# Extreme Precipitation Return Levels for Multiple Durations on a Global Scale

Gaby J Gründemann[1,1,1], Enrico Zorzetto[2,2,2], Hylke E Beck[3,3,3], Marc Schleiss[4,4,4], Nick Van de Giesen[1,1,1], Marco Marani[5,5,5], and Ruud J. van der Ent[1,1,1]

[1]Delft University of Technology
[2]Duke University
[3]Princeton Universtiy
[4]TU-Delft
[5]University of Padua and Duke University

November 30, 2022



**Abstract**

Quantifying the magnitude and frequency of extreme precipitation events is key in translating climate observations to planning and engineering design. Past efforts have mostly focused on the estimation of daily extremes using gauge observations. Recent development of high-resolution global precipitation products, now allow estimation of global extremes. This research aims to quantitatively characterize the spatiotemporal behavior of precipitation extremes, by calculating extreme precipitation return levels for multiple durations on the global domain using the Multi-Source Weighted-Ensemble Precipitation (MSWEP) dataset. Both classical and novel extreme value distributions are used to provide an insight into the spatial patterns of precipitation extremes. Our results show that the traditional Generalized Extreme Value (GEV) distribution and Peak-Over-Threshold (POT) methods, which only use the largest events to estimate precipitation extremes, are not spatially coherent. The recently developed Metastatistical Extreme Value (MEV) distribution, that includes all precipitation events, leads to smoother spatial patterns of local extremes. While the GEV and POT methods predict a consistent shift from heavy to thin tails with increasing duration, the heaviness of the tail obtained with MEV was relatively unaffected by the precipitation duration. The generated extreme precipitation return levels and corresponding parameters are provided as the Global Precipitation EXtremes (GPEX) dataset. These data can be useful for studying the underlying physical processes causing the spatiotemporal variations of the heaviness of extreme precipitation distributions.

# Highlights

**Extreme Precipitation Return Levels for Multiple Durations on a Global Scale**

Gaby J. Gründemann, Enrio Zorzetto, Hylke E. Beck, Marc Schleiss, Nick van de Giesen, Marco Marani, Ruud J. van der Ent

- Global precipitation return levels for 3-hour to 10-day durations are analysed
- Three different extreme value distributions are used to estimate the extremes
- The MEV distribution shows the most coherent spatiotemporal patterns
- Two distributions show a global shift from heavy to thin tails for longer durations

# Extreme Precipitation Return Levels for Multiple Durations on a Global Scale

Gaby J. Gründemann[a,b], Enrio Zorzetto[c], Hylke E. Beck[d], Marc Schleiss[e], Nick van de Giesen[a], Marco Marani[f], Ruud J. van der Ent[a,g,h]

[a] *Department of Water Management, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, Netherlands*
[b] *Centre for Hydrology, University of Saskatchewan, Canmore, Alberta, Canada*
[c] *Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, New Jersey, USA*
[d] *Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey, USA*
[e] *Department of Geoscience and Remote Sensing, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, Netherlands*
[f] *Dipartimento di Ingegneria Civile, Edile ed Ambientale, Universita' degli Studi di Padova, Padova, Italy*
[g] *Department of Physical Geography, Faculty of Geosciences, Utrecht University, Utrecht, Netherlands*
[h] *Water Research Centre, School of Civil and Environmental Engineering, University of New South Wales, Sydney, Australia*

**Abstract**

Quantifying the magnitude and frequency of extreme precipitation events is key in translating climate observations to planning and engineering design. Past efforts have mostly focused on the estimation of daily extremes using gauge observations. Recent development of high-resolution global precipitation products, now allow estimation of global extremes. This research aims to quantitatively characterize the spatiotemporal behavior of precipitation extremes, by calculating extreme precipitation return levels for multiple durations on the global domain using the Multi-Source Weighted-Ensemble

*Email address:* g.j.gruendemann@tudelft.nl (Gaby J. Gründemann )

Precipitation (MSWEP) dataset. Both classical and novel extreme value distributions are used to provide an insight into the spatial patterns of precipitation extremes. Our results show that the traditional Generalized Extreme Value (GEV) distribution and Peak-Over-Threshold (POT) methods, which only use the largest events to estimate precipitation extremes, are not spatially coherent. The recently developed Metastatistical Extreme Value (MEV) distribution, that includes all precipitation events, leads to smoother spatial patterns of local extremes. While the GEV and POT methods predict a consistent shift from heavy to thin tails with increasing duration, the heaviness of the tail obtained with MEV was relatively unaffected by the precipitation duration. The generated extreme precipitation return levels and corresponding parameters are provided as the Global Precipitation EXtremes (GPEX) dataset. These data can be useful for studying the underlying physical processes causing the spatiotemporal variations of the heaviness of extreme precipitation distributions.

*Keywords:* Precipitation extremes, MSWEP, Metastatistical extreme value distribution, Generalized extreme value distribution, Peaks-over-threshold, Global domain

*PACS:* 92.40.-t, 92.40.Ea, 92.70.Np, 02.50.-r

*2000 MSC:* 86A05

---

## 1. Introduction

Extreme precipitation events are a major contributor to natural disasters (CRED, 2019). Accurate estimates of the severity of intense precipitation events are needed for an enhanced disaster risk understanding, such as that of floods and landslides. The urgency of this is indicated as the first priority of the Sendai Framework for Disaster Risk Reduction (UNSIDR, 2015). The accurate quantification of extremes is also necessary for infrastructure planning and design. Some countries already provide spatiotemporal estimates of extreme precipitation based on extreme value distributions (EVDs), for example, for Australia (Ball et al., 2019), the Netherlands (Beersma et al., 2018), and the US (e.g., Perica et al., 2015, 2018). However, many countries and regions do not have sufficient local data available (Gründemann et al., 2018; Kidd et al., 2017; van de Giesen et al., 2014), such that spatially-distributed extreme precipitation estimates are not possible.

Several previous studies have developed global-scale datasets of extreme precipitation. Courty et al. (2019) calculated intensity-duration-frequency curves at the global domain and their scaling with different event durations using reanalysis data and the Generalized Extreme Value (GEV) distribution with fixed tail behavior. Dunn et al. (2020) produced the HadEX3 dataset, which contains 29 generic precipitation and temperature indices, although these indices are not based on EVDs. Furthermore, this dataset has a coarse 1.25° latitudinal $\times$ 1.875° longitudinal resolution, with data-gaps due to insufficient available gauge data. Other global studies mostly focused on examining which type of distribution is most suitable to capture the tail behavior of extreme precipitation (Cavanaugh and Gershunov, 2015; Cavanaugh et al., 2015; Papalexiou et al., 2013). In addition, the spatial patterns of the parameter that controls the tail decay have been studied for the GEV distribution (Papalexiou and Koutsoyiannis, 2013; Ragulina and Reitan, 2017), and the Generalized Pareto (GP) distribution (Serinaldi and

Kilsby, 2014). However, several issues remain to be addressed in order to obtain global-domain extreme precipitation return levels: 1) the choice of the dataset with associated uncertainties, 2) the focus on daily durations, 3) the choice of the time blocks over which block-maxima are determined, and 4) the exploration of possible alternatives to the classical EVDs and the associated uncertainty, especially with respect to the tail behavior.

1. Several (quasi-)global gridded precipitation datasets have been developed in recent years, each with strengths, weaknesses, and uncertainties. See Sun et al. (2018), Beck et al. (2019a) and Rajulapati et al. (2020) for recent overviews of available datasets and their associated uncertainties. Most of these datasets are based on gauge, reanalysis, or satellite sensor data. Notable examples of gauge-based datasets include GPCC-FDR (Becker et al., 2013; Schneider et al., 2011) and REGEN (Contractor et al., 2020). However, gauges are extremely unevenly distributed across the globe (Kidd et al., 2017; Schneider et al., 2014), and the number of active gauges has been declining in recent decades (Mishra and Coulibaly, 2009). Satellite-based products such as CMORPH (Joyce et al., 2004), GSMaP (Ushio et al., 2009), IMERG (Huffman et al., 2015), and PERSIANN (Hong et al., 2004) have a relatively high spatio-temporal resolution. However, they do not cover regions outside of 60°N/S, and are only available from 2000 onwards, which significantly hinders their use for extreme value analyses. Precipitation products with a true global coverage and long records are reanalyses, such as ERA-5 (Hersbach et al., 2020), JRA-55 (Kobayashi et al., 2015), and MERRA-2 (Gelaro et al., 2017). However, reanaly-

sis products tend to exhibit strong systematic biases in the magnitude and frequency of precipitation (Decker et al., 2012; Liu et al., 2018; Ménégoz et al., 2013).

2. Global-scale analyses of precipitation extremes are generally based on daily precipitation records (Cavanaugh et al., 2015; Koutsoyiannis, 2004a,b; Papalexiou and Koutsoyiannis, 2013; Papalexiou et al., 2013; Ragulina and Reitan, 2017; Serinaldi and Kilsby, 2014). In practice, however, multiple durations are needed for the design of infrastructure (e.g., Nissen and Ulbrich, 2017) or urban drainage networks (e.g., Mailhot and Duchesne, 2009). It is known that precipitation extremes of different durations scale differently with temperature (Wasko et al., 2015), but little is known about the variation of EVD properties (tail behavior) for different temporal resolutions. Studies that did derive extreme precipitation statistics for durations ranging from minutes to a few days have mostly focused on small regions (McGraw et al., 2019; Nissen and Ulbrich, 2017; Overeem et al., 2008).

3. Studies estimating return levels of extreme precipitation by using annual maxima typically use calendar years to delineate the annual periods from which maxima values are extracted (e.g., De Paola et al., 2018; Marani and Zanetti, 2015; Papalexiou and Koutsoyiannis, 2013; Ragulina and Reitan, 2017; Villarini et al., 2011). When the variable of interest is river discharge instead of precipitation, however, hydrological years are typically used instead of calendar years to determine the annual maxima (Ward et al., 2016). For discharge values this is important, since peak discharge and flooding could occur during 31

December to 1 January transition and one event would be included in two calendar years. Although not often considered, this could also happen for precipitation. The annual maxima method could pick multiple values from a single rainy season that may, for example, be highly influenced by the El Niño-/Southern Oscillation, which is known to impact precipitation extremes (Allan and Soden, 2008; Rasmusson and Arkin, 1993).

4. The Generalized Extreme Value (GEV) distribution, the most widely used EVD, is typically fitted through one of two approaches: a) using annual maximum precipitation series and maximum likelihood (Coles, 2001) or L-moment (Hosking, 1990) estimation approaches, or b) using a Peak-Over-Threshold (POT) method to fit a Generalized Pareto Distribution to excesses above the threshold and a Poisson process to the sequence of threshold exceedances (Coles, 2001). In contrast to GEV and POT, the recently developed Metastatistical Extreme Value (MEV) distribution is fitted using all events with recorded precipitation instead of only the most severe. The inclusion of more events reduces the uncertainty due to sampling effects, which is important when dealing with short time series (Hu et al., 2020; Marani and Ignaccolo, 2015; Marra et al., 2018, 2019a; Miniussi and Marani, 2020; Zorzetto et al., 2016; Zorzetto and Marani, 2019). This is particularly advantageous when analyzing short remote sensing precipitation products, as the commonly applied GEV requires many years of data to accurately estimate the tail of the distribution (Papalexiou and Koutsoyiannis, 2013). Additionally, GEV parameter estimation depends heavily on a

few large values, which makes it very sensitive to the possible presence of outliers, a relatively common occurrence in remote sensing estimates of precipitation amounts (Zorzetto and Marani, 2020). The GEV tail behavior is mostly controlled by its shape parameter, which is very sensitive to sampling effects and the choice of the method used for estimation. To overcome these problems, some studies have suggested to use one universal value of the shape parameter that is applicable to the whole world Koutsoyiannis (2004a,b), or a shape parameter value within a narrow range between exponential and heavy-tail behavior (Papalexiou and Koutsoyiannis, 2013), or one shape parameter per region, that is similar within climate types and elevation ranges (Ragulina and Reitan, 2017). The estimation of the shape parameter is particularly difficult with short data series, though crucial for the accurate estimation of extremes.

In this study we contribute to overcome these issues by 1) using a dataset that merges all three main sources of precipitation data, 2) estimating extremes for several event durations, 3) using hydrological years in our analyses, and 4) comparing results from three different extreme value methods (GEV, POT and MEV). Specifically, we are interested in quantitatively characterizing the behavior of extreme precipitation and the spatiotemporal variation of extreme value distributional tails at the global domain.

## 2. Material and Methods

### 2.1. Data

The global precipitation product used in this study is the Multi-Source Weighted-Ensemble Precipitation (MSWEP-V2.2) dataset. MSWEP is particularly suited for our purpose due to its global coverage, long temporal span, high spatial and temporal resolution. We used data from 1 January 1979 to 31 October 2017 at a 0.1° latitude × 0.1° longitude resolution at 3-hourly time steps. We selected all land-cells between 90°N and 58°S for our analysis. MSWEP precipitation estimates are derived by merging five different satellite- and reanalysis-based global precipitation datasets. The dataset is one of the few precipitation products with daily (as opposed to monthly) gauge corrections, applied using a scheme that accounts for gauge reporting times (Beck et al., 2019b). MSWEP has shown robust performance compared to other widely used precipitation datasets (e.g., Alijanian et al., 2017; Bai and Liu, 2018; Beck et al., 2017, 2019a; Casson et al., 2018; Hu et al., 2020; Sahlu et al., 2017; Satgé et al., 2019; Zhang et al., 2019), thus underlying its potential for improving the characterization of extreme precipitation worldwide. We refer to Beck et al. (2019b) for a comprehensive description of the dataset.

#### 2.1.1. Quality Control

The integration of erroneous gauge observations into MSWEP-V2.2 can occasionally result in implausible precipitation values. Therefore, we implemented a three-step quality control procedure of the 3-hourly data prior to the analysis. We first discarded negative values, which are physically impossible. The second step was to discard outliers, which we defined as values deviating from the mean by more than 30 standard deviations. We also discarded data surrounding the outliers for the same time step using a $11 \times 11$ grid-cell window, as erroneous gauge observations may have influenced surrounding cells in the production of the MSWEP dataset. The third step was to remove years with $> 30$ discarded days or $< 5$ 'wet' 3-hourly periods, identified using a threshold of 0.2 mm $3h^{-1}$ following Wasko et al. (2015). Finally, we only included in the analysis data from grid cells with at least 30 years of data remaining, as a minimum record length of 30 years is customary and recommended to obtain reliable results (Arguez and Vose, 2011; Kendon et al., 2018; Westra et al., 2013).

### *2.1.2. Durations and Identification of Independent Events*

The durations we selected for our analysis are 3, 6, 12 and 24 hours, and 2, 3, 5 and 10 days. In order to create statistically-independent precipitation events for multiple durations, we first separated 3-hourly events following the declustering method to limit the autocorrelation of the samples described in Marra et al. (2018, their Section 3.1). For longer durations, independent events are the maximum intensities within each independent event and non-overlapping period using moving windows (Marra et al., 2020).

### *2.1.3. Hydrological Year*

A common challenge in global-scale assessments is the delineation of the hydrological year, given the regional variability in the climatological precipitation seasonality. We therefore developed an uniform way to define the hydrological year. To avoid splitting one rainy season over two different years, we computed the median of the monthly precipitation for each grid-cell, and defined the start of the hydrological year to be the first day of the driest month. Supplementary Material Figure S1a shows the starting month of the hydrological year as determined by this method. These data are also available in the GPEX dataset (Gründemann et al., 2021). As MSWEP-V2.2 spans the interval from 1 January 1979 to 31 October 2017, we discarded the data prior to the start of the first hydrological year, thus keeping 38 complete years. Only where the hydrological year starts in December there are just 37 complete years, which occurs in 5.8 % of the grid cells.

We also investigated whether there is a significant difference between the use of calendar and hydrological years for the estimated daily extremes for GEV and MEV. The POT method is based on the values over a high threshold, irrespective of when they occurred. Therefore, there is by definition no difference in calculating the extremes using hydrological or calendar years for the POT method. To determine the difference for GEV and MEV, we first calculated the daily return levels for normal calendar years, using the MSWEP data from 1979 to 2016. Second, we calculated the return levels for the same distributions and the same years, by removing the months before the start of the hydrological year from the year 1979 and adding them to the year 2016. We did this in order to use the exact same data, so the differences in the return level estimates are solely due to a different starting month.

### *2.2. Extreme Value Distributions*

Three extreme value distributions were fitted to the MSWEP data to calculate extreme precipitation return levels: the GEV, POT, and MEV distributions. Annual (hydrological year) maxima were used to estimate the three parameters of the GEV using the L-moments approach, because of its robust performance for small samples (Hosking, 1990). The GEV cumulative distribution function (CDF) is given by:

$$G(z) = \begin{cases} \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}, \xi \neq 0 \\ \exp\left\{-\exp\left[-\left(\frac{z-\mu}{\sigma}\right)\right]\right\}, \xi = 0 \end{cases} \tag{1}$$

with location parameter $\mu \in (-\infty, \infty)$, scale parameter $\sigma > 0$, and shape parameter $\xi \in (-\infty, \infty)$. The annual extremes estimated by GEV are translated into those of the parent distribution, following Koutsoyiannis (2004a, equation 3).

As a second EV model we use a Peaks Over Threshold approach, describing precipitation accumulations exceeding a high threshold using a GP distribution, while modelling the frequency of threshold exceedances using a Poisson point process (Coles, 2001; Davison and Smith, 1990). This framework also yields GEV as the resulting extreme value distribution, which is then used to determine the quantile corresponding to a given return period. The GP CDF is given by:

$$H(y) = \begin{cases} 1-\left(1+\frac{\xi y}{\beta}\right)^{-\frac{1}{\xi}}, \xi \neq 0 \\ 1-\exp\left(-\frac{y}{\beta}\right), \xi = 0 \end{cases} \tag{2}$$

where $y > 0$ are precipitation excesses over the threshold, with $\beta > 0$ and $\xi \in (-\infty, \infty)$ the GP scale and shape parameters respectively. A relevant aspect in applying the POT model is a suitable choice of the threshold used to define precipitation exceedances. Our global-scale application requires studying the distribution of precipitation extremes across markedly different climatic regions, thus excluding the adoption of a constant threshold value.

We studied the effect of the threshold choice using multiple threshold selection methods on a global sample of grid cells (see Supplementary Material Section 2 and and Figure S3). Our results showed that this choice had a limited effect on the estimated return levels (Figure S3a). We chose to perform our global analysis by selecting for each cell a threshold value such that it is exceeded on average 3 times each hydrological year. As a consequence of this choice, the sample size available for fitting the GP distribution remains constant across different precipitation durations. The method used to fit the GP distributions is the Probability Weighted Moments (PWM; e.g., see Hosking and Wallis, 1987).

The third model applied here is the MEV distribution (Hosseini et al., 2020; Hu et al., 2020; Marani and Ignaccolo, 2015; Miniussi et al., 2020a,b; Zorzetto et al., 2016). In the MEV framework, all "ordinary" precipitation events, i.e. all events above a small threshold, are used to infer this EV distribution. The threshold we applied is 0.2 mm $3\text{h}^{-1}$, coinciding with the earlier defined 'wet event'. Weibull parameters were estimated for each hydrological year separately, based on all wet events using the PWM method (Greenwood et al., 1979) as done in Zorzetto et al. (2016). The MEV-Weibull CDF is given by:

$$\zeta_m(x) = \frac{1}{M} \sum_{j=1}^{M} \left\{ 1 - \exp\left[ - \left( \frac{\mathrm{x}}{\mathrm{C_j}} \right)^{\mathrm{w_j}} \right] \right\}^{n_j} \tag{3}$$

where $j$ is the hydrological year ($j = 1, 2, \ldots, M$), $C_j > 0$ is the Weibull scale parameter, $w_j > 0$ is the Weibull shape parameter, and $n_j$ is the number of wet events observed in hydrological year $j$ (Marani and Ignaccolo, 2015).

