Inferring heavy tails of flood distributions from common discharge dynamics

Hsing-Jui Wang[1], Ralf Merz[1,2], Soohyun Yang[3,4], and Stefano Basso[1,5]

[1]Department of Catchment Hydrology, Helmholtz Centre for Environmental Research – UFZ, Halle (Saale), 06120, Germany,
[2]Institute of Geosciences and Geography, Martin-Luther University Halle-Wittenberg, Halle (Saale), 06120, Germany,
[3]Department of Aquatic Ecosystem Analysis, Helmholtz Centre for Environmental Research – UFZ, Magdeburg, 39114, Germany
[4]Department of Civil and Environmental Engineering, Seoul National University, Seoul, 08826, Republic of Korea
[5]Norwegian Institute for Water Research (NIVA), Oslo, 0579, Norway

Correspondence to: Hsing-Jui Wang (hsing-jui.wang@ufz.de)

Abstract. Floods are often disastrous due to underestimation of the magnitude of rare events. Underestimation commonly happens when the occurrence of floods follow a heavy-tailed distribution, but this behavior is not recognized and thus neglected for flood hazard assessment. In fact, identifying heavy-tailed flood behavior is challenging because of limited data records and the lack of physical support for currently used indices. We address these issues by deriving a new index of heavy-tailed flood behavior from a physically-based description of streamflow dynamics. The proposed index, which is embodied by the hydrograph recession exponent, enables inferring heavy-tailed flood behavior from daily flow records, even of short length. We test the index in a large set of case studies across Germany encompassing a variety of climatic and physiographic settings. Our findings demonstrate that the new index enables reliable identification of cases with either heavy or nonheavy tailed flood behavior from daily flow records. Additionally, the index suitably estimates the severity of tail heaviness and ranks it across cases, achieving robust results even with short data records. The new index addresses the main limitations of currently used metrics, which lack physical support and require long data records to correctly identify tail behaviors, and provides valuable information on the tail behavior of flood distributions and the related flood hazard in river basins using commonly available discharge data.

1 Introduction

Floods remain the leading natural hazards worldwide, which directly threaten the livelihoods of at least one-fifth of the world's population (McDermott, 2022; Rentschler et al., 2022) and have caused enormous economic losses (Bevere and Remondi, 2022). To assess the flood hazards of extreme events, flood frequency analysis is the central and commonly used tool, which is usually achieved by parametrically fitting a selected probability distribution on flow maxima, e.g., the annual maximum flood (Villarini and Smith, 2010), or peak-over-threshold series (Pan et al., 2022). Selecting a suitable distribution that can properly describe (or predict) the extreme events is, however, often challenging due to the notable uncertainties caused by the lack of data in the maxima approach (Papalexiou and Koutsoyiannis, 2013; Hu et al., 2023). The upper-tailed behavior (which

we will refer to as 'tail behavior' throughout the paper for simplicity) of the underlying distribution critically determines the accuracy of the extreme events. If a catchment has the potential for heavy-tailed flood behavior but this characteristic is not accounted for in the selection of probability distributions, the probability of extreme floods may be significantly

35 underestimated (Merz et al., 2022). This can lead to disastrous floods and severe damages (Merz et al., 2021). Therefore, correctly identifying the tail behavior of flood distributions is crucial for avoiding potential underestimation of extreme floods. The tail heaviness of an empirical distribution is typically estimated through graphical or statistical methods, although both methods have limitations. Graphical methods, such as log-log plots (Beirlant et al., 2004), generalized Hill ratio plots (Resnick, 2007; El Adlouni et al., 2008), and mean excess functions (Embrechts et al., 1997; Nerantzaki and Papalexiou, 2019), are less

40 objective and efficient for large-scale analyses (Cooke et al., 2014). In contrast, statistical methods, such as parametric metrics that fit distributions to the observed data (Papalexiou et al., 2013; Seckin et al., 2011; Smith et al., 2018; Villarini and Smith, 2010) and non-parametric metrics like the upper tail ratio (Lu et al., 2017; Smith et al., 2018; Villarini et al., 2011; Wang et al., 2022), Gini index (Eliazar and Sokolov, 2010; Rajah et al., 2014), and obesity index (Cooke and Nieboer, 2011; Sartori and Schiavo, 2015), provide a more objective insight into tail behavior. However, obtaining reliable estimates from these

45 methods requires long data records (Papalexiou and Koutsoyiannis, 2013), which can be challenging globally (Lins, 2008) and may cause bias when comparing data across sites with different record lengths (Cunderlik and Burn, 2002; Wietzke et al., 2020). To reduce uncertainty, especially in estimating extremes, it is recommended to analyze ordinary dynamics instead of focusing solely on maximum values (Marani and Ignaccolo, 2015; Mushtaq et al., 2022), and to investigate the underlying factors that contribute to extreme events (Wilson and Toumi, 2005; Tarasova et al., 2020; Merz et al., 2022).

50 Floods are conventionally thought to be triggered by rainfall, and numerous studies have contributed to an improved understanding of rainfall extremes (e.g., Koutsoyiannis, 2004a,b; Martinez-Villalobos and Neelin, 2021; Koutsoyiannis, 2022). However, several studies have clarified that rainfall extremes do not necessarily translate into flood extremes (e.g., McCuen and Smith, 2008; Pall et al., 2011; Hall et al., 2014; Archfield et al., 2016; Rossi et al., 2016; Zhang et al., 2016; Hodgkins et al., 2017; Sharma et al., 2018). For instance, McCuen and Smith (2008) showed that skewed rainfall distributions do not always produce skewed flood distributions. They proposed that catchment responses and storage dynamics contribute to the generation of flood extremes. This view was supported by Sharma et al. (2018), who argued that despite a significant increase in rainfall extremes, a corresponding increase in flood extremes was not observed. The thorough review of Merz et al. (2022) concluded that while rainfall plays a primary role in generating runoff, the emergence of flood extremes is largely determined by catchment responses and water balance. Given these premises, an appropriate approach for describing runoff and its extremes

60 should be rooted in the dynamics of soil moisture and rainfall-runoff processes within catchments.

This study aims to investigate whether a suitable descriptor of the tail behavior of flood distributions exists by exploring the intrinsic hydrological dynamics of the flow regime. Currently, widely-used metrics for tail behavior estimation of flood distributions do not incorporate such a physical description, to the best of our knowledge. Using this descriptor as a proxy for estimating heavy-tailed flood behavior, rather than relying solely on statistical analysis of extreme events, we aim to bridge

65 this gap and improve the accuracy and reliability of tail behavior estimation for flood distributions. We begin the analysis with

a mechanistic description of hydrological processes. We subsequently distinguish between the key processes generating heavy and nonheavy tailed behavior of flood distributions and propose a physical descriptor for heavy-tailed flood behavior which is based on common streamflow dynamics. We verify its ability to identify heavy-tailed flood behavior and its robustness in datasets with decreasing lengths through numerous case studies across Germany, encompassing various climate and physiographic characteristics. This confirms the practical transferability and stability of the descriptor.

2 Identifying tail behavior from hydrological dynamics

We describe key hydrologic processes occurring at the catchment scale and the resulting probability distributions of streamflow and floods by means of the PHysically-based Extreme Value (PHEV) distribution of river flows (Basso et al., 2021). This framework is grounded on a well-established mathematical description of precipitation, soil moisture, and runoff generation in river basins (Laio et al., 2001; Porporato et al., 2004; Botter et al., 2007b, 2009). Rainfall is described as a marked Poisson process with frequency $\lambda_p [\text{T}^{-1}]$ and exponentially distributed depths with average α [L]. Soil moisture increases due to rainfall infiltration and decreases due to evapotranspiration. The latter is represented by a linear function of soil moisture between the wilting point and an upper critical value expressing the water holding capacity of the root zone. Runoff pulses occur with frequency $\lambda < \lambda_p$ when the soil moisture exceeds the critical value. These pulses replenish a single catchment storage, which drains according to a nonlinear storage-discharge relation. The related hydrograph recession is described via a power law function with exponent a $[-]$ and coefficient K $[\text{L}^{1-a}/\text{T}^{2-a}]$ (Brutsaert and Nieber, 1977), which allows for mimicking the joint effect of different flow components (Basso et al., 2015). Such a description of runoff generation and streamflow dynamics was successfully tested in a variety of hydro-climatic and physiographic conditions (Arai et al., 2020; Botter et al., 2007a; Botter et al., 2010; Ceola et al., 2010; Doulatyari et al., 2015; Mejía et al., 2014; Müller et al., 2014; Müller et al., 2021; Pumo et al., 2014; Santos et al., 2018; Schaefli et al., 2013).

PHEV provides a set of consistent expressions (Basso et al., 2021) for the probability distributions of daily streamflow, ordinary peak flows (i.e., local flow peaks occurring as a result of streamflow-producing rainfall events; *sensu* Miniussi et al., 2020), and floods (i.e., flow maxima in a certain timeframe; Basso et al., 2016). The probability distribution of daily streamflow q can be expressed as (Botter et al., 2009):

$$p(q) = C_1 \cdot q^{-a} \left(e^{\frac{-1}{\alpha K(2-a)} \cdot q^{2-a}} \right) \left(e^{\frac{\lambda}{K(1-a)} \cdot q^{1-a}} \right), \tag{1}$$

where C_1 is a normalization constant. The probability distribution of ordinary peak flows and flow maxima can be expressed as $p_j(q)$ and $p_M(q)$, respectively (Basso et al., 2016):

$$p_j(q) = C_2 \cdot q^{1-a} \cdot e^{-\frac{q^{2-a}}{\alpha K(2-a)}} \cdot e^{\frac{q^{1-a}}{K(1-a)}}, \tag{2}$$

$$p_M(q) = p_j(q) \cdot \lambda\tau \cdot e^{-\lambda\tau \cdot D_j(q)}, \tag{3}$$

where $D_j(q) = \int_q^\infty p_j(q)\,dq$, $\tau[day]$ is the duration of the considered time frame, C_2 is a normalization constant.

