EGUSPHERE-2023-660: Inferring heavy tails of flood distributions from common discharge dynamics

Response to Reviewers

We thank the Editor and the two Reviewers for providing valuable comments and suggestions. We have incorporated their comments in the revised version of the manuscript and answered each point below. The Reviewers' comments are in black font while our replies are in blue font. Text in the original version is reported in red, with revisions in dark-blue italics. L_o refers to the line number of the original version of the manuscript, whereas L_r refers to the revised version.

Editor:

Two expert reviewers have provided generous and substantial comments on this version of the manuscript, and the authors have provided responses that clearly indicate positive engagement in the review procedure. Both reviewers find the manuscript to be interesting but also, in places, lacking in clarity. My decision is therefore that a major revision is required paying particular attention to details highlighted in the discussion.

Thank you for the concise review summary. We have addressed all the comments provided by individual reviewers and enhanced clarity in the specific areas that were highlighted during the discussion, as outlined below.

Reviewer #1:

In this work, the authors investigate a new heavy-tail (i.e., extreme power-law tail) index a originating from the PHEV riverflow hydrograph model, which is compared to common practices in the literature of extreme-tail fitting, and it is applied to daily streamflow records in Germany. In my opinion, the manuscript needs multiple major and minor revisions before it can be evaluated for publication, and then, I believe it can serve as a review paper that compares several methods in the literature for extreme-tail oriented probability distribution fitting.

Thank you for the summarized review. In this study, we have identified a primary factor influencing heavy-tailed flood behavior (i.e., discharge dynamics embodied by the hydrograph recession exponent *a*) through theoretical analysis and substantiated our findings with empirical data. We have also discussed literature encompassing various methodologies for detecting heavy-tailed flood behavior. This discussion serves as a valuable point of reference to benchmark our novel results against standard practices and highlight the significance of our work, as it is typically done in any original research paper. We therefore kindly disagree with the Reviewer, who suggests our study as a review paper. In response to specific comments from the reviewer, we have provided our replies below each comment.

1) The a index seems to be based on some assumptions related to precipitation probability distribution and to rainfall-runoff model. Although the Poisson distribution has been used in the literature to fit

precipitation records (e.g., see Cox and Isham, 1988), rainfall extremes are shown to exhibit heavy-tail behaviour. Please see one of the first studies in the literature that suggest to use the EV2 distribution for rainfall extremes, through theoretical (i.e., physically-based) and empirical (I think is the first global-scale analyses on rainfall extremes) reasoning. This selection of distribution has been adopted by many researchers, and has been verified by many studies in the literature. Please see a recent extensive review on the rainfall-extremes in Koutsoyiannis (2022), where methods are also described how to adjust this to perform well in even short records.

Cox, D.R. and V. Isham, A simple spatial-temporal model of rainfall, Proceedings of the Royal Society London A415, 317–28, 1988.

Koutsoyiannis,D., Statistics of extremes and estimation of extreme rainfall, 1, Theoretical investigation, Hydrological Sciences Journal, 49 (4), 575-590, doi:10.1623/hysj.49.4.575.54430, 2004a.

Koutsoyiannis, D., Statistics of extremes and estimation of extreme rainfall, 2, Empirical investigation of long rainfall records, Hydrological Sciences Journal, 49 (4), 591-610, doi:10.1623/hysj.49.4.591.54424, 2004b.

Koutsoyiannis, D., Stochastics of Hydroclimatic Extremes - A Cool Look at Risk, Εκδοση 2, ISBN: 978-618-85370-0-2, 346 pages, doi:10.57713/kallipos-1, Kallipos Open Academic Editions, Athens, 2022.

We sincerely appreciate the reviewer for bringing attention to the significant contributions made by previous studies in advancing the simulation of observed rainfall extremes through the application of specific probability distributions. In contrast to simulating rainfall extremes, we aim to capture the characteristics of daily rainfall, and the reasons for this choice are as follows:

The object of this study is the ensemble of runoff events and floods in river basins, which determines their probability distributions. It is crucial to note that rainfall extremes do not necessarily translate into streamflow extremes, as demonstrated in numerous studies (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) identified this fact and proposed the influence of catchment responses and storages. Sharma et al. (2018) supported this notion, providing additional evidence and discussion that highlighted the limited correspondence between increased rainfall extremes and flooding. They found that factors such as decreased antecedent soil moisture and declining snowmelt work in conjunction with rainfall to modulate flood events. In summary, while rainfall is the primary contributor to runoff, the emergence of extreme floods is largely determined by catchment responses and water balance, as recently stressed by a thorough review of the state of knowledge (Merz et al., 2022). Given these premises, an appropriate approach for describing runoff and its extremes should be rooted in the dynamics of soil moisture within catchments. This necessitates the use of distributions which describe typical (e.g., daily) rainfall features (i.e., its frequency and magnitude) contributing to soil moisture dynamics , as employed in this study, rather than solely focusing on extreme rainfall events.

