SPECIAL
IPC12
COLLECTION

Assessment of Extremes in Global Precipitation Products: How Reliable Are They?

CHANDRA RUPA RAJULAPATI,[a] SIMON MICHAEL PAPALEXIOU,[a] MARTYN P. CLARK,[a] SAMAN RAZAVI,[a] GUOQIANG TANG,[a] AND JOHN W. POMEROY[a]

[a] *Centre for Hydrology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada*

(Manuscript received 13 February 2020, in final form 29 September 2020)

ABSTRACT: Global gridded precipitation products have proven essential for many applications ranging from hydrological modeling and climate model validation to natural hazard risk assessment. They provide a global picture of how precipitation varies across time and space, specifically in regions where ground-based observations are scarce. While the application of global precipitation products has become widespread, there is limited knowledge on how well these products represent the magnitude and frequency of extreme precipitation—the key features in triggering flood hazards. Here, five global precipitation datasets (MSWEP, CFSR, CPC, PERSIANN-CDR, and WFDEI) are compared to each other and to surface observations. The spatial variability of relatively high precipitation events (tail heaviness) and the resulting discrepancy among datasets in the predicted precipitation return levels were evaluated for the time period 1979–2017. The analysis shows that 1) these products do not provide a consistent representation of the behavior of extremes as quantified by the tail heaviness, 2) there is strong spatial variability in the tail index, 3) the spatial patterns of the tail heaviness generally match the Köppen–Geiger climate classification, and 4) the predicted return levels for 100 and 1000 years differ significantly among the gridded products. More generally, our findings reveal shortcomings of global precipitation products in representing extremes and highlight that there is no single global product that performs best for all regions and climates.

KEYWORDS: Risk assessment; Statistical techniques; Statistics; Diagnostics; Reanalysis data

1. Introduction

Global precipitation datasets, derived by combining data from various sources ranging from ground-based observations to radar and satellite data, are increasingly used by the Earth science community in applications such as land surface modeling, forcing and calibrating ecological and hydrological models, validation of climate models, trend analysis, water resources management, and extreme event characterization (MacKellar et al. 2007; New et al. 2000). Several gridded products [e.g., Climate Prediction Center (CPC) unified precipitation estimates, Global Precipitation Climatology Center (GPCC) precipitation dataset, NCEP–NCAR reanalysis data] have become available at various spatial and temporal resolutions based on different data sources (e.g., ground observations, satellites, radar, reanalysis) and data merging techniques. While such datasets are useful to investigate the spatial and temporal behavior in global precipitation (Fischer and Knutti 2014; Ghosh 2012; Trenberth et al. 2003), it is also important to quantify the differences in precipitation among the datasets and understand their reliability in estimates of extreme precipitation (Fischer and Knutti 2015; Ingram 2016; Min et al. 2011). Previous studies have shown that there are substantial differences between the precipitation estimates from different global precipitation datasets at a range of time scales from daily (Wong et al. 2017) to yearly (Sun et al. 2018), limiting understanding of the statistical behavior of regional and global precipitation. Therefore, the answer to the question of how reliable these datasets are in representing precipitation extremes remains still vague.

Precipitation is the main driver of terrestrial hydrology and therefore the most important input to hydrological models. Several sources of precipitation data exist, for example, ground measurements by precipitation gauges, remotely sensed data by radars and satellites, and reanalysis data that assimilate a myriad of observations into numerical weather prediction models. Ground measurements using precipitation gauges are the main source of information for point precipitation. However, observational records have limitations of sparse station network and/or gaps in records (Bell et al. 2015; Kidd et al. 2017). Satellite data, using infrared and microwave instruments, cover most parts of the globe overcoming the limitation of sparse network. Despite the limitations of the short record length, satellite data are widely used in hydrological studies, given the advantages in capturing extremes and poor gauged regions (Faridzad et al. 2018; Gado et al. 2017; Ombadi et al. 2018). Any systematic disturbances in the signal are usually corrected by using the ground observations. Reanalyses merge the ground observations and data from models that simulate physical and dynamic processes of climate system. Reanalysis data are highly dependent on the selected climate model (Trenberth et al. 2011),

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JHM-D-20-0040.s1.

Corresponding author: Simon Michael Papalexiou, sm.papalexiou@usask.ca

DOI: 10.1175/JHM-D-20-0040.1

leading to uncertainties in the precipitation estimates (Gehne et al. 2016).

Merged gridded precipitation products [e.g., Climate Hazards Center Infrared Precipitation with Station Data (CHIRPS) and Multi-Source Weighted Ensemble Precipitation (MSWEP)] have been developed in part to address these issues and been used to study the behavior of precipitation as well as to investigate changes in climatic means and extremes (Alexander et al. 2006; Beck et al. 2019; Rajah et al. 2014). Developing gridded products involve multiple layers of assimilation with various decisions (for example, choice of original data and of interpolation techniques, treating missing values) leading to differences in the final gridded products. A range of regional to global studies have compared gridded products and demonstrated inconsistencies (Akinsanola et al. 2017; Beck et al. 2019; Burton et al. 2018; Contractor et al. 2015; Dinku et al. 2008; Donat et al. 2014; Hu et al. 2018; Javanmard et al. 2010; Kidd et al. 2012; Sun et al. 2014; Wang and Zeng 2015; Yin et al. 2015; Zhang et al. 2013). Typically, differences in mean, total, and various statistics of precipitation are calculated at different time scales for comparing the datasets. Large differences in magnitude, with a deviation of 300 mm yr^{-1} in the annual precipitation among the datasets, have been noted at global scales (Sun et al. 2018). Large spatial variations have also been observed at daily and monthly temporal scales. Inconsistencies exist even between products derived purely based on gauge observations, owing to deficiencies in the data sources, wind undercatch of solid precipitation, recording of trace events, and different interpolation algorithms used to generate these products (Ahmed et al. 2019; Gehne et al. 2016; Newman et al. 2019; Pomeroy and Goodison 1997). Satellite-based datasets, adjusted typically by precipitation gauge observations (e.g., GPCC dataset) to increase their reliability (Adler et al. 2003), show also significant differences (Burton et al. 2018). While these studies expose the variation in average precipitation among global datasets, differences in extreme precipitation have not been adequately outlined.

In the past, many climate indices [e.g., Expert Team for Climate Change Detection Monitoring and Indices (ETCCDMI)] have been developed for understanding the statistical behavior of extremes across globe (Donat et al. 2013a,b; Papalexiou and Koutsoyiannis 2016; Zhang et al. 2011). Such indices, in general, have been used to investigate wide climate variations and included indicators such as R99p (annual total precipitation from days greater than 99th percentile), R20 (number of days with precipitation greater than 20 mm), CWD (maximum number of consecutive days when precipitation is greater than 1 mm), among other indices (Chen and Knutson 2008; Gehne et al. 2016). While these indices do provide a general perspective about the patterns of extremes and have been used mainly to assess changes in precipitation, a comprehensive characterization of the tail behavior of extremes at the global scale using gridded products has not been performed. While previous studies expose the differences in the aforementioned climate indices among global datasets (Contractor et al. 2015; Gehne et al. 2016), no study has yet scrutinized these datasets for practical applications where in the interest is a T-yr return level or the probability of exceeding a threshold for a given return period T (or frequency $1/T$) or the precipitation corresponding to a given return period. Understating the behavior of extremes in terms of tail heaviness, and exposing potential differences among the various products, is crucial for many practical applications including hydrological design and water resources management.

The study aims to evaluate how precisely different datasets represent the tail index and rare events at global scale. The objectives are to 1) quantify the tail index or the heaviness of the tails across different climate/geographical regions, 2) investigate the spatial pattern of extremes, and 3) compare the reliability of different datasets in describing extremes. The behavior of precipitation extremes in these data products is fully described by modeling the tails. This is done by fitting the power type and the stretched exponential tails, and estimating the probability of rare events such as 100- and 1000-yr return periods, which are in general used in the design of hydraulic infrastructure.

2. Methods and data

a. Datasets

Extreme precipitation is compared from five global datasets: 1) Climate Prediction Center (CPC) Unified Gauge-Based Analysis of Global Daily Precipitation, 2) Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks–Climate Data Record (PERSIANN-CDR, hereafter shortened to PERSN-CDR) v01r01, 3) Multi-Source Weighted-Ensemble Precipitation (MSWEP) v2.0, 4) Climate Forecast System Reanalysis (CFSR) v2, and 5) Water and Global Change (WATCH) Forcing Data–ERA-Interim (WFDEI) version 14 August 2018. The PERSN-CDR, MSWEP, and WFDEI datasets combine information from observations, satellites, and reanalysis. The CPC uses only observations and the CFSR is purely a reanalysis product. PERSN-CDR is derived from the satellite data (Gridsat-B1), adjusted using the precipitation data from Global Precipitation Climatology Project (Ashouri et al. 2015; Nguyen et al. 2018; Sorooshian et al. 2014). MSWEP v2.0 assimilates the ERA-Interim and JRA-55 reanalyses products as well as gauge [WorldClim, Global Historical Climatology Network Daily (GHCN-Daily), GPCC, CPC, and others] and satellite data (CMORPH, GridSat, GSMaP, and TMPA 3B42RT) (Beck et al. 2017, 2019). WFDEI is derived using the methodology of WATCH forcing data [based on the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-40 reanalysis with sequential elevation correction of meteorological variables and monthly bias correction from gridded observations] making use of the ERA-Interim reanalysis data (Weedon et al. 2014). The CPC combines precipitation from several in situ observation sources (from national and international agencies) and uses an optimal interpolation objective analysis technique (Chen et al. 2008). CFSR by the National Centers for Environmental Prediction (NCEP) is a coupled atmospheric–ocean–land surface–sea ice reanalysis product (Saha et al. 2014). Details such as the spatiotemporal resolution and time period of these datasets are given in Table 1.

More products are available, e.g., those using ground observations (e.g., GPCC), other reanalysis products (e.g., ERA5, MERRA), and a range of different satellite and radar products. Yet the spatial resolution and temporal coverage of these datasets differ in most cases. Therefore, five datasets were

TABLE 1. Details of the global precipitation datasets used in the work.

Dataset name	Short name	Temporal resolution	Spatial resolution	Temporal coverage	Record length (T)	Data source	Spatial coverage	References
National Centers for Environmental Prediction (NCEP) Climate Forecast System Reanalysis (CFSR)	CFSR	Subdaily	0.5°	1979–2017	39	Reanalysis	Global	Saha et al. (2014)
Multi-Source Weighted-Ensemble Precipitation	MSWEP	Daily	0.5°	1979–2016	38	Observations, satellite, reanalysis	Global	Beck et al. (2017, 2019)
CPC Unified Gauge-Based Analysis of Global Daily Precipitation	CPC	Daily	0.5°	1979–2017	39	Observations	Land	Xie et al. (2007), Chen et al. (2008)
Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) Climate Data Record (CDR)	PERSN-CDR	Subdaily	0.25°	1983–2017	35	Observations, satellite	60°S–60°N	Ashouri et al. (2015), Sorooshian et al. (2014)
WATCH Forcing Data ERA-Interim (WFDEI) corrected using Global Precipitation Climatology Centre (GPCC)	WFDEI	Daily	0.5°	1979–2016	38	Observations, reanalysis	Land	Weedon et al. (2014)

chosen that 1) cover approximately the same time period and have the same spatial resolution, 2) have more than 30 years of recent data (1979–2017), and 3) have same temporal resolution, i.e., at daily scale. These products are widely used in the scientific literature (Contractor et al. 2020; Satgé et al. 2020; Xu et al. 2020).