It should be noted that the methods we applied in this study do not use any parameter bounds. Although Papalexiou and Koutsoyiannis (2013) argued that the GEV shape parameter of daily precipitation lies between exponential and heavy-tail behavior, this is not used as additional information to constrain our fits. Doing so would cause artificial breaks in the obtained spatiotemporal patterns, the analysis of which is the main objective of this study. Moreover, the scientific debate on bounds is not settled, especially for durations longer than a day, and different bounds are used in different studies (e.g., Blanchet et al., 2016; Yilmaz et al., 2017). In order to avoid underestimation of extremes in practical settings, our dataset (Gründemann et al., 2021) also includes the Gumbel estimates which may be used as a lower bound (see Supplementary Material Section 6).

### *2.2.1. Observed Return Period*

The MSWEP dataset analyzed here has 38 complete years of data. Therefore, the empirical return period associated with the maximum value on record computed according to the Weibull empirical frequency estimate is $T_{\text{observed}} = 39$ years. However, only 91 % of all cells had 38 complete years of data, so the maximum observed return period is sometimes lower: for 7 % of the cells only 37 complete years were available, and for 2 % of the cells 36 years or less were available. However, for simplicity we still refer to the corresponding maximum return level as T39 in the results.

### *2.2.2. Tail Behavior*

Both the GEV and MEV distributions are flexible and can describe different tail behaviors. They are, therefore, appropriate models to study the characteristics of local precipitation extremes. The tail behavior of the two distributions differs, as illustrated in Figure S4 for different combinations of scale and shape parameters. The shape parameter $\xi$ of the GEV distribution, obtained either through the annual maxima or POT approach, encodes the nature of the tail of the distribution. Based on the value of $\xi$, the GEV can take one of three forms: a positive GEV shape parameter ($\xi > 0$, *"Fréchet"*) corresponds to a power-law tail, i.e., to a slowly-decaying probability of large events. This heavy-tail behavior contrasts with the case of an exponential tail ($\xi = 0$, *"Gumbel"*), and with the case of a distribution with an upper end point, which corresponds to negative values of the shape parameter ($\xi < 0$, *"inverse Weibull"*).

The MEV distribution assumes that precipitation events are Weibull-distributed. The tail decay of this distribution is controlled by its shape parameter: for $w < 1$ its tail behavior is "sub-exponential", i.e., heavier than that of an exponential (recovered for $w = 1$), albeit with a characteristic scale (Laherrere and Sornette, 1998; Wilson and Toumi, 2005). For $w > 1$ the Weibull tail is super-exponential, with a fast decaying tail, while still retaining an infinite upper end point. Hence, the shape parameter of the Weibull distribution encodes the propensity of a site to be subjected to large extreme events (Wilson and Toumi, 2005; Zorzetto et al., 2016). However, the tail decay of the MEV distribution is not only dependent on that of ordinary values (through $w$) but is also affected by the yearly number of events (Marra et al., 2018) and by the inter-annual variations of $C_j$, $w_j$ and $n_j$.

In an effort to compare the heaviness between the distributions, we have come up with a measure of heaviness that is based on the return levels them-

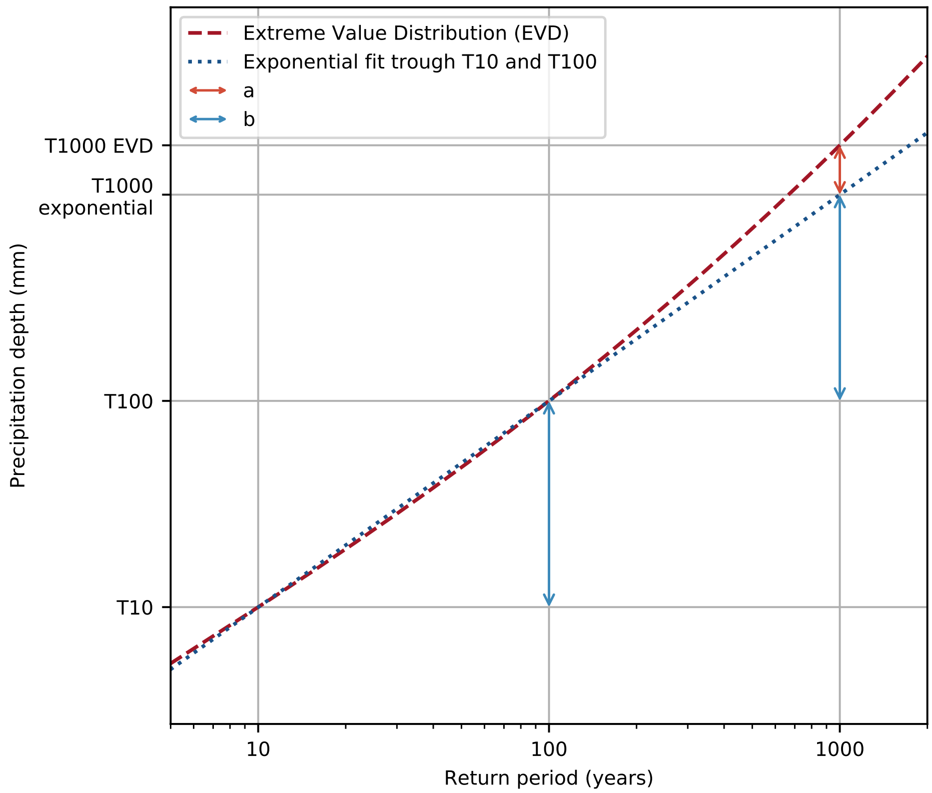


Figure 1: Illustration of our method to measure the tail heaviness for any distribution based on return levels only.

selves (Figure 1). The difference between the 1000-year return level and the 10-year return level can be described as follows:

$$\mathrm{T1000} = \mathrm{T10} + b + b + a \tag{4}$$

Where $b$ is the difference between the 100-year and 10-year return level, i.e.: $b = \mathrm{T100} - \mathrm{T10}$, and $a$ is the additional increase caused by the heaviness of the tail (Figure 1). A positive $a$ is indicative of heavy tails and a negative

$a$ of thin tails. For pure exponential tails it holds that $a = 0$. The value for $a$ is highly dependent on the local precipitation systems, so we defined the heaviness amplification factor $h_{\mathrm{T10-T100-T1000}}$ to be a normalization of $a$:

$$h_{T10-T100-T1000} = \frac{a}{b} = \frac{\mathrm{T1000} - 2 \times \mathrm{T100} + \mathrm{T10}}{\mathrm{T100} - \mathrm{T10}} \tag{5}$$

In words, the meaning of $h_{\mathrm{T10-T100-T1000}}$ is the fractional additional increase between T1000 and T100 that is more than the increase that could be expected from a pure exponentially tailed distribution. A distribution has a heavy tail for $h > 0$ and a thin tail for $h < 0$. Here, we chose a range for the heaviness metric over large return periods from 10 to 1000 years, since the 1000-year return levels are known to be influenced by the distribution choice (e.g., Rajulapati et al., 2020) and that is precisely what we wanted to compare. Yet, it should be noted that this metric may easily be adjusted to other return periods and other factors between the return periods. For GEV and POT the heaviness metric is independent of the return period range as long as the return periods are a factor 10 apart. Although for MEV this heaviness metric is only valid for the return period range over which it is computed, using other ranges (T2-T20-T200 and T5-T50-T500) did not yield significant differences (Figure S6).

## 3. Results and Discussion

### *3.1. Hydrological Year*

Figure 2 shows the frequency distribution of 1000-year return levels estimated using calendar and hydrological years for GEV and MEV. The spatial distribution of the T1000 differences is presented in Supplementary Material

Figure S1b for GEV and Figure S1c for MEV. We found that in the case of GEV quantiles, the fraction of sites characterized by differences within $\pm 0.5\,\%$ is larger than that observed for MEV. When the hydrological year starts in the winter months, the hydrological year is only shifted by a few months. In such instances, the annual maxima mostly stay the same between the calendar and hydrological years, though the included events could differ. For GEV this means that for many cells there is almost no difference in the T1000 estimates, whereas for MEV the difference is small.

On the other hand, when the offset with a calendar year is approximately 6 months, around June, there are many different events included in the hydrological years compared to the calendar years. This results in different annual maxima and large differences in the estimated extremes for GEV and MEV. The differences are most pronounced in the Southern hemisphere and in locations where the hydrological year starts around June, e.g., in the Mediterranean region, in the Middle-East, in Southern Africa, in Brazil, around Indonesia, and in the western US (Figure S1a). For MEV the overall sensitivity in T1000 estimates remains lower than that of GEV. In particular, the distribution of differences in Figure 2 exhibits thicker tails for GEV (e.g., as measured by the wider 5th to 95th percentile interval). This suggests that regional sensitivity to the definition of block maxima can be quite significant for the GEV approach.

Figure S2 in the supplementary material presents the frequency distributions of all analyzed return levels. The lower return levels are less impacted by the start of the hydrological year than the higher ones.

### 3.2. Extreme Precipitation Estimates

Figure 3 shows the 100-year precipitation return levels for a 24-hour duration. Extreme value estimates for other durations and return periods are featured in the Global Precipitation EXtremes (GPEX) dataset (Gründemann et al., 2021). The spatial patterns of the extremes estimated by GEV and MEV are similar to Zorzetto and Marani (2020, their Figure 9), while the spatial pattern of the underlying GEV parameters are consistent with Courty et al. (2019, their Figure 1). The global spatial pattern of return levels for the three EV methods is similar, although large regional differences can be observed. The GEV and POT results are similar in magnitude and show similar differences when compared to MEV. The estimated precipitation extremes are generally lower for both GEV and POT compared to MEV quantiles. MEV estimates exhibit smooth spatial patterns, whereas the spatial patterns using GEV and POT are more irregular, consistent with the results of Zorzetto and Marani (2020) for the conterminous US. The reduced spatial coherence in patterns of extremes for GEV and POT is particularly evident in the Great Plains of North America, and in Northern Russia, Southeast Asia, and Central Africa. Other extreme value approaches and distributions may also yield more coherent spatial patterns of precipitation extremes (e.g., Rajulapati et al., 2020), but comparison of all possible extreme value approaches was not the scope of this study. Furthermore, our analysis (Figure 3) reveals the presence of a large number of circular areas with heavier extremes, corresponding to the location of gauges used for correcting precipitation estimates in the MSWEP algorithm (Beck et al., 2019b). The effect of these local corrections is much larger for traditional EV models (POT and

GEV), while MEV appears less sensitive to these local corrections.

In order to study the ability of the three distributions to capture the spatial coherence of precipitation extremes, we selected several case study areas. They collectively cover a wide range of climates and domain sizes, the locations of which can be found in Figure 3a. Within a single case study area, we expect the precipitation estimates to be statistically homogeneous because of their precipitation generating mechanisms (Cavanaugh and Gershunov, 2015; Cavanaugh et al., 2015) or elevation (Ragulina and Reitan, 2017). Figure 4a shows the coefficient of variation (CV) of T100 extreme precipitation estimates for these case studies. The CV is the ratio of the standard deviation to the mean and is used to compare the relative variation between the study areas. The higher the CV, the higher the relative spread of the precipitation estimates within a spatial domain. This figure shows quite similar behavior for GEV and POT, though POT has a slightly lower spread. The CV for MEV is lower, which points to more spatially coherent T100 precipitation estimates based on single point time series (with 38 years of training data).

To further investigate the global differences in magnitude between the three methods, we examine the extremes for each distribution using a spatially weighted mean over the global land surface. This is displayed for multiple return periods and durations as depth-duration-frequency curves (Figure 5). We first compare the maximum precipitation observed in the dataset to the precipitation predicted from each distribution. As there are 38 complete years of MSWEP data, the maximum empirically observed return level is 39 years (T39 observed, the black dotted line in Figure 5). While locally the empirical T39 estimate could be very different from the true return level, we expect the global average of this value to be representative of the true T39. For GEV and POT, we expected the estimated T39 to be close to the observed value since only the largest values are used to fit these distributions. For MEV, we did not necessarily expect a good agreement for T39, but its performance should be better for return levels greater than the length of the observation time series (Marra et al., 2018, 2019b; Schellander et al., 2019; Zorzetto et al., 2016). The results in Figure 5 show that for the short duration events, the observed T39 is close to the T39 for all three distributions. For increasing durations, the deviation between empirically observed and EV modeled T39 quantiles increases, particularly for MEV. This could be because a smaller number of events per year is used for the fit of MEV-Weibull, whereas the number of events used for the fit of GEV and POT remains constant for all durations. Both GEV and POT show an underestimation and MEV an overestimation. This figure also shows again that the differences between GEV and POT are small. The global average estimated extremes for GEV and POT are notably lower than for MEV, as was already visible from Figure 3. This difference is more pronounced for larger return periods and longer durations.

One reason the quantiles estimated using MEV are higher than using GEV and POT is related to the increase in estimation uncertainty of Weibull parameters when the number of events per hydrological year is low. This is especially relevant in arid regions and for long durations. For instance, for 5 and 10-day durations the average annual number of events is 36 and 21 events respectively. It is therefore possible that this leads to an overestimation by

MEV. To overcome this, windows of two or more years could result in a better parameter estimation (Miniussi and Marani, 2020). A second factor which may be relevant relevant for MEV quantile estimates is the use of a fixed threshold for defining a precipitation event.

### *3.3. Tail Behavior*

To better understand the differences between extremes estimated using the three extreme value methods, we analyze their tail behavior using the heaviness amplification factor $h_{\mathrm{T10-T100-T1000}}$ (Eq. 5). Figure 6 presents $h_{\mathrm{T10-T100-T1000}}$ for a 24-hour duration worldwide for each of the three distributions. We refer to Figures S7-S13 in Section 4 in the supplementary material for maps of $h_{\mathrm{T10-T100-T1000}}$ for the other durations. Both GEV (Figure 6a) and POT (Figure 6b) exhibit a large spatial variability in addition to a low spatial coherence. This makes it difficult to discern clear spatial patterns with the exception of a few notable regions. For instance, in the Amazon, $h_{\mathrm{T10-T100-T1000}}$ is mostly negative, suggesting a tail with an upper limit, while in Eastern and Southern Australia $h_{\mathrm{T10-T100-T1000}}$ it is strongly positive, denoting strong heavy tail behavior. This map roughly corresponds to the spatial patterns of the daily GEV shape parameter shown by Papalexiou and Koutsoyiannis (2013, their Figure 13) and Ragulina and Reitan (2017, their Figure 4). We also find that for the GEV and POT methods, grid cells associated with heavy tails can be adjacent to cells with thin tails. Furthermore, GEV and POT do not always show the same type of tail, heavy or thin, in the same grid cells. In 72 % of the cases the sign of the underlying shape parameter agrees, while in 28 % of the cases the signs are different for daily precipitation. This highlights the large uncertainty associated with estimating reliable tail parameters from short time series and the sensitivity of the GEV and POT methods to sampling effects.

The heaviness of the MEV distribution (Figure 6c) shows a more coherent spatial pattern. At virtually all grid cells the heaviness amplification factor $h_{\mathrm{T10-T100-T1000}}$ (Eq. 5) indicates heavy tail behavior and there is a high consistency within geographic regions and for all durations (Figures S7-S13). Based on previous studies (Cavanaugh et al., 2015; Papalexiou and Koutsoyiannis, 2013; Papalexiou et al., 2013; Ragulina and Reitan, 2017), this predominantly heavy-tail behavior of daily precipitation was expected and is well captured by MEV. There are also topographical patterns visible in the heaviness amplification factor (Figure 6c), though they are not as clearly distinguishable as for the shape parameter itself (Figure S5). The heaviness tends to be higher in arid areas, and lower in mountainous areas. Examples of arid areas with high heaviness include the Sahara, the Namib and Kalahari in Africa, the Gobi, Thar and Taklamakan in Asia, the Atacama Desert in South America, large areas of Southwestern Australia, and the Arabian desert and other areas in the Middle East. This same pattern is to a lesser extent also visible for the heaviness of GEV (Figure 6a) and POT (Figure 6b).

At high elevations a small $h_{\mathrm{T10-T100-T1000}}$ is usually found for MEV (Figure 6c). Examples include the Rocky Mountains and the Sierra Madres in North America, the northern Andes and large areas of the Brazilian Highlands in South America, the Ethiopian Highlands, the Scandinavian Mountains, and the Tibetan Plateau. These spatial patterns are in contrast with what Papalexiou et al. (2018, their Figure 6) found for hourly Weibull tails in the USA, where the heaviest tails are in the mountainous areas, and the thin tails are in the south-east. However, our results correspond well to Ragulina and Reitan (2017, their Figure 4), who showed that heaviness decreases with elevation.

A comparison of the heaviness for different distributions and durations is presented as a boxplot in Figure 7. For spatial maps of the heaviness for the different durations we refer to Figures S7-S13. For GEV and POT, predominantly heavy tails are observed for short durations and thinner tails for long durations. Furthermore, GEV and POT both show a decreasing variability in the heaviness for longer durations, indicated by both shorter whiskers and boxes. The decrease of the heaviness of the tails for increasing durations is in line with the findings of Cavanaugh and Gershunov (2015), who found that longer duration extremes exhibit thinner tails. For GEV and POT the longer durations largely indicate tails with a finite upper end point. This occurs for instance in half of the cases for a duration of 10 days for GEV, and more than half for POT. One implication of this finding is that, when computing return levels for a single location (see Figures S3 and S12), it is possible for the very large return periods that the shorter duration quantiles are more intense than the longer duration quantiles. This is physically impossible (see Figure S14a,b,f and g), and we should thus be extremely careful when interpreting such results.

MEV, on the other hand, shows different heaviness patterns than GEV and POT (Figure 7 and Figures S7-S13). MEV shows almost entirely a heavy-tail behavior, which remains consistent across the range of durations examined. Furthermore, also the variability for MEV is constant across durations, though with a slight increase for longer durations. The MEV distribution thus produces a spatially and temporally coherent heavy tail behavior based on a 38 years calibration sample and a single grid-cell analysis. This is a promising result, as MEV, in contrast to the traditional methods analyzed, provides essential information on the spatial coherence of precipitation extremes without any prior hypothesis on its spatial structure, for example through a spatial clustering scheme (Demirdjian et al., 2018). In fact, the spatial structure of the tail heaviness obtained through the MEV analysis could be used as a measure of statistical homogeneity for regionalization studies.

In this work, we studied the global distribution of rainfall extremes based on stationary statistical models. It has long been recognized that climatic change as well as the inherent variability of the climate system modulate the frequency and intensity of heavy rainfall, prompting the adoption of non-stationary models in water resources management (Milly et al., 2008; Yilmaz and Perera, 2014; Gu et al., 2017). However, in addition to the uncertainty originating from inference on finite length measurements, adopting a non-stationary description of extreme rainfall leads to more complex statistical models, leading in turn to a potentially increased uncertainty of predicted return levels (Lins and Cohn, 2011; Montanari and Koutsoyiannis, 2014; Serinaldi and Kilsby, 2015; Milly et al., 2015). This issue can be mitigated by linking shifts in the rainfall process to mechanistic physical processes. For instance, this can be achieved by linking rainfall statistics to climate features that are more easily predicted by global circulation models (e.g., Grimm and Tedeschi, 2009; Whan et al., 2020; Zorzetto and Li, 2021; Fowler et al., 2021).

