti-  
841 mates from satellite observations and numerical models. *Bulletin of the American Meteorologi-*  
842 *cal Society*, **88** (1), 47–64.

843 Ellis, T. D., T. L’Ecuyer, J. M. Haynes, and G. L. Stephens, 2009: How often does it rain over  
844 the global oceans? The perspective from CloudSat. *Geophysical Research Letters*, **36** (3), doi:  
845 10.1029/2008GL036728.

846 Ensor, L. A., and S. M. Robeson, 2008: Statistical characteristics of daily precipitation: compar-  
847 isons of gridded and point datasets. *Journal of Applied Meteorology and Climatology*, **47** (9),  
848 2468–2476.

849 Fick, S. E., and R. J. Hijmans, 2017: WorldClim 2: new 1-km spatial resolution climate surfaces  
850 for global land areas. *International Journal of Climatology*, **37** (12), 4302–4315.

851 Funk, C., A. Verdin, J. Michaelsen, P. Peterson, D. Pedreros, and G. Husak, 2015a: A global  
852 satellite assisted precipitation climatology. *Earth System Science Data*, **7**, 275–287, doi:10.  
853 5194/essd-7-275-2015.

854 Funk, C., and Coauthors, 2015b: The climate hazards infrared precipitation with stations—a new  
855 environmental record for monitoring extremes. *Scientific Data*, **2**, 150 066, doi:10.1038/sdata.  
856 2015.66.

857 Goodison, B. E., P. Y. T. Louie, and D. Yang, 1998: WMO solid precipitation intercomparison.  
858 Tech. Rep. WMO/TD-872, World Meteorological Organization, Geneva.

859 Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez, 2009: Decomposition of the mean  
860 squared error and NSE performance criteria: Implications for improving hydrological mod-  
861 elling. *Journal of Hydrology*, **370** (1–2), 80–91.

862 Haberlandt, U., 2007: Geostatistical interpolation of hourly precipitation from rain gauges and  
863 radar for a large-scale extreme rainfall event. *Journal of Hydrology*, **332** (1), 144–157.

864 Habib, E., A. Henschke, and R. F. Adler, 2009: Evaluation of TMPA satellite-based research  
865 and real-time rainfall estimates during six tropical-related heavy rainfall events over Louisiana,  
866 USA. *Atmospheric Research*, **94** (3), 373–388.

867 Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. Jones, and M. New, 2008: A Eu-  
868 ropean daily high-resolution gridded data set of surface temperature and precipitation for 1950–  
869 2006. *Journal of Geophysical Research: Atmospheres*, **113** (D20), doi:10.1029/2008JD010201.

870 Herold, N., L. V. Alexander, M. G. Donat, S. Contractor, and A. Becker, 2016: How much does it  
871 rain over land? *Geophysical Research Letters*, **43** (1), 341–348.

872 Holleman, I., 2006: Bias adjustment of radar-based 3-hour precipitation accumulations. Tech. Rep.  
873 TR-290, Royal Netherlands Meteorological Institute (KNMI).

874 Hong, Y., K. Hsu, H. Moradkhani, and S. Sorooshian, 2006: Uncertainty quantification of satellite  
875 precipitation estimation and Monte Carlo assessment of the error propagation into hydrologic  
876 response. *Water Resources Research*, **42** (8), doi:10.1029/2005WR004398.

877 Hong, Y., K.-L. Hsu, S. Sorooshian, and X. Gao, 2004: Precipitation estimation from remotely  
878 sensed imagery using an artificial neural network cloud classification system. *Journal of Applied  
879 Meteorology*, **43** (12), 1834–1853.

880 Huffman, G. J., R. F. Adler, B. Rudolf, U. Schneider, and P. R. Keehn, 1995: Global precipitation  
881 estimates based on a technique for combining satellite-based estimates, rain gauge analysis, and  
882 NWP model precipitation information. *Journal of Climate*, **8** (5), 1284–1295.

883 Huffman, G. J., D. T. Bolvin, D. Braithwaite, K. Hsu, R. Joyce, C. Kidd, E. J. Nelkin, and P. Xie,  
884 2014: NASA global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals  
885 for GPM (IMERG). Algorithm Theoretical Basis Document (ATBD), NASA/GSFC, Greenbelt,  
886 MD 20771, USA.

887 Huffman, G. J., and Coauthors, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA):  
888 quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hy-  
889 drometeorology*, **8** (1), 38–55.

- Hutchinson, M. F., 1998: Interpolation of rainfall data with thin plate smoothing splines — part I: Two dimensional smoothing of data with short range correlation. *Journal of Geographic Information and Decision Analysis*, **2**, 168–185.
- Janowiak, J. E., 1992: Tropical rainfall: a comparison of satellite-derived rainfall estimates with model precipitation forecasts, climatologies, and observations. *Monthly Weather Review*, **120** (3), 448–462.
- 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. *Journal of Hydrometeorology*, **5** (3), 487–503.
- Kang, S., and J.-B. Ahn, 2015: Global energy and water balances in the latest reanalyses. *Asia-Pacific Journal of Atmospheric Sciences*, **51** (4), 293–302.
- Karbalaee, N., K. Hsu, S. Sorooshian, and D. Braithwaite, 2017: Bias adjustment of infrared based rainfall estimation using passive microwave satellite rainfall data. *Journal of Geophysical Research: Atmospheres*, **122** (7), 3859–3876.
- Kauffeldt, A., S. Halldin, A. Rodhe, C.-Y. Xu, and I. K. Westerberg, 2013: Disinformative data in large-scale hydrological modelling. *Hydrology and Earth System Sciences*, **17** (7), 2845–2013.
- Kenney, J. F., and E. S. Keeping, 1962: Linear regression and correlation. *Mathematics of statistics part I*, Van Nostrand, Princeton, NJ, chap. 15, 252–285.
- Kidd, C., and V. Levizzani, 2011: Status of satellite precipitation retrievals. *Hydrology and Earth System Sciences*, **15** (4), 1109–1116.

910 Kling, H., M. Fuchs, and M. Paulin, 2012: Runoff conditions in the upper Danube basin under an  
 911 ensemble of climate change scenarios. *Journal of Hydrology*, **424–425**, 264–277, doi:10.1016/  
 912 j.hydrol.2012.01.011.

913 Knapp, K. R., and Coauthors, 2011: Globally gridded satellite observations for climate studies.  
 914 *Bulletin of the American Meteorological Society*, **92 (7)**, 893–907.

915 Kobayashi, S., and Coauthors, 2015: The JRA-55 reanalysis: General specifications and basic  
 916 characteristics. *Journal of the Meteorological Society of Japan. Ser. I*, **93**, 5–48, doi:10.2151/  
 917 jmsj.2015-001.

918 Kuligowski, R. J., 1997: An overview of National Weather Service quantitative pre-  
 919 cipitation estimates. URL [https://repository.library.noaa.gov/gsearch?related\\_series=TDL%](https://repository.library.noaa.gov/gsearch?related_series=TDL%20office%20note%20%3B%2097-4)  
 920 [20office%20note%20%3B%2097-4](https://repository.library.noaa.gov/gsearch?related_series=TDL%20office%20note%20%3B%2097-4), United States, National Weather Service, Techniques De-  
 921 velopment Laboratory.

922 Kummerow, C., W. Barnes, T. Kozu, J. Shiue, and J. Simpson, 1998: The Tropical Rainfall Mea-  
 923 suring Mission (TRMM) sensor package. *Journal of Atmospheric and Oceanic Technology*,  
 924 **15 (3)**, 809–817.

925 Legates, D. R., 1988: A climatology of global precipitation. Ph.D. thesis, University of Delaware.

926 Legates, D. R., and C. J. Willmott, 1990: Mean seasonal and spatial variability in gauge-corrected  
 927 global precipitation. *International Journal of Climatology*, **10 (2)**, 111–127.

928 Levizzani, V., and Coauthors, 2018: The activities of the International Precipitation Working  
 929 Group. *Quarterly Journal of the Royal Meteorological Society*, doi:10.1002/qj.3214.

930 Liebmann, B., and D. Allured, 2005: Daily precipitation grids for South America. *Bulletin of the*  
 931 *American Meteorological Society*, **86 (11)**, 1567–1570.