In addition, ground observations (hereafter called InSitu) from GHCN-Daily data, an integrated database of precipitation from land surface stations across the globe, were considered to better understand the spatial variation in the tail index and return levels at global scale, and to provide a benchmark for the comparison of the products. Twenty-four thousand records were found from 100 000 stations that have 1) more than 35 years of data available, 2) percentage of missing values less than 20%, and 3) percentage of values with quality flags (such as failed gap check) less than 0.1%. Most of the selected stations are located in the United States, Europe, Australia and parts of Asia, southern South Africa, and central-eastern South America (Fig. S1 in the online supplemental material). The record lengths vary from 35 to 210 years. To increase the number of stations used, and derive more robust results, all available data and the full time period were used. This was tested over the continental United States and found that there was no impact on the tail index and its spatial variation between analyzing the full and the 1979–2017 period.

b. *Calculating the tail index*

The evaluation of tail heaviness is important for characterizing the extreme and rare events that often have adverse implications on the society. The tail of a distribution refers to the upper (or lower) part of the cumulative distribution function (CDF), and is linked with extremes, i.e., high (or low) values of the variable of interest. The upper part of CDF or the right tail of the distribution describes the behavior of extreme (rare) precipitation. The heaviness of tail represents the frequency and magnitude of extreme events, and quantifies the likelihood of extremes to occur. More frequent and larger extremes with respect to the average precipitation occur when a heavy tail is observed in a particular region. Therefore, assessment of a tail's heaviness is useful to understand the likelihood of extremes and thus guide risk management strategies.

The tail function $\overline{F}_X(x)$ of random variable X is the complimentary cumulative distribution function of X. Several classifications of tail functions exist, yet here two major tail types were selected based on their ability to represent precipitation extremes, that is, the Pareto II (PII) and the Weibull

(W) tail (Papalexiou et al. 2013, 2018). The tail functions of the PII and W distributions are given, respectively, by

$$\overline{F}_{\mathrm{PII}}(x) = \left(1 + \gamma\frac{x}{\beta}\right)^{-1/\gamma}, \tag{1}$$

$$\overline{F}_{\mathrm{W}}(x) = \exp\left[-\left(\frac{x}{\beta}\right)^{\gamma}\right], \tag{2}$$

where $\beta > 0$ is the scale parameter and $\gamma > 0$ is the shape (or the tail index) parameter which governs the asymptotic behavior of the tail. For PII, the range of shape parameter is $0 \leq \gamma \leq 0.5$; for $\gamma > 0.5$ the distribution has infinite variance and for $\gamma = 0$ it converges to exponential tail. As the value of γ increases, the tail becomes heavier. It is converse in the case of W, i.e., as γ decreases, the tail becomes heavier; the distribution is subexponential with heavy tail than the exponential for $\gamma < 1$, and for $\gamma > 1$, the distribution belongs to the hyperexponential class.

The tail sample, i.e., the data used to fit the tail function, cannot be uniquely defined. Different approaches exist, such as considering a fixed number of the most extreme events (e.g., equal to the numbers of years of data), considering all events above a fixed threshold, and considering a percentage of the largest values in the sample. Here, the largest 5% of the nonzero values were used to define the tail sample. The parameters were estimated by fitting the theoretical tails to the empirical ones by minimizing the probability root-mean-square error (PRMSE),

$$\mathrm{PRMSE} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\overline{F}_D(x_i)}{\overline{F}_E(x_i)} - 1\right]^2, \tag{3}$$

where n is the tail sample size; $\overline{F}_D(x_i)$ is the exceedance probability of x_i corresponding to the theoretical tail, i.e., PII or W; and $\overline{F}_E(x_i)$ is the empirical exceedance probability according to the Weibull plotting position given by $\overline{F}_E(x_i) = 1 - r(x_i)/(N+1)$, where $r(x_i)$ is the rank of x_i in an ascending ordered tail sample.

The PRMSE norm considers relative error between the theoretical and empirical values of the tail sample, thus weighing each point contributing to the sum (of the norm) equally. The PRMSE norm is unbiased, has low variance, and is efficient compared to the most commonly used norms such as the mean square error (Papalexiou et al. 2013). However, unbiased estimation of parameters crucial in considering tails because small biases in the tail index (or the shape parameter) can lead to substantial differences in return levels. To determine whether the proposed methodology of minimizing the PRMSE was truly unbiased, Monte Carlo (MC) simulations are performed—2000 random samples were generated from both distributions with randomly varying scale and shape parameters for a given sample size and the tails were fitted by minimizing the PRMSE, i.e., for a given distribution, 2000 time series are generated randomly for a given true parameter and sample parameter estimates are obtained for all the 2000 time series using PRMSE. Comparing the estimated and the true shape parameters showed that the shape parameter is indeed unbiased for both the tails (Fig. S2) and for all true shape parameter values.

The time series at some of the grids for CPC, PERSN-CDR, and WFDEI datasets have gaps. Only grid cells with 80% completeness of data were analyzed. This ensures that enough nonzero values were available at all locations around the globe including places where the precipitation is small (e.g., desert regions). Grid cells having less than 600 nonzero precipitation values were excluded to ensure that at least 30 values are included in the 5% tail sample.

To quantify spatial differences among the datasets, the Pearson cross-correlation coefficient ρ between tail index maps was calculated at continental and global scales. This coefficient describes the degree to which tail index values at two spatial locations (grids) are similar to each other. Calculating the spatial cross correlation is difficult because of the mismatched grid coordinates among the different products, even though the datasets have the same spatial resolution of 0.5° (except PERSN-CDR). Regridding is often performed in order to align grids, but this affects the calculation of extremes (Gervais et al. 2014). For example, regridding the MSWEP dataset, with starting coordinates (−89.75, −179.75), to match the CFSR grid [starting coordinates at (−90, −179.5)] results in a 20% reduction on an average in extreme values. To avoid changing the tail values, coarser-resolution maps (2° × 2°) were created. The tail index is averaged at the coarser grid instead of regridding the actual precipitation time series, thus avoiding alteration of extremes (Diaconescu et al. 2015). This is useful because initial coordinate difference becomes insignificant, and the mean value of the tail index in the coarser grid is more robust. In addition to the correlation coefficient, the return levels corresponding to various return periods were calculated by inverting the tail functions [Eqs. (1) and (2)], with $F(x) = 1/T$, where T is the return period. To understand the reliability of datasets in describing extremes, the probability of rare events (100 and 1000 years) is compared among the datasets at continental and global scales.

Gridded products are further compared with the InSitu data. Tail index at each station is calculated using both the distributions. To develop spatial visualization of the tail index across globe, spatial interpolation is required as the observations are point values. InSitu tail index values were averaged within a grid of 2° × 2° (used to calculate the cross correlation among tail indices across the datasets) instead of averaging, a priori, the time series and then estimating the tail index. This also helps in comparing the InSitu and gridded products.

3. Results

Both tails were fitted at all grids for all datasets. All products show subexponential tails, that is, heavier tails than that described by the exponential distribution that has been typically used in hydrological practice. In the case of the PII, there are grid cells where the shape parameter is less than zero, which implies a parent distribution with an upper bound. The percentage of cells with negative shape parameter for CPC, PERSN-CDR, MSWEP, CFSR, and WFDEI are 5%, 14%, 0.3%, 0.6%, and 0.7%, respectively. The distributions bounded from above are physically inconsistent (Papalexiou and Koutsoyiannis 2013), and a sample with negative shape parameter might be due to the sampling errors or variations. MC simulations show that negative shape parameter values

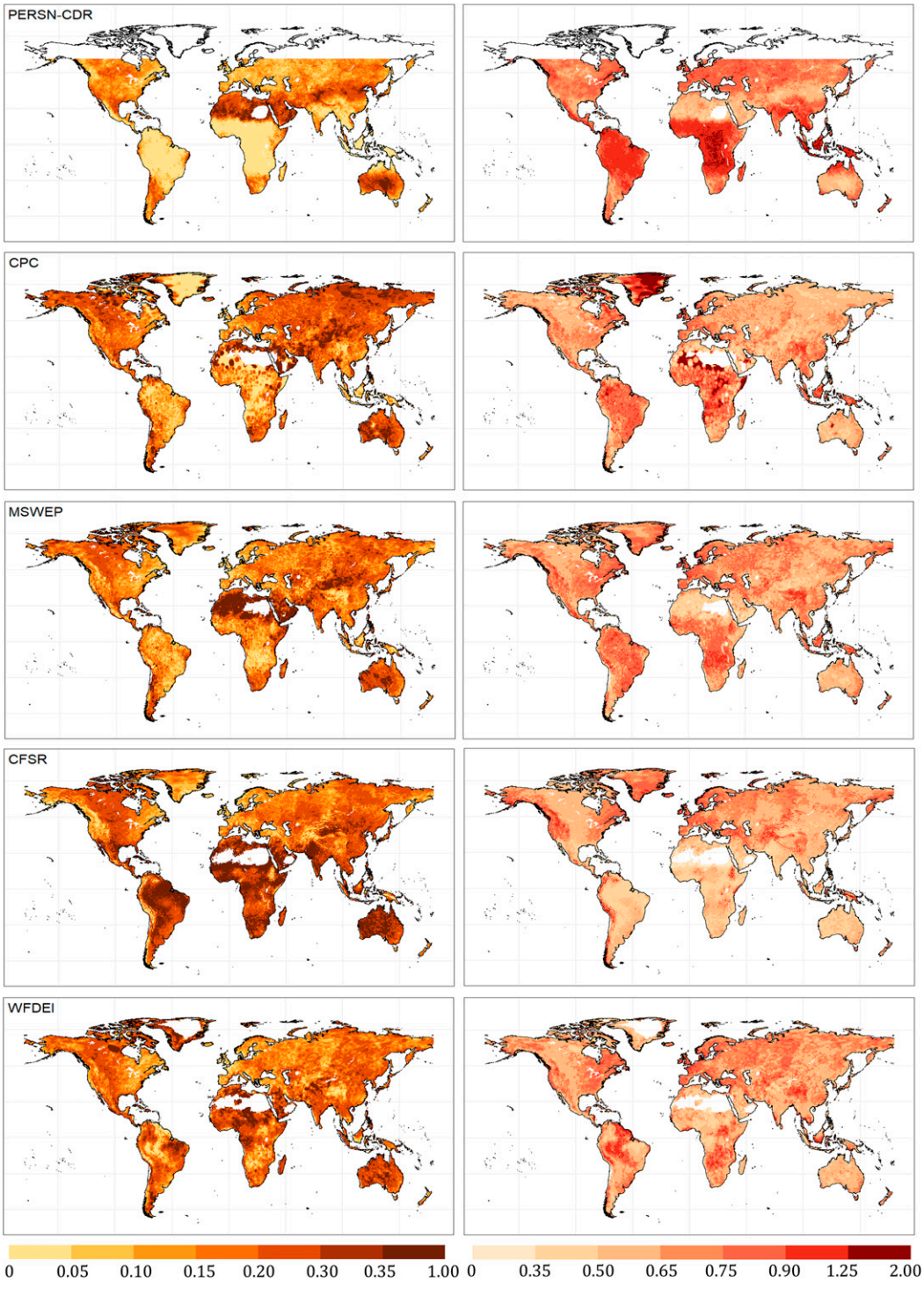

FIG. 1. Spatial variation of tail index using the (left) Pareto II and (right) Weibull distribution.

can be observed when the true tail is positive (Fig. S3). The number of negative values declines as the shape parameter increases and approach zero for shape parameters larger than 0.2. Therefore, for the PII tail, only positive shape parameter values were permitted.

The spatial pattern of the probability zero p_0 (dry days) was almost the same for all the datasets between 50°N and 50°S except in a few regions (Fig. S4). For example, in Southeast Asia (China, Myanmar, Cambodia, Vietnam), all the datasets have a low to medium p_0 ($0.25 < p_0 < 0.5$) except for PERSN-CDR.