Elucidating these links at the global scale and for the broad range of rainfall durations explored here remains a daunting task which we plan to explore in future work.

## 4. Conclusions

The aim of this research was to quantitatively characterize the spatiotemporal variation of global precipitation extremes and their associated extreme value distribution tails. We have fitted three different extreme value methods (GEV, POT, and MEV) to a global precipitation dataset, MSWEP V2.2, to estimate extreme precipitation return levels for several durations. In order to compare the tails of the three distributions, we introduced a novel heaviness amplification factor $h_{\mathrm{T10-T100-T1000}}$ (Eq. 5). Instead of using calendar years to delineate between different years, we used hydrological years, the start of which we defined as the driest month. To our knowledge, this is a novel approach for analyzing precipitation extremes on the global domain. We demonstrated that there is a substantial difference in the extremes depending on the definition of yearly blocks used in the extreme value analysis (Figure 2). These differences were most notable in the Southern hemisphere, and in locations where the driest month occurs around June (Figure S1). Although there is no systematic bias, we still recommend to apply the extreme value analyses for estimating extreme precipitation based on hydrological years in future studies. Our analysis indicates that this can be particularly relevant in the Southern hemisphere and in regions characterized by marked seasonal cycles.

It is well known that the traditional GEV and POT methods require very long data series for accurate estimation of the tail behavior, and our study confirms that there is a low spatial coherence for the tail properties of both distributions (Figure 6a and b) using just 38 years of training data. The tail properties of the MEV distribution are spatially more coherent (Figure 6c) and hence the estimated return levels are more spatially coherent as well (Figure 3c). This spatially coherent behavior, consistent with previous results obtained over the conterminous US (Zorzetto and Marani, 2020), shows that the MEV distribution is able to capture spatially consistent tail behavior from short time series and by a single grid-cell analysis, without any prior information on the spatial precipitation structures. The analysis of the MEV tail behavior reveals distinct spatial patterns, as the heaviness appears to be controlled by climate zones and orography. Heavier tails are observed in arid areas, and thinner tails in mountainous regions. More in-depth analyses are necessary to draw definite conclusions on what exactly controls the heaviness of extreme value distribution tails. The performance of MEV is promising for regions without long local precipitation records. Furthermore, our study shows that the tail behavior captured by MEV is coherent and heavy both spatially and temporally (Figures 6, 7 and S7-S13). For GEV and POT, on the other hand, the tail behavior decreases with increased event duration, resulting in a thin tail with a finite endpoint for about half of the cells for a duration of 10 days.

We also conclude that both GEV and POT generally underestimate the observed extremes, whereas MEV overestimates them (Figure 5). This occurs particularly for long-duration extremes and large return periods. We do consider it likely, however, that the results could be improved, for instance by changing the event threshold or by fitting the Weibull distribution over two or more years for dry areas (Miniussi and Marani, 2020), so as to reduce inter-annual variability of the parameters due to samples of limited length. Our results suggest that this issue is particularly relevant at the longest durations examined. For GEV and POT the results could also be improved by adopting spatial models (Davison et al., 2012; Huser and Wadsworth, 2020).

The data generated for this study are openly available as the GPEX dataset (Gründemann et al., 2021). These data include extreme precipitation return levels and extreme value distribution parameters for durations between 3 hours and 10 days at a global gridded 0.1° resolution. They could be used by engineers as a reference of precipitation extremes for data-scarce regions in particular. For scientific purposes, all underlying parameters are also available and can be used to answer several outstanding questions, such as: what are the controls on the tail behavior of extremes, and what is driving the different changes in tail heaviness with duration for GEV, POT, and MEV?.

## Declaration of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

We acknowledge the use of the `mevpy` Python package (`https://github.com/EnricoZorzetto/mevpy`) for the extreme value analysis. This work was carried out on the Dutch National e-Infrastructure (DNI) with support from the SURF cooperative. Enrico Zorzetto acknowledges support from the NASA Earth and Space Science Fellowship 80NSSC17K0364. Hylke Beck was supported in part by the U.S. Army Corps of Engineers' International Center for Integrated Water Resources Management (ICIWaRM). Nick van de Giesen acknowledges support of the European Commission's Horizon 2020 Programme under grant agreement number 776691 (TWIGA). Ruud van der Ent acknowledges funding from the Netherlands Organization for Scientific Research (NWO), project number 016.Veni.181.015.

The GPEX dataset is available at the 4TU repository Gründemann et al. (2021). The data included are the extremes estimated using the different distributions, the observed extremes, and the parameters to estimate the extremes. These data are available for all durations included in this study. The resolution of the dataset is 0.1°, the resolution of the MSWEP-V2.2 dataset. For more information we refer to the Dataset Usage Notes in Section 5 of the supplementary material.

## References

Alijanian, M., Rakhshandehroo, G.R., Mishra, A.K., Dehghani, M., 2017. Evaluation of satellite rainfall climatology using CMORPH, PERSIANN-CDR, PERSIANN, TRMM, MSWEP over Iran. International Journal of Climatology 37, 4896–4914. doi:10.1002/joc.5131.

Allan, R.P., Soden, B.J., 2008. Atmospheric warming and the amplification of precipitation extremes. Science 321, 1481–1484. doi:10.1126/science.1160787.

Arguez, A., Vose, R.S., 2011. The definition of the standard WMO climate normal: The key to deriving alternative climate normals. Bulletin of the American Meteorological Society 92, 699–704. doi:10.1175/2010BAMS2955.1.

Bai, P., Liu, X., 2018. Evaluation of five satellite-based precipitation products in two gauge-scarce basins on the Tibetan Plateau. Remote Sensing 10. doi:10.3390/RS10081316.

Ball, J., Babister, M., Nathan, R., Weeks, W., Weinmann, E., Retallick, M., Testoni, I., 2019. Australian Rainfall and Runoff: A guide to flood estimation. Commonwealth of Australia (Geoscience Australia).

Beck, H.E., Pan, M., Roy, T., Weedon, G.P., Pappenberger, F., van Dijk, A.I.J.M., Huffman, G.J., Adler, R.F., Wood, E.F., 2019a. Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. Hydrology and Earth System Sciences 23, 207–224. doi:10.5194/hess-23-207-2019.

Beck, H.E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A.I.M., Weedon, G.P., Brocca, L., Pappenberger, F., Huffman, G.J., Wood, E.F., 2017. Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. Hydrology and Earth System Sciences 21, 6201–6217. doi:10.5194/hess-21-6201-2017.

Beck, H.E., Wood, E.F., Pan, M., Fisher, C.K., Miralles, D.G., van Dijk, A.I.J.M., McVicar, T.R., Adler, R.F., 2019b. MSWEP V2 Global 3-Hourly 0.1° Precipitation: Methodology and Quantitative Assessment. Bulletin of the American Meteorological Society 100, 473–500. doi:10.1175/BAMS-D-17-0138.1.

Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Schamm, K., Schneider, U., Ziese, M., 2013. A description of the global land-surface precipitation data products of the global precipitation climatology centre with sample applications including centennial (trend) analysis from 1901-present. Earth System Science Data 5, 71. doi:10.5194/essd-5-71-2013.

Beersma, J., Versteeg, R., Hakvoort, H., 2018. Neerslagstatistieken voor korte duren actualisatie 2018. Technical Report. STOWA. Amersfoort.

Blanchet, J., Ceresetti, D., Molinié, G., Creutin, J.D., 2016. A regional gev scale-invariant framework for intensity–duration–frequency analysis. Journal of hydrology 540, 82–95. doi:10.1016/j.jhydrol.2016.06.007.

Casson, D.R., Werner, M., Weerts, A., Solomatine, D., 2018. Global re-analysis datasets to improve hydrological assessment and snow water equivalent estimation in a sub-arctic watershed. Hydrology and Earth System Sciences 22, 4685–4697. doi:10.5194/hess-22-4685-2018.

Cavanaugh, N.R., Gershunov, A., 2015. Probabilistic tail dependence of intense precipitation on spatiotemporal scale in observations, reanalyses, and GCMs. Climate Dynamics 45, 2965–2975. doi:10.1007/s00382-015-2517-1.

Cavanaugh, N.R., Gershunov, A., Panorska, A.K., Kozubowski, T.J., 2015. The probability distribution of intense daily precipitation. Geophysical Research Letters 42, 1560–1567. doi:10.1002/2015GL063238.

Coles, S., 2001. An introduction to statistical modeling of extreme values. Springer Series in Statistics; Springer: London, UK.

Contractor, S., Donat, M., Alexandre, L.V., Ziese, M., Meyer-Christoffer, A., Schneider, U., Rustemeier, E., Becker, A., Durre, I., Vose, R.S., 2020. Rainfall estimates on a gridded network (regen)–a global land-based gridded dataset of daily precipitation from 1950 to 2016. Hydrology and Earth System Sciences 24, 919–943. doi:10.5194/hess-24-919-2020.

Courty, L.G., Wilby, R.L., Hillier, J.K., Slater, L.J., 2019. Intensity-duration-frequency curves at the global scale. Environmental Research Letters 14, 084045. doi:10.1088/1748-9326/ab370a.

CRED, 2019. Natural disasters 2018. Technical Report. Institute Health and Society UClouvain. Brussels, Belgium. URL: https://www.cred.be/sites/default/files/CREDNaturalDisaster2018.pdf.

Davison, A.C., Padoan, S.A., Ribatet, M., et al., 2012. Statistical modeling of spatial extremes. Statistical science 27, 161–186.

Davison, A.C., Smith, R.L., 1990. Models for Exceedances over High Thresholds. Journal of the Royal Statistical Society. Series B (Methodological) 52, 393–442. URL: https://www.jstor.org/stable/2345667.

De Paola, F., Giugni, M., Pugliese, M., Annis, A., Nardi, F., 2018. GEV parameter estimation and stationary vs. non-stationary analysis of extreme rainfall in African test cities. Hydrology 5. doi:10.3390/hydrology5020028.

Decker, M., Brunke, M.A., Wang, Z., Sakaguchi, K., Zeng, X., Bosilovich, M.G., 2012. Evaluation of the Reanalysis Products from GSFC, NCEP, and ECMWF Using Flux Tower Observations. Journal of Climate 25, 1916–1944. doi:10.1175/JCLI-D-11-00004.1.

Demirdjian, L., Zhou, Y., Huffman, G.J., 2018. Statistical modeling of extreme precipitation with trmm data. Journal of applied meteorology and climatology 57, 15–30. doi:10.1175/JAMC-D-17-0023.1.

Dunn, R.J., Alexander, L.V., Donat, M.G., Zhang, X., Bador, M., Herold, N., Lippmann, T., Allan, R., Aguilar, E., Barry, A.A., et al., 2020. Development of an updated global land in situ-based data set of temperature and precipitation extremes: Hadex3. Journal of Geophysical Research: Atmospheres 125, e2019JD032263.

Fowler, H.J., Lenderink, G., Prein, A.F., Westra, S., Allan, R.P., Ban, N., Barbero, R., Berg, P., Blenkinsop, S., Do, H.X., Guerreiro, S., Haerter, J.O., Kendon, E.J., Lewis, E., Schaer, C., Sharma, A., Villarini, G., Wasko, C., Zhang, X., 2021. Anthropogenic intensification of short-duration rainfall extremes. Nature Reviews Earth & Environment 2, 107–122. doi:10.1038/s43017-020-00128-6.

Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., Da Silva, A.M., Wei, G., Kim, G.K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J.E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S.D., Sienkiewicz, M., Zhao, B., 2017. The modern-era retrospective analysis for research and applications, version 2 (merra-2). Journal of Climate 30, 5419–5454. doi:10.1175/JCLI-D-16-0758.1.

van de Giesen, N., Hut, R., Selker, J., 2014. The Trans-African Observatory (TAHMO). WIREs Water 1, 341–348. doi:10.1002/wat2.1034.

Greenwood, J.A., Landwehr, J., Matalas, N., Wallis, J., 1979. Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form. Water Resources Research 15, 1049–1054. doi:10.1029/WR015i005p01049.

Grimm, A.M., Tedeschi, R.G., 2009. ENSO and extreme rainfall events in South America. Journal of Climate 22, 1589–1609. doi:10.1175/2008JCLI2429.1.

Gründemann, G.J., Werner, M., Veldkamp, T.I.E., 2018. The potential of global reanalysis datasets in identifying flood events in southern africa. Hydrology and Earth System Sciences 22, 4667–4683. doi:10.5194/hess-22-4667-2018.

Gründemann, G.J., Zorzetto, E., Beck, H.E., Schleiss, M., van de Giesen, N., Marani, M., van der Ent, R.J., 2021. Global Precipitation EXtremes dataset. doi:10.4121/uuid:12b5c941-cd54-45db-8d7b-fefaacecaa69.

Gu, X., Zhang, Q., Singh, V.P., Shi, P., 2017. Non-stationarities in the occurrence rate of heavy precipitation across china and its relationship to climate teleconnection patterns. International Journal of Climatology 37, 4186–4198.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horanyi, A., Munoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G.D., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Holm, E., Janiskova, M., Keeley, S., Laloyaux, P., Lopez, P., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thepaut, J.N., 2020. The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society , 1–51doi:10.1002/qj.3803.

Hong, Y., Hsu, K.L., Sorooshian, S., Gao, X., 2004. Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. Journal of Applied Meteorology 43, 1834–1853. doi:10.1175/JAM2173.1.

Hosking, J.R.M., 1990. L-Moments: Analysis and estimation of distributions using linear combinations of order statistics. Journal of the Royal Statistical Society. Series B (Methodological) 52, 105–124.

Hosking, J.R.M., Wallis, J.R., 1987. Parameter and Quantile Estimation for the Generalized Pareto Distribution Parameter and Quantile Estimation Generalized Pareto Distribution. Technometrics 23, 339–349. doi:10.1080/00401706.1987.10488243.

Hosseini, S.R., Scaioni, M., Marani, M., 2020. Extreme Atlantic Hurricane Probability of Occurrence Through the Metastatistical Extreme Value Distribution. Geophysical Research Letters 47, 1–9. doi:10.1029/2019GL086138.

Hu, L., Nikolopoulos, E.I., Marra, F., Morin, E., Marani, M., Anagnostou, E.N., 2020. Evaluation of MEVD-based precipitation frequency analyses from quasi- global precipitation datasets against dense rain gauge networks. Journal of Hydrology 590, 125564. doi:10.1016/j.jhydrol.2020.125564.

Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E.J., Xie, P., 2015. NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG): Algorithm Theoretical Basis Document (ATBD) Version 4.5. Technical Report. NASA/GSFC, Greenbelt, MD 20771, USA.

Huser, R., Wadsworth, J.L., 2020. Advances in statistical modeling of spatial extremes. Wiley Interdisciplinary Reviews: Computational Statistics , e1537.

Joyce, R.J., Janowiak, J.E., Arkin, P.A., Xie, P., 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, 487–503.

Kendon, E.J., Blenkinsop, S., Fowler, H.J., 2018. When will we detect changes in short-duration precipitation extremes? Journal of Climate 31, 2945–2964. doi:10.1175/JCLI-D-17-0435.1.

Kidd, C., Becker, A., Huffman, G.J., Muller, C.L., Joe, P., Skofronick-Jackson, G., Kirschbaum, D.B., 2017. So, how much of the Earth's surface is covered by rain gauges? Bulletin of the American Meteorological Society 98, 69–78. doi:10.1175/BAMS-D-14-00283.1.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., et al., 2015. The jra-55 reanalysis: General specifications and basic characteristics. Journal of the Meteorological Society of Japan 93, 5–48. doi:10.2151/jmsj.2015-001.

Koutsoyiannis, D., 2004a. Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation. Hydrological Sciences Journal 49, 575–590. doi:10.1623/hysj.49.4.575.54430.

Koutsoyiannis, D., 2004b. Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records. Hydrological Sciences Journal 49, 591–610. doi:10.1623/hysj.49.4.591.54424.

Laherrere, J., Sornette, D., 1998. Stretched exponential distributions in nature and economy: "fat tails" with characteristic scales. The European Physical Journal B-Condensed Matter and Complex Systems 2, 525–539.

Lins, H.F., Cohn, T.A., 2011. Stationarity: wanted dead or alive? 1. JAWRA Journal of the American Water Resources Association 47, 475–480.

Liu, Z., Liu, Y., Wang, S., Yang, X., Wang, L., Baig, M.H.A., Chi, W., Wang, Z., 2018. Evaluation of Spatial and Temporal Performances of

ERA-Interim Precipitation and Temperature in Mainland China. Journal of Climate 31, 4347–4365. doi:10.1175/JCLI-D-17-0212.1.

Mailhot, A., Duchesne, S., 2009. Design criteria of urban drainage infrastructures under climate change. Journal of Water Resources Planning and Management 136, 201–208. doi:10.1061/(asce)wr.1943-5452.0000023.

Marani, M., Ignaccolo, M., 2015. A metastatistical approach to rainfall extremes. Advances in Water Resources 79, 121–126. doi:10.1016/j.advwatres.2015.03.001.

Marani, M., Zanetti, S., 2015. Long-term oscillations in rainfall extremes in a 268 year daily time series. Water Resources Research 51, 639–647. doi:10.1002/2014WR015885.

Marra, F., Borga, M., Morin, E., 2020. A unified framework for extreme subdaily precipitation frequency analyses based on ordinary events. Geophysical Research Letters 47, e2020GL090209.

Marra, F., Nikolopoulos, E.I., Anagnostou, E.N., Bárdossy, A., Morin, E., 2019a. Precipitation frequency analysis from remotely sensed datasets: A focused review. Journal of Hydrology 574, 699–705. doi:10.1016/j.jhydrol.2019.04.081.

Marra, F., Nikolopoulos, E.I., Anagnostou, E.N., Morin, E., 2018. Metastatistical Extreme Value analysis of hourly rainfall from short records: Estimation of high quantiles and impact of measurement errors. Advances in Water Resources 117, 27–39. doi:10.1016/j.advwatres.2018.05.001.

Marra, F., Zoccatelli, D., Armon, M., Morin, E., 2019b. A simplified MEV formulation to model extremes emerging from multiple nonstationary underlying processes. Advances in Water Resources 127, 280–290. doi:10.1016/j.advwatres.2019.04.002.

McGraw, D., Nikolopoulos, E.I., Marra, F., Anagnostou, E.N., 2019. Precipitation frequency analyses based on radar estimates: An evaluation over the contiguous United States. Journal of Hydrology 573, 299–310. doi:10.1016/j.jhydrol.2019.03.032.

Ménégoz, M., Gallée, H., Jacobi, H.W., 2013. Precipitation and snow cover in the himalaya: from reanalysis to regional climate simulations. Hydrology and Earth System Sciences 17, 3921–3936. doi:10.5194/hess-17-3921-2013.

Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P., Stouffer, R.J., 2008. Stationarity is dead: Whither water management? Earth 4. doi:10.1126/science.1151915.

Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P., Stouffer, R.J., Dettinger, M.D., Krysanova, V., 2015. On critiques of "stationarity is dead: Whither water management?". Water Resources Research 51, 7785–7789. doi:10.1002/2015WR01740.