932 Lin, Y., and K. E. Mitchell, 2005: The NCEP stage II/IV hourly precipitation analyses: devel-  
 933 opment and applications. *19th Conf. Hydrology*, URL [https://ams.confex.com/ams/pdfpapers/](https://ams.confex.com/ams/pdfpapers/83847.pdf)  
 934 [83847.pdf](https://ams.confex.com/ams/pdfpapers/83847.pdf).

935 Liu, J., T. Xiao, and L. Chen, 2011: Intercomparisons of air-sea heat fluxes over the Southern  
 936 Ocean. *Journal of Climate*, **24** (4), 1198–1211.

937 Liu, Q., T. R. McVicar, Z. Yang, R. J. Donohue, L. Liang, and Y. Yang, 2016: The hydrological  
 938 effects of varying vegetation characteristics in a temperate water-limited basin: Development  
 939 of the dynamic Budyko-Choudhury-Porporato (dBCP) model. *Journal of Hydrology*, **543**, 595–  
 940 611.

941 Lopez, P., 2007: Cloud and precipitation parameterizations in modeling and variational data as-  
 942 simulation: a review. *Journal of the Atmospheric Sciences*, **64** (11), 3766–3784.

943 Maggioni, V., P. C. Meyers, and M. D. Robinson, 2016: A review of merged high resolution satel-  
 944 lite precipitation product accuracy during the Tropical Rainfall Measuring Mission (TRMM)-  
 945 era. *Journal of Hydrometeorology*, **17**, 1101–1117, doi:10.1175/JHM-D-15-0190.1.

946 Martens, B., and Coauthors, 2017: GLEAM v3: satellite-based land evaporation and root-zone  
 947 soil moisture. *Geoscientific Model Development*, **10** (5), 1903–1925.

948 Massari, C., W. Crow, and L. Brocca, 2017: An assessment of the accuracy of global rainfall  
 949 estimates without ground-based observations. *Hydrology and Earth System Sciences*, **21** (9),  
 950 4347–4361, doi:10.5194/hess-2017-163.

951 Mega, T., T. Ushio, T. Kubota, M. Kachi, K. Aonashi, and S. Shige, 2014: Gauge adjusted global  
 952 satellite mapping of precipitation (GSMaP\_Gauge). *2014 XXXIth URSI General Assembly and*  
 953 *Scientific Symposium (URSI GASS)*, 1–4, doi:10.1109/URSIGASS.2014.6929683.

954 Menne, M. J., I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston, 2012: An overview of the  
 955 Global Historical Climatology Network-Daily database. *Journal of Atmospheric and Oceanic*  
 956 *Technology*, **29** (7), 897–910.

957 Michaelides, S., V. Levizzani, E. Anagnostou, P. Bauer, T. Kasparis, and J. E. Lane, 2009: Precipi-  
 958 tation: measurement, remote sensing, climatology and modeling. *Atmospheric Research*, **94** (4),  
 959 512–533.

960 Nair, A. S., and J. Indu, 2017: Performance assessment of Multi-Source Weighted-Ensemble  
 961 Precipitation (MSWEP) product over India. *Climate*, **5** (1), doi:10.3390/cli5010002.

962 Oki, T., and S. Kanae, 2006: Global hydrological cycles and world water resources. *Science*,  
 963 **313** (5790), 1068–1072.

964 Osborn, T. J., and M. Hulme, 1997: Development of a relationship between station and grid-box  
 965 rainday frequencies for climate model evaluation. *Journal of Climate*, **10** (8), 1885–1908.

966 Papagiannopoulou, C., D. G. Miralles, W. A. Dorigo, N. E. C. Verhoest, M. Depoorter, and  
 967 W. Waegeman, 2017a: Vegetation anomalies caused by antecedent precipitation in most of the  
 968 world. *Environmental Research Letters*, **12** (7), doi:10.1088/1748-9326/aa7145.

969 Papagiannopoulou, C., D. G. Miralles, N. E. C. Verhoest, W. Dorigo, and W. Waegeman, 2017b:  
 970 A non-linear Granger causality framework to investigate climate-vegetation dynamics. *Geosci-*  
 971 *entific Model Development*, **10** (5), 1945–1960.

972 Prein, A. F., and A. Gobiet, 2017: Impacts of uncertainties in European gridded precipitation  
 973 observations on regional climate analysis. *International Journal of Climatology*, **37** (1), 305–  
 974 327.

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. *Journal of Climate*, **30** (5), 1643–1664.

Roe, G. H., 2005: Orographic precipitation. *Annual Review of Earth and Planetary Sciences*,  
**33** (1), 645–671.

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 A*, **62** (4), 435–452.

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 Sensing*, **9** (3),  
doi:10.3390/rs9030218.

Schellekens, J., and Coauthors, 2017: A global water resources ensemble of hydrological models:  
the earthH2Observe Tier-1 dataset. *Earth System Science Data*, **9**, 389–413.

Schlosser, C. A., and P. R. Houser, 2007: Assessing a satellite-era perspective of the global water  
cycle. *Journal of Climate*, **20** (7), 1316–1338.

Schneider, T., T. Bischoff, and G. H. Haug, 2014a: Migrations and dynamics of the intertropical  
convergence zone. *Nature*, **513**, 45–53, doi:10.1038/nature13636.

Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, M. Ziese, and B. Rudolf, 2014b:  
GPCC’s new land surface precipitation climatology based on quality-controlled in situ data  
and its role in quantifying the global water cycle. *Theoretical and Applied Climatology*, **115** (1),  
15–40.

Schneider, U., 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** (3), 52.

- 997 Scofield, R. A., and R. J. Kuligowski, 2003: Status and outlook of operational satellite precipita-  
998 tion algorithms for extreme-precipitation events. *Weather and Forecasting*, **18** (6), 1037–1051.
- 999 Sevruk, B., M. Ondrás, and B. Chvíla, 2009: The WMO precipitation measurement intercompar-  
1000 isons. *Atmospheric Research*, **92** (3), 376–380.
- 1001 Skok, G., N. Žagar, L. Honzak, R. Žabkar, J. Rakovec, and A. Ceglar, 2015: Precipitation inter-  
1002 comparison of a set of satellite- and raingauge-derived datasets, ERA Interim reanalysis, and a  
1003 single WRF regional climate simulation over Europe and the North Atlantic. *Theoretical and*  
1004 *Applied Climatology*, **123** (1), 217–232.
- 1005 Sorooshian, S., K.-L. Hsu, X. Gao, H. V. Gupta, B. Imam, and D. Braithwaite, 2000: Evaluation  
1006 of PERSIANN system satellite-based estimates of tropical rainfall. *Bulletin of the American*  
1007 *Meteorological Society*, **81** (9), 2035–2046.
- 1008 Stephens, G. L., and C. D. Kummerow, 2007: The remote sensing of clouds and precipitation from  
1009 space: a review. *Journal of the Atmospheric Sciences*, **64** (11), 3742–3765.
- 1010 Stephens, G. L., and Coauthors, 2010: Dreary state of precipitation in global models. *Journal of*  
1011 *Geophysical Research: Atmospheres*, **115** (D24), doi:10.1029/2010JD014532.
- 1012 Sun, Q., C. Miao, Q. Duan, H. Ashouri, S. Sorooshian, and K.-L. Hsu, 2018: A review of global  
1013 precipitation datasets: data sources, estimation, and intercomparisons. *Reviews of Geophysics*,  
1014 **56** (1), 79–107.
- 1015 Sun, Y., S. Solomon, A. Dai, and R. W. Portmann, 2006: How often does it rain? *Journal of*  
1016 *Climate*, **19** (6), 916–934.
- 1017 Tarnavsky, E., D. Grimes, R. Maidment, E. Black, R. P. Allan, M. Stringer, R. Chadwick, and  
1018 F. Kayitakire, 2014: Extension of the TAMSAT satellite-based rainfall monitoring over Africa

and from 1983 to present. *Journal of Applied Meteorology and Climatology*, **53** (12), 2805–2822.

Tian, Y., and C. D. Peters-Lidard, 2010: A global map of uncertainties in satellite-based precipitation measurements. *Geophysical Research Letters*, **37** (24), doi:10.1029/2010GL046008.

Tian, Y., and Coauthors, 2009: Component analysis of errors in satellite-based precipitation estimates. *Journal of Geophysical Research: Atmospheres*, **114** (D24), doi:10.1029/2009JD011949.