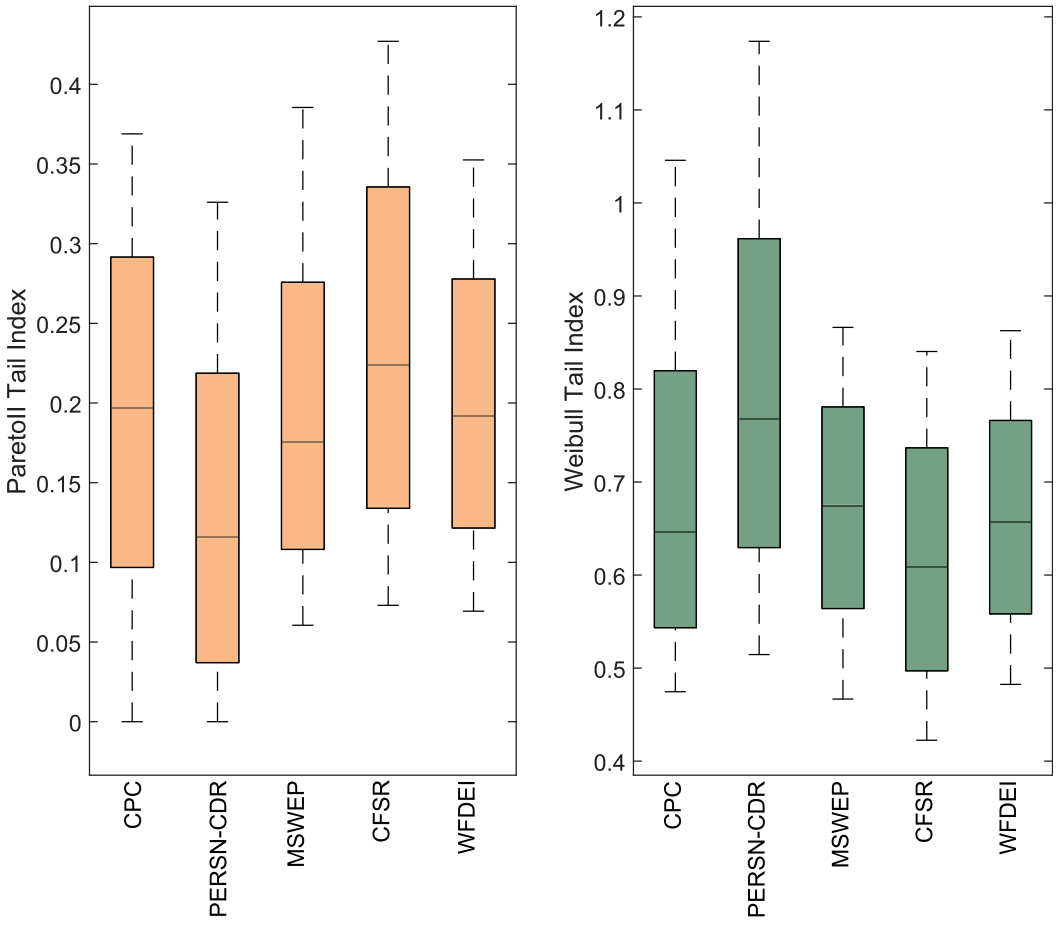

FIG. 2. Boxplots of tail index for Pareto II and Weibull at global scale. Whiskers denote the 95% empirical confidence interval. The boxplots were constructed using 61 556, 184 195, 61 556, 61 636, and 61 556 grids for CPC, PERSN-CR, MSWEP, CFSR, and WFDEI, respectively.

In equatorial Africa, however, all the datasets have a low p_0 ($0.02 < p_0 < 0.25$) except those from the CPC, which have high values over Africa. This difference in the results of the two regions may be due to the sparse gauge network and interpolation effects between distant gauges in equatorial Africa. For regions north of 50°N the WFDEI has high p_0 while the rest of datasets (except PERSN-CDR) have low p_0, except in Greenland.

a. *Tail index across different climate/geographic regions*

The spatial pattern of the tail index (Fig. 1) shows a close association with the Köppen–Geiger climate classification (Kottek et al. 2006). Light tails ($\gamma_{\text{PII}} < 0.05$ and $\gamma_{\text{W}} > 0.65$) were observed in humid equatorial climates, where the minimum temperature is >18°C and monsoonal precipitation occurs. Hot, arid climates with winter minima in precipitation have heavy tails ($\gamma_{\text{PII}} > 0.3$ and $\gamma_{\text{W}} < 0.45$). Warm, humid, temperate climates with higher precipitation show light to medium tails. The polar regions with temperature maxima < 10°C, characteristic of tundra, generally have medium tails. While northern North America and Asia, where there is a seasonal snow cover and a humid continental climate, have medium to heavy tails, the subhumid continental climate interior regions have heavy to medium heavy tails. In the arid and semiarid regions of Australia and the southwestern United States, heavy to medium tails are observed. Also, southwestern Africa has an arid climate with dry summers and so exhibits heavy to medium tails. In general, light tails are observed in the coastal regions of North America, Australia, Europe, and South America.

The causes of extremes that result in specific types of precipitation (e.g., convective or orographic) cannot be identified from the tail index, and a high tail index value does not necessarily mean high precipitation. For example, Brazilian and African tropical forests typically receive high precipitation, but volumes are consistently high and therefore a light tail occurs. However, in the arid and semiarid parts of Africa and Australia, there is a heavy tail as a result of extreme precipitation events that are much larger than the average. Therefore, an understanding of the tail behavior helps in risk management for extreme precipitation events.

b. *Differences in tail index among datasets*

Though the spatial pattern of the probability zero p_0 (dry days) is almost the same for all the datasets (Fig. S4),

TABLE 2. Summary statistics of tail index at global scale.

Distribution	Dataset	Mean	Median	SD	Min	Max	Q_5	Q_{95}
Pareto II	CPC	0.191	0.197	0.128	0	0.937	0	0.369
	PERSN-CDR	0.121	0.116	0.114	0	0.817	0	0.326
	MSWEP	0.192	0.176	0.102	0	1.000	0.061	0.385
	CFSR	0.236	0.224	0.118	0	1.000	0.073	0.427
	WFDEI	0.200	0.192	0.090	0	0.795	0.069	0.353
Weibull	CPC	0.693	0.646	0.237	0.232	2.000	0.475	1.046
	PERSN-CDR	0.793	0.768	0.191	0.254	1.665	0.515	1.174
	MSWEP	0.673	0.675	0.120	0.190	1.228	0.467	0.866
	CFSR	0.617	0.609	0.133	0.179	2.000	0.422	0.840
	WFDEI	0.663	0.657	0.119	0.265	1.589	0.482	0.863

considerable differences are noted in the spatial patterns of tail index among the datasets (Fig. 1). Summary statistics of tail index also show large deviations among the datasets with mean values varying from 0.12 to 0.24 and 0.62 to 0.8 for PII and W, respectively (Fig. 2; Table 2). The heaviest tails (global mean) were observed in the CFSR product and the thinnest tails in the PERSN-CDR. There is a 0.115 difference in the mean value of the PII tail index between these two datasets emphasizing the large differences. For example, assuming a scale parameter value equal to the global mean results in estimating 80 mm more precipitation for the 100-yr return level for the CFSR product, which is almost double the PERSN-CDR estimate. The CPC dataset derived from ground observations shows a higher standard deviation in the estimated tail index values than the rest of the data products.

The spatial patterns of CPC, PERSN-CDR, and MSWEP closely match each other (Fig. 1), whereas the pattern in the tail index for the CFSR and WFDEI datasets could not be discerned. The possible reason for consistency in the spatial pattern among the three datasets (CPC, PERSN-CDR, and MSWEP) is that they use ground observations. A light tail ($\gamma_{PII} < 0.05$ and $\gamma_W > 0.65$) was observed in the tropical forest biomes of South America, equatorial Africa, and Southeast Asia in the CPC, PERSN-CDR, and MSWEP datasets. Typically, a light tail was found near the equator compared to the polar regions, except Greenland, where light tails were observed. In the Alaska and northern Russia, heavy to medium-heavy tails were observed in all datasets, except CFSR. Heavy tails ($\gamma_{PII} > 0.3$ and $\gamma_W < 0.45$) were found in the CPC, PERSN-CDR, and MSWEP datasets from the deserts of Africa, and in all datasets from central and western parts of Australia (rangelands). Tropical and east coast Australia have heavy tails in CFSR and WFDEI datasets and light to medium-heavy tails in CPC, PERSN-CDR, and MSWEP datasets. In the eastern United States and in the Appalachian Mountains, light tails were found ($\gamma_{PII} < 0.15$ and $0.65 < \gamma_W < 1.0$) in all the

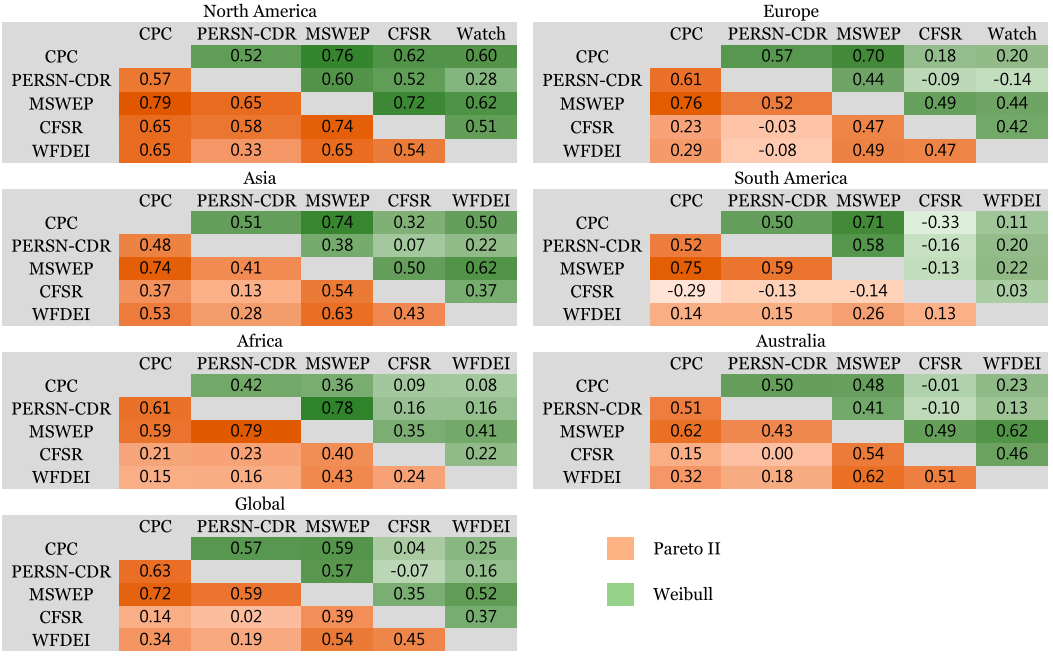

FIG. 3. Pearson cross correlations of tail index among the gridded precipitation products at global and continental scale. Higher values imply a stronger agreement in the spatial patterns of the tail index between two products.