Notably, the mathematical expression of flow distributions provided by the PHEV framework are composed of a power law and two stretched exponential distributions, although it's important to note that PHEV doesn't assume a specific probability distribution for streamflow representation. The use of stretched exponential distributions introduces greater flexibility in capturing tail behavior compared to the exponential distribution. Depending on its parameter values, the stretched exponential distribution can display either light-tailed or heavy-tailed behavior, whereas the exponential distribution consistently exhibits a light-tailed behavior. In fact, recent studies (Basso et al., 2016; 2021; 2023) have substantiated and documented PHEV's efficacy in representing high flow behaviors.

Taking the limit of Eq. (1) for $q \to +\infty$ provides indications on the tail behavior of the flow distribution (Basso et al., 2015). This is determined by the three terms in the equation, namely, one power law and two exponential functions, which behave differently depending on the value of the hydrograph recession exponent a (Eq. (4); notice that $a > 1$ in most river basins; Biswal and Kumar, 2014; Tashie et al., 2020b).

$$\lim_{q\to+\infty} p(q) = \lim_{q\to+\infty} \left\{ C_1 \cdot q^{-a} \left(e^{\frac{-1}{\alpha K(2-a)} \cdot q^{2-a}} \right) \left(e^{\frac{\lambda}{K(1-a)} \cdot q^{1-a}} \right) \right\}, \tag{4}$$

with annotations: over the terms, $\mapsto 0$, $\mapsto 0$, $\mapsto e^0 = 1$ for $1 < a < 2$; under the terms, $\mapsto 0$, $\mapsto e^0 = 1$, $\mapsto e^0 = 1$ for $a > 2$.

When $1 < a < 2$, the last term on the right-hand side converges to a constant value of one as q increases, thereby no more influencing how the distribution decreases toward zero. The first two terms instead decrease toward zero, affecting how the probability decreases for increasing values of q. The tail behavior is in this case determined by both a power law and a stretched exponential function, indicating that the probability decreases faster than a stretched exponential but slower than a power law. When $a > 2$, both the stretched exponential terms converge to a constant value of one as q increases, and thus no more influence how the probability decreases toward zero. In this case the tail of the distribution is solely determined by the power law function. Despite being aware that several definitions of heavy-tailed distribution exist (El Adlouni et al., 2008; Vázquez et al., 2006), in the remainder of the manuscript we refer to distributions which exhibit a power law tail as heavy-tailed.

From the above derivations, the hydrograph recession exponent emerges as a key index of the tail behavior of streamflow distributions, which shall be heavy-tailed for values of $a > 2$. We apply the same analyses to infer the tail behavior of the probability distributions of ordinary peak flows and floods by taking the limit of $q \to +\infty$ for both Eq. (2) and (3). Because $\lim_{q\to\infty} D_j(q) = \int_\infty^\infty p_j(q)\,dq = 0$, the Eq. (2) and (3) can be transformed into: (set $C_3 = \lambda\tau C_2$)

$$\lim_{q\to+\infty} p_j(q) = \lim_{q\to+\infty} \left\{ C_2 \cdot \overbrace{q^{1-a}}^{\mapsto 0} \left(\overbrace{e^{\frac{-1}{\alpha K(2-a)} \cdot q^{2-a}}}^{\mapsto 0} \right) \right\}, \quad \begin{array}{l} \text{for } 1 < a < 2 \\ \\ \text{for } a > 2 \end{array}$$

$$\underbrace{\phantom{q^{1-a}}}_{\mapsto 0} \quad \underbrace{\phantom{e^{\frac{-1}{\alpha K}}}}_{\mapsto e^0 = 1}$$

(5)

$$\lim_{q\to+\infty} p_M(q) = \lim_{q\to+\infty} \left\{ C_3 \cdot \overbrace{q^{1-a}}^{\mapsto 0} \left(\overbrace{e^{\frac{-1}{\alpha K(2-a)} \cdot q^{2-a}}}^{\mapsto 0} \right) \right\}, \quad \begin{array}{l} \text{for } 1 < a < 2 \\ \\ \text{for } a > 2 \end{array}$$

$$\underbrace{\phantom{q^{1-a}}}_{\mapsto 0} \quad \underbrace{\phantom{e^{\frac{-1}{\alpha K}}}}_{\mapsto e^0 = 1}$$

(6)

Notably, we observe that the same critical value of the recession exponent equal to 2 separate the absence and presence of heavy-tailed behavior also in these cases. Therefore, we propose the hydrograph recession exponent a as a suitable indicator of heavy-tailed flood behavior, based on the description of hydrological processes embedded in the physically-based extreme value model. We test its capability to correctly predict such behavior in Sec. 4, and discuss the results in Sec. 5.

3 Data and parameter estimation

To test the proposed index of heavy-tailed flood behavior (i.e., the hydrograph recession exponent a), we use daily streamflow records of 98 gauges across Germany (Appx. B). The analyzed river basins encompass a variety of climate and physiographic settings (Tarasova et al., 2020). Their areas range from 110 to 23,843 km^2 with a median value of 1,195 km^2. The length of the streamflow records range from 35 to 63 years with a median value of 58 years (inbetween 1951 – 2013).. We perform all analyses on a seasonal basis (winter: December–February, spring: March–May, summer: June–August, fall: September–November) to account for the seasonality of the hydrograph recessions (Tashie et al., 2020b) and flood distributions (Durrans et al., 2003). We term the analysis of a given river gauge during a season a case study. We select gauges for which processes driving streamflow dynamics are reasonably consistent with the adopted theoretical framework. Hence, we discard gauges affected by large dams, reservoirs (Lehner et al., 2011) and anthropogenic flow disturbances (based on visual examination; Tarasova et al., 2018). Case studies with strong snowfall (during a season), for which the average daily temperature is below zero degrees during precipitation events for over 50% of a season, are also discarded (i.e., only the affected season is removed from the analyses). This results in an overall number of 386 case studies, including 97 case studies in spring, 96 in summer, 98 in autumn and 95 in the winter season.

The proposed index is derived from hydrograph recession analysis. The hydrograph recession is typically described by a power law relationship between the rate of change of streamflow in time, dq/dt, and the magnitude of streamflow q (Brutsaert and

Nieber, 1977). Recent studies have suggested estimating this power law relationship for individual recession events rather than aggregating them, enhancing the representation of observed recession behavior (Biswal and Marani, 2010; Basso et al., 2015; Karlsen et al., 2019; Jachens et al., 2020; Tashie et al., 2020a; Biswal, 2021). In line with these studies, we calculate the recession exponent for each individual event and then take the median exponent across all events as the representative value for a given case study. In particular, a power law is used to represent hydrograph recessions of a single event i, $dq/dt = -K_i \cdot q^{a_i}$, where t denotes the unit time, K_i and a_i denote the estimated coefficient and exponent of hydrograph recessions for event i, respectively. The median value of all the a_i is the estimated value of a considered in this study and here used to represent the average nonlinearity of catchment response. Hydrograph recessions are composed of ordinary peak flows and the following streamflow values decreasing for a minimum duration of five days. The proposed index of heavy-tailed flood behavior can thus be estimated based on commonly available daily discharge observations.

To validate the identification of tail behavior obtained by means of the proposed index, we benchmark it against data by fitting a power law distribution to the empirical data distribution. A case study is considered to be heavy-tailed according to the observations if the fitted power law reliably describe the tail behavior of the data distribution. This is evaluated by means of a state-of-the-art method proposed by Clauset et al. (2009). The exponent b of the empirical power law is first computed by fitting a power law to the upper tail of the data distribution. An optimized lower boundary is determined by considering the best fit according to the Kolmogorov-Smirnov (KS) statistic, one of the most common measures of the distance between two non-normal distributions. The method then assesses whether the fitted power law reliably represents the observed data by using statistical tests, such as the Kolmogorov-Smirnov statistic and a Monte Carlo procedure, to verify that the residual errors between the data and the power law distribution fall within the range of fluctuations expected from random sampling. If the residual errors are found to be within the range of fluctuations expected from random sampling, the power law is deemed a reliable representation of the empirical data distribution (Appx. A). We use the python package plfit 1.0.3 to implement these computations and refer to Clauset et al. (2009) for further details concerning the approach.

We analyze three types of empirical data, namely daily streamflow, ordinary peaks, and monthly maxima , and obtain estimates of the fitted exponent b for each case. We use these results to validate the capabilities of the proposed index to infer heavy-tailed flood behavior from common hydrological dynamics, i.e., from the analysis of hydrograph recessions. We acknowledge that the benchmark we use, i.e., the empirical power law, may be influenced by fitting uncertainty due to data scarcity in some cases (i.e., especially when we analyse maxima; we indeed considered monthly maxima (Fischer and Schumann, 2016; Malamud and Turcotte, 2006) instead of the seasonal maxima previously used in the literature (e.g., Basso et al., 2021) to extend the sample size). The parallel analyses for cases with larger sample size (i.e., daily streamflow and ordinary peaks) provide more robust validation and support the interpretation of results for maxima. The topic is further discussed in Secs. 4 and 5.

4 Results

We examine if power law distributions fitted to the empirical distributions of daily streamflow, ordinary peaks, and monthly maxima well describe the observed data for case studies identified as having heavy-tailed behavior (i.e., a > 2) according to the proposed index. First, we identify the case studies with either heavy- (a > 2) or nonheavy (a < 2) -tailed behavior based on the proposed index. Then, we utilize the KS statistic κ to measure the distance between the frequency distributions of observations and a power law distribution (specifically, on the tail of the distribution). This assessment gauges the effectiveness of the fitted power law distribution in characterizing the dataset (with $\kappa \in [0,\infty]$, where $\kappa=0$ represents the utmost reliability). The KS test is a common nonparametric method suitable for non-normal distributions. Low values of the KS statistic κ indicate that the empirical data are likely to be drawn from a power law. Figs. 1a-1c show that the histograms of the number of case studies for decreasing values of the KS statistic are significantly skewed (i.e., the skewness is significantly different from zero) toward lower values of κ for all cases of daily streamflows, ordinary peak flows, and monthly flow maxima with a > 2 (red histograms), whereas this is not true for cases with a < 2 (green histograms) (i.e., the skewness is not significantly different from zero in these cases). Statistical significance of the skewness was evaluated through the Jarque–Bera test at a significance level of 0.05. The result indicates that data from case studies which are identified with heavy-tailed behavior according to the proposed index (a>2, red) are indeed more likely to come from power law distributions.