Trenberth, K. E., J. T. Fasullo, and J. Mackaro, 2011: Atmospheric moisture transports from ocean to land and global energy flows in reanalyses. *Journal of Climate*, **24** (18), 4907–4924.

Trenberth, K. E., 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. *Journal of Hydrometeorology*, **8** (4), 758–769.

Trenberth, K. E., and Y. Zhang, 2018: How often does it really rain? *Bulletin of the American Meteorological Society*, **99** (2), 289–298.

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. *Journal of the Meteorological Society of Japan*, **87A** (II), 137–151.

Vicente, G. A., R. A. Scofield, and W. P. Menzel, 1998: The operational GOES infrared rainfall estimation technique. *Bulletin of the American Meteorological Society*, **79** (9), 1883–1898.

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. *Journal of Hydrometeorology*, **10** (2), 533–543.

- 1041 Villarini, G., and W. F. Krajewski, 2007: Evaluation of the research version TMPA three-hourly  
1042  $0.25^\circ \times 0.25^\circ$  rainfall estimates over Oklahoma. *Geophysical Research Letters*, **34** (5), doi:  
1043 10.1029/2006GL029147.
- 1044 Viney, N. R., and B. C. Bates, 2004: It never rains on Sunday: the prevalence and implications of  
1045 untagged multi-day rainfall accumulations in the Australian high quality data set. *International*  
1046 *Journal of Climatology*, **24** (9), 1171–1192.
- 1047 Wang, Y., Z. Guo, and G. Li, 2017: Precipitation estimation and analysis of the Three Gorges Dam  
1048 region (1979–2014) by combining gauge measurements and MSWEP with generalized additive  
1049 model. *Acta Geographica Sinica*, **72** (7), 1207–1220.
- 1050 Weatherhead, E. C., and Coauthors, 1998: Factors affecting the detection of trends: Statistical  
1051 considerations and applications to environmental data. *Journal of Geophysical Research: At-*  
1052 *mospheres*, **103** (D14), 17 149–17 161.
- 1053 Weedon, G. P., G. Balsamo, N. Bellouin, S. Gomes, M. J. Best, and P. Viterbo, 2014: The WFDEI  
1054 meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim  
1055 reanalysis data. *Water Resources Research*, **50** (9), 7505–7514.
- 1056 Weedon, G. P., and Coauthors, 2011: Creation of the WATCH forcing data and its use to assess  
1057 global and regional reference crop evaporation over land during the twentieth century. *Journal*  
1058 *of Hydrometeorology*, **12** (5), 823–848.
- 1059 Wolff, D. B., and B. L. Fisher, 2008: Comparisons of instantaneous TRMM ground validation  
1060 and satellite rain-rate estimates at different spatial scales. *Journal of Applied Meteorology and*  
1061 *Climatology*, **47** (8), 2215–2237.

Wood, E. F., and Coauthors, 2011: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth’s terrestrial water. *Water Resources Research*, **47** (5), doi: 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. *Bulletin of the American Meteorological Society*, **78** (11), 2539–2558.

Xie, P., M. Chen, S. Yang, A. Yatagai, T. Hayasaka, Y. Fukushima, and C. Liu, 2007: A gauge-based analysis of daily precipitation over East Asia. *Journal of Hydrometeorology*, **8** (3), 607–626.

Xie, P., 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. *Journal of Hydrometeorology*, **18** (6), 1617–1641.

Yang, Y., R. J. Donohue, and T. R. McVicar, 2016: Global estimation of effective plant rooting depth: Implications for hydrological modeling. *Water Resources Research*, **52** (10), 8260–8276.

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. *Bulletin of the American Meteorological Society*, **93** (9), 1401–1415.

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. *Journal of Hydrometeorology*, **5** (6), 1207–1222.

- 1083 Zambrano-Bigiarini, M., A. Nauditt, C. Birkel, K. Verbist, and L. Ribbe, 2017: Temporal and spa-  
1084 tial evaluation of satellite-based rainfall estimates across the complex topographical and climatic  
1085 gradients of Chile. *Hydrology and Earth System Sciences*, **21** (2), 1295–1320.
- 1086 Zhang, L., W. R. Dawes, and G. R. Walker, 2001: Response of mean annual evapotranspiration  
1087 to vegetation changes at catchment scale. *Water Resources Research*, **37** (3), 701–708, doi:  
1088 10.1029/2000WR900325.
- 1089 Zhang, W., M. Brandt, F. Guichard, Q. Tian, and R. Fensholt, 2017: Using long-term daily satellite  
1090 based rainfall data (1983–2015) to analyze spatio-temporal changes in the Sahelian rainfall  
1091 regime. *Journal of Hydrology*, **550**, 427–440.
- 1092 Zhang, X., E. N. Anagnostou, and C. S. Schwartz, 2018: NWP-based adjustment of IMERG pre-  
1093 cipitation for flood-inducing complex terrain storms: evaluation over CONUS. *Remote Sensing*,  
1094 **10** (4), 642.
- 1095 Zhu, Y., and Y. Luo, 2015: Precipitation calibration based on the frequency-matching method.  
1096 *Weather and Forecasting*, **30** (5), 1109–1124.
- 1097 Zohaib, M., H. Kim, and M. Choi, 2017: Evaluating the patterns of spatiotemporal trends of root  
1098 zone soil moisture in major climate regions in East Asia. *Journal of Geophysical Research:*  
1099 *Atmospheres*, **122** (15), 7705–7722.
- 1100 Zolina, O., A. Kapala, C. Simmer, and S. K. Gulev, 2004: Analysis of extreme precipitation over  
1101 Europe from different reanalyses: a comparative assessment. *Global and Planetary Change*,  
1102 **44** (1–4), 129–161.



1103

1104

1105

1106

1107

1108

1109

1110

1111

LIST OF TABLES

Table 1. Overview of the gridded  $P$  datasets incorporated in MSWEP V2 and used for comparison. Abbreviations: G=gauge; S=satellite; R=reanalysis; N=radar; NRT=near real-time. 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. . . . . 51

Table 2. Long-term mean annual  $P$  estimates ( $\text{mm y}^{-1}$ ) for global, land, and ocean domains from various sources. . . . . 52

TABLE 1. Overview of the gridded  $P$  datasets incorporated in MSWEP V2 and used for comparison. Abbreviations: G=gauge; S=satellite; R=reanalysis; N=radar; NRT=near real-time. 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.

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 minutes	1998–NRT <sup>1</sup>	Joyce et al. (2004)
Daily gauge data	Compiled from GHCN-D, GSOD, and other sources	G	—	Land	Daily	1979–2017	This study (Appendix j)
ERA-Interim	European Centre for Medium-range Weather Forecasts ReAnalysis Interim (ERA-Interim)	R	~80 km	Global	3 hourly	1979–NRT <sup>3</sup>	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 <sup>2</sup>	Schneider et al. (2014b, 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 (Appendix c)
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 <sup>2</sup>	Ushio et al. (2009)
JRA-55	Japanese 55-year ReAnalysis (JRA-55)	R	~60 km	Global	3 hourly	1959–NRT <sup>2</sup>	Kobayashi et al. (2015)
TMPA 3B42RT	TRMM Multi-satellite Precipitation Analysis (TMPA) 3B42RT V7	S	0.25°	≤ 50°N/S	3 hourly	2000–NRT <sup>1</sup>	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 (Appendix f)
<b>Datasets used for comparison</b>							
CHIRPS V2.0	Climate Hazards group Infrared Precipitation with Stations (CHIRPS) V2.0	G, S, R	0.05°	Land, ≤ 50°N/S	Daily	1981–NRT <sup>2</sup>	Funk et al. (2015b)
CMAP V1707	CPC Merged Analysis of Precipitation (CMAP) V1707	G, S, R	2.5°	Global	5 days	1979–2017	Xie and Arkin (1997)
GPCC V2015	Global Precipitation Climatology Centre (GPCC) Climatology V2015	G	0.25°	Land	Monthly	1951–2000	Schneider et al. (2014b, 2017)
GPCP V2.3	Global Precipitation Climatology Project (GPCP) Monthly Analysis Product V2.3	G, S	2.5°	Global	Monthly	1996–NRT <sup>2</sup>	Adler et al. (2003, 2018)
HOAPS V3.2	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 <sup>3</sup>	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 <sup>1</sup>	Lin and Mitchell (2005)

<sup>1</sup> Available until the present with a delay of several hours.