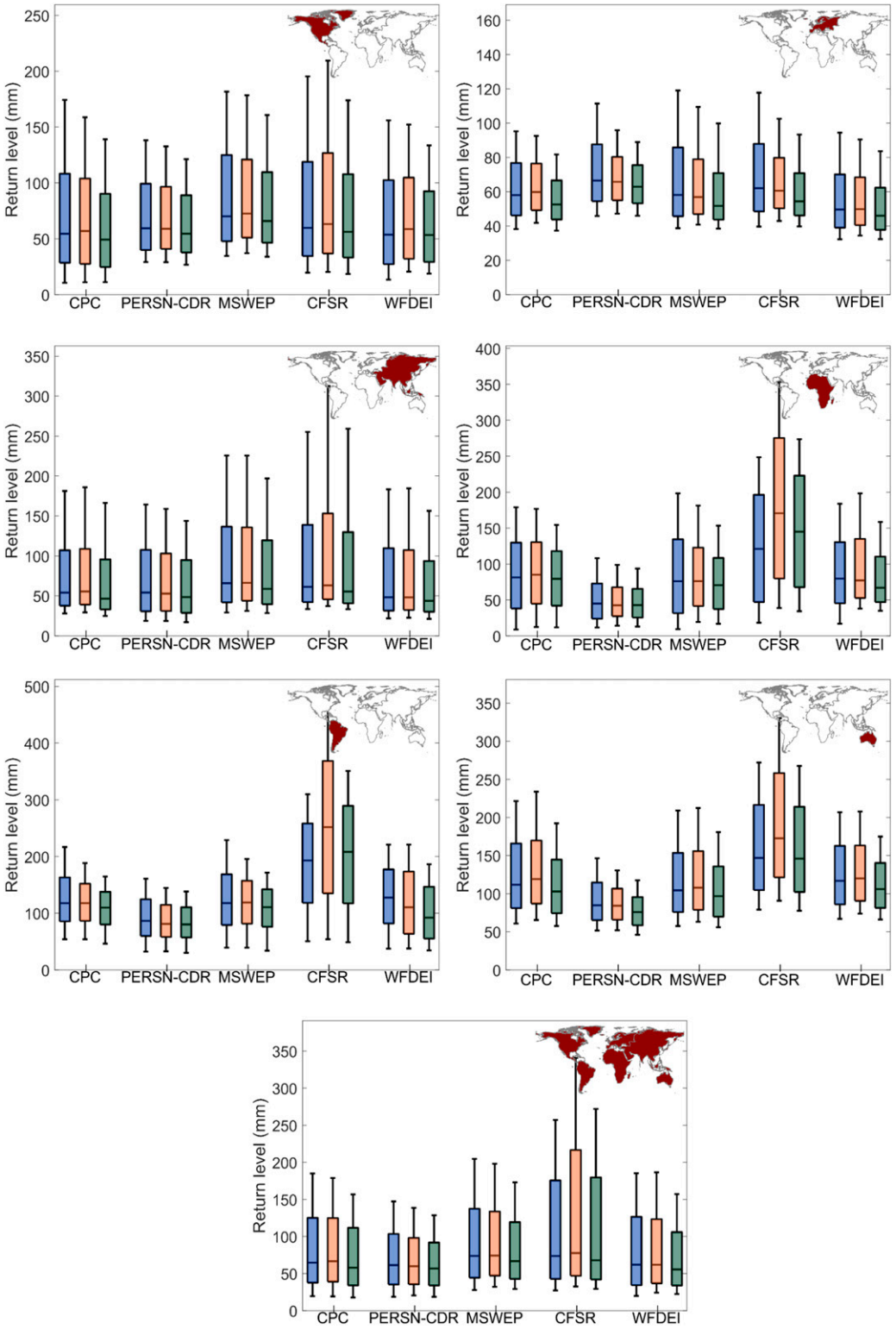

FIG. 4. Comparison of observed (blue) and calculated depths of precipitation (return levels) using Pareto II (orange) and Weibull (green) for T-yr return period (T = 39 years for CPC and CFSR; 38 years for MSWEP and WFDEI; 35 years for PERSN-CDR) at global and continental scale. Observed return levels are the maximum of observed precipitation. Whiskers denote the 95% empirical confidence interval.

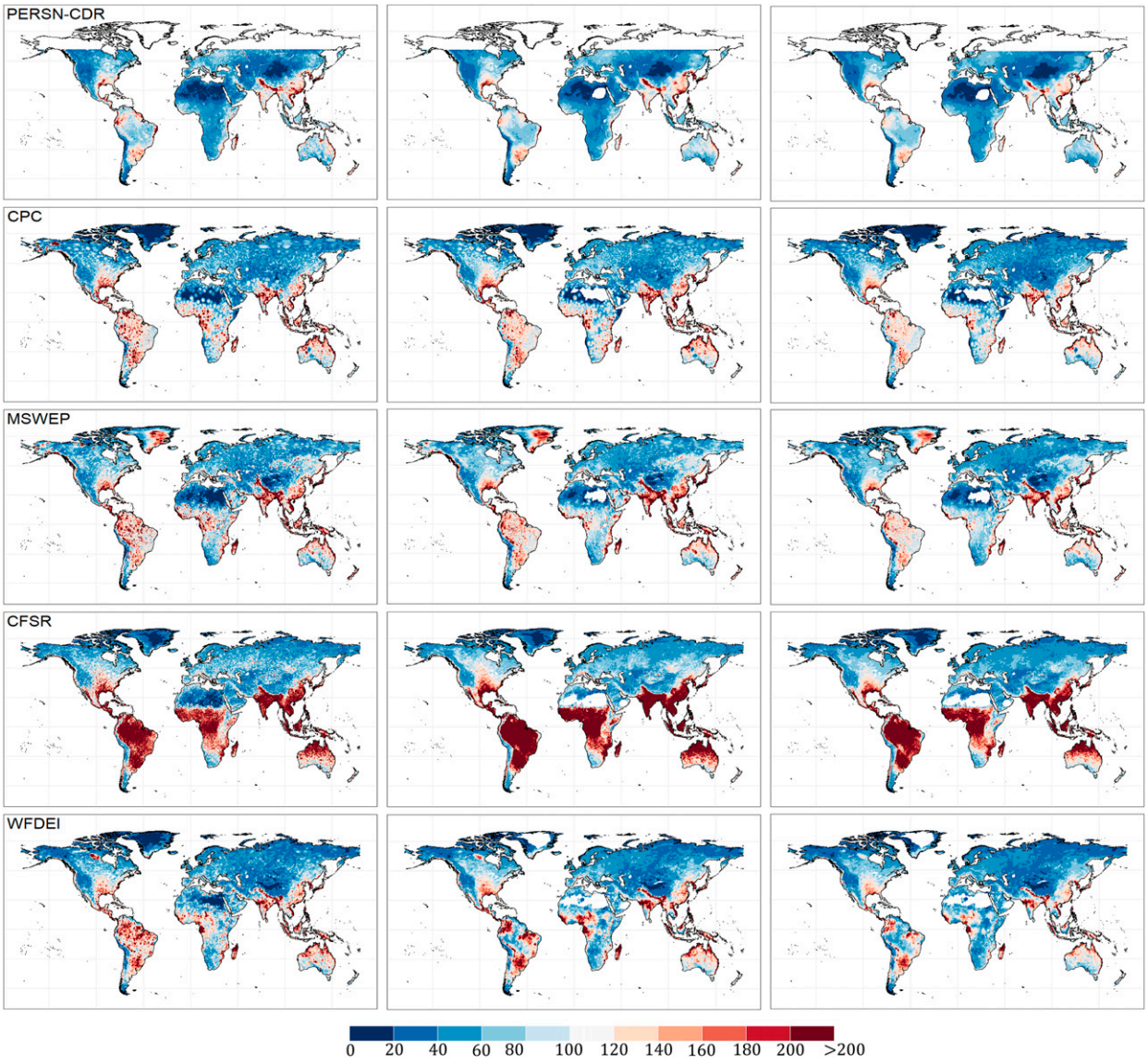

FIG. 5. Spatial patterns of (right) observed and calculated precipitation depths (return levels) using (center) Pareto II and (left) Weibull for T-yr return period (T = 39 years for CPC and CFSR; 38 years for MSWEP and WFDEI; 35 years for PERSN-CDR) at global and continental scale. Observed return levels are the maximum of observed precipitation.

datasets except CFSR. Yet in the western United States, medium tails ($0.2 < \gamma_{PII} < 0.3$ and $0.35 < \gamma_W < 0.5$) were found only in CPC and the other datasets showed light tails. Medium tails were found in the central northern United States and in the Rocky Mountains in MSWEP, CFSR, and WFDEI. Eastern Canada is consistently characterized by light tails in all datasets, while medium tails were found in western and northern Canada. In Asia, the tail index changes from region to region. In India, including the high-mountain Hindu Kush–Himalaya, the CPC, MSWEP, and WFDEI products indicate medium tails, but the CFSR indicates heavy tails. PERSN-CDR differs with light tails in the north and medium tails in the south of India. All datasets show heavy tails ($\gamma_{PII} > 0.35$ and $\gamma_W < 0.5$) in the Tibetan Plateau, except in CFSR. In the United Kingdom and Europe, light tails ($\gamma_{PII} < 0.15$ and $0.7 < \gamma_W < 2.0$) were observed in all the datasets. European high mountain regions show medium tails in the CPC, WFDEI, and CFSR datasets and light tails in PERSN-CDR and MSWEP.

The spatial cross correlation calculated using the Pearson cross-correlation coefficient ρ between tail index maps, shows the interaction among the mapped tail index patterns. Striking variations in the correlations of the tail index are evident among the five datasets at global and continental scales (Fig. 3). The CPC, PERSN-CDR, and MSWEP datasets are highly correlated with each other, while the CFSR and WFDEI are the least correlated with the others at global scale. At the continental scale, CPC and MSWEP are highly correlated,

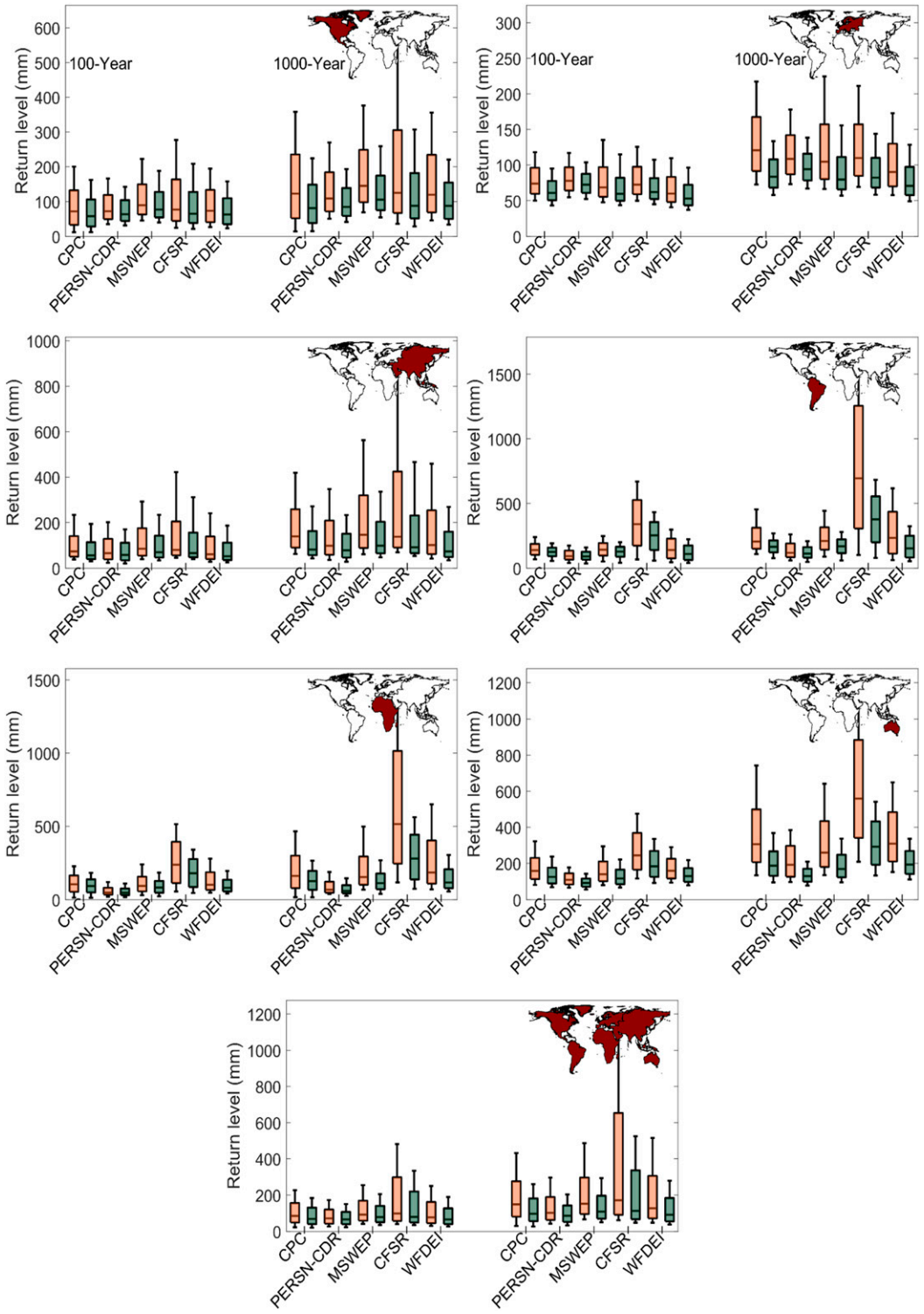

FIG. 6. Predicted 100- and 1000-yr return levels using both Pareto-II (orange) and Weibull (green) distributions at global and continental scale. The left side and right side of the plot shows 100-yr and 1000-yr return period, respectively. Whiskers denote the 95% empirical confidence interval.