We further estimate the accuracy of the proposed index based on the fraction of case studies that are identified as heavy-tailed by the proposed index among all cases that are heavy-tailed according to the available observations. To define the latter, we set a threshold value of κ: the power law is a reliable representation of the data for cases with κ below the threshold. Mathematically, the accuracy can be expressed as $P(a > 2 | \kappa < \kappa_r) = N_c(a > 2 | \kappa < \kappa_r) / N_c(\kappa < \kappa_r)$, where κ_r is the imposed threshold of κ, $N_c(\kappa < \kappa_r)$ is the number of case studies whose $\kappa < \kappa_r$, and $N_c(a > 2 | \kappa < \kappa_r)$ is the number of case studies with $a > 2$ among the $N_c(\kappa < \kappa_r)$ case studies. Higher accuracy essentially means that a higher fraction of heavy-tailed cases is correctly identified by means of the proposed index. To achieve this, we systematically reduce the threshold of KS statistic κ_r (imposing a more stringent criterion for incorporating cases in the computation of conditional probability of accuracy) along the x-axis in Fig. 1, progressing from left to right. It's important to note that as the κ_r threshold becomes smaller, the reliability of describing the data using power law distributions increases (as denoted by the second axis legend of Fig.1).

Figs. 1d-1f display the accuracy of the proposed index as a function of the reliability threshold κ_r. In all three cases (daily streamflows, ordinary peak flows, and monthly flow maxima), the accuracy values increase with the reliability level of the power law distribution fitted on observed data. This means that the proposed index shows high accuracy for case studies where the empirical distributions of observed data are more consistent with power laws. In other words the proposed index, which is

205 estimated from common streamflow dynamics as the hydrograph recession exponent, accurately identifies heavy-tailed behavior of streamflow and flood distributions displayed by the available observations.

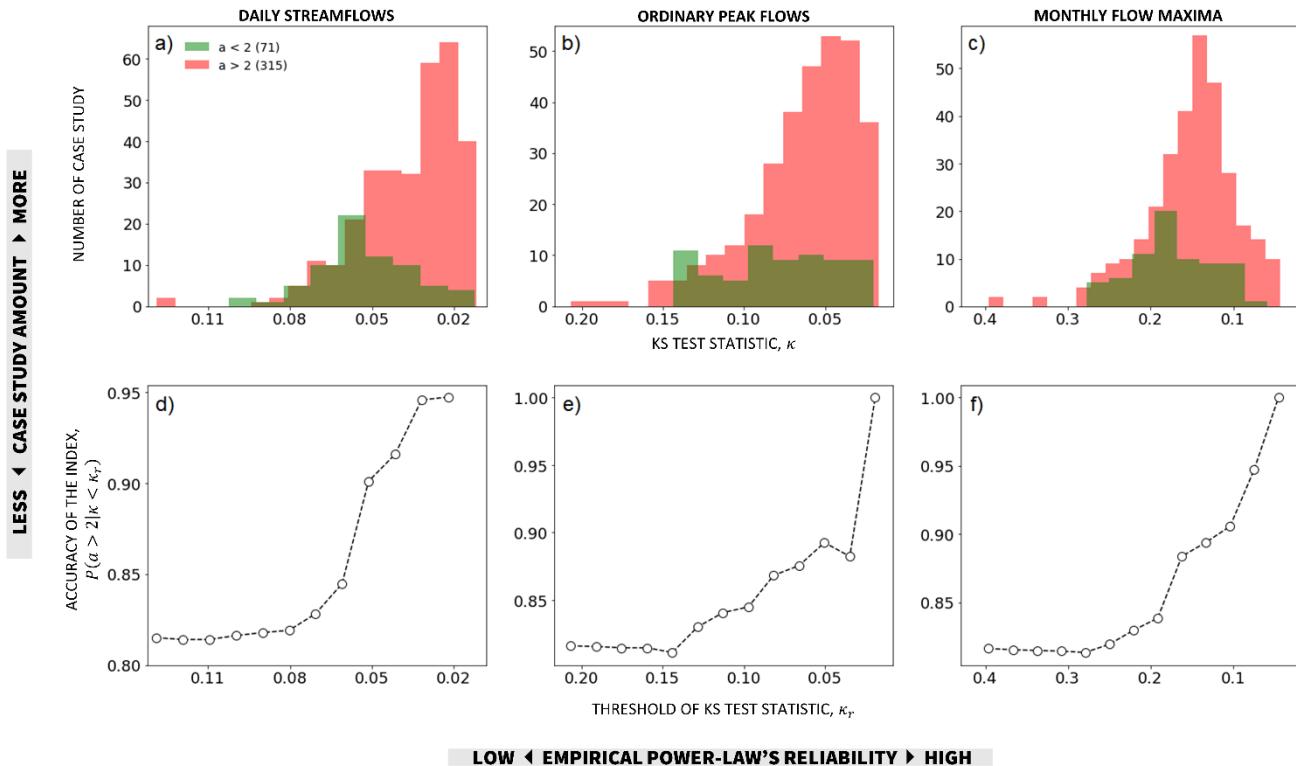



Figure 1. Accuracy of the proposed index. (a)-(c) Number of analyzed case studies as a function of the KS statistic κ of empirically fitted power law distributions (the latter is a measure of how reliable the power law is as a model for the given data: the lower κ, the more reliable the power law model). Case studies are identified as having either heavy- ($a > 2$, red histograms) or nonheavy ($a < 2$, green histograms) – tailed behavior based on the hydrograph recession exponent a estimated from daily flow records, which is proposed as an index of heavy-tailed streamflow and flood behavior. (d)-(f) Accuracy of the proposed index as a function of decreasing thresholds of κ_r (i.e., increasing reliability of empirical power laws). The values of the KS statistic κ are derived from records of (a, d) daily streamflows, (b, e) ordinary peak flows, and (c, f) monthly flow maxima.

215 We further employ the goodness-of-fit testing procedure proposed by Clauset et al. (2009) (Appx. A) to identify case studies for which the representation of daily streamflow, ordinary peak flows, and monthly maxima by means of power law distributions is convincingly supported by the available data. We refer to these case studies as 'confirmed heavy-tailed cases' (Fig. 2, black dots). Conversely, we term the remaining ones as 'uncertain cases' (Fig. 2, gray). The latter label denotes the fact that it cannot be determined with certainty whether the distributions underlying the available observations in these cases are

220 or not power laws due to scarcity of data.

Fig. 2 shows the empirical power law exponent b as a function of the proposed index of heavy-tailed behavior a. Red markers display the median values of a and b (squares), the interquartile intervals of b (vertical bars), and the binning ranges of a (horizontal bars, equal number of case studies in each bin), highlighting the correlation between the empirical power law

exponent *b* and the hydrograph recession exponent *a* for confirmed heavy-tailed cases (black dots) in all three cases (i.e., daily streamflows, ordinary peak flows, and monthly flow maxima). We test the correlation by calculating their distance (Székely et al., 2007) and Spearman (Spearman, 1904) correlations, which are valid for both linear and nonlinear associations between random variables. We find that *a* and *b* are significantly correlated at a significance level of 0.05 in all three cases with distance (Spearman) correlation coefficients of 0.45, 0.44, and 0.81 (0.42, 0.46, and 0.60) for daily streamflows, ordinary peak flows, and monthly flow maxima. The high values of the correlation coefficients for monthly flow maxima are likely affected by the existence of two clusters in Fig. 2c. Nonetheless, the existence of a statistically significant correlation between the empirical power law exponent and the proposed index, obtained for panels a, b, and c, confirms that the proposed index not only can be used to identify heavy-tailed flood behavior (as Fig. 1 shows) but also to evaluate the degree of the tail heaviness of the underlying distributions.

Fig. 2c is of particular interest because it shows an example of the typical limitations of methods that rely solely on observations to determine the tail behavior of the distribution of maxima (e.g., Papalexiou and Koutsoyiannis, 2013) and, at the same time, highlights the power of the proposed index. Large values of the recession exponent *a*, in agreement with corresponding large values of b, are found for all confirmed heavy-tailed cases (black dots in Fig. 2c) where the power law provides a plausible representation of the empirical distribution of monthly maxima. For uncertain cases (gray dots in Fig. 2c) the values of the empirical power law exponents are unreliable (according to the applied method; Clauset et al., 2009) since it cannot be determined with certainty whether the empirical distributions are or not power laws due to data scarcity. Conversely, the hydrograph recession exponent is calculated from daily streamflow data. We can therefore identify cases with heavy-tailed behavior and evaluate their tail heaviness based on the values of *a*. This estimate is deemed robust, provided that the predictions of the proposed index are confirmed by observations in cases (panels a and b) where data size is not a limitation (i.e., for daily streamflow and ordinary peak flows).

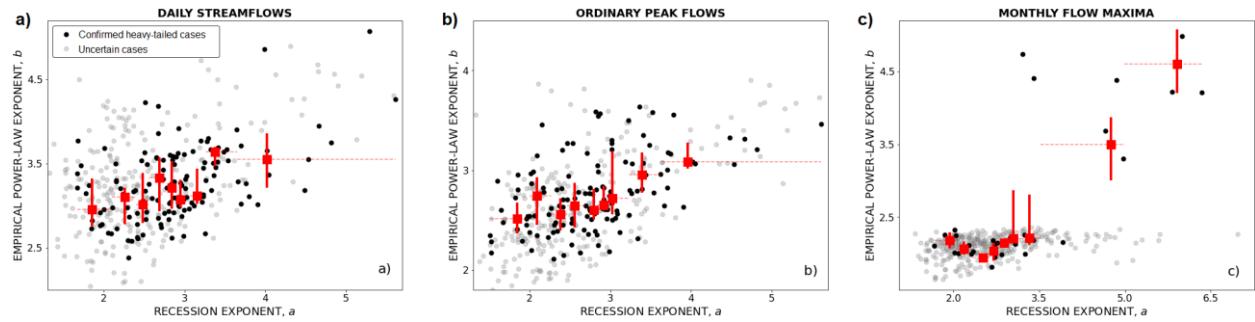



Figure 2. Empirical power law exponent *b* as a function of the proposed index of heavy-tailed behavior *a*. Case studies are classified into groups of confirmed heavy-tailed (black dots) and uncertain (gray dots) cases on the basis of the hypothesis test (Appx. A; Clauset et al., 2009). The former denotes cases for which a power law provides a reliable description of the empirical data distribution, while the latter denotes cases whose data cannot convincingly support such a distribution. Red markers highlight the correlation between the empirical power law exponent *b* and the hydrograph recession exponent *a* for confirmed heavy-tailed cases in the case of (a) daily streamflows (*n*=121 case studies), (b) ordinary peak flows (*n*=116), and (c) monthly flow maxima (*n*=34). Red markers display the median values of *a* and *b*

(squares), the interquartile intervals of b (vertical bars), and the binning ranges of a (horizontal bars, equal number of case studies in each bin).