Miniussi, A., Marani, M., 2020. Estimation of daily rainfall extremes through the metastatistical extreme value distribution: Uncertainty minimization and implications for trend detection. Water Resources Research 56, e2019WR026535. doi:10.1029/2019WR026535.

Miniussi, A., Marani, M., Villarini, G., 2020a. Metastatistical Extreme Value Distribution applied to floods across the continental United States. Advances in Water Resources 136, 103498. doi:10.1016/j.advwatres.2019.103498.

Miniussi, A., Villarini, G., Marani, M., 2020b. Analyses Through the Metastatistical Extreme Value Distribution Identify Contributions of Tropical Cyclones to Rainfall Extremes in the Eastern United States Geophysical Research Letters. Geophysical Research Letters 47, e2020GL087238. doi:10.1029/2020GL087238.

Mishra, A.K., Coulibaly, P., 2009. Developments in hydrometric network design: A review. Reviews of Geophysics 47. doi:10.1029/2007RG000243.

Montanari, A., Koutsoyiannis, D., 2014. Modeling and mitigating natural hazards: Stationarity is immortal! Water resources research 50, 9748–9756. doi:10.1002/2014WR016092.

Nissen, K.M., Ulbrich, U., 2017. Increasing frequencies and changing characteristics of heavy precipitation events threatening infrastructure in Europe under climate change. Natural Hazards and Earth System Sciences 17, 1177–1190. doi:10.5194/nhess-17-1177-2017.

Overeem, A., Buishand, A., Holleman, I., 2008. Rainfall depth-duration-frequency curves and their uncertainties. Journal of Hydrology 348, 124–134. doi:10.1016/j.jhydrol.2007.09.044.

Papalexiou, S.M., Aghakouchak, A., Foufoula-Georgiou, E., 2018. A diagnostic framework for understanding climatology of tails of hourly precipitation extremes in the United States. Water Resources Research 54, 6725–6738. doi:10.1029/2018WR022732.

Papalexiou, S.M., Koutsoyiannis, D., 2013. Battle of extreme value distributions: A global survey on extreme daily rainfall. Water Resources Research 49, 187–201. doi:10.1029/2012WR012557.

Papalexiou, S.M., Koutsoyiannis, D., Makropoulos, C., 2013. How extreme is extreme? An assessment of daily rainfall distribution tails. Hydrology and Earth System Sciences 17, 851–862. doi:10.5194/hess-17-851-2013.

Perica, S., Pavlovic, S., St. Laurent, M., Trypaluk, C., Unruh, D., Martin, D., Wilhite, O., 2015. NOAA Atlas 14 Volume 10, Precipitation-Frequency Atlas of the United States, Northeastern States. Technical Report. NOAA, National Weather Service.

Perica, S., Pavlovic, S., St. Laurent, M., Trypaluk, C., Unruh, D., Wilhite, O., 2018. NOAA Atlas 14: Precipitation-frequency atlas of the United States, Texas. Technical Report. NOAA, National Weather Service.

Ragulina, G., Reitan, T., 2017. Generalized extreme value shape parameter and its nature for extreme precipitation using long time series and the Bayesian approach precipitation. Hydrological Sciences Journal 62, 863–879. doi:10.1080/02626667.2016.1260134.

Rajulapati, C.R., Papalexiou, S.M., Clark, M.P., Razavi, S., Tang, G., Pomeroy, J.W., 2020. Assessment of extremes in global precipitation products: How reliable are they? Journal of Hydrometeorology 21, 2855–2873. doi:10.1175/JHM-D-20-0040.1.

Rasmusson, E.M., Arkin, P.A., 1993. A global view of large-scale precipitation variability. Journal of Climate 6, 1495–1522.

Sahlu, D., Moges, S.A., Nikolopoulos, E.I., Anagnostou, E.N., Hailu, D., 2017. Evaluation of high-resolution multisatellite and reanalysis rainfall products over East Africa. Advances in Meteorology 2017. doi:10.1155/2017/4957960.

Satgé, F., Ruelland, D., Bonnet, M.P., Molina, J., Pillco, R., 2019. Consistency of satellite-based precipitation products in space and over time compared with gauge observations and snow-hydrological modelling in the Lake Titicaca region. Hydrology and Earth System Sciences 23, 595–619. doi:10.5194/hess-23-595-2019.

Schellander, H., Lieb, A., Hell, T., 2019. Error structure of metastatistical and generalized extreme value distributions for modeling extreme rainfall in austria. Earth and Space Science 6, 1616–1632. doi:10.1029/2019EA000557.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Ziese, M., 2011. GPCC full data reanalysis version 6.0 at 0.5°: Monthly land-surface precipitation from rain-gauges built on GTS-based and historic data. doi:10.5676/DWD_GPCC/FD_M_V6_050.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., Rudolf, B., 2014. 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, 15–40. doi:10.1007/s00704-013-0860-x.

Serinaldi, F., Kilsby, C.G., 2014. Rainfall extremes: Toward reconciliation after the battle of distributions. Water Resources Research 50, 336–352. doi:10.1002/2013WR014211.

Serinaldi, F., Kilsby, C.G., 2015. Stationarity is undead: Uncertainty dominates the distribution of extremes. Advances in Water Resources 77, 17–36. doi:10.1016/j.advwatres.2014.12.013.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., Hsu, K., 2018. A review of global precipitation data sets: Data sources, estimation, and intercomparisons. Review of Geophysics 56, 79–107. doi:10.1002/2017RG000574.

UNSIDR, 2015. Sendai Framework for Disaster Risk Reduction 2015 - 2030. Technical Report. URL: www.unisdr.org/we/inform/publications/43291.

Ushio, T., Sasashige, K., Kubota, T., Shige, S., Okamoto, K., Aonashi, K., Inoue, T., Takahashi, N., Iguchi, T., Kachi, M., Oki, R., Morimoto, T., Kawasaki, Z.I., 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 87 A, 137–151. doi:10.2151/jmsj.87A.137.

Villarini, G., Smith, J.A., Lynn, M., Vitolo, R., Stephenson, D.B., Krajewski, W.F., 2011. On the frequency of heavy rainfall for the Midwest of the United States. Journal of Hydrology 400, 103–120. doi:10.1016/j.jhydrol.2011.01.027.

Ward, P.J., Kummu, M., Lall, U., 2016. Flood frequencies and durations and their response to El Niño Southern Oscillation: Global analysis. Journal of Hydrology 539, 358–378. doi:10.1016/j.jhydrol.2016.05.045.

Wasko, C., Sharma, A., Johnson, F., 2015. Does storm duration modulate the extreme precipitation-temperature scaling relationship? Geophysical Research Letters 42, 8783–8790. doi:10.1002/2015GL066274.

Westra, S., Alexander, L.V., Zwiers, F.W., 2013. Global increasing trends in annual maximum daily precipitation. Journal of Climate 26, 3904–3918. doi:10.1175/JCLI-D-12-00502.1.

Whan, K., Sillmann, J., Schaller, N., Haarsma, R., 2020. Future changes in atmospheric rivers and extreme precipitation in Norway. Climate Dynamics 54, 2071–2084. doi:10.1007/s00382-019-05099-z.

Wilson, P.S., Toumi, R., 2005. A fundamental probability distribution for heavy rainfall. Geophysical Research Letters 32. doi:10.1029/2005GL022465.

Yilmaz, A., Imteaz, M., Perera, B., 2017. Investigation of non-stationarity of extreme rainfalls and spatial variability of rainfall intensity–frequency–duration relationships: a case study of Victoria, Australia. International Journal of Climatology 37, 430–442. doi:10.1002/joc.4716.

Yilmaz, A., Perera, B., 2014. Extreme rainfall nonstationarity investigation and intensity–frequency–duration relationship. Journal of Hydrologic Engineering 19, 1160–1172. doi:10.1061/(ASCE)HE.1943-5584.0000878.

Zhang, D., Liu, X., Bai, P., Li, X.H., 2019. Suitability of satellite-based precipitation products for water balance simulations using multiple observations in a humid catchment. Remote Sensing 11. doi:10.3390/rs11020151.

Zorzetto, E., Botter, G., Marani, M., 2016. On the emergence of rainfall extremes from ordinary events. Geophysical Research Letters 43, 8076–8082. doi:10.1002/2016GL069445.

Zorzetto, E., Li, L., 2021. Impacts of the North Atlantic subtropical high on daily summer precipitation over the conterminous United States. Journal of Hydrometeorology doi:10.1175/JHM-D-20-0242.1.

Zorzetto, E., Marani, M., 2019. Downscaling of rainfall extremes from satellite observations. Water Resources Research 55, 156–174. doi:10.1029/2018WR022950.

Zorzetto, E., Marani, M., 2020. Extreme value metastatistical analysis of remotely sensed rainfall in ungauged areas: Spatial downscaling and error modelling. Advances in Water Resources 135, 103483. doi:10.1016/j.advwatres.2019.103483.

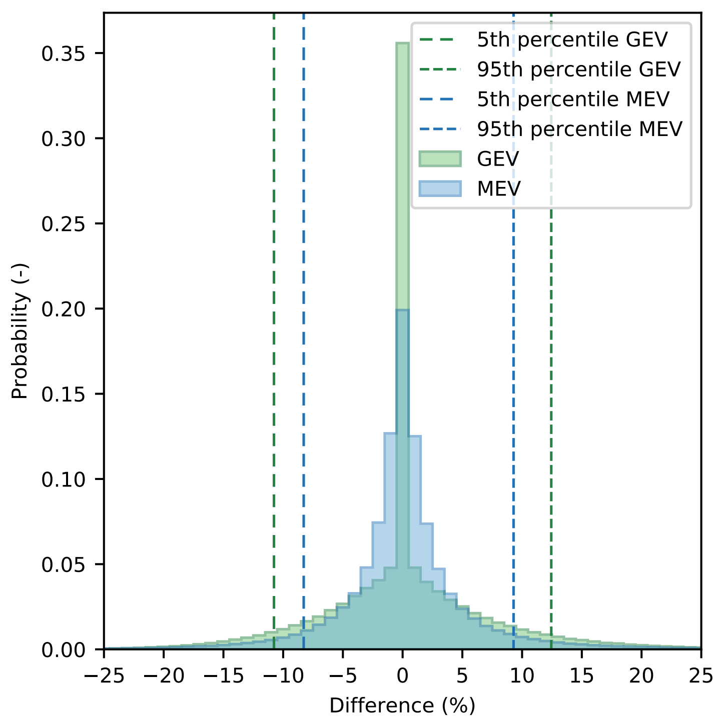


Figure 2: Weighted histogram showing the percentage difference in the values of T1000 quantiles calculated using calendar years and hydrological years. Included in the figure are all cells where the start of the hydrological year is different than the calendar year (i.e., the hydrological year does not start in January, see Supplementary Material Figure 1a). A negative difference indicates that the T1000 estimate is larger using hydrological years, whereas a positive difference indicates that the T1000 estimate is larger using calendar years.

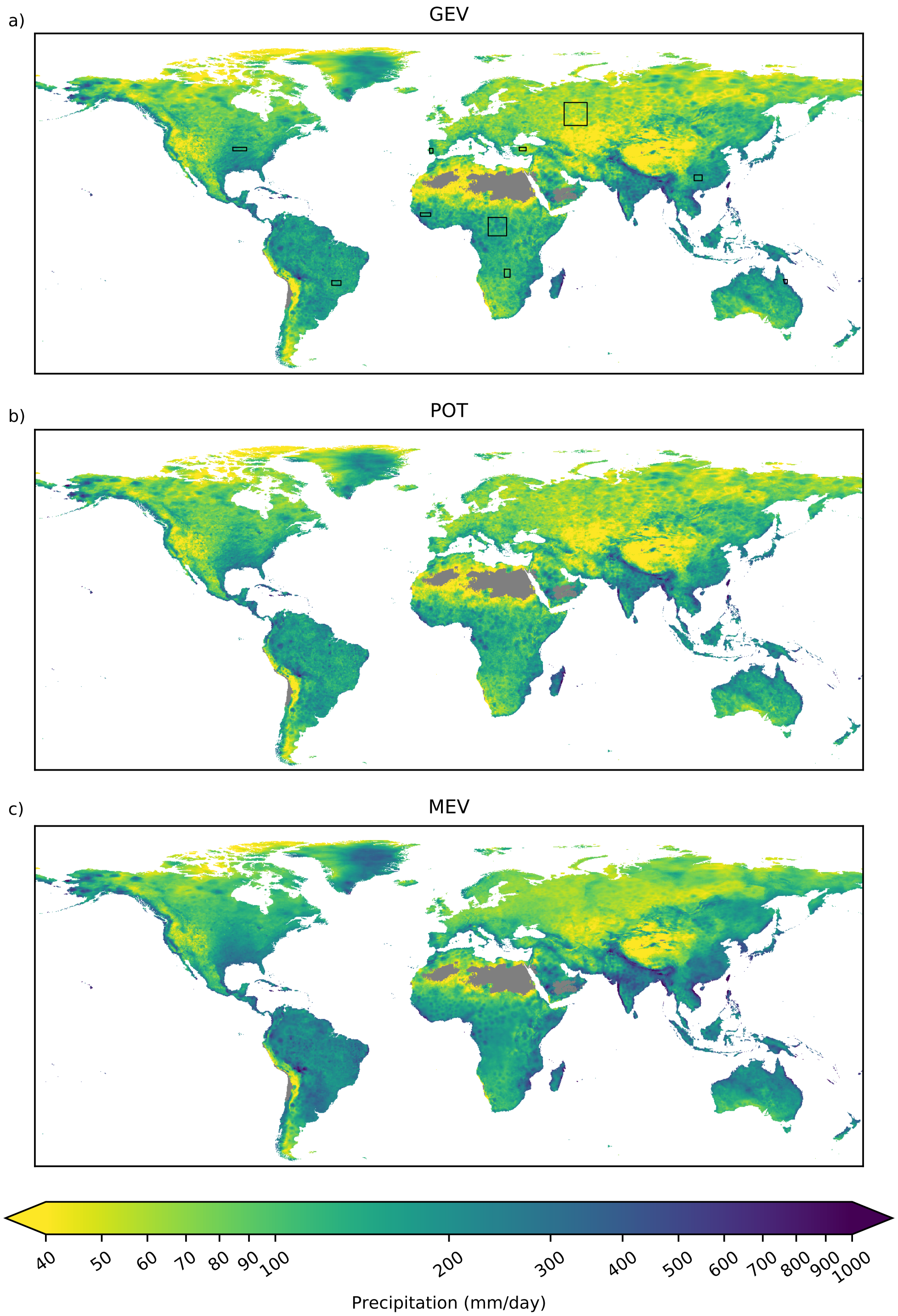


Figure 3: Precipitation return levels with a duration of 24-hours for a 100-year return period for different extreme value distributions: (a) the Generalized Extreme Value (GEV) distribution, (b) the Peak Over Threshold (POT) method, and (c) the Metastatistical Extreme Value (MEV) distribution. The black rectangles in panel a are the case studies corresponding to the areas in Figure 4.

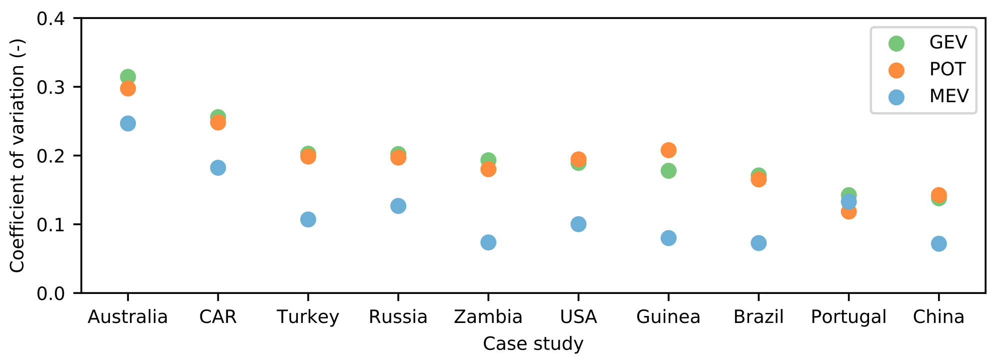


Figure 4: Coefficient of variation for the difference in estimated T100 quantiles for the three extreme value methods for 24-hour precipitation at selected case study areas. The coefficient of variation is the standard deviation of the precipitation divided by the mean precipitation. The locations of the case study areas are displayed in Fig 3a.

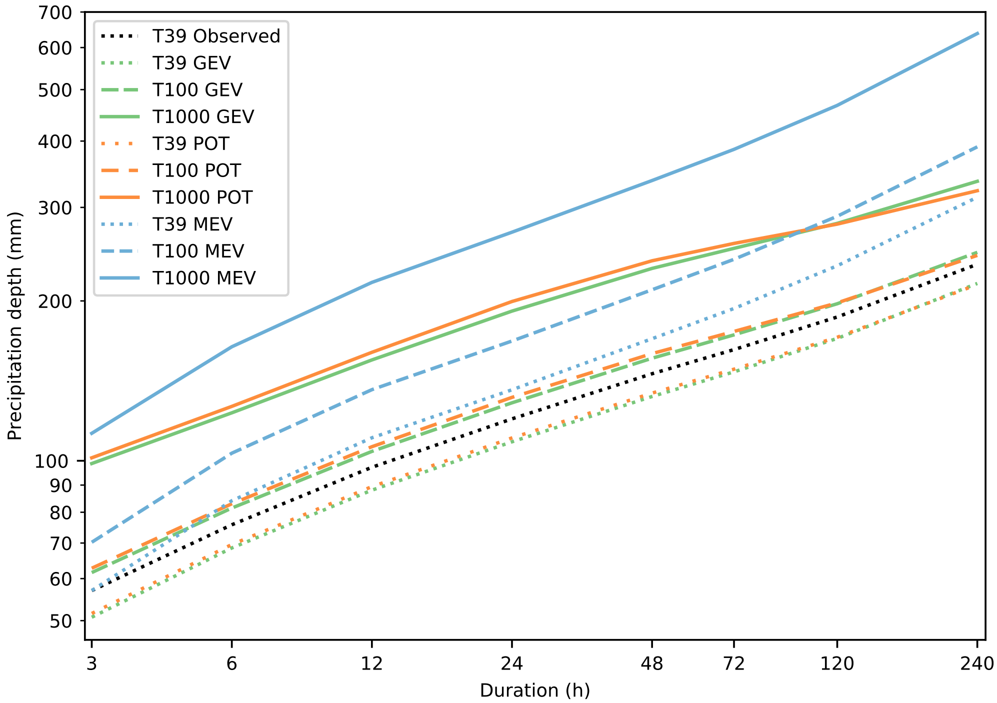


Figure 5: Area-weighted average depth-duration-frequency curves for the global land surface. T39 Observed is the mean spatially weighted maximum precipitation observed in the MSWEP-V2.2 dataset.