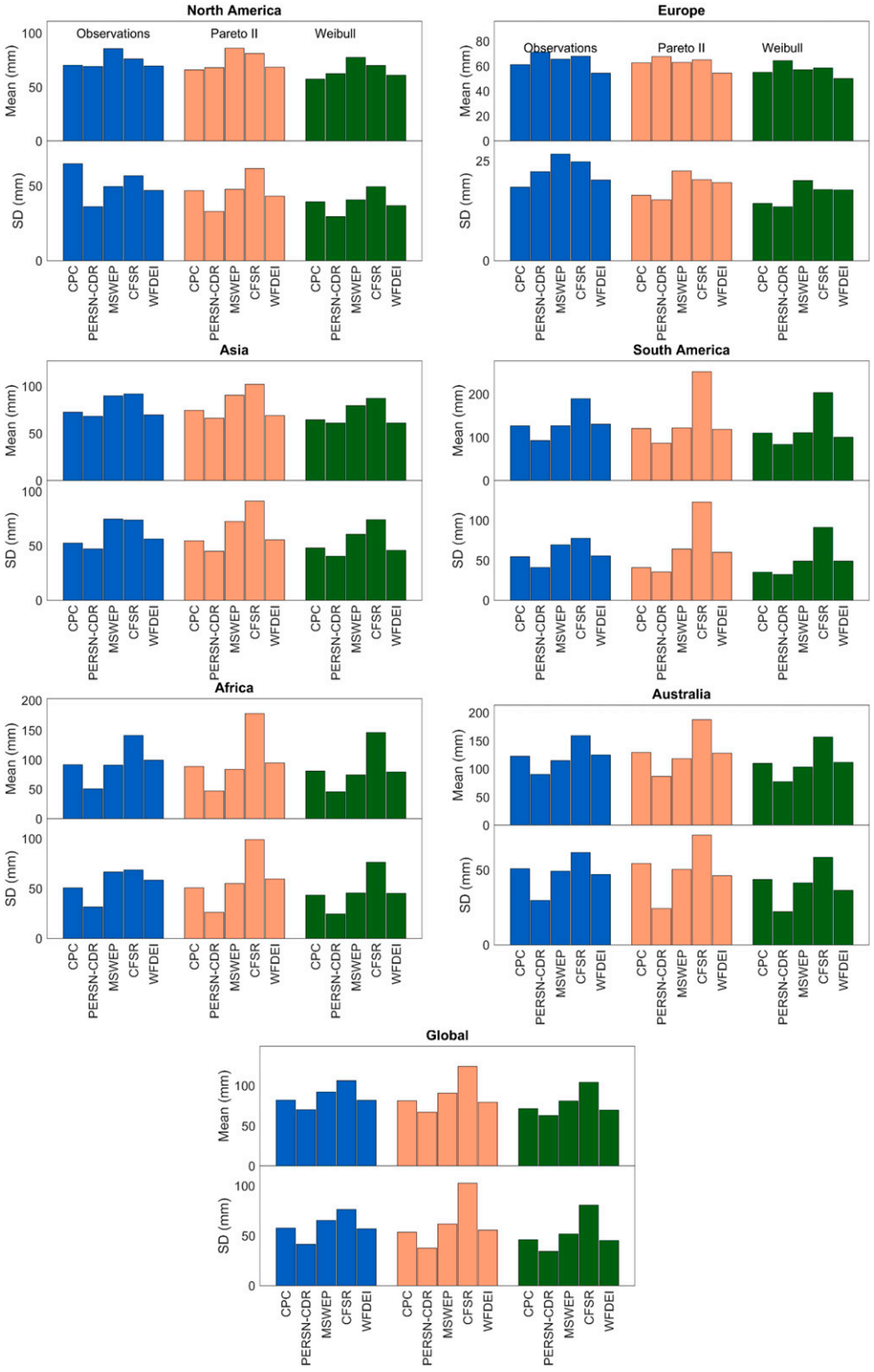

FIG. 7. Mean and standard deviation of the observed and calculated (using the PII and W distributions) precipitation return levels for T-yr return period at global and continental scale (T = 39 years for CPC and CFSR; 38 years for MSWEP and WATCH; 35 years for PERSN-CDR).

TABLE 3. Summary statistics of tail index for InSitu observations (InSitu) at global scale.

	Pareto II	Weibull
Mean	0.147	0.734
SD	0.075	0.118
Minimum	0.000	0.241
Maximum	1.000	1.896
Q_5	0.029	0.553
Median	0.142	0.730
Q_{95}	0.278	0.934

in Asia, Europe, and South America ($\rho > 0.7$). PERSN-CDR is moderately correlated with CPC in Asia, Europe, and North America ($\rho > 0.5$). However, the PERSN-CDR has a good correlation with MSWEP ($\rho > 0.78$) in Africa. The CFSR reanalysis dataset has very low to negative correlations with all of the other products in many regions. The reason for the low to negative correlation in Africa may be related to the observation network. While low correlations are observed in Africa for both tails among CFSR and other products, high correlations ($\rho > 0.6$ approximately) are observed in North America among all the datasets, including CFSR. This could be due to the existence of a higher-quality and denser observation network in North America, compared to the sparser African networks.

c. *Discrepancies among datasets in describing extremes*

Hydrological design and risk management use precipitation depths corresponding to specific return periods. To investigate how the spatial variation of the tail index relates to return levels of specific return periods, the observed maxima corresponding to T years (record length) was compared with the predicted using the fitted tails (Fig. 4). The boxplots with blue colors show the observed return levels (maxima of observed precipitation record); orange and green colors show the T-yr return levels calculated using the fitted PII and W tails, respectively. Note that record length varies from 35 to 39 years among the different datasets (Table 1). There is a close agreement between the observed maxima and the calculated T-yr return levels. The PII tail slightly overestimates while the W tail underestimates all datasets (Fig. 4). Histograms clearly show the difference between or the ratio of the observed and calculated maxima (see Fig. S5). Figure 4 also reveals differences in the observed and calculated maxima among datasets.

The spatial patterns of the calculated return levels using the PII and W distributions match well with observations, but there is a considerable variation across the datasets. Figure 5 shows the observed and the calculated return levels using the PII distribution and the W distribution. The spatial patterns of CPC and MSWEP closely match except in Greenland. The reanalysis-based CFSR dataset has higher values of maxima in the tropics and the satellite-based PERSN-CDR dataset has consistently lower values across the globe compared to the other products. While the spatial patterns of all datasets match well in North America, CFSR has slightly higher values compared to the rest in the southeastern United States.

For hydrological infrastructure designs, return levels for 100- and 1000-yr return periods are typically used. The difference in the predicted 100- and 1000-yr return levels between the datasets is very high (Fig. 6), specifically in Africa, South America, and Australia. The difference between the 100- and 1000-yr tails is high for CFSR (~41 and 165 mm, respectively) at the global scale, and low for PERSN-CDR both at global and continental scales (~8.7 and 29 mm, respectively, at global scale). The average differences for the 1000-yr return period are 396 and 340 mm respectively for South America and Africa. In some of the South American grids these differences are as high as 3000 and 9000 mm at 100- and 1000-yr return periods, respectively. However, in North America and Europe, differences were relatively small, on average 8 and 24 mm, respectively for the 100- and 1000-yr return periods. Though these differences highlight the critical importance in choosing the right tail, they also emphasize that uncertainty is reduced when using dense networks of high quality observations.

There is a 35 mm difference in the mean value of the observed maxima among the global datasets (Fig. 7). CFSR has consistent higher values and PERSN-CDR has lower values among the products at global and continental scales, except in North America and Europe. CFSR compares closely with the other datasets in North America, Asia, and Europe with differences in the mean values of less than 20 mm. The differences become larger in Africa and South America (90 mm) and Australia (70 mm). Even the 100- and 1000-yr return levels of the CFSR have high values compared to the other datasets at both the global and continental scales, with comparable values in North America (Fig. S6). The mean 100-yr return levels of PERSN-CDR are approximately twice the CFSR levels, and become 4 times larger for the 1000-yr levels at the global scale. The differences are large at continental scales as well, specifically in Africa and South America. The 1000-yr return level in CFSR is almost 8 and 6 times higher than PERSN-CDR in Africa and South America, respectively. Similarly, large differences in the standard deviations of observed maxima, T-, 100-, and 1000-yr return levels. Again, PERSN-CDR has consistently lower values compared to all the datasets at both global and continental scales, and CFSR has higher values except in Europe. The differences are high in South America and low in Europe and North America compared to other continents.

d. *Comparison with in situ observations (InSitu)*

The InSitu data show an average value of 0.146 for the PII tail and 0.734 for the W tail (Table 3). These values are close to the tail index of the PERSN-CDR for both tails. The InSitu tail index spatial patterns match those of CPC, MSWEP, and PERSN-CDR very well (Fig. 8). Light tails ($0.05 < \gamma_{PII} < 0.1$) occur in the eastern and western United States and medium tails ($0.15 < \gamma_{PII} < 0.3$) occur in the central United States, which is comparable with MSWEP and PERSN-CDR patterns. InSitu data show light tails along the eastern Brazilian coast similar to the CPC, PERSN-CDR, and MSWEP products. The observed light tails in India and in the eastern coast of Australia are only consistent with PERSN-CDR. The heavy to medium tails in the dry rangelands of Australia are consistent with the InSitu in all the datasets.

The spatial patterns of the observed maximum precipitation (Fig. 9a) and the calculated T-yr return levels (Figs. 9b,c) are

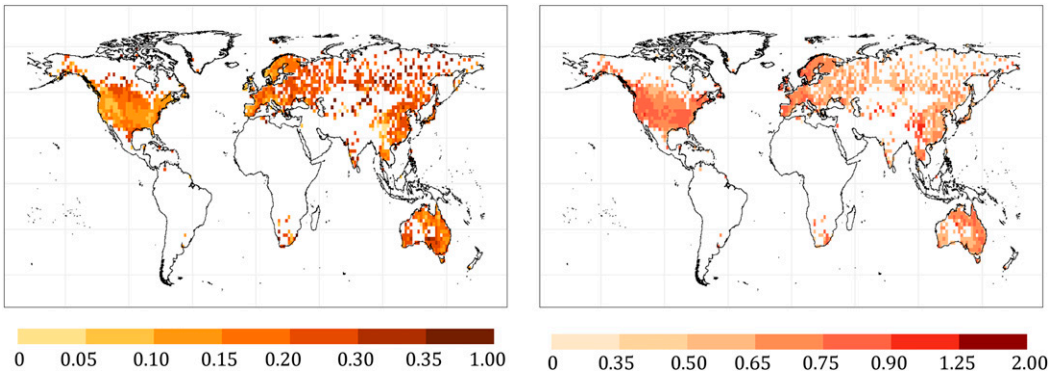

FIG. 8. Spatial patterns of (left) Pareto II and (right) Weibull tail index for in situ observations (InSitu).

similar. Figure 9 and show that InSitu spatial patterns closely match the CPC and MSWEP patterns, except at a few locations (e.g., southern India). Spatial patterns are generally similar in the continental United States for all datasets; however, InSitu data show high return levels (>160 mm) in the eastern United States that match the CFSR, and lower return levels (<60 mm) in the western United States that match the MSWEP. The high return levels (of order 120–180 mm) observed in the InSitu on the eastern Brazilian coast are not observed in any of the gridded datasets. Similarly, the higher return levels in the southern tip of Africa do not match those indicated by the products; the exception is the eastern coast of South Africa in CPC and MSWEP. To some extent, the high return levels (>180 mm) in the Indian subcontinent are consistent with the CPC, MSWEP, and CFSR. However, the return levels in CFSR are much higher (around 20%) compared to the InSitu, specifically in southern India. The low return levels noted in the Europe match all the datasets. Similar differences are found between the InSitu and the gridded products in both the 100- and 1000-yr return levels (see Fig. 10 and Fig. S7 for the 100- and 1000-yr return level maps, respectively).