In Fig. 3 we test the stability of the categorization of case studies into heavy/nonheavy-tailed flood behavior provided by the proposed index (i.e., the hydrograph recession exponent a) for decreasing data lengths. We compare results for the proposed index against two other frequently used metrics of heavy tails in hydrological studies: (1) the upper tail ratio (UTR) (Lu et al., 2017; Smith et al., 2018; Villarini et al., 2011) and (2) the shape parameter ξ of the GEV distribution (Morrison and Smith, 2002; Papalexiou et al., 2013; Villarini and Smith, 2010). The UTR is defined as the ratio of the flood of record to the 0.9 quantile of floods (Smith et al., 2018) here represented by monthly flow maxima, while ξ is estimated by fitting a GEV distribution on the sample of monthly maxima using the python package OpenTURNS 1.16 (Baudin et al., 2017). For all three indices ($a, UTR,$ and ξ), we estimate their values for data lengths decreasing from 35 (i.e., the shortest entire record length in the dataset) to 2 years. We acknowledge that estimating parameters of extreme value distributions from such short records is not recommended. However, the exercise highlights the perks of the proposed index that, as it will be shown, is able to provide robust results also when short data series only are available. For each case study, we obtain 30 samples with the assigned test length from the entire data series using resampling without substitution. For each test length, we calculate the median values of the indices estimated from these samples, and plot them in Fig. 3 together with their variability across case studies (vertical shaded bars and lines in Fig. 3 show the 0.25–0.75 and 0.05–0.95 quantile ranges of the index estimates across case studies).

To evaluate the consistency of the categorization of tail behavior across different data lengths we proceed as follows. For each case we first compute the hydrograph recession exponent and GEV shape parameter from the entire data record and denote them with an asterisk superscript (i.e., a^* or ξ^*). Heavy-tailed cases are defined as having $a^*>2$ or $\xi^*>0$ (Godrèche et al., 2015), while non-heavy-tailed cases have values below these thresholds. To visualize heavy-tailed and non-heavy-tailed behaviors, we mark them in Fig. 3 in red and green colors, respectively, based on the reference values obtained from the entire data record. We then recalculate the indices from shorter samples and evaluate whether their values are consistent with the above categorization. For the UTR, we cannot implement this approach because there is no specific threshold for the identification of heavy/nonheavy tails. We therefore directly compare the stability of the UTR's values across data lengths (a larger value indicates a heavier tail).

The proposed index provides consistent categorization of heavy/nonheavy-tailed flood behavior across varying data lengths (Fig. 3a). The index estimates remain above 2 for most heavy-tailed cases (red) and below 2 for most nonheavy-tailed cases (green) (as defined according to the reference value a^* computed using the entire data record) when the data length decreases. The index estimates demonstrate the consistency throughout the test data length, as evidenced by the narrow range of variation in the median values of the estimates. For heavy-tailed cases, the median values ranged from 2.64 to 2.92, while for nonheavy-tailed cases, they ranged from 1.84 to 2.0. Additionally, the coefficient of variation for the estimates remained relatively constant, ranging from 0.29 to 0.33 for both heavy and nonheavy-tailed cases. This indicates that the variability of the results

(vertical shaded bars and lines in Fig. 3) is mostly due to pooling together different case studies belonging to the same category

285 (heavy or nonheavy-tailed), and does not increase as a result of decreasing length of the available data.

In contrast, the upper tail ratio shows pronounced instability for decreasing data lengths (Fig. 3b). The median value of the index estimates varies between 1.32 and 2.36, with a coefficient of variation ranging from 0.15 to 0.64. These values indicate uncertain assessments based on the UTR and its tendency to underestimate the tail heaviness as the data length decreases.

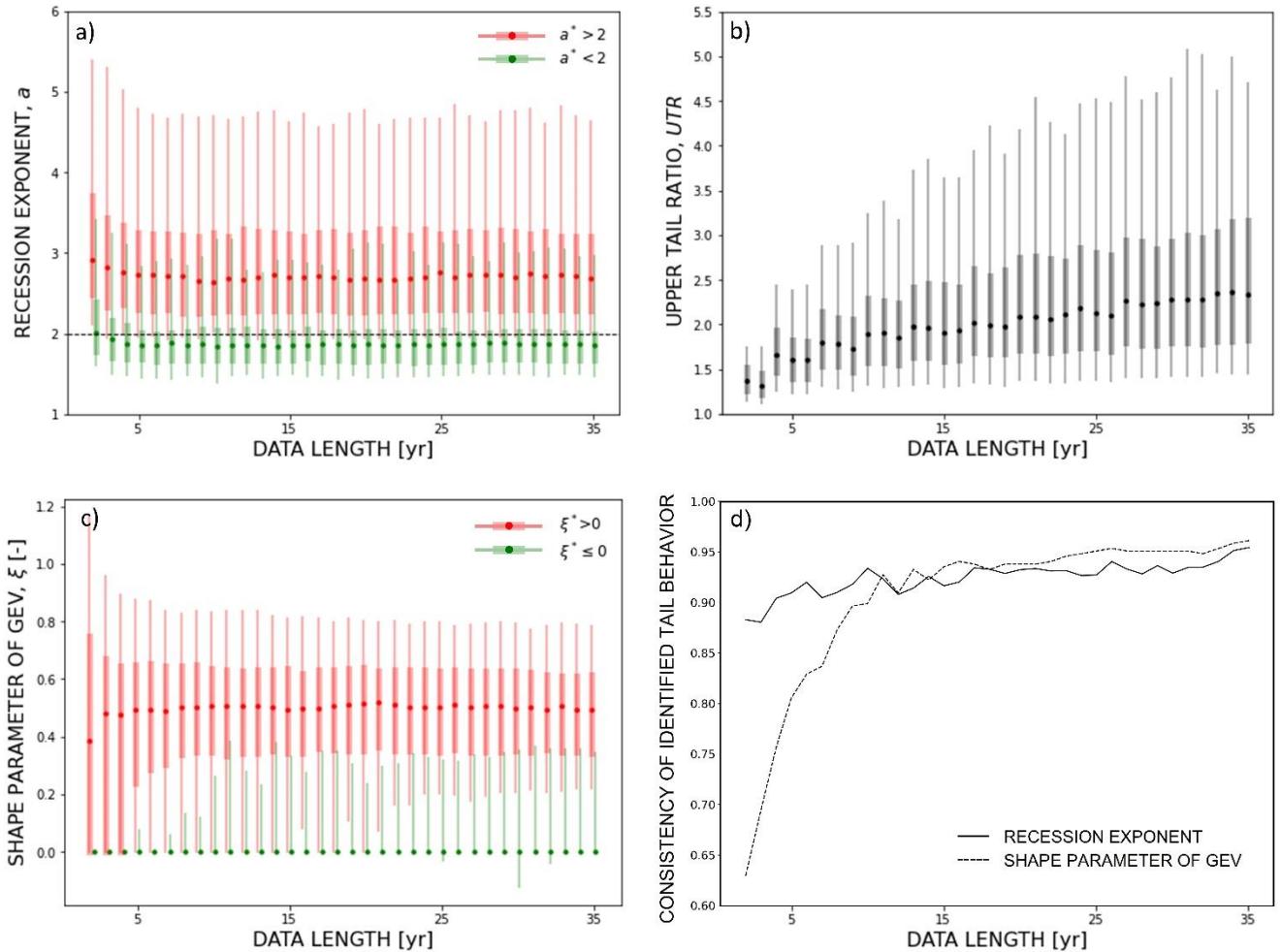
Fig. 3c illustrates the categorization of tail behavior using GEV shape parameter estimates. The results indicate that ξ estimates

290 are stable with longer data series, yet their variability increases — leading to both underestimation and overestimation of tail heaviness — when data length is short. To ensure a stable categorization of flood tail behavior using this method data series spanning more than 10 years (for seasonal analyses and monthly maxima, i.e., sample sizes of around 30 values) are needed, in line with the findings of previous studies (Cai and Hames, 2010; Németh et al., 2019). The median values of ξ range from 0.39 to 0.52 for heavy-tailed cases and remain at 0 for nonheavy-tailed cases. Furthermore, the coefficient of variation

295 demonstrates relatively higher variation across different test data lengths, ranging from 0.37 to 1.03 for heavy-tailed cases.

Figure 3d presents a summary of the consistency in identifying tail behavior (either heavy or nonheavy) compared to the identification based on the complete data record (i.e., fraction of cases for which categorization based on shorter data series provides the same result obtained with the complete data record). This assessment is conducted for both the methods of recession exponents and GEV shape parameters (unfortunately, this approach is inapplicable to the UTR due to the absence of

300 a specific threshold for distinguishing heavy/nonheavy tails). The comparison underscores that discrepancies in consistency between the two indices (ξ and a) are predominantly noticeable when analyzing data series shorter than 10 years in this study.

Conversely, for data series longer than 10 years, both indices exhibit comparable consistency and display an ascending trend, with the performance of the GEV shape parameters slightly higher than the one of the recession exponents.