<sup>2</sup> Available until the present with a delay of several days.

<sup>3</sup> Available until the present with a delay of several months.

1116      TABLE 2. Long-term mean annual  $P$  estimates ( $\text{mm y}^{-1}$ ) for global, land, and ocean domains from various  
1117      sources.

Domain	MSWEP V2	MSWEP V1	GPCC V2015	GPCP V2.3	HOAPS V3.2	MERRA-2	Behrangi et al. (2014)
Global	955	—	—	982	—	946	—
Land (excl. Antarctica)	839	858	793	853	—	785	—
Land (incl. Antarctica)	781	798	—	798	—	735	—
Ocean <sup>1</sup>	1025	—	—	1057	—	1031	1074 <sup>3</sup>
Ocean (HOAPS mask <sup>2</sup> )	1068	—	—	1101	1037	1066	—

<sup>1</sup>Includes the Hudson Bay, Baltic Sea, Mediterranean Sea, Black Sea, Caspian Sea, and Red Sea.

<sup>2</sup>Smaller mask that excludes coastal and high-latitude regions for which HOAPS does not provide continuous data.

<sup>3</sup>Based on the 2007–2009 period.

## LIST OF FIGURES

1119	<b>Fig. 1.</b>	(a) Flowchart outlining the main processing steps implemented to produce MSWEP V2. For each step, the reference to the Appendix subsection that provides detail is provided between parentheses. (b) Example of the wet-day bias correction for ERA-Interim. Time series of the satellite and reanalysis $P$ datasets and MSWEP V2 are presented in (c) and (d). The importance of accounting for reporting times when applying gauge corrections is illustrated in (e) and (f).	55
1125	<b>Fig. 2.</b>	(a) Daily $P$ measured at GSOD station 038660 (50.58°N, 1.30°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-hour period starting at 0600 UTC of the current day and ending at 0600 UTC of the next day.	56
1132	<b>Fig. 3.</b>	Temporal correlations ( $r_{3\text{ day}}$ ) for (a) CMORPH, (b) ERA-Interim, and (c) GridSat. The difference in $r_{3\text{ day}}$ values between CMORPH and ERA-Interim is presented in (d). Also shown is the bias in the number of wet days ( $\beta_{\text{WD}}$ ) for (e) CMORPH and (f) ERA-Interim (note the non-linear 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.	57
1138	<b>Fig. 4.</b>	Relative weights assigned to the gauge-, satellite, and reanalysis-based $P$ estimates shown using a barycentric color map for the periods (a) 1979–1982, (b) 1983–1999, and (c) 2000–2017. 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.	58
1143	<b>Fig. 5.</b>	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.	59
1146	<b>Fig. 6.</b>	The long-term mean $P$ ( $\text{mm y}^{-1}$ ) for MSWEP V2 is presented in (a). Also shown are the differences in long-term mean $P$ between MSWEP V2 and (b) MSWEP V1, (c) GPCC V2015, (d) GPCP V2.3, (e) HOAPS V3.2, and (f) MERRA-2. The values represent 1979–2015 for MSWEP V1, 1987–2008 for HOAPS V3.2, 1980–2017 for MERRA-2, and 1979–2017 for the other datasets. Areas with no data are shown in white. For HOAPS V2.3, only grid-cells with continuous data are displayed.	60
1152	<b>Fig. 7.</b>	Global maps of 99.99th percentile 3-hourly $P$ amounts ( $\text{mm } 3\text{h}^{-1}$ ) for (a) MSWEP V2, (b) CMORPH, and (c) ERA-Interim for 2000–2017. CMORPH is limited to latitudes $\leq 60^\circ$ . Note the non-linear color scale.	61
1155	<b>Fig. 8.</b>	The percentage of time without $P$ for (a) MSWEP V2, (b) CMORPH, and (c) ERA-Interim for 2000–2017. 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 non-linear color scale.	62
1158	<b>Fig. 9.</b>	Linear trends in mean annual $P$ ( $\text{mm y}^{-2}$ ) for (a) MSWEP V2, (b) MSWEP V1, (c) CHIRPS V2.0, (d) CMAP V1707, (e) GPCC FDR V7, (f) GPCP V2.3, (g) HOAPS V3.2, and (h) MERRA-2. The trends represent 1979–2015 for MSWEP V1, 1981–2017 for	

1161	CHIRPS V2.0, 1987–2008 for HOAPS V3.2, 1980–2017 for MERRA-2, and 1979–2017	
1162	for the other datasets. Areas with no data are shown in white. . . . .	63

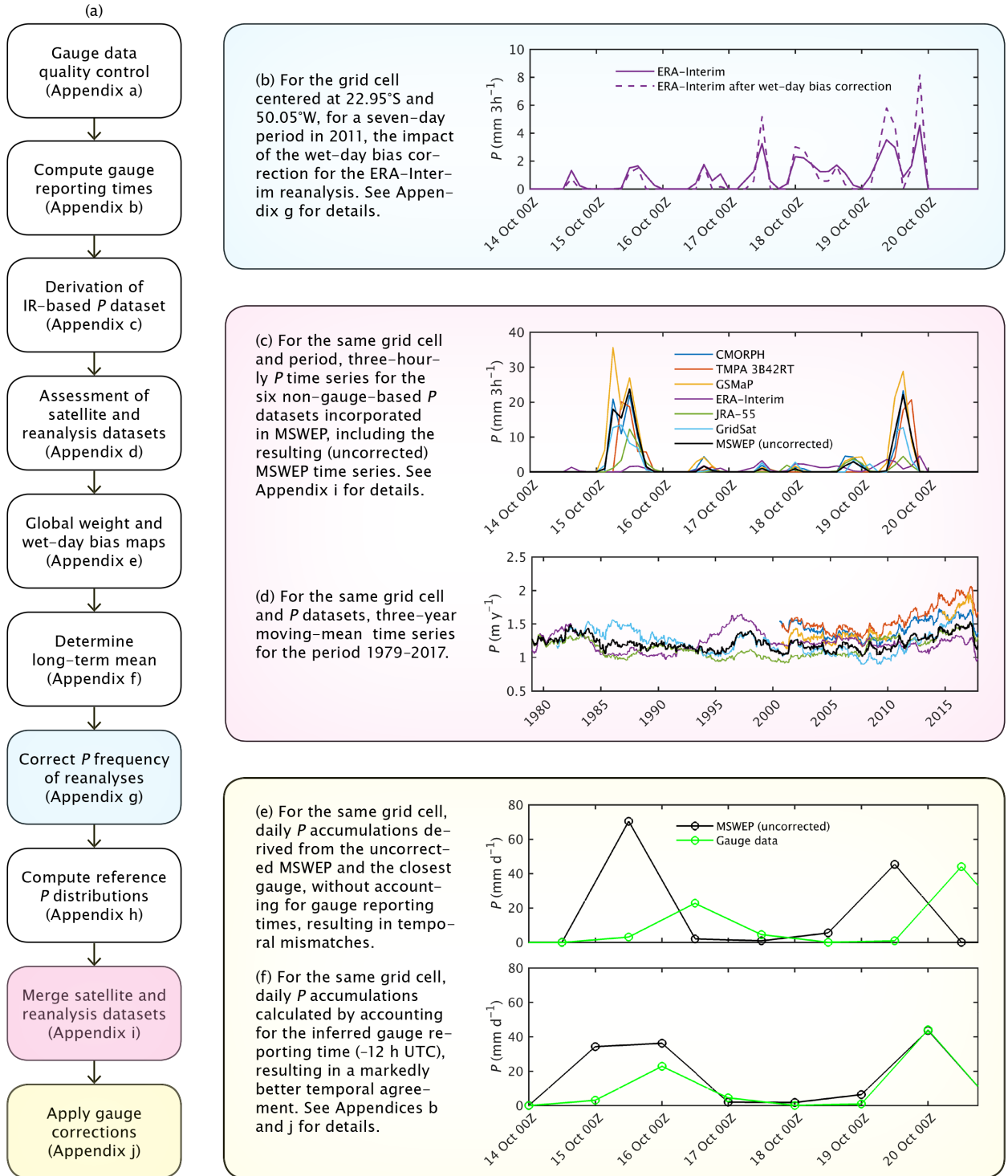


FIG. 1. (a) Flowchart outlining the main processing steps implemented to produce MSWEP V2. For each step, the reference to the Appendix subsection that provides detail is provided between parentheses. (b) Example of the wet-day bias correction for ERA-Interim. Time series of the satellite and reanalysis  $P$  datasets and MSWEP V2 are presented in (c) and (d). The importance of accounting for reporting times when applying gauge corrections is illustrated in (e) and (f).