Interestingly, the differences and ratios between the InSitu and the gridded products return levels are not same for the return periods considered (Fig. 11 and Table 4). Figure 11 shows the histograms of differences and ratios, and Table 4. gives the average values at global scale. Albeit high cross correlations and return levels in CFSR, on average, the differences between the return levels of InSitu and CFSR are low compared to other datasets for observations and T-yr return period. The differences and ratios are high for 100- and 1000-yr return periods in the CFSR. Low values for the 100- and

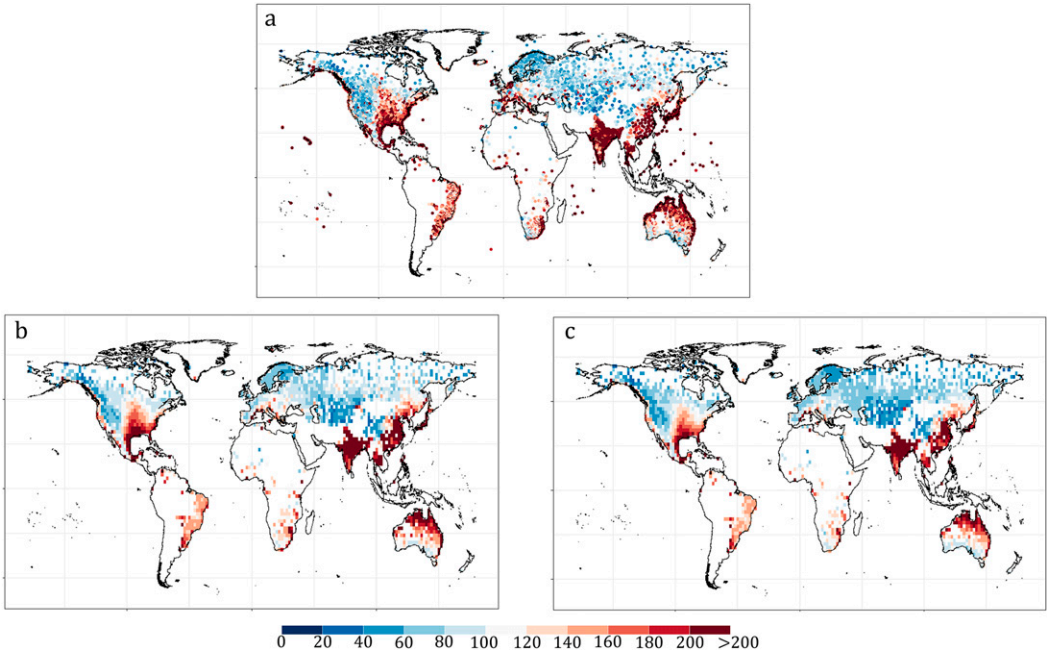

FIG. 9. Spatial pattern of T-yr return levels (mm) of (a) empirical (InSitu) and calculated return levels using (b) Pareto II and (c) Weibull distributions.

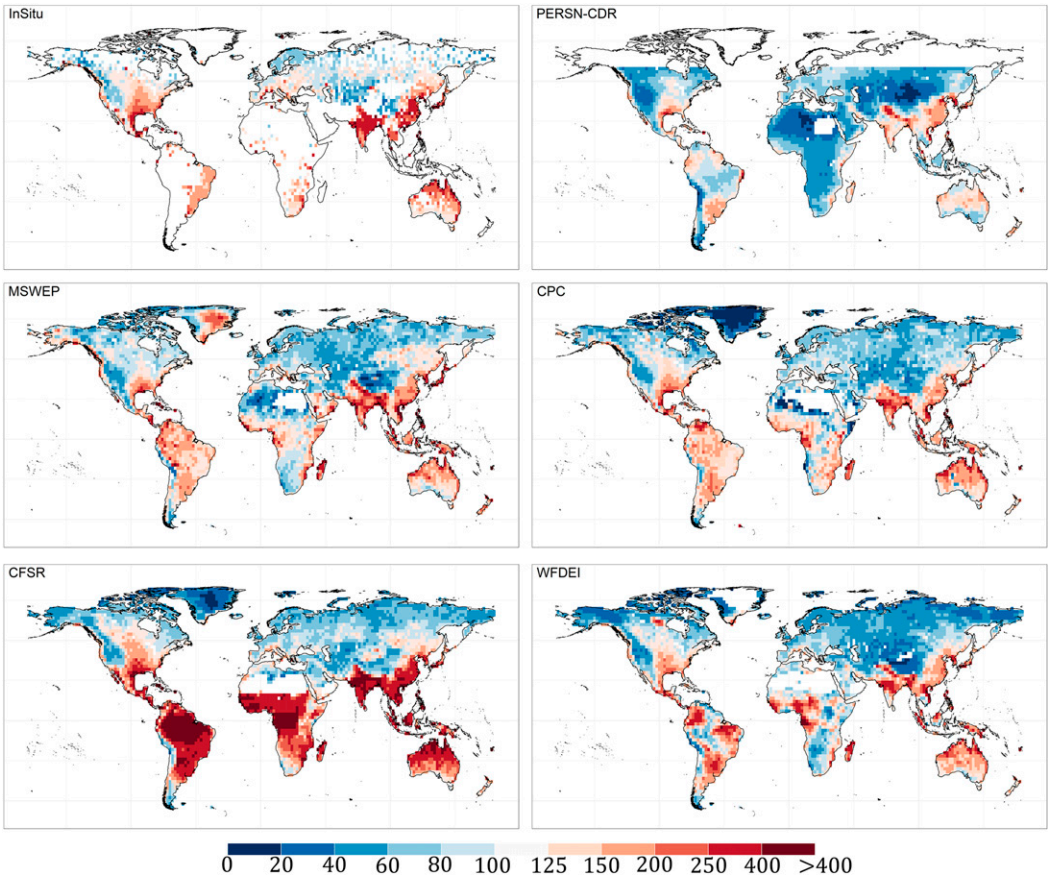

FIG. 10. Spatial pattern of predicted 100-yr return levels (in mm) using the PII tail for the InSitu observations (InSitu) and the gridded products. The return levels are averaged to a common resolution of 2° × 2°.

1000-yr return periods are noted in the CPC and WFDEI, but PERSN-CDR has high values at all the return periods. In the continental United States these differences and ratios are small for all datasets.

4. Discussion

Quantifying the tail heaviness was challenging because of the high variability of precipitation extremes. Gridded products are expected to be helpful in understanding the spatial variation of precipitation extremes globally, yet significant differences, both in the spatial pattern and in the absolute values of the tail index, were found among the different products. Notably, the return levels are in general underestimated by the satellite-based PERSN-CDR (Shah and Mishra 2014) and overestimated by the reanalysis-based CFSR, when compared with CPC dataset. Spatial patterns of InSitu tail index match with those of MSWEP and PERSN-CDR, which is expected as PERSN-CDR is adjusted using GHCN-Daily data, and MSWEP also uses GHCN-Daily in addition to other observational products (e.g., GPCC). However, this comparison is limited to North America, Europe, Australia, and some parts of Asia, where InSitu data exist.

In the tropics, where precipitation is mainly convective, the satellite-based PERSN-CDR, the observations-based CPC, and the optimally merged MSWEP show light tails. The reanalysis datasets, CFSR and WFDEI, differ substantially in their spatial patterns from the other products. One reason for this difference may be that the simulated precipitation extremes are affected by the parameterizations of deep convection in the tropics (Kharin et al. 2005). In mid- and high latitudes, while the other products show medium to heavy tails, the satellite-based PERSN-CDR indicates light tails, limiting its credibility (Gehne et al. 2016). At high latitudes, CPC differs from the other products, which could be due to substantial measurement errors (sometimes up to 100%) in gauge data as a result of wind undercatch especially for solid precipitation, lack of recording of trace events, and/or the sparse solid precipitation gauging network (Gervais et al. 2014; Goodison et al. 1998; Pomeroy and Goodison 1997; Yang et al. 2001). In contrast reanalysis products analyzing atmospheric and surface fields at high latitudes have been able to produce credible precipitation fields (Barrett et al. 2020; Krogh et al. 2015; Krogh and Pomeroy 2018; Lindsay et al. 2014; Serreze et al. 2005).

The MSWEP seems to perform reasonably well at all latitudes with spatial patterns of tail index and return

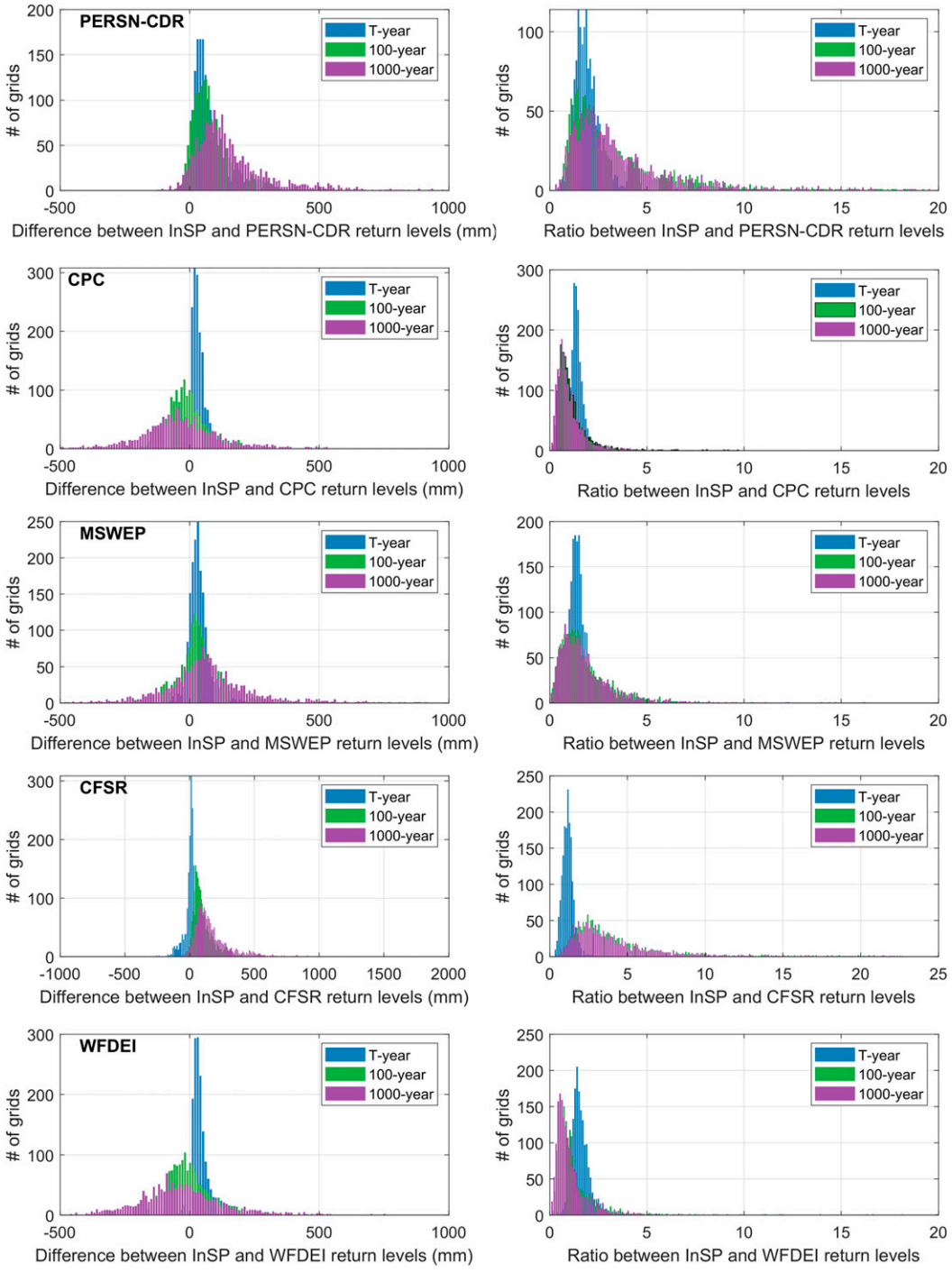

FIG. 11. Histograms of (left) differences and (right) ratios between the in situ observations (InSitu) and the gridded products for different return periods. The return levels are averaged to a common resolution of 2° × 2°.

levels closely matching with observations and also highly correlated with CPC and InSitu at both global and continental scales. Good correlations among all the datasets, including the reanalyses, were observed in the continental United States owing to a dense, high quality gauge network. Though this reiterates the necessity of appropriate gauge observations, the number of gauges available worldwide has been declining (Sun et al. 2018), reducing the ability to obtain reliable gridded products based on ground observations (Herrera et al. 2019). Reduction in uncertainty in estimation of precipitation extremes from gridded products that depend on observations therefore relies on the preservation

TABLE 4. Average differences and ratios between the return levels of InSitu observations (InSitu) and gridded products at global scale.