305 **Figure 3. Stability of the categorization of case studies into heavy/nonheavy-tailed flood behavior for decreasing data lengths.**
Estimates of three different indices of tail behavior as a function of data length. (a) Hydrograph recession exponent a (i.e., the proposed index of this study). Two frequently used metrics of heavy tails in hydrological studies: (b) the upper tail ratio UTR, and (c) the shape parameter ξ of the GEV distribution. Dots display the median values of the estimates for 386 case studies; vertical shaded bars and lines show the 0.25-0.75 and 0.05-0.95 quantile ranges of the estimates, respectively. The entire data record was used for computing the reference
310 values of the hydrograph recession exponent a^* and the GEV shape parameter ξ^* and categorizing each case study as either having (red) or not (green) heavy-tailed behavior. (d) Consistency of identified tail behavior (either heavy or nonheavy) as a function of available data length for the indices recession exponent and shape parameter of GEV.

5 Discussion

Assessment of flood tail behavior is challenging due to high levels of uncertainty arising from the scarcity of data on floods,
315 which are by definition rare events. This issue is particularly prominent when maxima are used in the analysis as in the annual

maximum approach. Despite the widespread use of this method, its limitations for what concerns the reliability of flood tail estimates are well recognized. Very large sample sizes are indeed essential for obtaining accurate prediction of tail behavior (Papalexiou and Koutsoyiannis, 2013).

To address the challenge of obtaining reliable estimates, alternative methods have been proposed. A frequently used approach

320   is the peak-over-threshold analysis, which uses the information content of a larger sample of data (Lang et al., 1999; Pan et al., 2022). Previous studies have demonstrated that this method leads to lower uncertainty in estimating high floods (Kumar et al., 2020). Volpi et al. (2019) also showed the advantage of using all the available observations (i.e., not only the peaks over a certain threshold) for estimating extreme events. In summary, all these methods suggest that discharge values other than maxima can provide information about the characteristics of extreme events. Specifically, incorporating information from less

325   extreme (but more numerous) observations can reduce the uncertainty in the estimation of extreme events and lead to improved accuracy. Furthermore, non-asymptotic methods suggest that extremes are realizations of the underlying ordinary events (Marani and Ignaccolo, 2015; Lombardo et al., 2019), which can thus be used to assess rare events. These methods have significantly improved the estimation of extreme values with lower uncertainty (Marra et al., 2018; Miniussi and Marani, 2020; Mushtaq et al., 2022; Hu et al., 2023).

330   Similarly to the latter approaches, the index introduced in this study (i.e., the hydrograph recession exponent) leverages information on ordinary discharge dynamics to infer the tail behavior of flood distributions. This approach entails some advantages: firstly, it effectively extracts information from a larger amount of available streamflow data. Secondly, estimating the hydrograph recession exponent requires significantly less data than conventional approaches that involve fitting probability distributions to hydrological samples, while providing reliable results. But most importantly, the proposed index offers a

335   mechanistic approach to understand the emergence of heavy-tailed flood behavior, thus providing a process-based alternative to methods that solely rely on statistical analysis of observations. The importance of understanding intrinsic watershed dynamics which promote the occurrence of extreme events and contributing factors that lead to heavy-tailed flood behavior (Tarasova et al., 2020) was recently highlighted in a comprehensive review by Merz et al. (2022). Identify reliable proxies for inferring such behavior (*sensu* Wilson and Toumi, 2005) is as well important. The proposed index, which represents such a

340   proxy grounded on intrinsic hydrologic dynamics of the river basin, is thus especially useful in the very common cases when the tail of the flood distribution cannot be known from limited available observations.

The hydrograph recession exponent (which is the identified index of heavy-tailed flood behavior) essentially represents the nonlinearity of the storage-discharge response in catchments (Wittenberg, 1999; Biswal and Marani, 2010). A higher degree of nonlinearity leads to higher peak flows and heavier tail of the streamflow distribution (Basso et al., 2015). In agreement

345   with these findings, former simulation-based and field studies have shown that high nonlinearity of the catchment hydrological response linked to an increase of the runoff contributing area results in a marked increase of the slope of flood frequency curves (Fiorentino et al., 2007; Rogger et al., 2012), which may be indicative of a heavy-tailed flood behavior. Gioia et al. (2012) also demonstrated that a nonlinear catchment response can convert light-tailed rainfall inputs into flood distributions with heavy tails, further confirming the role of nonlinear storage-discharge responses in producing heavy-tailed flood behavior.

Merz et al. (2022) established, based on a comprehensive review, that the nonlinearity of the catchment response is a plausible contributor to the emergence of heavy-tailed flood behavior. Additionally, Basso et al. (2023) demonstrated that the hydrograph recession exponent aids in predicting the propensity of rivers for generating extreme floods. In line with these studies, our research further highlights that the hydrograph recession exponent, which provides a description of catchment nonlinearity obtained from common streamflow dynamics, is capable of robustly identifying heavy-tailed flood behavior.

The findings in Fig. 2 showcase the drawbacks of relying on purely statistical data analyses (which supply the empirical power law exponents *b*) to identify flood tail behaviors and the advantages of adopting the mechanistic approach proposed in this study (which yield the hydrograph recession exponent *a*). The gray markers in Fig. 2 indicate uncertainty in determining whether the distribution has a power law tail, which is shown to be more prevalent when the sample size is reduced, based on statistical analyses according to the Clauset (2009) method (69%, 70%, and 91% of the case studies for daily streamflow, ordinary peaks, and monthly maxima, respectively). The proposed index finds a solution to these limitations through a mathematical description of hydrological processes. Such an index is shown to perform well in cases where statistical methods may be limited due to a lack of data, as confirmed by the significant correlations between the recession exponent and the reliably empirical power law exponent in all three panels (represented by black dots in Fig. 2). Even in cases where the statistical method is unable to confirm the underlying distribution (e.g., monthly maxima in panel c), our proposed index can still provide robust estimates of tail heaviness based on the values of recession exponents. This is supported by the analyses of daily streamflows and ordinary peaks, where sample size is not a limitation and the predictions of the proposed index are confirmed by observations. Overall, our proposed index offers a promising solution for accurately characterizing the tail behavior of flood distributions, especially when traditional statistical methods may be limited due to a lack of data.

Data scarcity is a major challenge for reliable flood hazard assessment, mainly because of relatively short hydrological data records worldwide (Lins, 2008). The availability of a robust index of heavy-tailed flood behavior that work even with short data records is desirable. We test three indices, namely the recession exponent (the proposed index), the upper tail ratio (UTR), and the shape parameter of the Generalized Extreme Value (GEV) distribution (ξ), for categorizing tail behavior for decreasing data lengths. The results (Fig. 3a) show that the recession exponent provides stable estimates and categorizes cases consistently into heavy or non-heavy tails for decreasing data lengths. Furthermore, the slight variation in the estimates of the recession exponent for each test data length implies that variation in estimates primarily arises from case study heterogeneity rather than decreasing data length. Conversely, UTR significantly underestimates both the tail heaviness and the variation across cases for decreasing data lengths (Fig. 3b). In agreement with previous studies, underestimation of tail heaviness occurred using UTR when the sample size was small (Smith et al., 2018; Wietzke et al., 2020). Meanwhile, the categorization of tail behavior was stable for cases with dataset longer than 10 years using the GEV shape parameter. However, high uncertainty in the variation of estimates across cases is observed when available data is relatively short as also highlighted by previous studies (e.g., Wietzke et al., 2020) (Fig. 3c). Implied by this observation is that the estimates are biased by the short analyzed data and a longer data record is desirable for a more reliable fitting of a GEV on data (Papalexiou and Koutsoyiannis, 2013). In summary, both the recession exponent and the GEV shape parameter exhibit greater stability across data lengths than the UTR, which is

highly dependent on the available amount of data. When comparing the first two indices (recession exponent and GEV shape parameter) (Fig. 3d), the recession exponent demonstrates a high level of stability across all data lengths, even those shorter than 10 years based on this study's analyses. On the other hand, the GEV shape parameter displays lower stability when the available data are shorter than 10 years, but this stability significantly improves as the data length exceeds 10 years. Beyond the 10-year threshold, both indices show comparable consistency and an upward trend, with GEV shape parameters slightly outperforming recession exponents.

The hydrograph recession exponent allows at least two significant applications as a proxy for heavy-tailed flood behavior. Firstly, it can be directly used to improve comparability across catchments and provide a fair assessment of mapping regional patterns of flood hazards (Merz et al., 2022). Traditionally, assessing flood behavior across catchments using the same record length has been preferred (Cunderlik and Burn, 2002), but this is often not possible due to differences in data availability. The proposed index can robustly estimate heavy-tailed flood behavior from data with different record lengths, overcoming this limitation. Secondly, it can be applied as a preliminary step to correctly identify whether a considered catchment exhibits heavy-tailed flood behavior or not, and to select an appropriate probability distribution to be used in flood frequency analysis. This prior identification of tail behavior is crucial to avoid potential underestimation of flood extremes. (Miniussi et al., 2020; Mushtaq et al., 2022).

6 Conclusions

A new index of heavy-tailed flood behavior is identified from a physically-based description of streamflow dynamics. The new index is embodied by the hydrograph recession exponent and can be readily estimated from daily streamflow records. Our findings demonstrate that this index enables the identification of heavy or nonheavy tailed flood behaviors in a large set of case studies across Germany. Importantly, it provides an evaluation of the tail heaviness (i.e., the severity of flood risks) based on analyses of common discharge dynamics, and remarkably, the results remain robust even with limited data records.

The proposed index addresses the main limitations of current approaches, including the lack of physical support and low reliability in cases with limited data records. By extracting more information from available data and manifesting the nonlinearity of catchment response, it represents a reliable method to select suitable underlying distributions for flood frequency analyses and assess the peril of extreme floods in data poor areas.

Acknowledgments

This work is funded by the Deutsche Forschungsgemeinschaft-Project 421396820 "Propensity of rivers to extreme floods: climate-landscape controls and early detection (PREDICTED)" and FOR 2416 "Space-Time Dynamics of Extreme Floods (SPATE)". The financial support of the Helmholtz Centre for Environmental Research and the Norwegian Institute for Water Research is as well acknowledged. SY (the 3rd author) acknowledges the support of the Helmholtz Climate Initiative Project

funded by the Helmholtz Association. The manuscript and supporting information provide all the information needed to replicate the results.