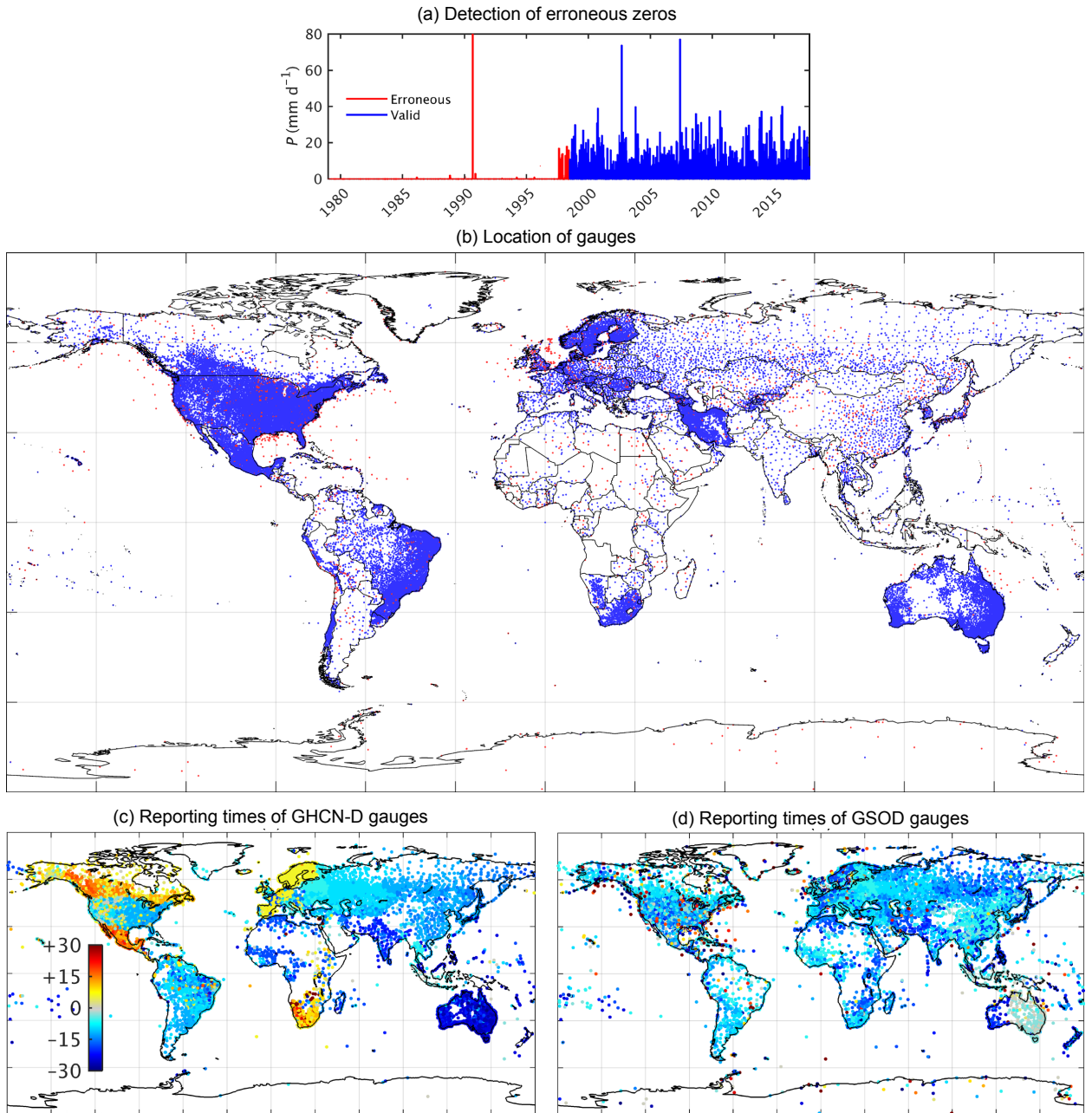


FIG. 2. (a) Daily  $P$  measured at GSOD station 038660 (50.58°N, 1.30°W) with the automatically detected erroneous zeros indicated in red. (b) The gauges used to produce MSWEP V2 in blue ( $n = 76747$ ) and the gauges that did not pass the quality control in red ( $n = 4300$ ). Also shown are the inferred 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-hour period starting at 0600 UTC of the current day and ending at 0600 UTC of the next day.

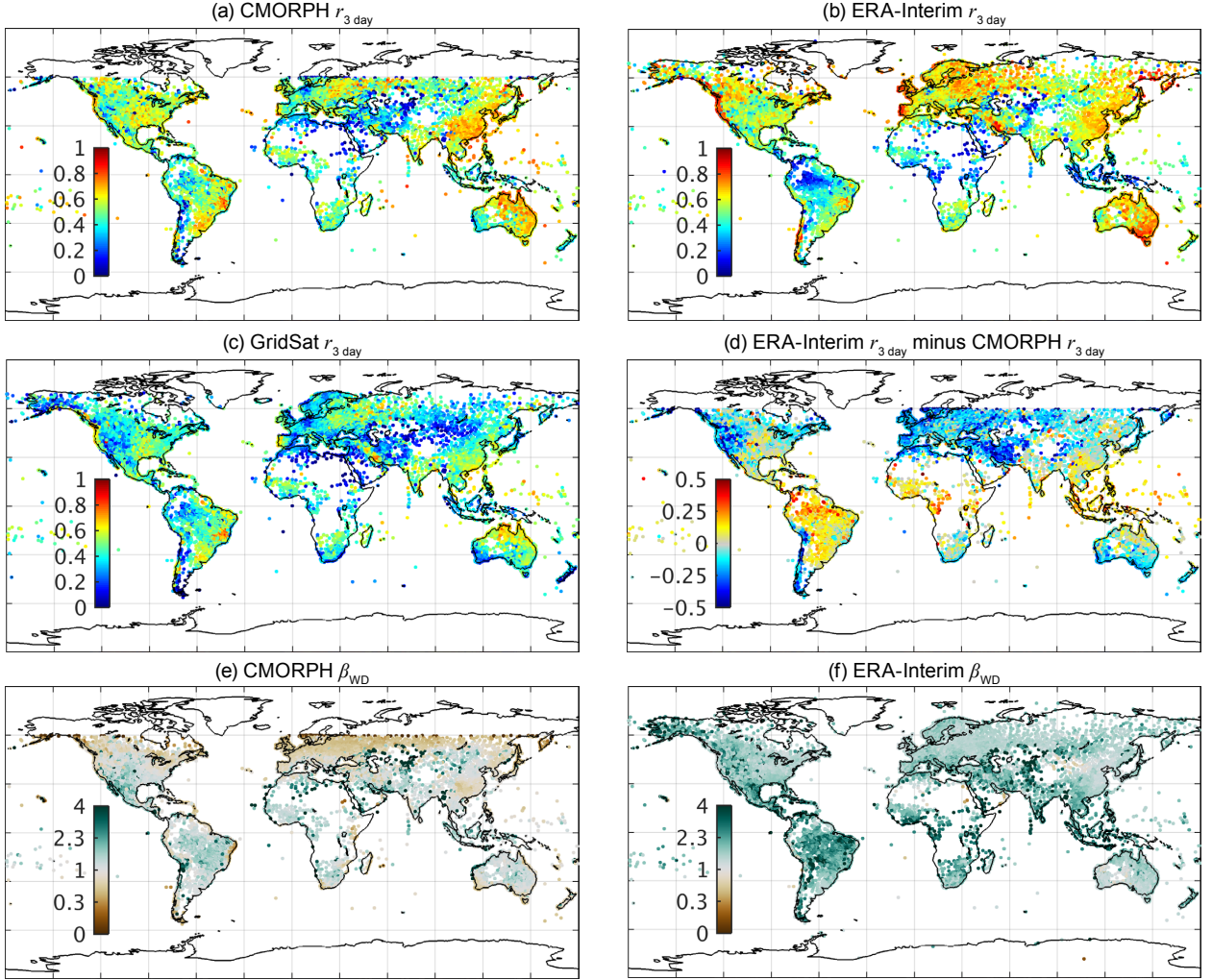


FIG. 3. Temporal correlations ( $r_{3 \text{ day}}$ ) for (a) CMORPH, (b) ERA-Interim, and (c) GridSat. The difference in  $r_{3 \text{ day}}$  values between CMORPH and ERA-Interim is presented in (d). Also shown is the bias in the number of wet days ( $\beta_{\text{WD}}$ ) for (e) CMORPH and (f) ERA-Interim (note the non-linear 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.