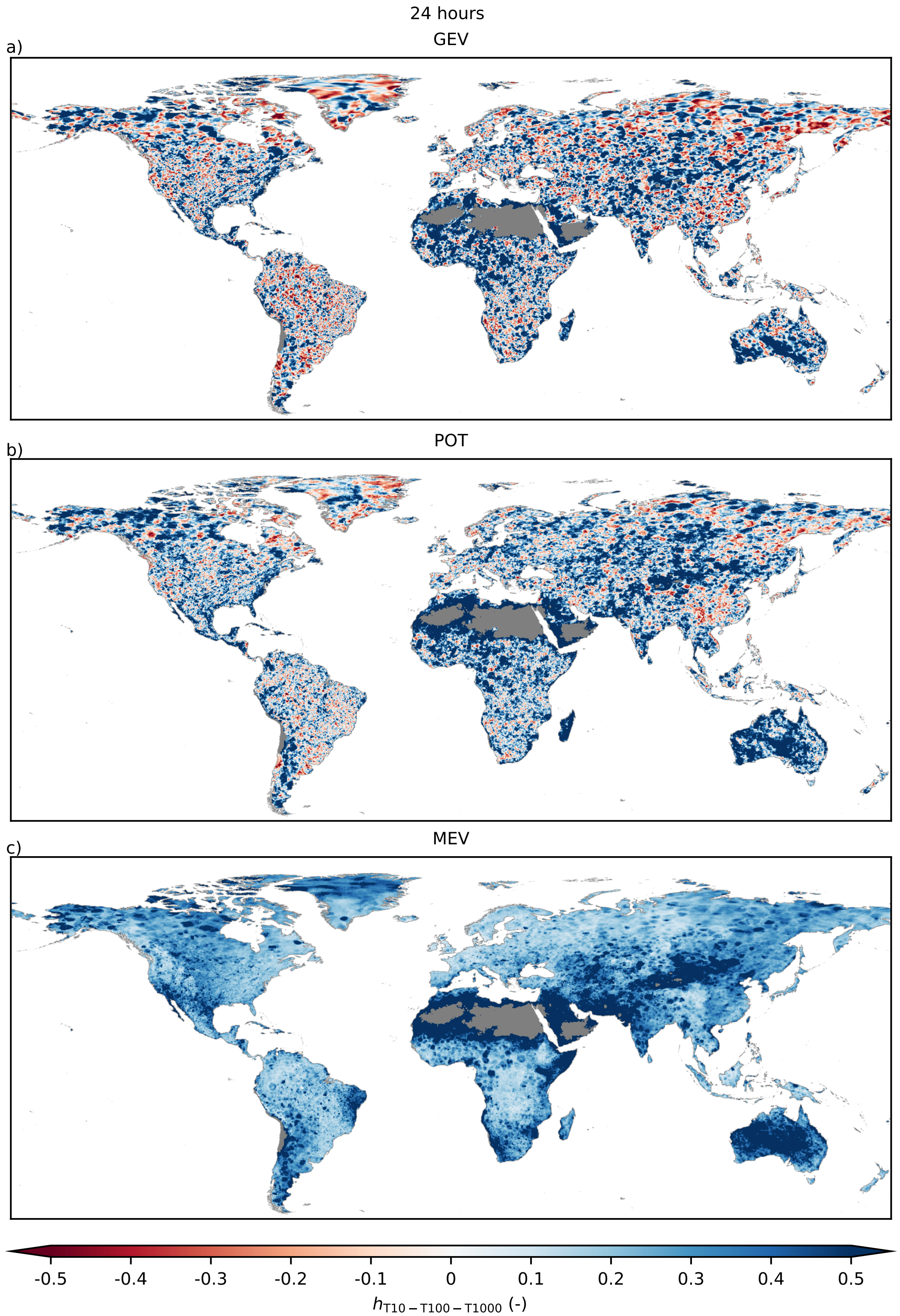




Figure 6: The heaviness amplification factor $h_{\mathrm{T10-T100-T1000}}$ (Eq. 5) for daily precipitation calculated for different extreme value methods: (a) GEV, (b) POT, (c) MEV. Red indicates a thin tail, white an exponential tail, and blue a heavy tail. See section 2.2.2 for more information on the heaviness metric, and Figures S7-S13 for maps of $h_{\mathrm{T10-T100-T1000}}$ for the other durations.

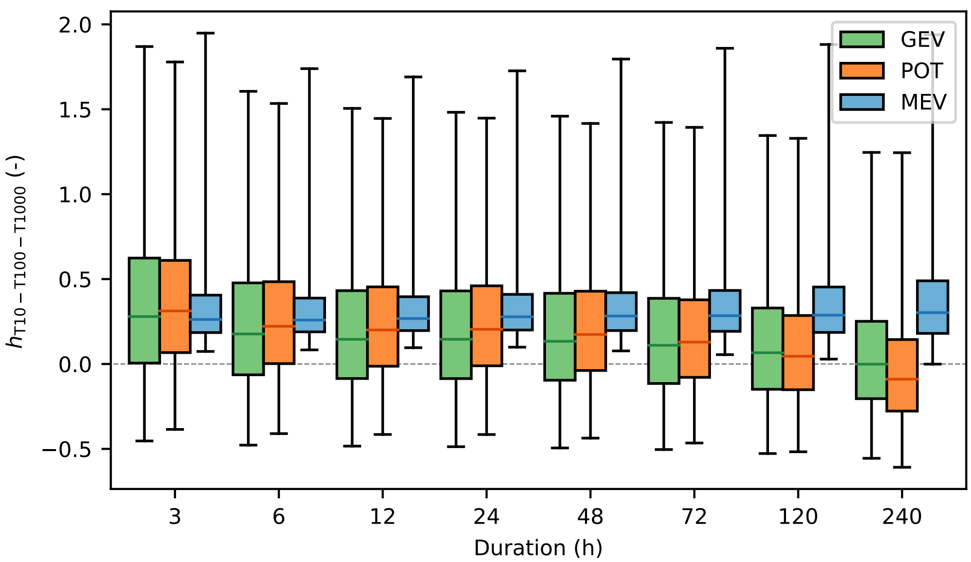


Figure 7: Boxplots showing the distribution of the heaviness amplification factor $h_{\mathrm{T10-T100-T1000}}$ (-) for different durations and extreme value methods: (a) GEV and POT, and (b) MEV. The whiskers denote the 1st and 99th percentiles. The top and bottom of the boxes represent the 75th and 25th percentiles, respectively. The dashed gray horizontal lines indicate exponential tails. See section 2.2.2 for more information on the heaviness metric.

# Supplementary Material for “Extreme Precipitation Return Levels for Multiple Durations on a Global Scale”

August 25, 2021

## Contents

# 1 Calendar and Hydrological Year

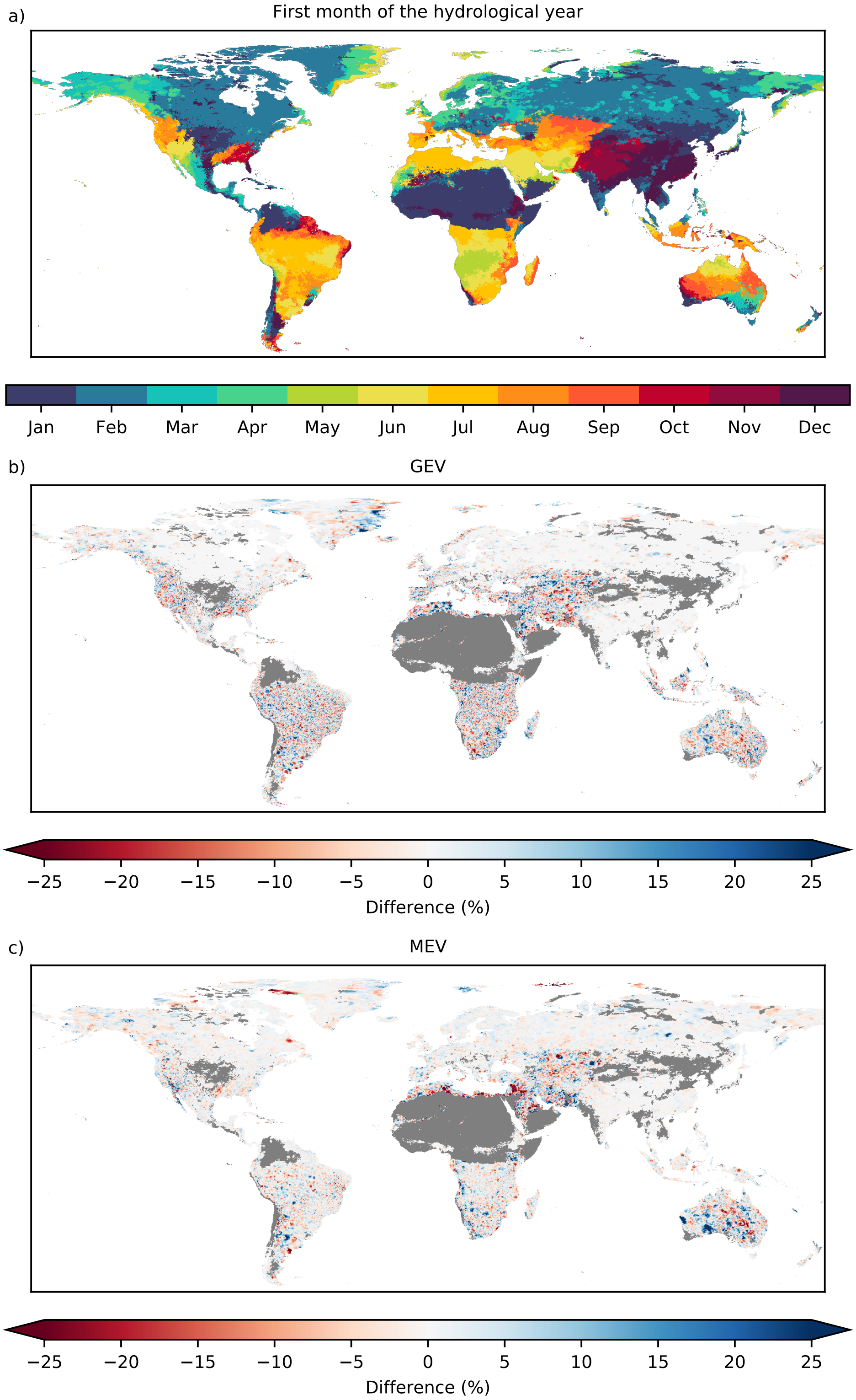


Figure S1: (a) The month indicating the start of the hydrological year. (b) The percentage difference of the daily 1000-year return levels of calendar and hydrological years for GEV. (c) The percentage difference of the daily 1000-year return levels of calendar and hydrological years for MEV. A negative difference indicates that the T1000 estimate is larger using hydrological years, whereas a positive difference indicates that the T1000 estimate is larger using calendar years.

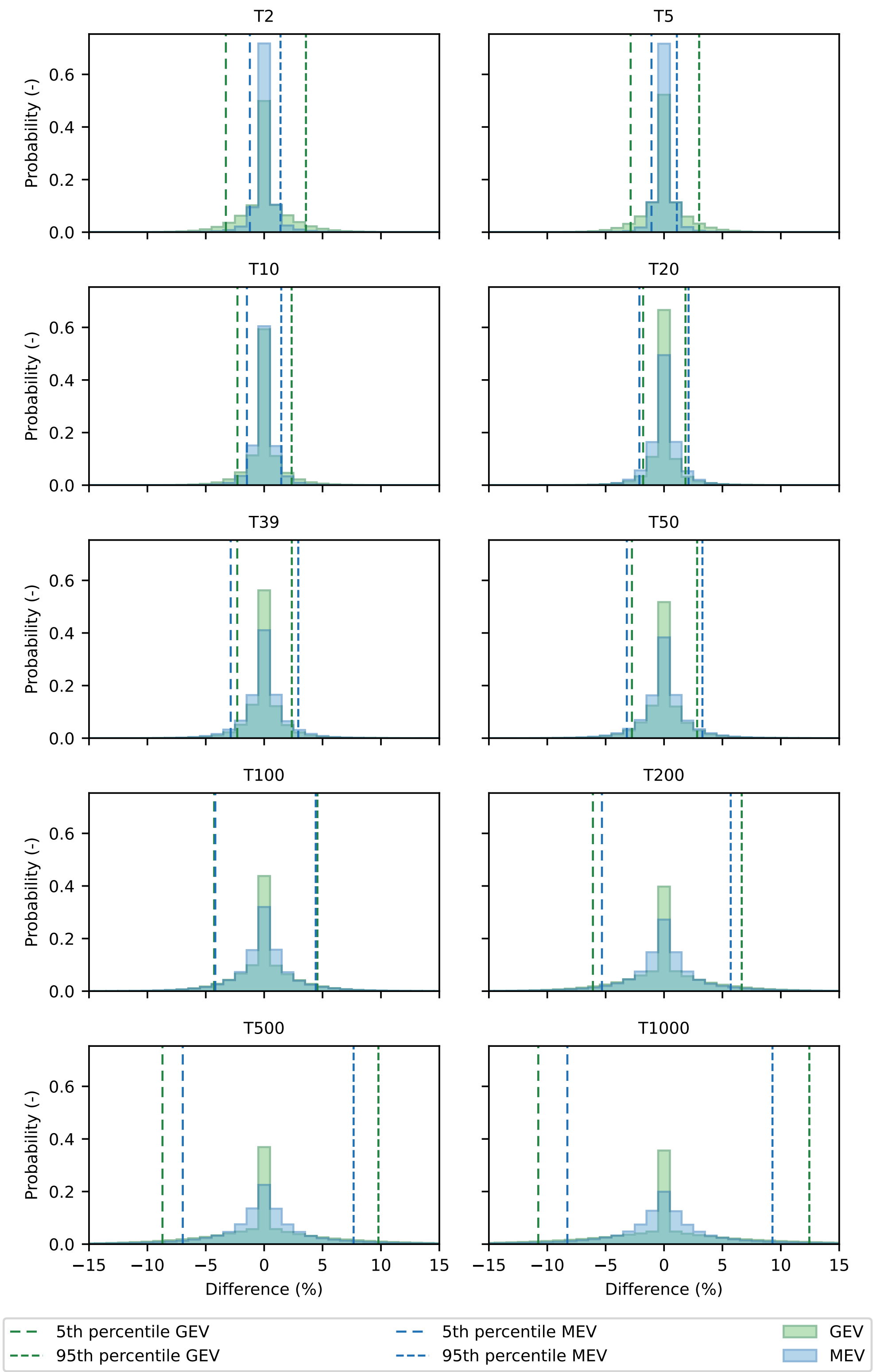


Figure S2: Weighted histograms showing the percentage difference in the values of several return periods (T, see titles) calculated using calendar years and hydrological years. Included in the figure are all cells where the start of the hydrological year is different from the calendar year (i.e., the hydrological year does not start in January, see Figure S1a). A negative difference indicates that the return period estimate is larger using hydrological years, whereas a positive difference indicates that the return period estimate is larger using calendar years.

# 2 Threshold Analysis for POT

There are many different ways to select an appropriate threshold for the Peak Over Threshold (POT) analysis, such as a value over a specific threshold, a percentage, or an average number of events per year. We refer to Caeiro and Gomes (2016); Langousis, Mamalakis, Puliga, and Deidda (2016) for recent overviews of such threshold selection. Which method is the most effective remains to be an unanswered question. For our global-scale application, we have analyzed to take on average 1 to 5 events per year. This fixed number of events, as opposed to the events over a predefined threshold, ensures that the number of events per year remains constant for all durations, and allows for analyses on the global domain.

The 100-year return levels for all durations and thresholds are shown in Figure S3a. The boxplot shows minimal difference in the $T100$ estimates for the different thresholds for the short durations, and a minor difference for the longer durations, namely that the estimated $T100$ are slightly larger for inclusion of less events per year. Overall the figure shows only slight differences, so more information has to be considered for accurate threshold selection.

A comparison of the three parameters for all durations and thresholds is shown in Figure S3b-d. The location and scale parameters in S3b and c show very similar results for the different thresholds. The shape parameter in S3d, however, has more variation for the different POT-thresholds. For all durations, the variability of the shape parameter decreases for the inclusion of more events per year, so more events to fit the distribution to. For durations between 3 and 48 hours, the shape slightly increases for the inclusion of more events per year. In other words, the shape is lower for on average 1 event per year, and higher for on average 5 events per year for durations between 3 and 48 hours. For 72 hours, the median shape parameter of the different thresholds is constant, though the variability decreases for the inclusion of more events. For the 5 and 10-day durations (120 and 240 hours, respectively), the shape parameter slightly decreases for the inclusion of more (non-extreme) events. The general pattern of the GP shape parameter is similar to that of the GEV, as they both show a decreasing shape parameter for increasing durations.

Previous studies that have looked into the GP shape parameter on the global domain focused on daily durations. Papalexiou and Koutsoyiannis (2013) estimated the mode of the shape for daily precipitation as 0.134, but displaying a large spread. Serinaldi and Kilsby (2014) estimated the shape for four different seasons based on 1898 stations with more than 100 years of data. They found the shape depends on the season and varies between 0.061 and 0.097, also displaying a large spread. Furthermore, they found that if you lower the threshold and include more non-extreme events, the shape parameter is higher. Our results are similar to that of Serinaldi and Kilsby (2014), as the median for daily precipitation for the different thresholds varies between 0.711 and 0.918, and we also found that the inclusion of more non-extreme events leads to a higher shape parameter for daily precipitation.

In short, the above analysis shows that the threshold selection is of minor influence on the estimation of the 100-year return level. Of the three underlying parameters, the threshold selection has minimal influence on the location and scale parameter, though a much greater influence on the shape parameter. The variability of the shape parameter is the highest for the threshold of on average 1 event per year, and decreases when more non-extreme events are used to fit the distribution to. Furthermore, the shape parameter increases for the inclusion of more events for the shorter durations (3-hours to 2-days), remains constant for a 3-day duration, and decreases for the 5 and 10-day durations. Based on these analyses, we chose to show the threshold of on average 3 events per year in the main manuscript, as that threshold has the right balance of a low variability and not as much of an underestimation of the 100-year return levels for the longest durations.