Data Availability Statement

For providing the discharge data for Germany, we are grateful to the Bavarian State Office of Environment (LfU, https://www.gkd.bayern.de/de/fluesse/abfluss) and the Global Runoff Data Centre (GRDC) prepared by the Federal Institute for Hydrology (BfG, http://www.bafg.de/GRDC). Climatic data can be obtained from the German Weather Service (DWD; ftp://ftp-cdc.dwd.de/pub/CDC/). The digital elevation model can be retrieved from Shuttle Radar Topography Mission (SRTM; https://cgiarcsi.community/data/srtm-90m-digital-elevation-database-v4-1/).

Appendix A Goodness-of-fit tests for the empirical power laws

To test if the empirical power law is a plausible underlying distribution of the observed data, we follow the hypothesis test proposed by Clauset et al. (2009). The null hypothesis is 'The empirical power law is a plausibly underlying distribution of the observed data.' Residual errors exist between the empirical power law and the observed data, which can be estimated by the error distance ε_d by means of the Kolmogorov-Smirnov statistic. The Kolmogorov-Smirnov test is selected because it is one of the most common measures for non-normal data. The core of the hypothesis test is to statistically prove that the errors between the data and the power law (i.e., ε_d) are rational fluctuation of sampling randomness rather than being drawn from an incorrect underlying distribution. To determine the rationality of the sampling randomness, a Monte Carlo procedure is introduced: (1) a large number of groups n of synthetic data (with the same size as the observed data) are randomly generated from the empirical power law; (2) the error distance ε_{s_i} of each synthetic group to the empirical power law is calculated for $i = 1,2,\cdots,n$; (3) the frequency of $\varepsilon_s > \varepsilon_d$ defines the p-value of the hypothesis test, which indicates the probability that the residual errors between the empirical power law and the observed data locates within the range of sampling randomness fluctuations; and (4) the rationality is determined by $p > 0.1$ using this package.

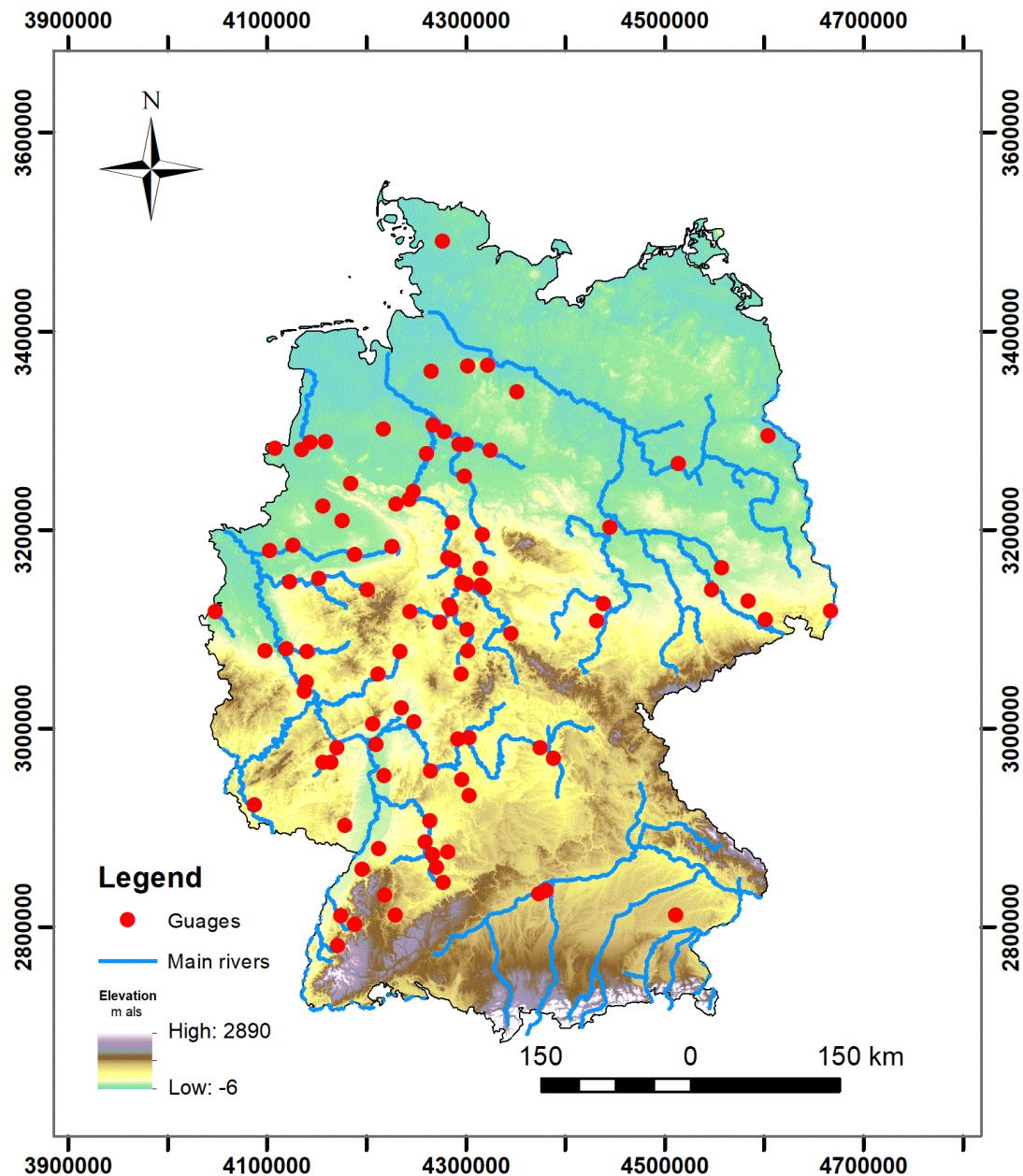
When $p \leq 0.1$, the null hypothesis is rejected; that is, the observed data are not plausibly drawn from the empirical power law. On the contrary, the empirical power law is considered a plausible distribution for the observed data for their residual errors are statistically rational fluctuation of sampling randomness when $p > 0.1$. Notice that a greater p-value is better in this case because the aim is to verify the null hypothesis rather than to indicate it is unlikely to be correct, as others often considered. Thus $p > 0.1$ is a more rigorous setting than $p > 0.05$ in this case.

The setting of $n = 1000$ is used as an adequate (great enough) number of iterations in this framework to distinguish underlying distributions that are commonly mixed (as suggested by Clauset et al. (2009)).

The hypothesis test of the empirical power law including all the above procedures can be implemented via the function test_pl in the python package plfit 1.0.3 (https://pypi.org/project/plfit/).

It is worth mentioning that, statistically, we cannot say those who do not pass the hypothesis test 'are not' power law distributions. There are at least two potential reasons for this result: (1) they are indeed not power law functions, or (2) the underlying distribution cannot be concluded due to the high uncertainty in the empirical data with small sample sizes. We thus use the term 'uncertain cases' to indicate this awareness in the main manuscript.

445

Appendix B A reference map of gauges across Germany used in this study

450 **Figure B1 A reference map of gauges across Germany used in this study.** These river basins encompass a variety of climate and physiographic settings without strong impact from dams and snowfall. Their areas range from 110 to 23,843 km² with a median value of 1,195 km². The minimum, median, and maximum lengths of the daily streamflow records are 35, 58, and 63 years (inbetween 1951 – 2013).

Author contributions

Hsing-Jui Wang: Conceptualization (lead); methodology (lead); investigation (lead); formal analysis (lead); writing – original draft (lead); writing – review and editing (equal).

Ralf Merz: Conceptualization (supporting); methodology (supervision); investigation (supervision); formal analysis (supervision); writing – review and editing (equal).

Soohyun Yang: Investigation (supporting); methodology (supervision); formal analysis (supervision); writing – review and editing (equal).

Stefano Basso: Conceptualization (supporting); methodology (supervision); investigation (supervision); formal analysis (supervision); writing – review and editing (equal).

Declaration of interests

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.

References

Arai, R., Toyoda, Y., & Kazama, S. (2020). Runoff recession features in an analytical probabilistic streamflow model. Journal of Hydrology, 597, 125745. https://doi.org/10.1016/j.jhydrol.2020.125745

Archfield, S. A., Hirsch, R. M., Viglione, A., & Blöschl, G. (2016). Fragmented patterns of flood change across the United States. Geophysical Research Letters, 43(10), 10232–10239. https://doi.org/10.1002/2016GL070590

Basso, S., Botter, G., Merz, R., & Miniussi, A. (2021). PHEV! The PHysically-based Extreme Value distribution of river flows. Environmental Research Letters, 16(12). https://doi.org/10.1088/1748-9326/ac3d59

Basso, S., Schirmer, M., & Botter, G. (2015). On the emergence of heavy-tailed streamflow distributions. Advances in Water Resources, 82, 98–105. https://doi.org/10.1016/j.advwatres.2015.04.013

Basso, S., Schirmer, M., & Botter, G. (2016). A physically based analytical model of flood frequency curves. Geophysical Research Letters, 43(17), 9070–9076. https://doi.org/10.1002/2016GL069915

Basso, S., Merz, R., Tarasova, L., & Miniussi, A. (2023). Extreme flooding controlled by stream network organization and flow regime. Nature Geoscience, 16(April), 339–343. https://doi.org/10.1038/s41561-023-01155-w

Baudin, M., Dutfoy, A., Iooss, B., & Popelin, A.-L. (2017). OpenTURNS: An Industrial Software for Uncertainty Quantification in Simulation BT - Handbook of Uncertainty Quantification. In R. Ghanem, D. Higdon, & H. Owhadi (Eds.) (pp. 2001–2038). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-12385-1_64

Beirlant, J., Goegebeur, Y., Teugels, J., Segers, J., De Waal, D., & Ferro, C. (2004). Statistics of extremes: Theory and applications. Wiley. https://doi.org/https://doi.org/10.1002/0470012382

Bevere, L., & Remondi, F. (2022). Natural catastrophes in 2021: the floodgates are open. Swiss Re Institute sigma research.