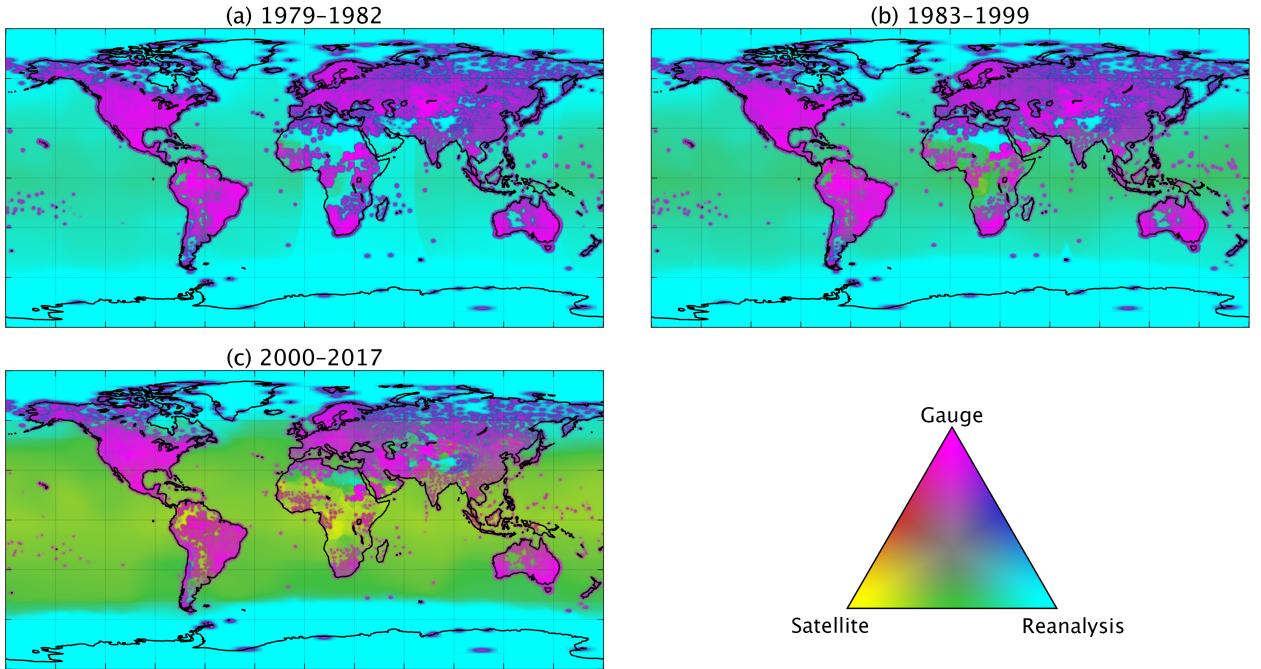


FIG. 4. Relative weights assigned to the gauge-, satellite, and reanalysis-based  $P$  estimates shown using a barycentric color map for the periods (a) 1979–1982, (b) 1983–1999, and (c) 2000–2017. The weights represent averages over the respective periods. The satellite and reanalysis weights represent cumulative weights assigned to the respective satellite and reanalysis  $P$  datasets.

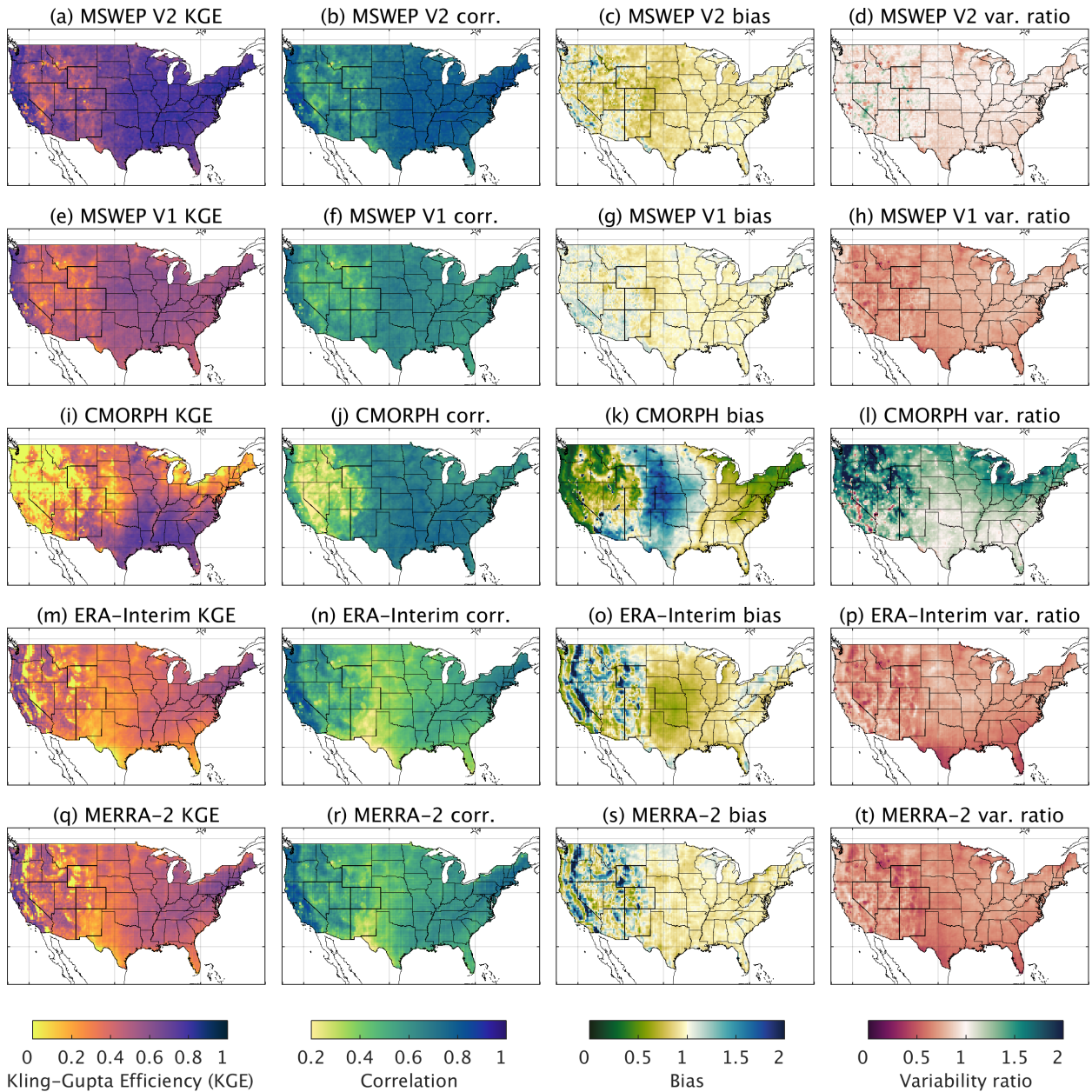


FIG. 5. KGE, correlation, bias, and variability ratio scores for the CONUS calculated from 3-hourly  $P$  time series using the Stage-IV gauge-radar dataset as reference. Regions without Stage-IV coverage are shown in white.