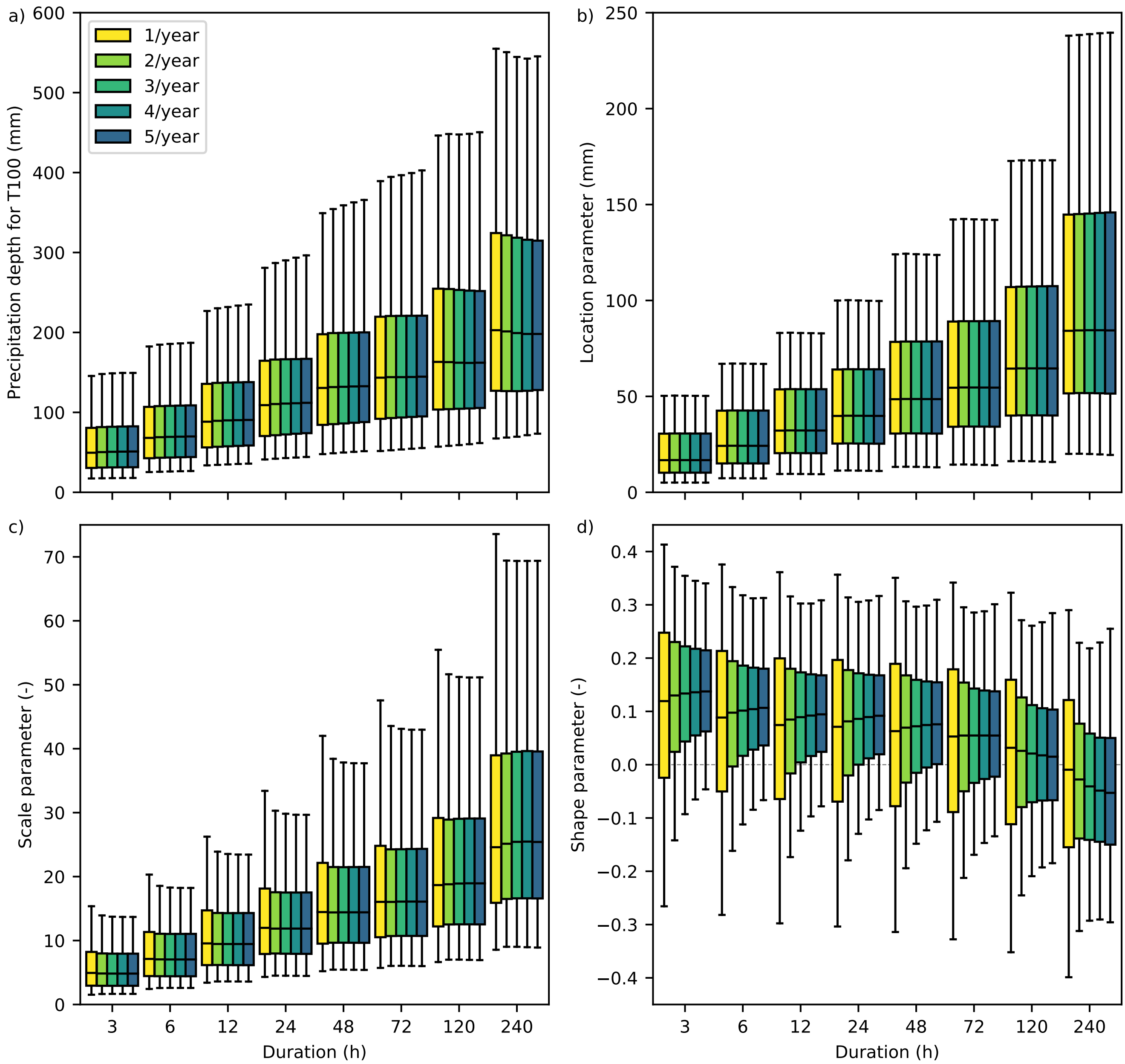


Figure S3: Weighted boxplot showing the distribution of (a) the 100-year precipitation return levels, and the (b) location, (c) scale, and (d) shape parameters for several durations and thresholds for the Peak Over Threshold method. Thresholds range from on average one to five events per year. The top and bottom of the boxes represent the 75th and 25th percentiles, respectively. The whiskers denote the 1st and 99th percentiles. The dashed gray horizontal line in (d) indicates exponential tails. A shape parameter smaller than zero indicates a thin tail with an upper limit, a shape parameter larger than zero indicates a heavy power-law tail.

## 3 Shape Parameter

The GEV and MEV distributions are both flexible and able to describe different tail behaviors. The tail behavior the two distributions varies, see Section 2.2.2 in the main manuscript for an overview and Figure S4 for different combinations of scale and shape parameters.

Maps of the shape parameter for the three distributions for daily duration are shown in Figure S5. The spatial patterns of the shape parameters are largely similar to those of the heaviness $h_{\mathrm{T10-T100-T1000}}$, described in Section 3.3 and Figures 5 and 6 of the main manuscript. There are, however, some notable differences. For MEV, the mountainous areas are more defined, indicated by higher shape values (less heavy tail behavior). Looking at the shape parameter, we find, however, that the relationship of the Weibull shape parameter with elevation is more complicated. Lower shapes (heavier tails) are generally observed on the leeward side of large mountain ranges, and higher shapes (though still indicative of heavy-tail behavior) on the windward side that is dominated by orographically enhanced frontal precipitation. This is for example visible in the Rocky Mountains, Indonesia, and Norway, and corresponds to findings of Cavanaugh and Gershunov (2015, their Figure 5), who showed that exponential tails are observed in regions where extreme precipitation is predominantly generated by one type of system.

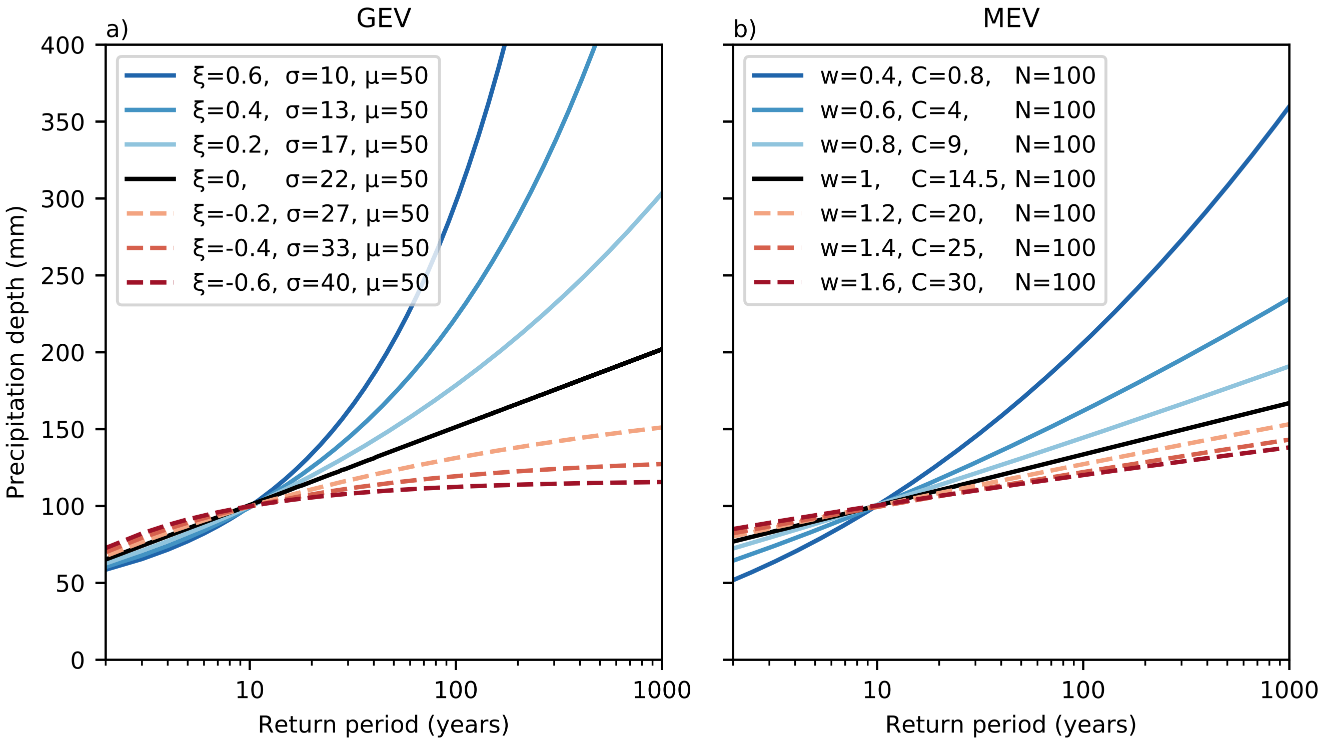


Figure S4: (a) Behavior of the shape ($\xi$) and scale ($\sigma$) parameters of the GEV distribution, with a constant location parameter ($\mu$), and (b) behavior of the shape ($w$) and scale ($C$) parameters of the MEV-Weibull distribution, with a constant number of events ($N$). The results for MEV have been obtained with constant $w$, $C$ and $N$ parameters for each year. The values of the shape and scale parameter pairs have been chosen such that they all have a precipitation depth of approximately 100 mm for a 10-year return period.

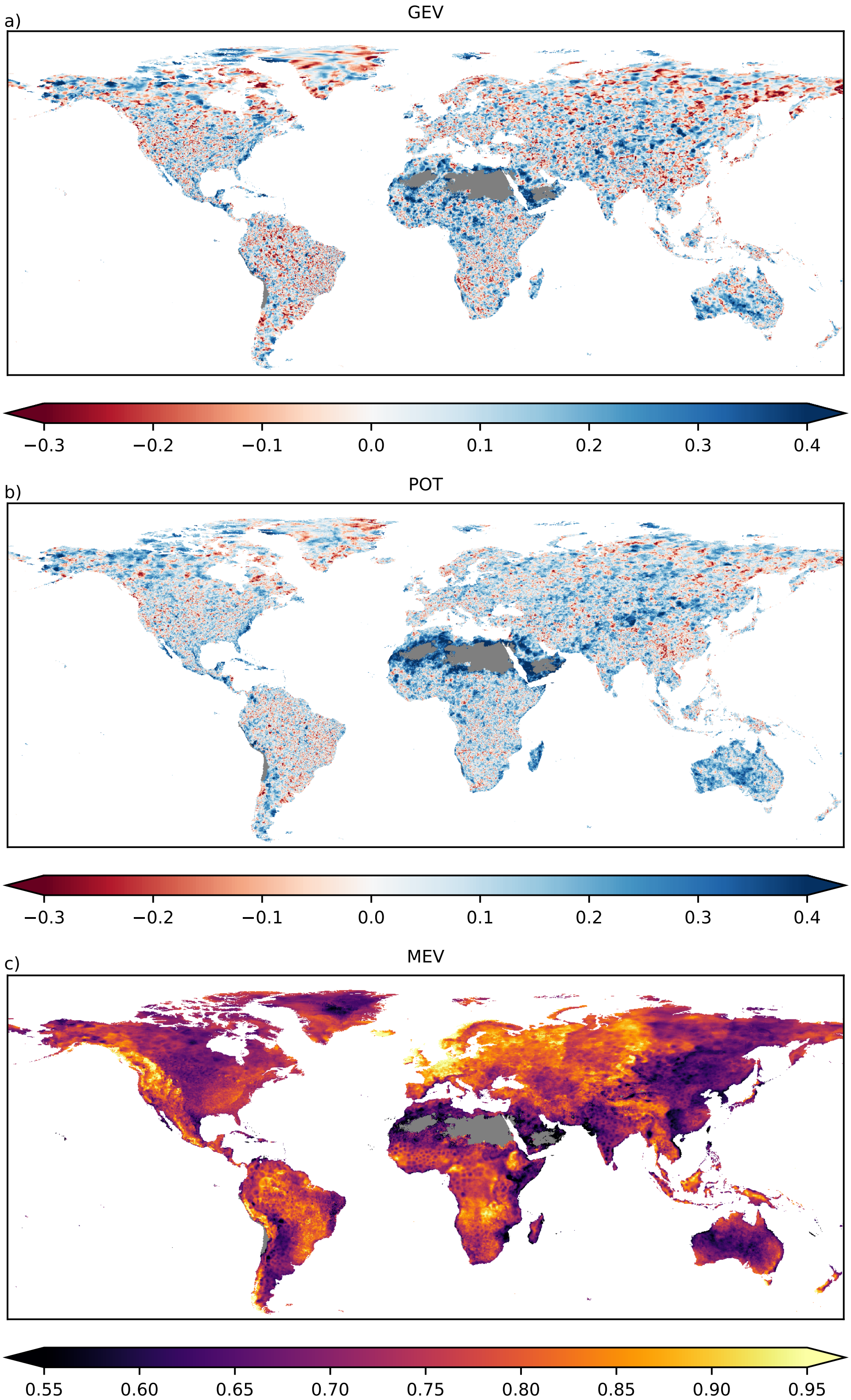


Figure S5: The shape parameter (−) for daily precipitation calculated for different extreme value methods: (a) GEV, equation 1 — $\xi_{\mathrm{GEV}}$, (b) POT, equation 2 — $\xi_{\mathrm{GP}}$ and (c) MEV, equation 3 — $w$. For MEV, the mean shape parameter of all yearly Weibull distributions is displayed. The colorbar min and max are based on the 1st and 99th percentile. For GEV and POT, red indicates a thin shape with an upper limit, white an exponential shape, and blue a heavy power-law shape. For MEV all median shapes are indicative of sub-exponential heavy shapes, though darker colors are heavier than lighter colors.

# 4 Tail Behavior MEV for Different Return Level Combinations

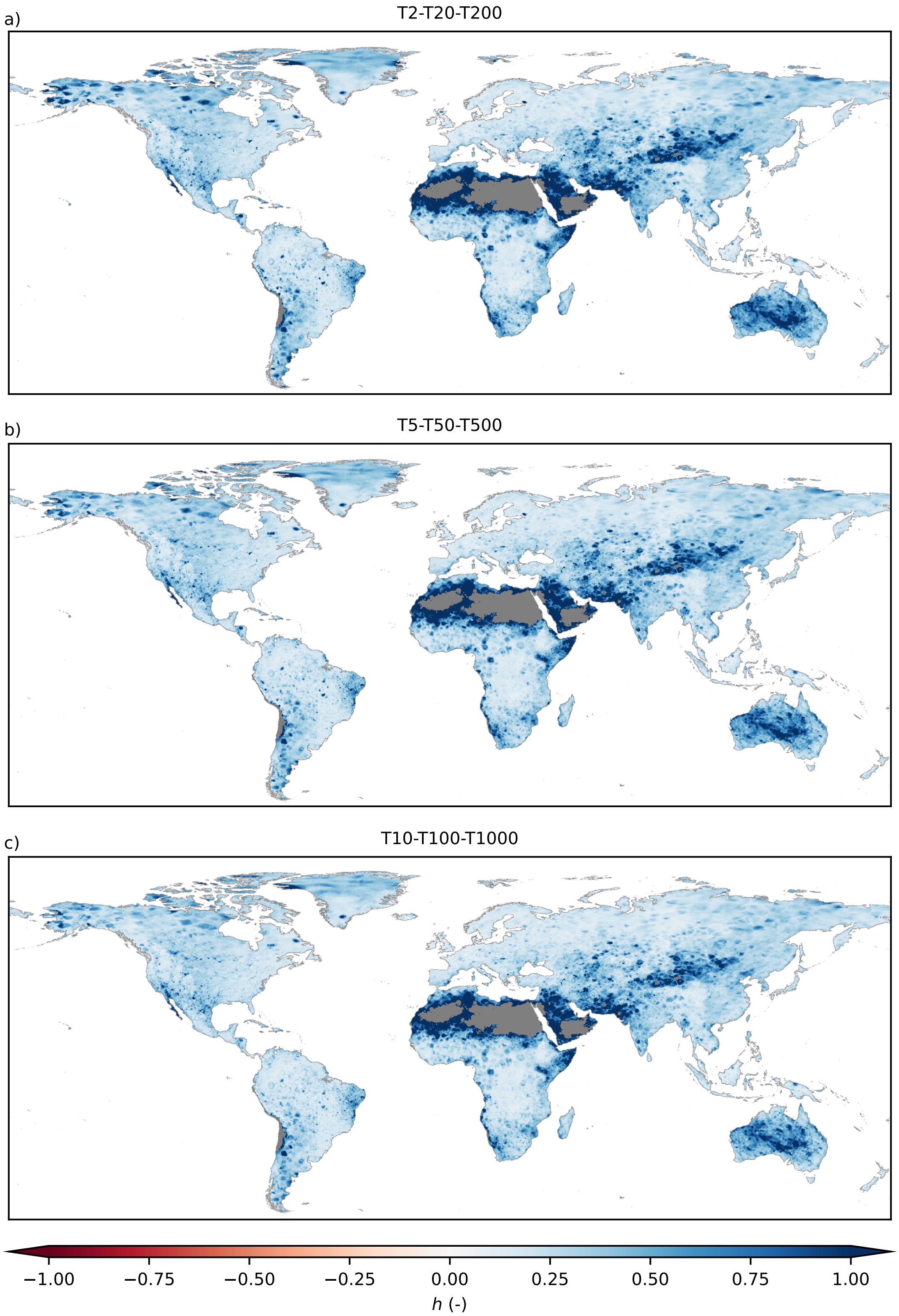


Figure S6: The heaviness amplification factor $h$ (Eq. 5) for daily precipitation calculated for different combinations of return levels for the MEV distribution: (a) $h_{\mathrm{T2-T20-T200}}$, (b) $h_{\mathrm{T5-T50-T500}}$, (c) $h_{\mathrm{T10-T100-T1000}}$. Red indicates a thin tail, white an exponential tail, and blue a heavy tail. See section 2.2.2 for more information on the heaviness metric.

# 5 Tail Behavior for Multiple Durations

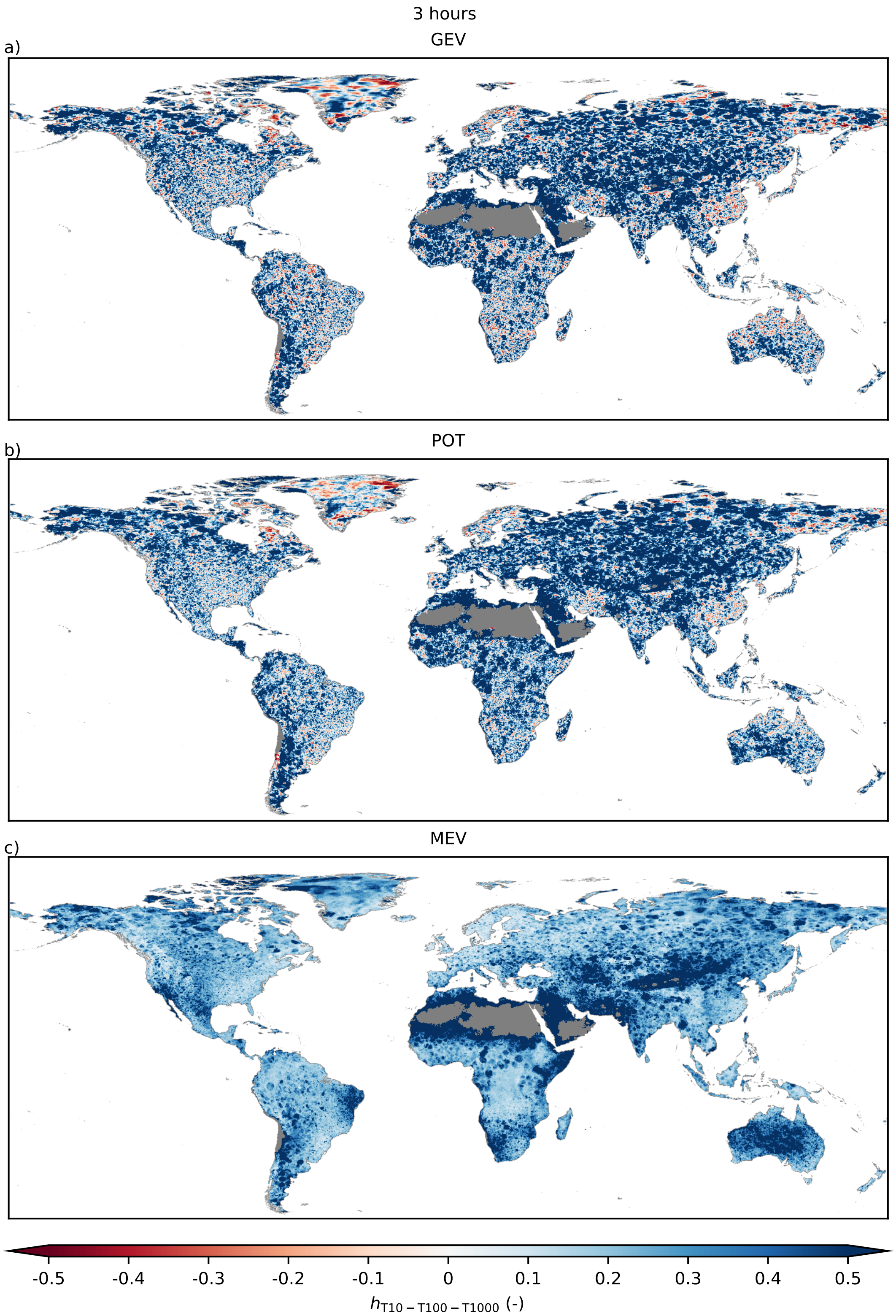


Figure S7: The heaviness amplification factor $h_{\mathrm{T10-T100-T1000}}$ (Eq. 5) for 3-hourly precipitation calculated for different extreme value methods: (a) GEV, (b) POT, (c) MEV. Red indicates a thin tail, white an exponential tail, and blue a heavy tail. See section 2.2.2 for more information on the heaviness metric.