Biswal, B. (2021). Decorrelation is not dissociation: There is no means to entirely decouple the Brutsaert-Nieber parameters in streamflow recession analysis. Advances in Water Resources, 147, 103822. https://doi.org/https://doi.org/10.1016/j.advwatres.2020.103822

Biswal, B., & Kumar, D. N. (2014). Study of dynamic behaviour of recession curves. Hydrological Processes, 792(November 2012), 784–792. https://doi.org/10.1002/hyp.9604

Biswal, B., & Marani, M. (2010). Geomorphological origin of recession curves. Geophysical Research Letters, 37(24), 1–5. https://doi.org/10.1029/2010GL045415

Botter, G., Basso, S., Porporato, A., Rodriguez-Iturbe, I., & Rinaldo, A. (2010). Natural streamflow regime alterations: Damming of the Piave river basin (Italy). Water Resources Research, 46(6), 1–14. https://doi.org/10.1029/2009WR008523

Botter, G., Peratoner, F., Porporato, A., Rodriguez-Iturbe, I., & Rinaldo, A. (2007). Signatures of large-scale soil moisture dynamics on streamflow statistics across U.S. climate regimes. Water Resources Research, 43(11), 1–10. https://doi.org/10.1029/2007WR006162

Botter, G., Porporato, A., Rodriguez-Iturbe, I., & Rinaldo, A. (2007). Basin-scale soil moisture dynamics and the probabilistic characterization of carrier hydrologic flows: Slow, leaching-prone components of the hydrologic response. Water Resources Research, 43(2), 1–14. https://doi.org/10.1029/2006WR005043

Botter, G., Porporato, A., Rodriguez-Iturbe, I., & Rinaldo, A. (2009). Nonlinear storage-discharge relations and catchment streamflow regimes. Water Resources Research, 45(10), 1–16. https://doi.org/10.1029/2008WR007658

Brutsaert, W., & Nieber, J. L. (1977). Regionalized drought flow hydrographs from a mature glaciated plateau. Water Resources Research, 13(3), 637–643. https://doi.org/10.1029/WR013i003p00637

Cai, Y., & Hames, D. (2010). Minimum sample size determination for generalized extreme value distribution. Communications in Statistics: Simulation and Computation, 40(1), 87–98. https://doi.org/10.1080/03610918.2010.530368

Ceola, S., Botter, G., Bertuzzo, E., Porporato, A., Rodriguez-Iturbe, I., & Rinaldo, A. (2010). Comparative study of ecohydrological streamflow probability distributions. Water Resources Research, 46(9), 1–12. https://doi.org/10.1029/2010WR009102

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. SIAM Review, 51(4), 661–703. https://doi.org/10.1137/070710111

Cooke, R. M., Nieboer, D., & Misiewicz, J. (2014). Fat-Tailed Distributions: Data, Diagnostics and Dependence (volume 1). John Wiley & Sons.

Cooke, R. M., & Nieboer, D. (2011). Heavy-Tailed Distributions: Data, Diagnostics, and New Developments. Resources for the Future Discussion Paper, No. 11-19. https://doi.org/dx.doi.org/10.2139/ssrn.1811043

Cunderlik, J. M., & Burn, D. H. (2002). The use of flood regime information in regional flood frequency analysis. Hydrological Sciences Journal, 47(1), 77–92. https://doi.org/10.1080/02626660209492909

Doulatyari, B., Betterle, A., Basso, S., Biswal, B., Schirmer, M., & Botter, G. (2015). Predicting streamflow distributions and flow duration curves from landscape and climate. Advances in Water Resources, 83, 285–298. https://doi.org/10.1016/j.advwatres.2015.06.013

Durrans, S. R., Eiffe, M. A., Thomas, W. O., & Goranflo, H. M. (2003). Joint Seasonal /Annual Flood Frequency Analysis. Journal of Hydrologic Engineering, 8(4), 181–189. https://doi.org/10.1061/(asce)1084-0699(2003)8:4(181)

El Adlouni, S., Bobée, B., & Ouarda, T. B. M. J. (2008). On the tails of extreme event distributions in hydrology. Journal of Hydrology, 355(1–4), 16–33. https://doi.org/10.1016/j.jhydrol.2008.02.011

Eliazar, I., & Sokolov, I. (2010). Gini characterization of extreme-value statistics. Physica A-Statistical Mechanics and Its Applications - PHYSICA A, 389, 4462–4472. https://doi.org/10.1016/j.physa.2010.07.005

Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). Modelling extreme events for insurance and finance. Springer Berlin Heidelberg.

Fiorentino, M., Manfreda, S., & Iacobellis, V. (2007). Peak runoff contributing area as hydrological signature of the probability distribution of floods. Advances in Water Resources, 30(10), 2123–2134. https://doi.org/10.1016/j.advwatres.2006.11.017

Fischer, S., & Schumann, A. (2016). Robust flood statistics: comparison of peak over threshold approaches based on monthly maxima and TL-moments. Hydrological Sciences Journal, 61(3), 457–470. https://doi.org/10.1080/02626667.2015.1054391

Gioia, A., Iacobellis, V., Manfreda, S., & Fiorentino, M. (2012). Influence of infiltration and soil storage capacity on the skewness of the annual maximum flood peaks in a theoretically derived distribution. Hydrology and Earth System Sciences, (16), 937–951. https://doi.org/10.5194/hess-16-937-2012

Godrèche, C., Majumdar, S. N., & Schehr, G. (2015). Statistics of the longest interval in renewal processes. Journal of Statistical Mechanics: Theory and Experiment, 2015(3). https://doi.org/10.1088/1742-5468/2015/03/P03014

Hall, J., Arheimer, B., Borga, M., Brázdil, R., Claps, P., Kiss, A., et al. (2014). Understanding flood regime changes in Europe: A state-of-the-art assessment. Hydrology and Earth System Sciences, 18(7), 2735–2772. https://doi.org/10.5194/hess-18-2735-2014

Hodgkins, G. A., Whitfield, P. H., Burn, D. H., Hannaford, J., Renard, B., Stahl, K., et al. (2017). Climate-driven variability in the occurrence of major floods across North America and Europe. Journal of Hydrology, 552, 704–717. https://doi.org/10.1016/j.jhydrol.2017.07.027

Hu, L., Nikolopoulos, E. I., Marra, F., & N., A. E. (2023). Toward an improved estimation of flood frequency statistics from simulated flows. Journal of Flood Risk Management, 1–13. https://doi.org/10.1111/jfr3.12891

Jachens, E. R., Rupp, D. E., Roques, C., & Selker, J. S. (2020). Recession analysis revisited: Impacts of climate on parameter estimation. Hydrology and Earth System Sciences, 24(3), 1159–1170. https://doi.org/10.5194/hess-24-1159-2020

Karlsen, R. H., Bishop, K., Grabs, T., Ottosson-Löfvenius, M., Laudon, H., & Seibert, J. (2019). The role of landscape properties, storage and evapotranspiration on variability in streamflow recessions in a boreal catchment. Journal of Hydrology, 570(2019), 315–328. https://doi.org/10.1016/j.jhydrol.2018.12.065

Koutsoyiannis, D. (2022). Stochastics of Hydroclimatic Extremes – A Cool Look at Risk. Ka llipos (2nd ed.). Athens: Open Academic Editions. https://doi.org/10.57713/kallipos-1

Koutsoyiannis, D. (2004). Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records. Hydrological Sciences Journal, 49(4), 591–610. https://doi.org/10.1623/hysj.49.4.591.54424

Koutsoyiannis, D. (2004). Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation. Hydrological Sciences Journal, 49(4), 575–590. https://doi.org/10.1623/hysj.49.4.575.54430

Kumar, M., Sharif, M., & Ahmed, S. (2020). Flood estimation at Hathnikund Barrage, River Yamuna, India using the Peak-Over-Threshold method. ISH Journal of Hydraulic Engineering, 26(3), 291–300. https://doi.org/10.1080/09715010.2018.1485119

Laio, F., Porporato, A., Fernandez-Illescas, C. P., & Rodriguez-Iturbe, I. (2001). Plants in water-controlled ecosystems: Active role in hydrologic processes and responce to water stress IV. Discussion of real cases. Advances in Water Resources, 24(7), 745–762. https://doi.org/10.1016/S0309-1708(01)00007-0

Lang, M., Ouarda, T. B. M. J., & Bobée, B. (1999). Towards operational guidelines for over-threshold modeling. Journal of Hydrology, 225(3–4), 103–117. https://doi.org/10.1016/S0022-1694(99)00167-5

Lehner, B., Liermann, C. R., Revenga, C., Vörömsmarty, C., Fekete, B., Crouzet, P., et al. (2011). High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. Frontiers in Ecology and the Environment, 9(9), 494–502. https://doi.org/10.1890/100125

Lins, H. F. (2008). Challenges to hydrological observations. WMO Bulletin, 57(January), 55–58.

Lombardo, F., Napolitano, F., Russo, F., & Koutsoyiannis, D. (2019). On the Exact Distribution of Correlated Extremes in Hydrology. Water Resources Research, 55(12), 10405–10423. https://doi.org/10.1029/2019WR025547

Lu, P., Smith, J. A., & Lin, N. (2017). Spatial characterization of flood magnitudes over the drainage network of the Delaware river basin. Journal of Hydrometeorology, 18(4), 957–976. https://doi.org/10.1175/JHM-D-16-0071.1

Malamud, B. D., & Turcotte, D. L. (2006). The applicability of power-law frequency statistics to floods. Journal of Hydrology, 322(1–4), 168–180. https://doi.org/10.1016/j.jhydrol.2005.02.032

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

580   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. https://doi.org/10.1016/j.advwatres.2018.05.001

Martinez-Villalobos, C., & Neelin, J. D. (2021). Climate models capture key features of extreme precipitation probabilities across regions. Environmental Research Letters, 16(2). https://doi.org/10.1088/1748-9326/abd351

585   McCuen, R. H., & Smith, E. (2008). Origin of Flood Skew. Journal of Hydrologic Engineering, 13(9), 771–775. https://doi.org/10.1061/(asce)1084-0699(2008)13:9(771)

McDermott, T. K. J. (2022). Global exposure to flood risk and poverty. Nature Communications, 13(1), 6–8. https://doi.org/10.1038/s41467-022-30725-6

Mejía, A., Daly, E., Rossel, F., Javanovic, T., & Gironás, J. (2014). A stochastic model of streamflow for urbanized basins.
590   Water Resources Research, 50, 1984–2001. https://doi.org/10.1002/2013WR014834