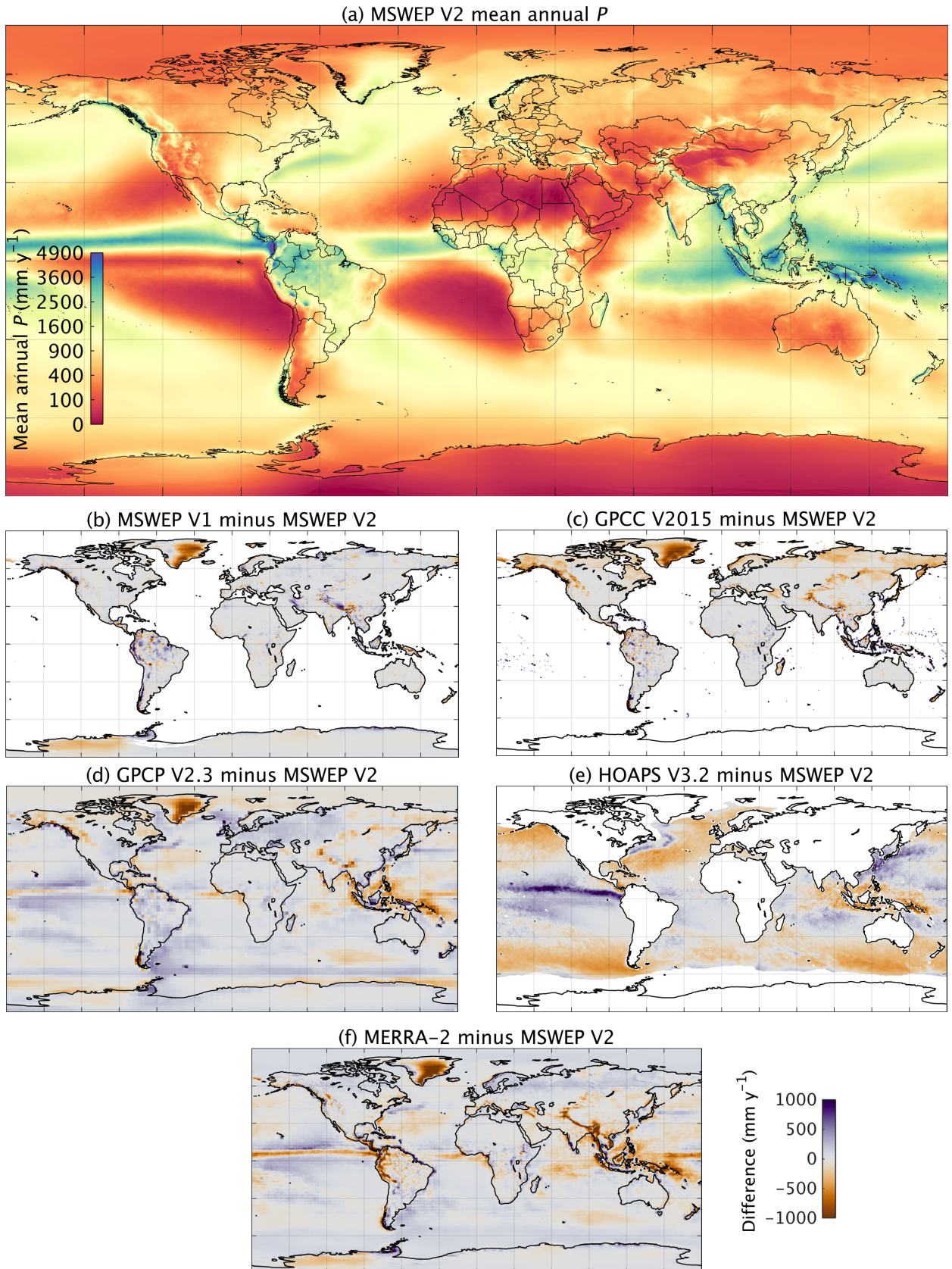


FIG. 6. The long-term mean  $P$  ( $\text{mm y}^{-1}$ ) for MSWEP V2 is presented in (a). Also shown are the differences in long-term mean  $P$  between MSWEP V2 and (b) MSWEP V1, (c) GPCC V2015, (d) GPCP V2.3, (e) HOAPS V3.2, and (f) MERRA-2. The colors represent 1979–2015 for MSWEP V1, 1987–2008 for



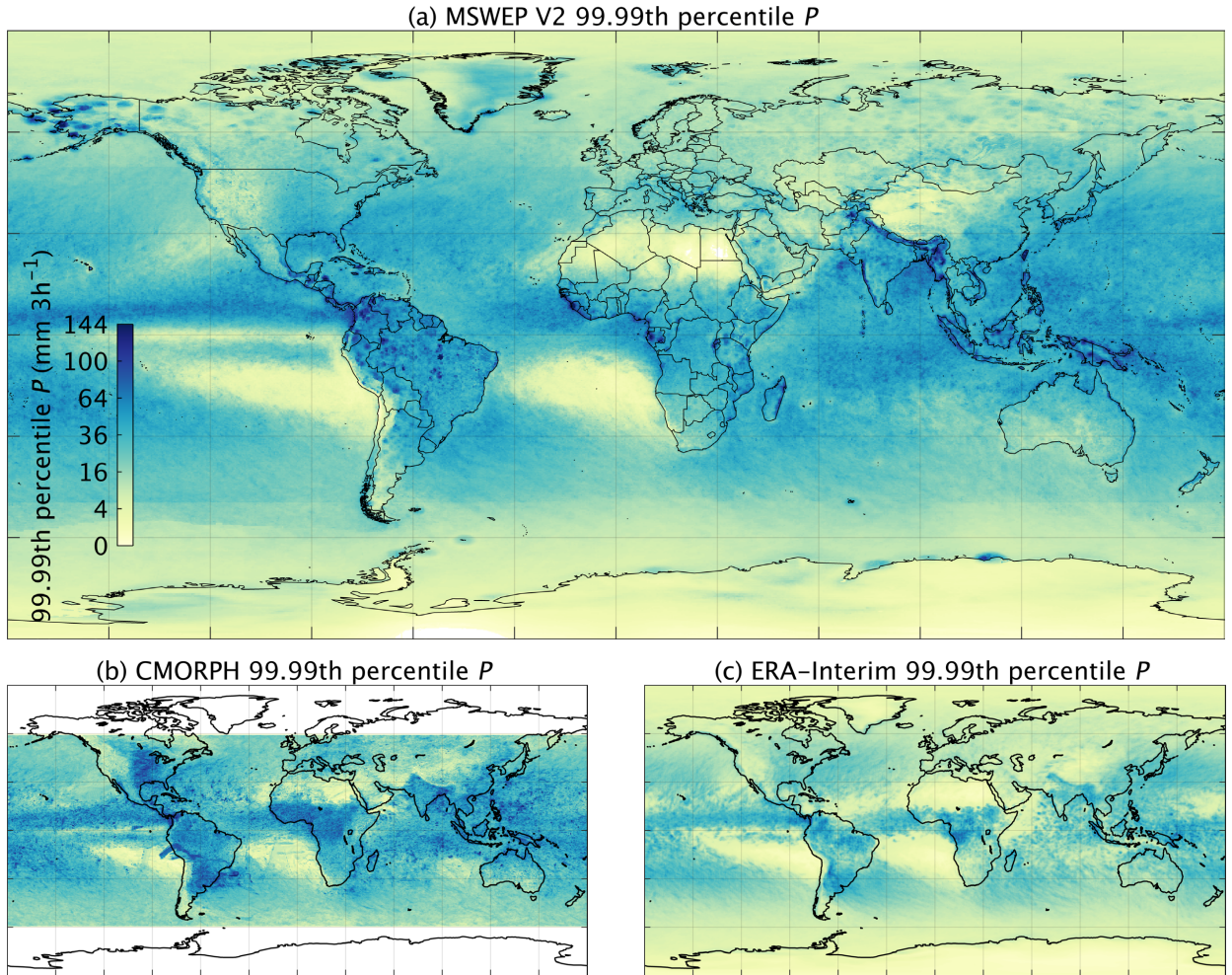


FIG. 7. Global maps of 99.99th percentile 3-hourly  $P$  amounts ( $\text{mm } 3\text{h}^{-1}$ ) for (a) MSWEP V2, (b) CMORPH, and (c) ERA-Interim for 2000–2017. CMORPH is limited to latitudes  $\leq 60^\circ$ . Note the non-linear color scale.