Dataset	Observations	Pareto II			Weibull		
		T-yr	100-yr	1000-yr	*T*-yr	100-yr	1000-yr
		Differences (mm)					
CPC	53.53	38.93	1.79	−17.25	35.88	10.66	11.29
PERSN-CDR	76.94	64.36	93.59	175.76	55.33	77.95	113.93
MSWEP	49.33	36.39	43.71	77.66	31.96	39.60	56.88
CFSR	31.67	6.72	104.87	187.62	11.09	91.37	130.43
WFDEI	57.96	44.60	−0.47	−10.89	40.06	5.12	4.45
		Ratios					
CPC	1.67	1.46	1.14	1.11	1.47	1.22	1.19
PERSN-CDR	2.15	1.98	3.62	3.91	1.89	3.58	3.58
MSWEP	1.62	1.46	1.84	1.99	1.43	1.87	1.90
CFSR	1.39	1.19	4.28	4.78	1.20	4.30	4.47
WFDEI	1.81	1.64	1.20	1.22	1.60	1.25	1.23

and restoration of high quality precipitation observing networks.

5. Summary and conclusions

Daily extreme precipitation was assessed in gridded precipitation products by quantifying the heaviness of the tail through the tail index. Five popular global datasets, CPC, PERSN-CDR, MSWEP, CFSR, and WFDEI, having more than 35 years of recent data were used to examine: 1) the spatial variability of the tail index across different climate and geographical regions, 2) the patterns of observed and predicted extremes around the globe, and 3) the differences in predicted return levels among the gridded products. These datasets were compared with point surface observations (InSitu) that were gridded so as to examine the differences in tail index and extreme precipitation return levels.

The spatial patterns of the tail index largely match climate classifications by Köppen–Geiger. Considerable differences were found among all the datasets. The differences between the gauge-based datasets and the reanalysis products are particularly large. Very large differences in 100- and 1000-yr return levels were found among the gridded products. The degree of discrepancy varied from region to region, with low differences in North America and Europe, and huge differences in Africa and South America. Notably, the returns levels are in general underestimated by the satellite-based PERSN-CDR (Shah and Mishra 2014) and overestimated by the reanalysis-based CFSR, when compared with CPC and InSitu.

Two theoretical tails, asymptotically equivalent with the tails of many well-known distributions employed in practice, were used to represent two major families of distributions: the power (Pareto II) and exponential form (Weibull). A large difference between the predicted return levels resulting from the two tails is noted. Differences of about 10, 22, and 80 mm were found between the prediction of the two tails in the global mean return levels for the observed period (approximately 38 years), 100- and 1000-yr return periods, respectively. Again, these differences vary among the datasets, high for the CFSR and low for the PERSN-CDR, and at continental scale, high for South America and Africa, and low for North America and Europe. In some grids, the return levels predicted by the PII tail were almost 2 and 3 times higher than those by the W tail for 100 and 1000 years, respectively. Such variations emphasize the importance of choosing the correct tail.

From the analysis, it should be noted that for small return periods (e.g., up to 100 years) the data play crucial role in estimating the tail. Any inconsistencies in the data (like gaps, error in measurement or reporting instrument, drizzle effects or very low events etc.) lead to unreliable estimates of tail index and differences in the return levels. For large return periods, like 1000 years, the distribution choice might also influence return levels in addition to the data. This is in view of the fact that both tails fit well to the observed sample (~38 years here) and thus their asymptotic difference is not revealed for small return periods, and as we extrapolate to large return periods the heavier tail of the power type (PII) could result in much larger return levels compared to the stretched exponential (W). Typically it is recommended to perform an uncertainty analysis to obtain a confidence interval for such high return levels (Shamir et al. 2013). It should be noted that the 1000-yr return levels are not offered for operational use (since their estimates are based on small samples); yet, these estimates highlight how strongly the return levels can be affected by the choice of the probability model.

Though the InSitu data are considered based on quality flags provided by GHCN-Daily, the gridded products are considered to be quality controlled by the data providers. The comparison between the products and the InSitu data is simplified by averaging the tail index of stations within a coarser grid. However, it should be noted that the InSitu is point data and gridded products represent areal precipitation. As these two are different processes, comparison is not straightforward requiring spatial interpolation techniques, where each technique has its own advantages and disadvantages, to obtain gridded data from point data. At a coarse scale, the extremes are considered to be smoothened as a precipitation value at a particular grid is the average of whole grid, and therefore averaging the tail index in a coarse grid is considered to be valid

for comparison. Another issue with InSitu data is the timing of daily accumulations, which is neglected in the current study.

Gridded precipitation products are extensively used in research and practice. However, these results show that such products do not provide a consistent representation of the extreme precipitation. Data products that rely on observations are better where observation network is denser and more reliable. Those that are largely model driven has utility in poorly gauged regions. Climate is also associated with data product performance in describing extremes. Thus, the question that naturally arises is ''which product is the best for studying extremes?'' The results presented here suggest that MSWEP performs better, yet there is no single global product that works best for all regions and climates. As such, it is still advisable to include multiple products in a study in order to gain some insights into the uncertainties in extremes.

Acknowledgments. The authors wish to acknowledge funding from the Canada First Excellence Research Fund's Global Water Futures programme, NSERC Discovery Grants, Canada Research Chair programme, and the PIMS-GWF postdoctoral fellowship programme. The support by NSF (Grant EAR-1928724) and NASA (Grant 80NSSC19K0726) to organize the 12th International Precipitation Conference (IPC12), Irvine, California, June 2019, and produce the IPC12 special collection of papers is gratefully acknowledged.

REFERENCES

Adler, R. F., and Coauthors, 2003: The Version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeor.*, **4**, 1147–1167, https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2.

Ahmed, K., S. Shahid, X. Wang, N. Nawaz, and N. Khan, 2019: Evaluation of gridded precipitation datasets over arid regions of Pakistan. *Water*, **11**, 210, https://doi.org/10.3390/w11020210.

Akinsanola, A. A., K. O. Ogunjobi, V. O. Ajayi, E. A. Adefisan, J. A. Omotosho, and S. Sanogo, 2017: Comparison of five gridded precipitation products at climatological scales over West Africa. *Meteor. Atmos. Phys.*, **129**, 669–689, https://doi.org/10.1007/s00703-016-0493-6.

Alexander, L. V., and Coauthors, 2006: Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res.*, **111**, D05109, https://doi.org/10.1029/2005JD006290.

Ashouri, H., K.-L. Hsu, S. Sorooshian, D. K. Braithwaite, K. R. Knapp, L. D. Cecil, B. R. Nelson, and O. P. Prat, 2015: PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bull. Amer. Meteor. Soc.*, **96**, 69–83, https://doi.org/10.1175/BAMS-D-13-00068.1.

Barrett, A. P., J. C. Stroeve, and M. C. Serreze, 2020: Arctic Ocean precipitation from atmospheric reanalyses and comparisons with North Pole drifting station records. *J. Geophys. Res. Oceans*, **125**, e2019JC015415, https://doi.org/10.1029/2019JC015415.

Beck, H. E., and Coauthors, 2017: Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrol. Earth Syst. Sci.*, **21**, 6201–6217, https://doi.org/10.5194/hess-21-6201-2017.

——, and Coauthors, 2019: Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrol. Earth Syst. Sci.*, **23**, 207–224, https://doi.org/10.5194/hess-23-207-2019.

Bell, S., D. Cornford, and L. Bastin, 2015: How good are citizen weather stations? Addressing a biased opinion. *Weather*, **70**, 75–84, https://doi.org/10.1002/wea.2316.

Burton, C., S. Rifai, and Y. Malhi, 2018: Inter-comparison and assessment of gridded climate products over tropical forests during the 2015/2016 El Niño. *Philos. Trans. Roy. Soc. London*, **373B**, 20170406, https://doi.org/10.1098/rstb.2017.0406.

Chen, C.-T., and T. Knutson, 2008: On the verification and comparison of extreme rainfall indices from climate models. *J. Climate*, **21**, 1605–1621, https://doi.org/10.1175/2007JCLI1494.1.

Chen, M., W. Shi, P. Xie, V. B. S. Silva, V. E. Kousky, R. W. Higgins, and J. E. Janowiak, 2008: Assessing objective techniques for gauge-based analyses of global daily precipitation. *J. Geophys. Res.*, **113**, D04110, https://doi.org/10.1029/2007JD009132.

Contractor, S., L. V. Alexander, M. G. Donat, and N. Herold, 2015: How well do gridded datasets of observed daily precipitation compare over Australia? *Adv. Meteor.*, **2015**, 325718, https://doi.org/10.1155/2015/325718.

——, and Coauthors, 2020: Rainfall Estimates on a Gridded Network (REGEN) – A global land-based gridded dataset of daily precipitation from 1950 to 2016. *Hydrol. Earth Syst. Sci.*, **24**, 919–943, https://doi.org/10.5194/hess-24-919-2020.

Diaconescu, E. P., P. Gachon, and R. Laprise, 2015: On the remapping procedure of daily precipitation statistics and indices used in regional climate model evaluation. *J. Hydrometeor.*, **16**, 2301–2310, https://doi.org/10.1175/JHM-D-15-0025.1.

Dinku, T., S. J. Connor, P. Ceccato, and C. F. Ropelewski, 2008: Comparison of global gridded precipitation products over a mountainous region of Africa. *Int. J. Climatol.*, **28**, 1627–1638, https://doi.org/10.1002/joc.1669.

Donat, M. G., L. V. Alexander, H. Yang, I. Durre, R. Vose, and J. Caesar, 2013a: Global land-based datasets for monitoring climatic extremes. *Bull. Amer. Meteor. Soc.*, **94**, 997–1006, https://doi.org/10.1175/BAMS-D-12-00109.1.

——, and Coauthors, 2013b: Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset. *J. Geophys. Res. Atmos.*, **118**, 2098–2118, https://doi.org/10.1002/jgrd.50150.

——, J. Sillmann, S. Wild, L. V. Alexander, T. Lippmann, and F. W. Zwiers, 2014: Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis datasets. *J. Climate*, **27**, 5019–5035, https://doi.org/10.1175/JCLI-D-13-00405.1.

Faridzad, M., T. Yang, K. Hsu, S. Sorooshian, and C. Xiao, 2018: Rainfall frequency analysis for ungauged regions using remotely sensed precipitation information. *J. Hydrol.*, **563**, 123–142, https://doi.org/10.1016/j.jhydrol.2018.05.071.

Fischer, E. M., and R. Knutti, 2014: Detection of spatially aggregated changes in temperature and precipitation extremes. *Geophys. Res. Lett.*, **41**, 547–554, https://doi.org/10.1002/2013GL058499.

——, and ——, 2015: Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes. *Nat. Climate Change*, **5**, 560–564, https://doi.org/10.1038/nclimate2617.

Gado, T. A., K. Hsu, and S. Sorooshian, 2017: Rainfall frequency analysis for ungauged sites using satellite precipitation products. *J. Hydrol.*, **554**, 646–655, https://doi.org/10.1016/j.jhydrol.2017.09.043.

Gehne, M., T. M. Hamill, G. N. Kiladis, and K. E. Trenberth, 2016: Comparison of global precipitation estimates across a range of

temporal and spatial scales. *J. Climate*, **29**, 7773–7795, https://doi.org/10.1175/JCLI-D-15-0618.1.