Merz, B., Basso, S., Fischer, S., Lun, D., Blöschl, G., Merz, R., et al. (2022). Understanding heavy tails of flood peak distributions. Water Resources Research, 1–37. https://doi.org/10.1029/2021wr030506

Merz, B., Blöschl, G., Vorogushyn, S., Dottori, F., Aerts, J. C. J. H., Bates, P., et al. (2021). Causes, impacts and patterns of disastrous river floods. Nature Reviews Earth and Environment, 2(9), 592–609. https://doi.org/10.1038/s43017-021-
595   00195-3

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(7), 1–18. https://doi.org/10.1029/2019WR026535

Miniussi, A., Marani, M., & Villarini, G. (2020). Metastatistical Extreme Value Distribution applied to floods across the
600   continental United States. Advances in Water Resources, 136(December 2019), 103498. https://doi.org/10.1016/j.advwatres.2019.103498

Morrison, J. E., & Smith, J. A. (2002). Stochastic modeling of flood peaks using the generalized extreme value distribution. Water Resources Research, 38(12), 41-1-41–12. https://doi.org/10.1029/2001wr000502

Müller, M. F., Dralle, D. N., & Thompson, S. E. (2014). Analytical model for flow duration curves in seasonally dry
605   climates. Water Resources Research, 50, 5510–5531. https://doi.org/10.1002/2014WR015301

Müller, M. F., Roche, K. R., & Dralle, D. N. (2021). Catchment processes can amplify the effect of increasing rainfall variability. Environmental Research Letters, 16(8). https://doi.org/10.1088/1748-9326/ac153e

Mushtaq, S., Miniussi, A., Merz, R., & Basso, S. (2022). Reliable estimation of high floods: A method to select the most suitable ordinary distribution in the Metastatistical extreme value framework. Advances in Water Resources,
610   161(September 2021), 104127. https://doi.org/10.1016/j.advwatres.2022.104127

Németh, L., Hübnerová, Z., & Zempléni, A. (2019). Trend detection in GEV models, 1–13. Retrieved from http://arxiv.org/abs/1907.09435

Nerantzaki, S. D., & Papalexiou, S. M. (2019). Tails of extremes: Advancing a graphical method and harnessing big data to assess precipitation extremes. Advances in Water Resources, 134. https://doi.org/10.1016/j.advwatres.2019.103448

615   Pall, P., Aina, T., Stone, D. A., Stott, P. A., Nozawa, T., Hilberts, A. G. J., et al. (2011). Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. Nature, 470(7334), 382–385. https://doi.org/10.1038/nature09762

Pan, X., Rahman, A., Haddad, K., & Ouarda, T. B. M. J. (2022). Peaks-over-threshold model in flood frequency analysis: a scoping review. Stochastic Environmental Research and Risk Assessment, 36(9), 2419–2435.
620   https://doi.org/10.1007/s00477-022-02174-6

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(2), 851–862. https://doi.org/10.5194/hess-17-851-2013

Papalexiou, S. M., & Koutsoyiannis, D. (2013). Battle of extreme value distributions : A global survey on extreme daily rainfall. Water Resources Research, 49(1), 187–201. https://doi.org/10.1029/2012WR012557

625   Porporato, A., Daly, E., & Rodriguez-Iturbe, I. (2004). Soil water balance and ecosystem response to climate change. American Naturalist, 164(5), 625–632. https://doi.org/10.1086/424970

Pumo, D., Viola, F., La Loggia, G., & Noto, L. V. (2014). Annual flow duration curves assessment in ephemeral small basins. Journal of Hydrology, 519(PA), 258–270. https://doi.org/10.1016/j.jhydrol.2014.07.024

Rajah, K., O'Leary, T., Turner, A., Petrakis, G., Leonard, M., & Westra, S. (2014). Changes to the temporal distribution of
630   daily precipitation. Geophysical Research Letters, 41(24), 8887–8894. https://doi.org/10.1002/2014GL062156

Rentschler, J., Salhab, M., & Jafino, B. A. (2022). Flood exposure and poverty in 188 countries. Nature Communications, 13(1), 3527. https://doi.org/10.1038/s41467-022-30727-4

Resnick, S. I. (2007). Heavy-Tail Phenomena: Probabilistic and Statistical Modeling. New York: Springer US.

Rogger, M., Pirkl, H., Viglione, A., Komma, J., Kohl, B., Kirnbauer, R., & Merz, R. (2012). Step changes in the flood
635   frequency curve : Process controls. Water Resources Research, 48, 1–15. https://doi.org/10.1029/2011WR011187

Rossi, M. W., Whipple, K. X., & Vivoni, E. R. (2016). Precipitation and evapotranspiration controls on daily runoff variability in the contiguous United States and Puerto Rico. Journal of Geophysical Research: Earth Surface, 128–145. https://doi.org/doi:10.1002/2015JF003446

Santos, A. C., Portela, M. M., Rinaldo, A., & Schaefli, B. (2018). Analytical flow duration curves for summer streamflow in
640   Switzerland. Hydrology and Earth System Sciences, 22(4), 2377–2389. https://doi.org/10.5194/hess-22-2377-2018

Sartori, M., & Schiavo, S. (2015). Connected we stand: A network perspective on trade and global food security. Food Policy, 57, 114–127. https://doi.org/https://doi.org/10.1016/j.foodpol.2015.10.004

Schaefli, B., Rinaldo, A., & Botter, G. (2013). Analytic probability distributions for snow-dominated streamflow. Water Resources Research, 49(5), 2701–2713. https://doi.org/10.1002/wrcr.20234

645   Seckin, N., Haktanir, T., & Yurtal, R. (2011). Flood frequency analysis of Turkey using L-moments method. Hydrological Processes, 25(22), 3499–3505. https://doi.org/10.1002/hyp.8077

Sharma, A., Wasko, C., & Lettenmaier, D. P. (2018). If Precipitation Extremes Are Increasing, Why Aren't Floods? Water Resources Research, 54(11), 8545–8551. https://doi.org/10.1029/2018WR023749

Smith, J. A., Cox, A. A., Baeck, M. L., Yang, L., & Bates, P. (2018). Strange Floods: The Upper Tail of Flood Peaks in the
650   United States. Water Resources Research, 54(9), 6510–6542. https://doi.org/10.1029/2018WR022539

Spearman, C. (1904). The proof and measurement of association between two things. American Journal of Psychology, 15(1), 72–101. https://doi.org/10.2307/1412159

Székely, G. J., Rizzo, M. L., & Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. Annals of Statistics, 35(6), 2769–2794. https://doi.org/10.1214/009053607000000505

655   Tarasova, L., Basso, S., & Merz, R. (2020). Transformation of Generation Processes From Small Runoff Events to Large Floods. Geophysical Research Letters, 47(22). https://doi.org/10.1029/2020GL090547

Tarasova, L., Basso, S., Zink, M., & Merz, R. (2018). Exploring Controls on Rainfall-Runoff Events: 1. Time Series-Based Event Separation and Temporal Dynamics of Event Runoff Response in Germany. Water Resources Research, 54(10), 7711–7732. https://doi.org/10.1029/2018WR022587

660   Tashie, A., Pavelsky, T., & Band, L. E. (2020). An Empirical Reevaluation of Streamflow Recession Analysis at the Continental Scale. Water Resources Research, 56(1), 1–18. https://doi.org/10.1029/2019WR025448

Tashie, A., Pavelsky, T., & Emanuel, R. E. (2020). Spatial and Temporal Patterns in Baseflow Recession in the Continental United States. Water Resources Research, 56(3), 1–18. https://doi.org/10.1029/2019WR026425

Vázquez, A., Oliveira, J. G., Dezsö, Z., Goh, K. Il, Kondor, I., & Barabási, A. L. (2006). Modeling bursts and heavy tails in
665   human dynamics. Physical Review E - Statistical, Nonlinear, and Soft Matter Physics, 73(3), 1–19. https://doi.org/10.1103/PhysRevE.73.036127

Villarini, G., & Smith, J. A. (2010). Flood peak distributions for the eastern United States. Water Resources Research, 46(6), 1–17. https://doi.org/10.1029/2009WR008395

Villarini, G., Smith, J. A., Baeck, M. L., Marchok, T., & Vecchi, G. A. (2011). Characterization of rainfall distribution and
670   flooding associated with U.S. landfalling tropical cyclones: Analyses of Hurricanes Frances, Ivan, and Jeanne (2004). Journal of Geophysical Research Atmospheres, 116(23). https://doi.org/10.1029/2011JD016175

Volpi, E., Fiori, A., Grimaldi, S., Lombardo, F., & Koutsoyiannis, D. (2019). Save hydrological observations! Return period estimation without data decimation. Journal of Hydrology, 571(February), 782–792. https://doi.org/10.1016/j.jhydrol.2019.02.017

675   Wang, H., Merz, R., Yang, S., Tarasova, L., & Basso, S. (2022). Emergence of heavy tails in streamflow distributions: the role of spatial rainfall variability. Advances in Water Resources Journal, 171(104359). https://doi.org/10.1016/j.advwatres.2022.104359

Wietzke, L. M., Merz, B., Gerlitz, L., Kreibich, H., Guse, B., Castellarin, A., & Vorogushyn, S. (2020). Comparative analysis of scalar upper tail indicators. Hydrological Sciences Journal, 65(10), 1625–1639. https://doi.org/10.1080/02626667.2020.1769104

Wilson, P. S., & Toumi, R. (2005). A fundamental probability distribution for heavy rainfall. Geophysical Research Letters, 32(14), 1–4. https://doi.org/10.1029/2005GL022465

Wittenberg, H. (1999). Baseflow recession and recharge as nonlinear storage processes, 726(January 1998), 715–726.

Zhang, X. S., Amirthanathan, G. E., Bari, M. A., Laugesen, R. M., Shin, D., Kent, D. M., et al. (2016). How streamflow has changed across Australia since the 1950s: Evidence from the network of hydrologic reference stations. Hydrology and Earth System Sciences, 20(9), 3947–3965. https://doi.org/10.5194/hess-20-3947-2016