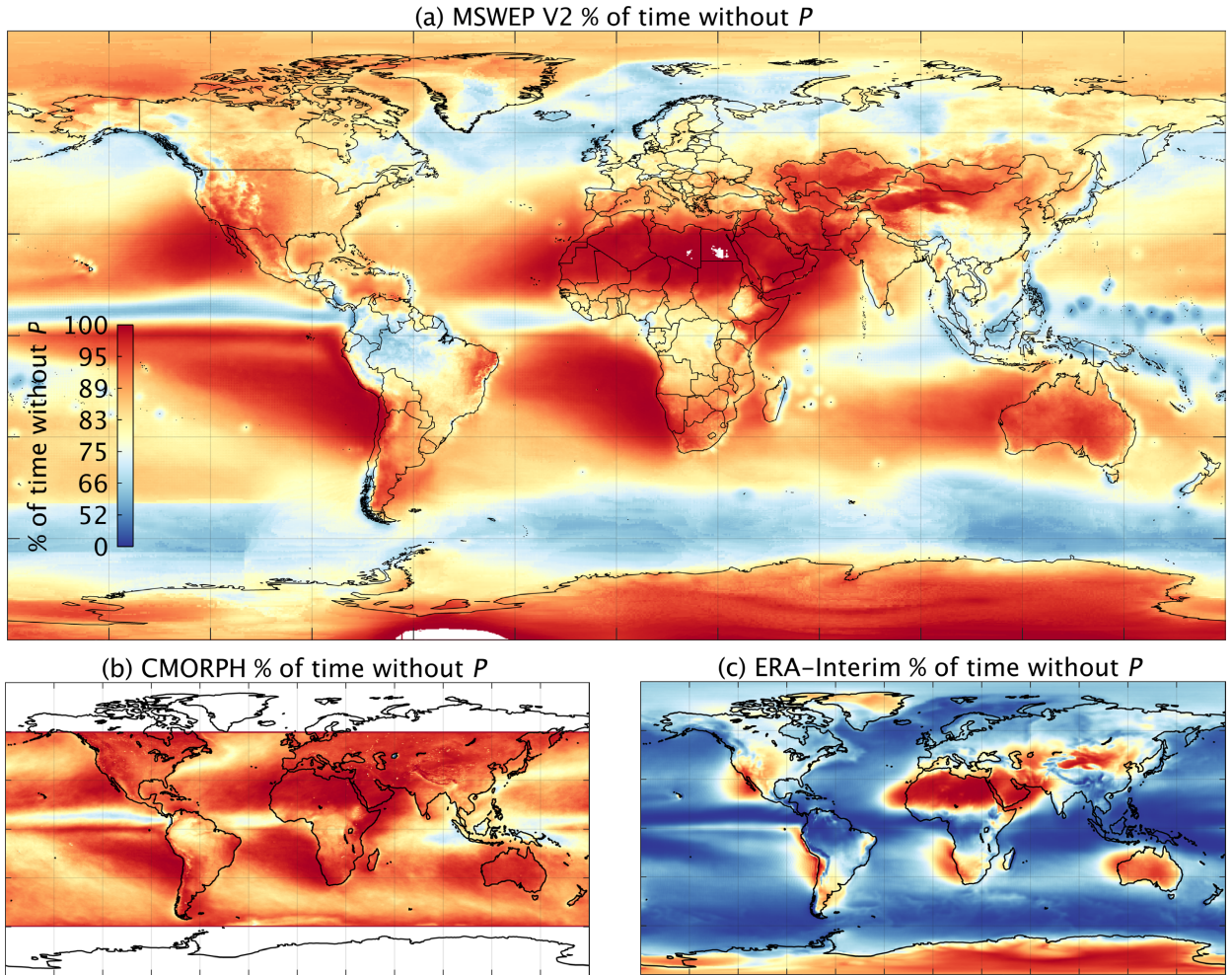


FIG. 8. The percentage of time without  $P$  for (a) MSWEP V2, (b) CMORPH, and (c) ERA-Interim for 2000–2017. A  $0.06 \text{ mm } 3\text{h}^{-1}$  threshold was used to identify 3-hourly intervals with  $P$ . Areas with no data are shown in white. Note the highly non-linear color scale.



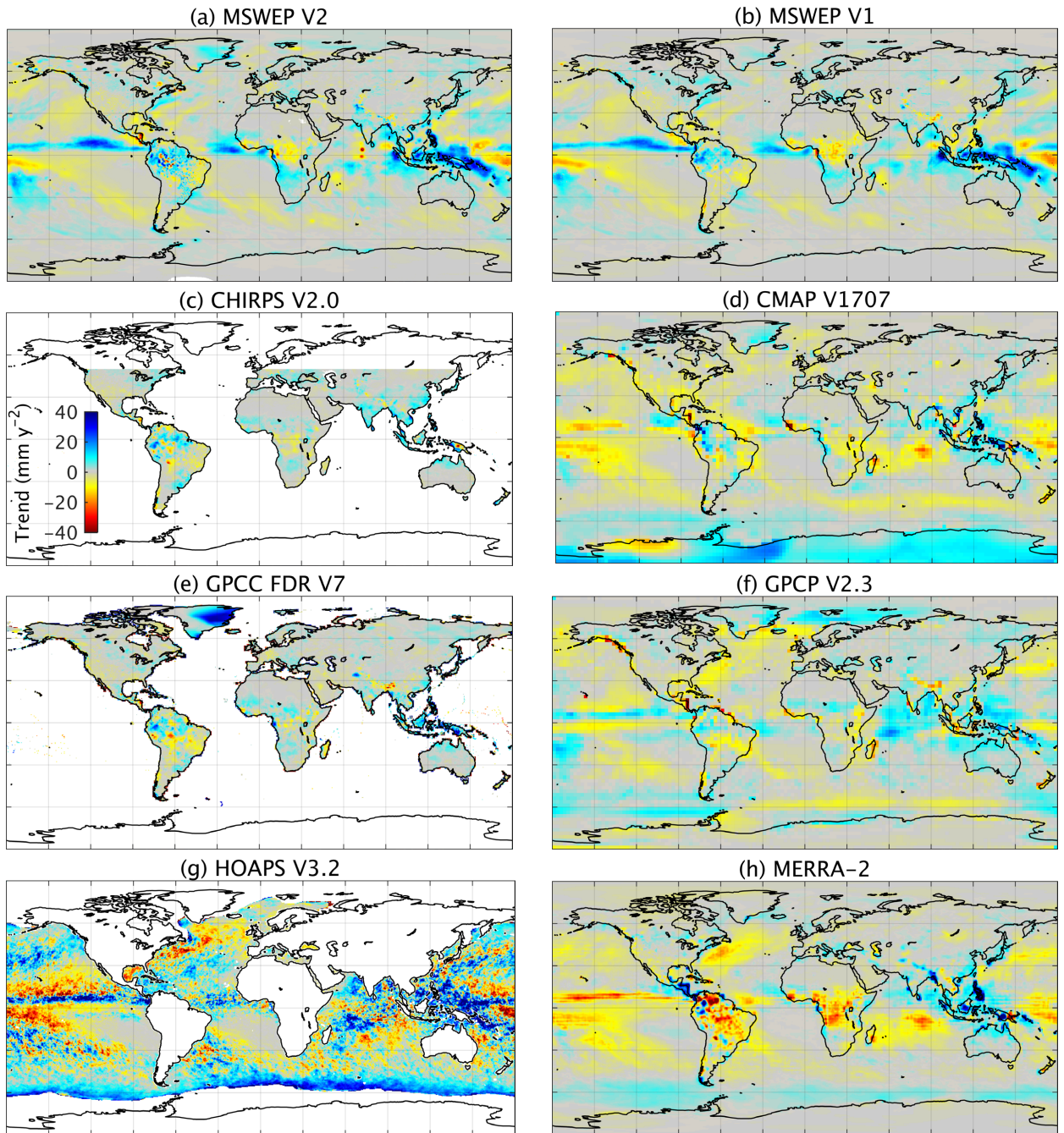


FIG. 9. Linear trends in mean annual  $P$  ( $\text{mm y}^{-2}$ ) for (a) MSWEP V2, (b) MSWEP V1, (c) CHIRPS V2.0, (d) CMAP V1707, (e) GPCC FDR V7, (f) GPCP V2.3, (g) HOAPS V3.2, and (h) MERRA-2. The trends represent 1979–2015 for MSWEP V1, 1981–2017 for CHIRPS V2.0, 1987–2008 for HOAPS V3.2, 1980–2017 for MERRA-2, and 1979–2017 for the other datasets. Areas with no data are shown in white.