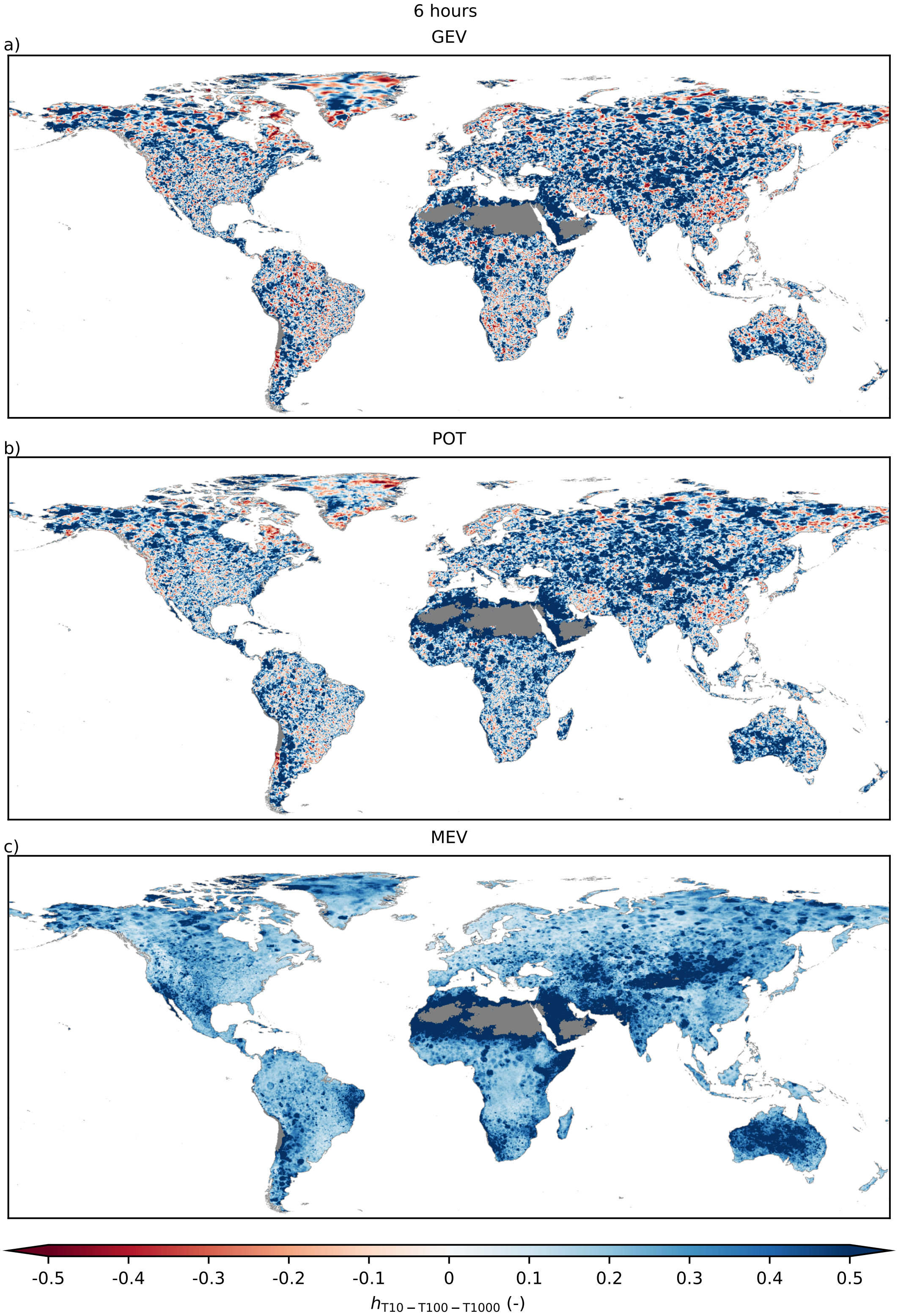


Figure S8: The heaviness amplification factor $h_{\mathrm{T10-T100-T1000}}$ (Eq. 5) for 6-hourly precipitation calculated for different extreme value methods: (a) GEV, (b) POT, (c) MEV. Red indicates a thin tail, white an exponential tail, and blue a heavy tail. See section 2.2.2 for more information on the heaviness metric.

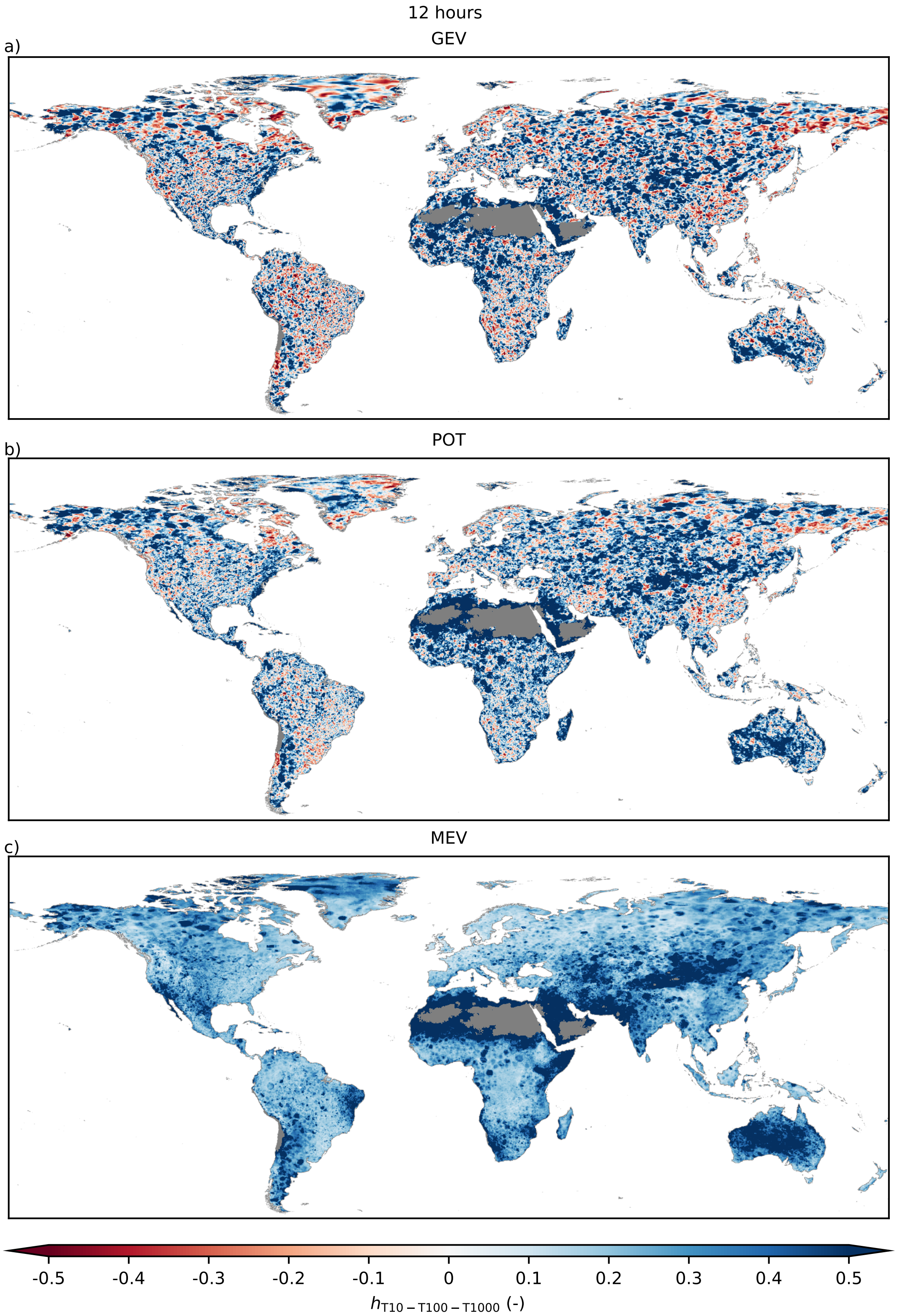


Figure S9: The heaviness amplification factor $h_{\mathrm{T10-T100-T1000}}$ (Eq. 5) for 12-hourly precipitation calculated for different extreme value methods: (a) GEV, (b) POT, (c) MEV. Red indicates a thin tail, white an exponential tail, and blue a heavy tail. See section 2.2.2 for more information on the heaviness metric.

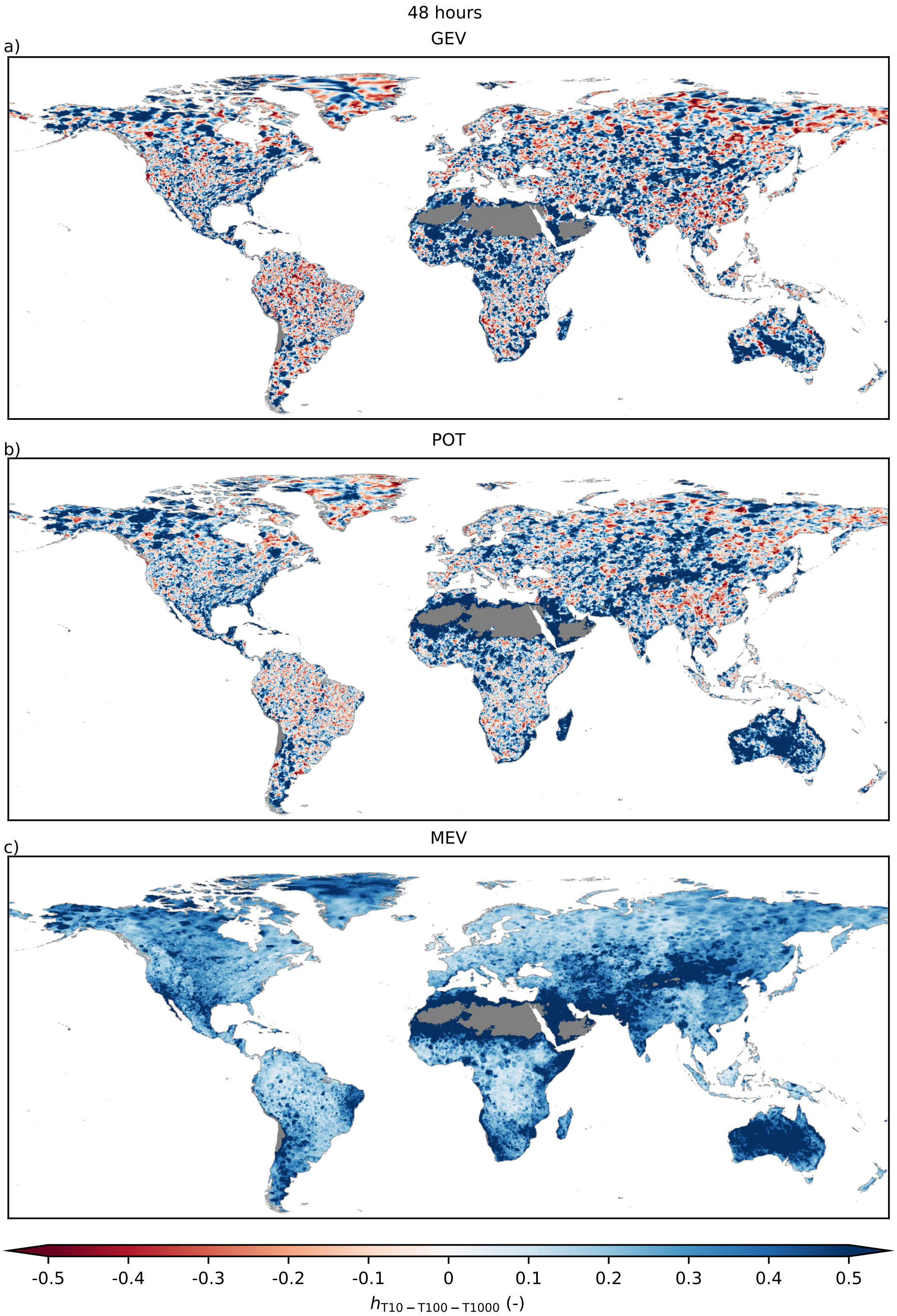


Figure S10: The heaviness amplification factor $h_{\mathrm{T10-T100-T1000}}$ (Eq. 5) for 2-day precipitation calculated for different extreme value methods: (a) GEV, (b) POT, (c) MEV. Red indicates a thin tail, white an exponential tail, and blue a heavy tail. See section 2.2.2 for more information on the heaviness metric.

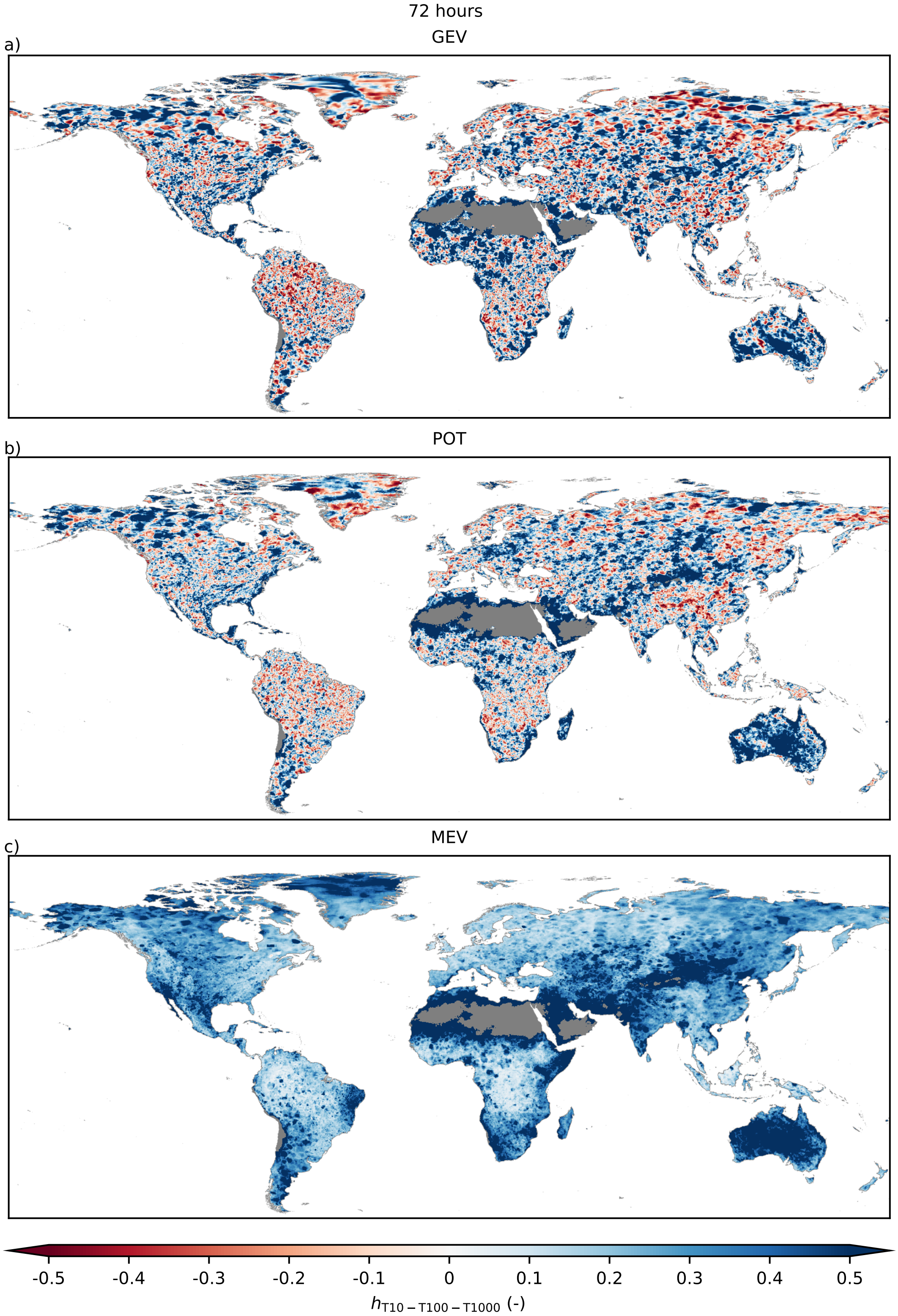


Figure S11: The heaviness amplification factor $h_{\mathrm{T10-T100-T1000}}$ (Eq. 5) for 3-day precipitation calculated for different extreme value methods: (a) GEV, (b) POT, (c) MEV. Red indicates a thin tail, white an exponential tail, and blue a heavy tail. See section 2.2.2 for more information on the heaviness metric.

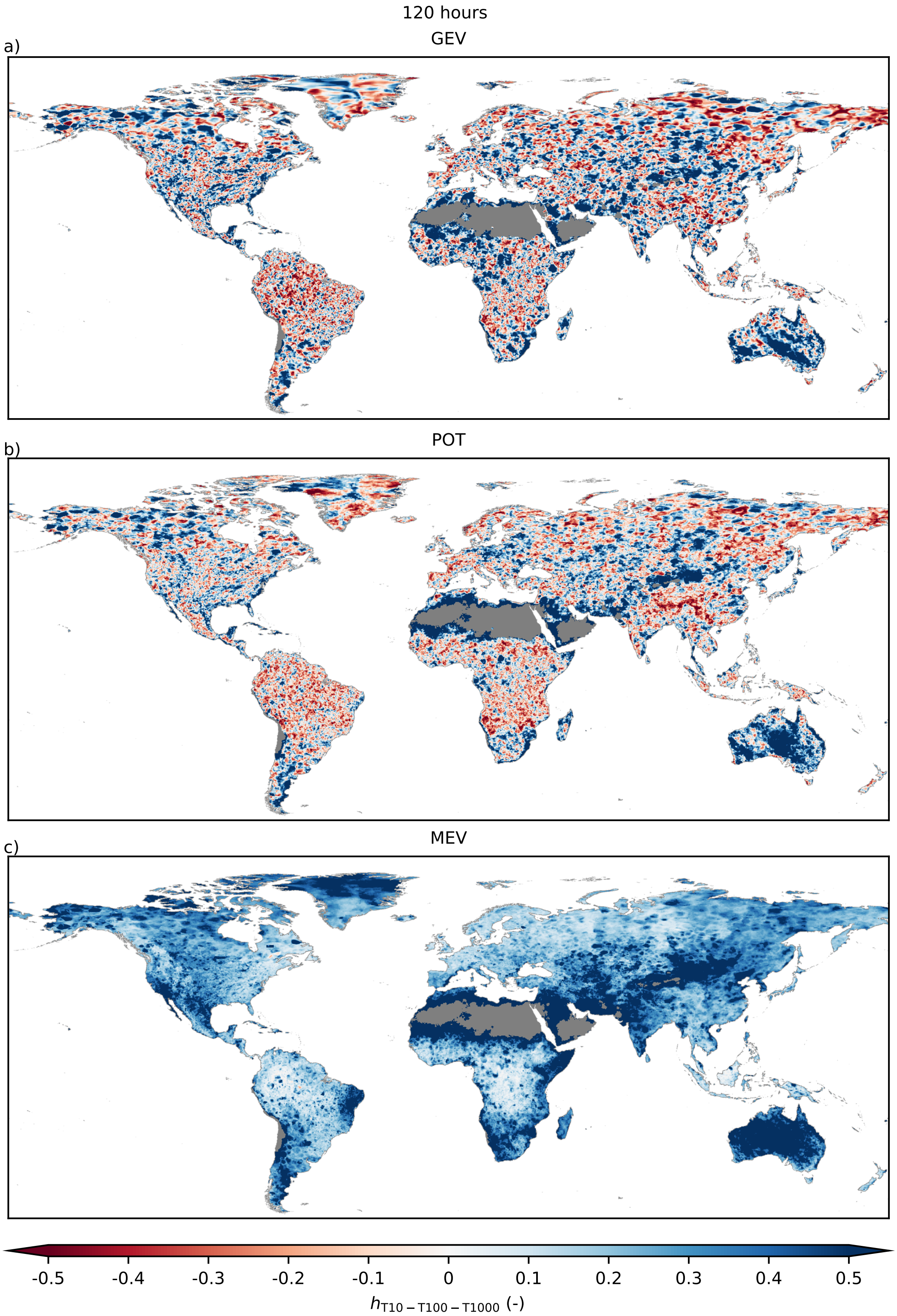


Figure S12: The heaviness amplification factor $h_{\mathrm{T10-T100-T1000}}$ (Eq. 5) for 5-day precipitation calculated for different extreme value methods: (a) GEV, (b) POT, (c) MEV. Red indicates a thin tail, white an exponential tail, and blue a heavy tail. See section 2.2.2 for more information on the heaviness metric.