Gervais, M., L. B. Tremblay, J. R. Gyakum, and E. Atallah, 2014: Representing extremes in a daily gridded precipitation analysis over the United States: Impacts of station density, resolution, and gridding methods. *J. Climate*, **27**, 5201–5218, https://doi.org/10.1175/JCLI-D-13-00319.1.

Ghosh, S., D. Das, S.-C. Kao, and A. R. Ganguly, 2012: Lack of uniform trends but increasing spatial variability in observed Indian rainfall extremes. *Nat. Climate Change*, **2**, 86–91, https://doi.org/10.1038/nclimate1327.

Goodison, B. E., P. Y. T. Louie, and D. Yang, 1998: WMO Solid Precipitation Measurement Intercomparison. Instruments and Observing Methods Rep. 67, 318 pp., https://www.wmo.int/pages/prog/www/IMOP/publications/IOM-67-solid-precip/WMOtd872.pdf.

Herrera, S., S. Kotlarski, P. M. M. Soares, R. M. Cardoso, A. Jaczewski, J. M. Gutiérrez, and D. Maraun, 2019: Uncertainty in gridded precipitation products: Influence of station density, interpolation method and grid resolution. *Int. J. Climatol.*, **39**, 3717–3729, https://doi.org/10.1002/joc.5878.

Hu, Z., Q. Zhou, X. Chen, J. Li, Q. Li, D. Chen, W. Liu, and G. Yin, 2018: Evaluation of three global gridded precipitation data sets in central Asia based on rain gauge observations. *Int. J. Climatol.*, **38**, 3475–3493, https://doi.org/10.1002/joc.5510.

Ingram, W., 2016: Increases all round. *Nat. Climate Change*, **6**, 443–444, https://doi.org/10.1038/nclimate2966.

Javanmard, S., A. Yatagai, M. I. Nodzu, J. BodaghJamali, and H. Kawamoto, 2010: Comparing high-resolution gridded precipitation data with satellite rainfall estimates of TRMM_3B42 over Iran. *Adv. Geosci.*, **25**, 119–125, https://doi.org/10.5194/adgeo-25-119-2010.

Kharin, V. V., F. W. Zwiers, and X. Zhang, 2005: Intercomparison of near-surface temperature and precipitation extremes in AMIP-2 simulations, reanalyses, and observations. *J. Climate*, **18**, 5201–5223, https://doi.org/10.1175/JCLI3597.1.

Kidd, C., P. Bauer, J. Turk, G. J. Huffman, R. Joyce, and D. Braithwaite, 2012: Intercomparison of high-resolution precipitation products over northwest Europe. *J. Hydrometeor.*, **13**, 67–83, https://doi.org/10.1175/JHM-D-11-042.1.

——, A. Becker, G. J. Huffman, C. L. Muller, P. Joe, G. Skofronick-Jackson, and D. B. Kirschbaum, 2017: So, how much of the Earth's surface is covered by rain gauges? *Bull. Amer. Meteor. Soc.*, **98**, 69–78, https://doi.org/10.1175/BAMS-D-14-00283.1.

Kottek, M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel, 2006: World Map of the Köppen-Geiger climate classification updated. *Meteor. Z.*, **15**, 259–263, https://doi.org/10.1127/0941-2948/2006/0130.

Krogh, S. A., and J. W. Pomeroy, 2018: Recent changes to the hydrological cycle of an Arctic basin at the tundra–taiga transition. *Hydrol. Earth Syst. Sci.*, **22**, 3993–4014, https://doi.org/10.5194/hess-22-3993-2018.

——, ——, and J. McPhee, 2015: Physically based mountain hydrological modeling using reanalysis data in Patagonia. *J. Hydrometeor.*, **16**, 172–193, https://doi.org/10.1175/JHM-D-13-0178.1.

Lindsay, R., M. Wensnahan, A. Schweiger, and J. Zhang, 2014: Evaluation of seven different atmospheric reanalysis products in the Arctic. *J. Climate*, **27**, 2588–2606, https://doi.org/10.1175/JCLI-D-13-00014.1.

MacKellar, N. C., B. C. Hewitson, and M. A. Tadross, 2007: Namaqualand's climate: Recent historical changes and future scenarios. *J. Arid Environ.*, **70**, 604–614, https://doi.org/10.1016/j.jaridenv.2006.03.024.

Min, S.-K., X. Zhang, F. W. Zwiers, and G. C. Hegerl, 2011: Human contribution to more-intense precipitation extremes. *Nature*, **470**, 378–381, https://doi.org/10.1038/nature09763.

New, M., M. Hulme, and P. Jones, 2000: Representing twentieth-century space–time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. *J. Climate*, **13**, 2217–2238, https://doi.org/10.1175/1520-0442(2000)013<2217:RTCSTC>2.0.CO;2.

Newman, A. J., M. P. Clark, R. J. Longman, and T. W. Giambelluca, 2019: Methodological intercomparisons of station-based gridded meteorological products: Utility, limitations, and paths forward. *J. Hydrometeor.*, **20**, 531–547, https://doi.org/10.1175/JHM-D-18-0114.1.

Nguyen, P., M. Ombadi, S. Sorooshian, K. Hsu, A. AghaKouchak, D. Braithwaite, H. Ashouri, and A. R. Thorstensen, 2018: The PERSIANN family of global satellite precipitation data: A review and evaluation of products. *Hydrol. Earth Syst. Sci.*, **22**, 5801–5816, https://doi.org/10.5194/hess-22-5801-2018.

Ombadi, M., P. Nguyen, S. Sorooshian, and K. Hsu, 2018: Developing Intensity-Duration-Frequency (IDF) curves from satellite-based precipitation: Methodology and evaluation. *Water Resour. Res.*, **54**, 7752–7766, https://doi.org/10.1029/2018WR022929.

Papalexiou, S. M., and D. Koutsoyiannis, 2013: Battle of extreme value distributions: A global survey on extreme daily rainfall. *Water Resour. Res.*, **49**, 187–201, https://doi.org/10.1029/2012WR012557.

——, and ——, 2016: A global survey on the seasonal variation of the marginal distribution of daily precipitation. *Adv. Water Resour.*, **94**, 131–145, https://doi.org/10.1016/j.advwatres.2016.05.005.

——, ——, and C. Makropoulos, 2013: How extreme is extreme? An assessment of daily rainfall distribution tails. *Hydrol. Earth Syst. Sci.*, **17**, 851–862, https://doi.org/10.5194/hess-17-851-2013.

——, A. AghaKouchak, and E. Foufoula-Georgiou, 2018: A diagnostic framework for understanding climatology of tails of hourly precipitation extremes in the United States. *Water Resour. Res.*, **54**, 6725–6738, https://doi.org/10.1029/2018WR022732.

Pomeroy, J. W., and B. E. Goodison, 1997: Winter and snow. *The Surface Climates of Canada*, McGill-Queen's University Press, 68–100.

Rajah, K., T. O'Leary, A. Turner, G. Petrakis, M. Leonard, and S. Westra, 2014: Changes to the temporal distribution of daily precipitation. *Geophys. Res. Lett.*, **41**, 8887–8894, https://doi.org/10.1002/2014GL062156.

Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, https://doi.org/10.1175/JCLI-D-12-00823.1.

Satgé, F., D. Defrance, B. Sultan, M.-P. Bonnet, F. Seyler, N. Rouché, F. Pierron, and J.-E. Paturel, 2020: Evaluation of 23 gridded precipitation datasets across West Africa. *J. Hydrol.*, **581**, 124412, https://doi.org/10.1016/j.jhydrol.2019.124412.

Serreze, M. C., A. P. Barrett, and F. Lo, 2005: Northern high-latitude precipitation as depicted by atmospheric reanalyses and satellite retrievals. *Mon. Wea. Rev.*, **133**, 3407–3430, https://doi.org/10.1175/MWR3047.1.

Shah, R., and V. Mishra, 2014: Evaluation of the reanalysis products for the monsoon season droughts in India. *J. Hydrometeor.*, **15**, 1575–1591, https://doi.org/10.1175/JHM-D-13-0103.1.

Shamir, E., K. P. Georgakakos, and M. J. Murphy, 2013: Frequency analysis of the 7–8 December 2010 extreme precipitation in

the Panama canal watershed. *J. Hydrol.*, **480**, 136–148, https://doi.org/10.1016/j.jhydrol.2012.12.010.

Sorooshian, S., K. Hsu, D. Braithwaite, H. Ashouri, and NOAA CDR Program, 2014: NOAA Climate Data Record (CDR) of Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN-CDR), version 1 revision 1. NOAA National Centers for Environmental Information, accessed 12 December 2018, https://doi.org/10.7289/V51V5BWQ.

Sun, Q., D. Kong, C. Miao, Q. Duan, T. Yang, A. Ye, Z. Di, and W. Gong, 2014: Variations in global temperature and precipitation for the period of 1948 to 2010. *Environ. Monit. Assess.*, **186**, 5663–5679, https://doi.org/10.1007/s10661-014-3811-9.

——, C. Miao, Q. Duan, H. Ashouri, S. Sorooshian, and K.-L. Hsu, 2018: A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Rev. Geophys.*, **56**, 79–107, https://doi.org/10.1002/2017RG000574.

Trenberth, K. E., A. Dai, R. M. Rasmussen, and D. B. Parsons, 2003: The changing character of precipitation. *Bull. Amer. Meteor. Soc.*, **84**, 1205–1218, https://doi.org/10.1175/BAMS-84-9-1205.

——, J. T. Fasullo, and J. Mackaro, 2011: Atmospheric moisture transports from ocean to land and global energy flows in reanalyses. *J. Climate*, **24**, 4907–4924, https://doi.org/10.1175/2011JCLI4171.1.

Wang, A., and X. Zeng, 2015: Global hourly land surface air temperature datasets: Inter-comparison and climate change. *Int. J. Climatol.*, **35**, 3959–3968, https://doi.org/10.1002/joc.4257.

Weedon, G. P., G. Balsamo, N. Bellouin, S. Gomes, M. J. Best, and P. Viterbo, 2014: The WFDEI meteorological forcing data set: WATCH forcing data methodology applied to ERA-interim reanalysis data. *Water Resour. Res.*, **50**, 7505–7514, https://doi.org/10.1002/2014WR015638.

Wong, J. S., S. Razavi, B. R. Bonsal, H. S. Wheater, and Z. E. Asong, 2017: Inter-comparison of daily precipitation products for large-scale hydro-climatic applications over Canada. *Hydrol. Earth Syst. Sci.*, **21**, 2163–2185, https://doi.org/10.5194/hess-21-2163-2017.

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. *J. Hydrometeor.*, **8**, 607–626, https://doi.org/10.1175/JHM583.1.

Xu, L., N. Chen, H. Moradkhani, X. Zhang, and C. Hu, 2020: Improving global monthly and daily precipitation estimation by fusing gauge observations, remote sensing, and reanalysis data sets. *Water Resour. Res.*, **56**, e2019WR026444, https://doi.org/10.1029/2019WR026444.

Yang, D., and Coauthors, 2001: Compatibility evaluation of national precipitation gage measurements. *J. Geophys. Res.*, **106**, 1481–1491, https://doi.org/10.1029/2000JD900612.

Yin, H., M. G. Donat, L. V. Alexander, and Y. Sun, 2015: Multi-dataset comparison of gridded observed temperature and precipitation extremes over China. *Int. J. Climatol.*, **35**, 2809–2827, https://doi.org/10.1002/joc.4174.

Zhang, Q., H. Körnich, and K. Holmgren, 2013: How well do reanalyses represent the southern African precipitation? *Climate Dyn.*, **40**, 951–962, https://doi.org/10.1007/s00382-012-1423-z.

Zhang, X., L. Alexander, G. C. Hegerl, P. Jones, A. K. Tank, T. C. Peterson, B. Trewin, and F. W. Zwiers, 2011: Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdiscip. Rev.: Climate Change*, **2**, 851–870, https://doi.org/10.1002/wcc.147.