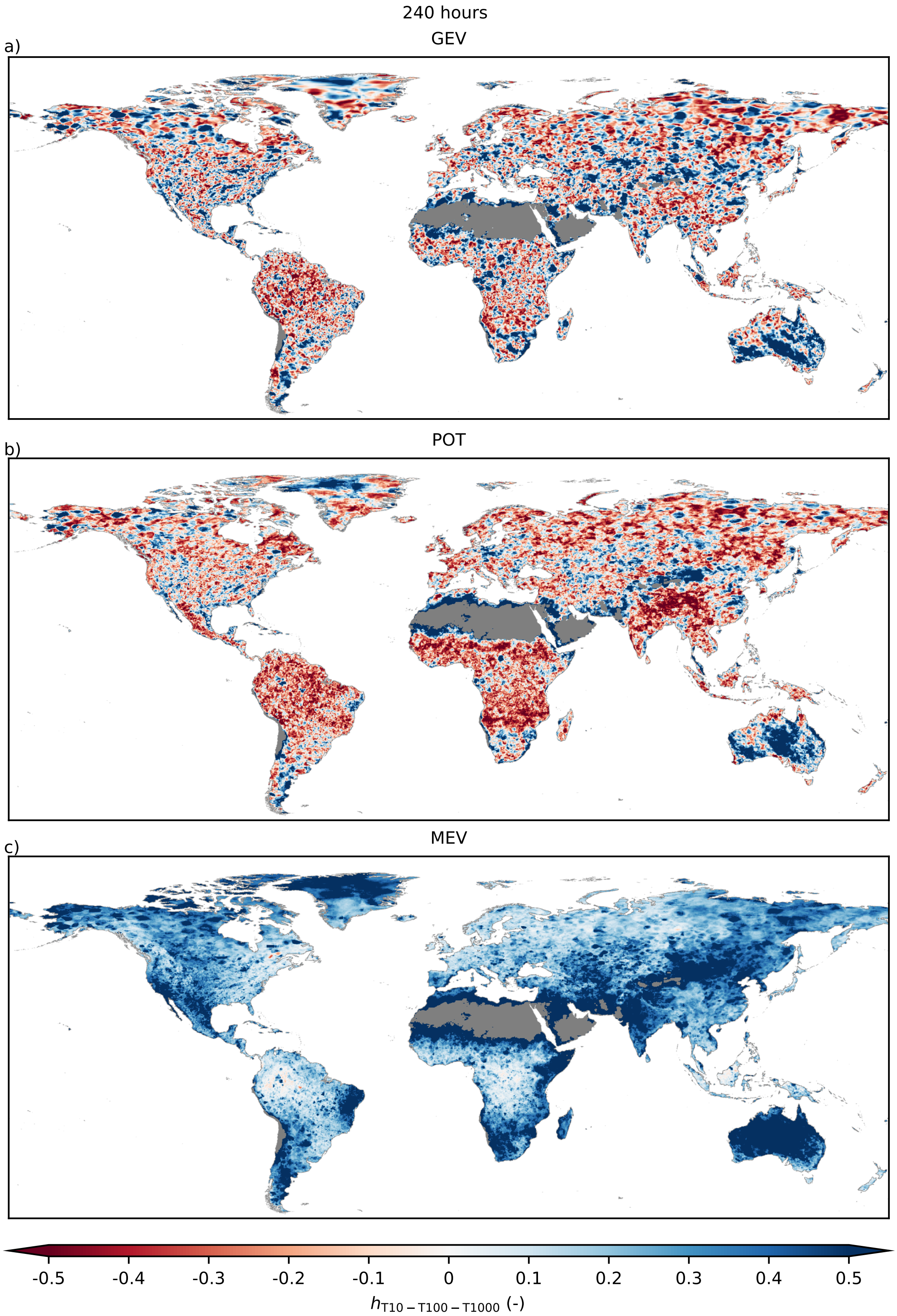


Figure S13: The heaviness amplification factor $h_{\mathrm{T10-T100-T1000}}$ (Eq. 5) for 10-day precipitation calculated for different extreme value methods: (a) GEV, (b) POT, (c) MEV. Red indicates a thin tail, white an exponential tail, and blue a heavy tail. See section 2.2.2 for more information on the heaviness metric.

## 6 Dataset Usage Notes

The GPEX dataset created in this study is available at the 4TU repository (Gründemann et al., 2021). It provides openly accessible and readily available hydrologically relevant return levels of extreme precipitation estimates worldwide. It contains the precipitation estimates of the three extreme values distributions, the observed estimates, the parameters of the three distributions, as well as a few other variables used in this study (Table S1). Furthermore, the extreme precipitation estimates and parameters of the Gumbel distribution are included in the dataset. In this section we provide some possible uses of the dataset, and instructions and disclaimers for proper use, both for large or regional-scale usage as well as for a single cell or point-scale.

Table S1: Variables included in the GPEX dataset.

| Variable | Description |
|---|---|
| GEV estimate | Extreme precipitation return levels estimated using GEV (mm) |
| POT estimate | Extreme precipitation return levels estimated using POT (mm) |
| MEV estimate | Extreme precipitation return levels estimated using MEV (mm) |
| Gumbel estimate | Extreme precipitation return levels estimated using Gumbel (mm) |
| Observed estimate | Observed extreme precipitation return levels (mm) |
| GEV location parameter | Location parameter of the GEV distribution |
| GEV scale parameter | Scale parameter of the GEV distribution |
| GEV shape parameter | Shape parameter of the GEV distribution |
| GEV heaviness | Heaviness amplification factor ($h_{\mathrm{T10-T100-T1000}}$) of the GEV distribution |
| POT location parameter | Location parameter for a GEV distribution estimated by fitting the GP distribution |
| POT scale parameter | Scale parameter for a GEV distribution estimated by fitting the GP distribution |
| POT shape parameter | Shape parameter for a GEV distribution estimated by fitting the GP distribution |
| POT heaviness | Heaviness amplification factor ($h_{\mathrm{T10-T100-T1000}}$) of the POT distribution |
| MEV number of events | $n$ parameter of the MEV distribution, number of events per hydrological year |
| MEV scale parameter | Scale parameter of the MEV distribution |
| MEV shape parameter | Shape parameter of the MEV distribution |
| MEV mean number of events | Mean of the $n$ parameter of the MEV distribution, mean number of events per hydrological year |
| MEV mean scale parameter | Mean of the scale parameter of the MEV distribution |
| MEV mean shape parameter | Mean of the shape parameter of the MEV distribution |
| MEV heaviness | Heaviness amplification factor ($h_{\mathrm{T10-T100-T1000}}$) of the MEV distribution |
| Gumbel scale parameter | Scale parameter of the Gumbel distribution |
| Gumbel shape parameter | Shape parameter of the Gumbel distribution |
| Annual maxima | Annual maximum precipitation for each hydrological year (mm) |
| Start hydrological year | Number indicating the month in which the hydrological year starts |
| Running parameter | Running parameter (hours) of the declustering method by Marra et al. (2018) |
| Land mask | Mask used for this study to indicate land cells and ocean cells |

### 6.1 Large-Scale Applications

The GPEX dataset contains global scale extreme precipitation estimates and its parameters at a spatial resolution of 0.1°, covering 3-hourly to 10-day durations. The dataset contains information about precipitation extremes for the entire Earth's land surface except Antarctica. The estimates of three distributions described in the main manuscript, as well as the return levels as observed, and estimated using the Gumbel distribution are included in the dataset. Among the three distributions, the traditional GEV and POT provide comparable large-scale average extremes, although differences can be substantial at smaller scales. When using the dataset at regional scales, we advise taking the average of the precipitation estimates, as neighboring cells could differ. Note that since only 38 years of data were available, the fitted model parameters and associated return values are subject to considerable uncertainty. Furthermore, we acknowledge that the use of just one dataset does not represent the true uncertainty in the generation of the dataset created. We do not think this affects our results for observed global spatial patterns significantly, but in a practical setting we recommend verifying the estimates with local observations if available, and to reproduce the precipitation return level estimates with a full uncertainty range estimation.

The novel MEV distribution provides more spatially coherent patterns of the extremes. Its mean shape parameter for daily events captures the (heavy-)tail behavior, and follows orographic patterns. The extremes estimated by MEV are higher than those estimated by GEV and POT. However, for large return periods and long durations, MEV can overestimate the extremes, due to the small number of events available for the fitting. We, therefore, recommend analyzing the extremes of all distributions in this dataset to obtain an indication of the uncertainty.

## 6.2 Small-Scale Applications

The dataset is also suitable for small-scale applications either in comparative studies or for direct use in data sparse regions, but one should be aware of the different statistical characteristics of point-scale and grid-scale. Due to averaging effects in gridded datasets, precipitation extremes of point-scale observations are higher (Cavanaugh & Gershunov, 2015; De Michele, Kottegoda, & Rosso, 2001; Ensor & Robeson, 2008; Hu et al., 2020; Sivapalan & Blöschl, 1998; Zorzetto & Marani, 2019). Illustrative examples of two locations, Vienna and San Francisco, are included in Figure S14. Analysis of the return level plots shows the estimates of the three distributions discussed in the main manuscript, as well as the estimates using the Gumbel distribution compared to the observed ones. We converted the annual maximum precipitation to 'observed' return levels (Figure S14e+j). It should be kept in mind though that these 'observed' return levels are also different from the 'true' return levels. For (sub-)daily durations and low return periods, there is generally a good agreement between the observed return levels and the estimates of the three EVDs. For longer durations and return periods, however, the estimated extremes deviate from the observed extremes. This is seen in San Fransisco (Figure S14f-j) where MEV overestimates and GEV, POT and Gumbel underestimate the extremes.

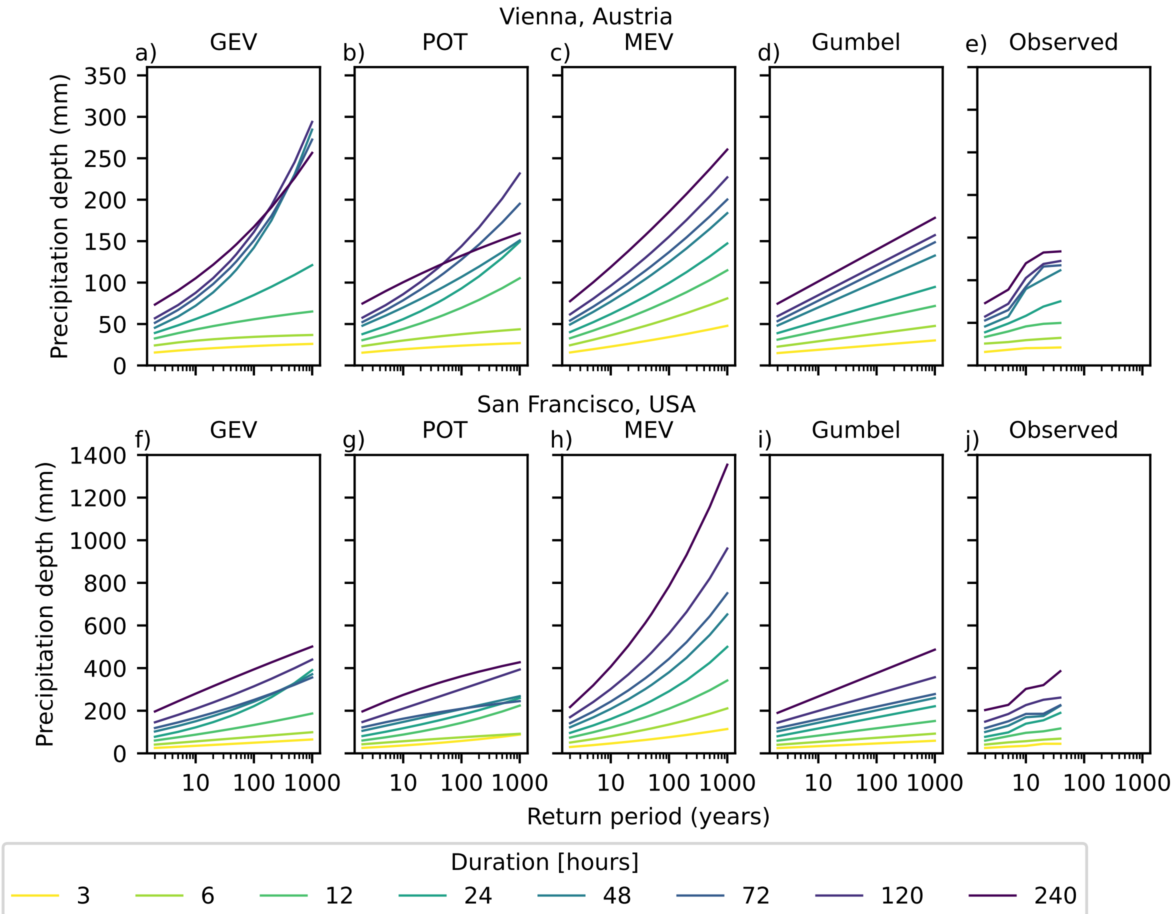


Figure S14: Return level plots for specific locations and different distributions. a–e) Vienna, Austria (48.234°N, 16.415°E) and f–j) San Francisco, California, USA (37.784°N, 122.400°W). Observed (e & j) are the annual maxima converted to return periods.

Furthermore, increasing event durations result in lower shape parameters (less heavy tails), which was seen for all three distributions discussed in the main manuscript. An implication of this is that for long-durations the shape parameter indicates a finite endpoint (GEV and POT), or a very thin tail (MEV), while heavier tails are generally observed for short-durations. When estimating very large return periods (e.g., $T500$), it is therefore possible for shorter duration estimates to be more intense than the corresponding quantiles computed for longer durations, which is physically impossible (see also Figure S14a,b,f and g). Additionally, Papalexiou and Koutsoyiannis (2013) argued based on long time series that the true population GEV shape parameter of daily precipitation lies between exponential and heavy-tail behavior, and thin-tails lead to an underestimation of the extremes. For applications where underestimation is undesirable, we have included the extreme precipitation estimates using the Gumbel distribution as well. The Gumbel distribution is an exponential distribution, and equal to the GEV distribution where the shape parameter equals zero. Therefore, if the shape parameter of GEV or POT is negative, the Gumbel estimates could be used instead in order to avoid underestimation.

To get a better understanding of the range and uncertainty of a single cell location, we recommend to look at return level plots of the four distributions at the cell of interest in combination with its neighboring cells. This is particularly important for GEV and POT, due to the absence of coherent spatial patterns and the erratic manifestation of the tail behaviors. Previous results (Zorzetto, Botter, & Marani, 2016) show that the benefits of MEV over GEV are greater for large return periods relative to the sample size available for the fit. Hence, for the estimation of large quantiles, MEV may be presumed to be more accurate. Depending on the practical application one could then choose to use the most extreme value, use the MEV value, use the Gumbel value in case of a negative shape for GEV or POT, or use a spatial average of the GEV and POT estimates.

## References


Caeiro, F., & Gomes, M. I. (2016). Threshold selection in extreme value analysis. *Extreme value modeling and risk analysis: Methods and applications*, *1*, 69–86.

Cavanaugh, N. R., & Gershunov, A. (2015). Probabilistic tail dependence of intense precipitation on spatiotemporal scale in observations, reanalyses, and GCMs. *Climate Dynamics*, *45*(11-12), 2965–2975. doi: 10.1007/s00382-015-2517-1

De Michele, C., Kottegoda, N. T., & Rosso, R. (2001). The derivation of areal reduction factor of storm rainfall from its scaling properties. *Water Resources Research*, *37*(12), 3247–3252. doi: 10.1029/2001WR000346

Ensor, L. A., & Robeson, S. M. (2008). Statistical characteristics of daily precipitation: comparisons of gridded and point datasets. *Journal of Applied Meteorology and Climatology*, *47*(9), 2468–2476. doi: 10.1175/2008JAMC1757.1

Gründemann, G. J., Zorzetto, E., Beck, H. E., Schleiss, M., van de Giesen, N., Marani, M., & van der Ent, R. J. (2021). *Global Precipitation EXtremes dataset.* doi: 10.4121/uuid:12b5c941-cd54-45db-8d7b-fefaacecaa69

Hu, L., Nikolopoulos, E. I., Marra, F., Morin, E., Marani, M., & Anagnostou, E. N. (2020). Evaluation of MEVD-based precipitation frequency analyses from quasi- global precipitation datasets against dense rain gauge networks. *Journal of Hydrology*, *590*, 125564. doi: 10.1016/j.jhydrol.2020.125564

Langousis, A., Mamalakis, A., Puliga, M., & Deidda, R. (2016). Threshold detection for the generalized pareto distribution: Review of representative methods and application to the noaa ncdc daily rainfall database. *Water Resources Research*, *52*(4), 2659–2681.

Papalexiou, S. M., & Koutsoyiannis, D. (2013). Battle of extreme value distributions: A global survey on extreme daily rainfall. *Water Resources Research*, *49*, 187–201. doi: 10.1029/2012WR012557

Serinaldi, F., & Kilsby, C. G. (2014). Rainfall extremes: Toward reconciliation after the battle of distributions. *Water Resources Research*, *50*, 336–352. doi: 10.1002/2013WR014211

Sivapalan, M., & Blöschl, G. (1998). Transformation of point rainfall to areal rainfall: Intensity-duration-frequency curves. *Journal of Hydrology*, *204*(1-4), 150–167.

Zorzetto, E., Botter, G., & Marani, M. (2016). On the emergence of rainfall extremes from ordinary events. *Geophysical Research Letters*, *43*(15), 8076–8082. doi: 10.1002/2016GL069445

Zorzetto, E., & Marani, M. (2019). Downscaling of rainfall extremes from satellite observations. *Water Resources Research*, *55*, 156–174. doi: 10.1029/2018WR022950