Particularly, the framework used in this work is accomplished by stochastically representing the effective contribution of rainfall, which refers to the fraction of rainfall that infiltrates into the hillslopes, eventually reaching the outlet through the channel network. The representation incorporates classical hydrological processes, i.e., soil moisture dynamics in the top layer storage primarily involve rainfall-related processes, including infiltration as a positive contribution and loss through evapotranspiration, direct surface flow, and/or deep percolation, which reduce rainfall's contribution to the storage. The Poisson distribution has been commonly employed to model precipitation records, as noted by the reviewer. Consistent with this, we characterize daily precipitation (not rainfall extremes) as a Poisson process. Specifically, we utilize a Poisson process to simulate the temporal occurrence of rainfall events, while modeling rainfall depths as exponentially distributed, following the approach proposed by Botter et al. (2007). The effective contribution from rainfall to runoff is determined by applying an exceedance threshold for the soil moisture.

This method offers a physically-based description of the role played by nonlinear hydrological responses and catchment water balance, which are highly relevant to the occurrence of heavy-tailed floods (Merz et al., 2022). As a result, it facilitates the investigation of the objectives outlined in this study.

Inspired by the Reviewer's comment, we have enhanced the introductory section by introducing the following paragraph after $L_0$49, ahead of delving into the presentation of the physically-based method employed in this study:

L_r50-60: "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 should be rooted in the dynamics of soil moisture and rainfall-runoff processes within catchments."

2) Additionally, the proposed distribution by the authors for the streamflow process (that combines exponential and power-type expressions) resembles the Pearson-III distribution, which has been used to fit streamflow records (e.g., Buckett and Oliver, 1977), but again it is not always recommended for the streamflow extreme-tail (e.g., Anghel and Ilinca, 2023), whereas, it has been shown that, for the parent distribution of streamflow (and by accounting for the impact of correlation among streamflow records through higher-order moments), a pareto-distribution type (i.e., the so-called Pareto-Burr-Feller probability distribution, which is a generalized form of the Pareto IV or Burr XII distribution) seems to adequately fit streamflow records even in a global-scale (please see such analysis with thousand of streamflow gauges in Dimitriadis et al., 2021).

Anghel, C.G., and C. Ilinca, Evaluation of Various Generalized Pareto Probability Distributions for Flood Frequency Analysis, Water, 15, 1557, 2023.

Buckett, J., and F.R. Oliver, F.R., Fitting the Pearson type 3 distribution in practice, Water Resources Research, 13(5), 851–852, 1977.

Dimitriadis, P., D. Koutsoyiannis, T. Iliopoulou, and P. Papanicolaou, A global-scale investigation of stochastic similarities in marginal distribution and dependence structure of key hydrological-cycle processes, Hydrology, 8 (2), 59, doi:10.3390/hydrology8020059, 2021.

Thank you for the comment. We would like to emphasize that the adopted distribution of streamflow (Equation (1)) is distinct from the Pearson-III distribution. While the Pearson-III distribution comprises a power law and an exponential distribution, Equation (1) includes a power law and two stretched exponential distributions. It is important to note that the stretched exponential distribution offers greater flexibility in terms of tail behavior compared to the exponential distribution. Depending on its parameters, the stretched exponential distribution can exhibit either a light-tailed or heavy-tailed behavior, whereas the exponential distribution always exhibits a light-tailed behavior. This constitutes a significant difference between the distribution of streamflow used in our study (PHEV equation) and the Pearson-III distribution, particularly regarding tail behavior, which is a focal point of this analysis.

Regardless of these differences, it is essential to clarify that this study does not rely on assuming a standard probability distribution to represent streamflow (as the Pearson-III, Pareto IV or Burr XII may be). Instead, starting from a stochastic description of daily rainfall as a Poisson process and a mathematical representation of key hydrological processes, we derive the resulting probability distributions of daily flows, ordinary peak flows and flow maxima, which assume the forms described in equations (1), (2), (3) and summarized in the PHEV framework. In this regard, the PHEV approach is similar to the results of large simulation ensembles of deterministic hydrological models forced by diverse realizations of hydro-meteorological scenarios and boundary/antecedent catchment conditions.

We acknowledge that the description of precipitation and runoff generation mechanisms incorporated in PHEV does not encompass the entirety of potential rainfall-runoff processes. However, the chosen representation is firmly rooted in established scientific frameworks that have undergone extensive testing through numerous case studies over the past decades (see references at L_r83-L_r85). Its application to high flows has also been validated and documented in previous studies, e.g., Basso et al., *Geophysical Research Letters* (2016); *Environmental Research Letters* (2021); and *Nature Geoscience* (2023).

To clarify this matter and highlight its representation of high flows, we have added the following paragraph in the revision:

L_r96-102: "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."

Meanwhile, we have recognized that some of the terms used in the original manuscript might be misleading. Therefore, we have made the necessary corrections to accurately reflect the intended descriptions as follows:

$L_0$88-$L_0$90: "The tail behavior is in this case determined by both a power law and an exponential function, indicating that the probability decreases faster than an exponential but slower than a power law. When a > 2, both the exponential terms converge to a constant value of one as q increases…"

L_r110-L_r112: "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…"

3) By observing the results from the fitting illustrated in the Figure 3, it can be observed that there is a smaller variability on the ξ-index than in the a-index for longer lengths; please consider discussing this observation and in what cases the a-index can offer a higher statistically significant extreme-tail fitting (perhaps only in the small-length records?).

Thank you for your comment. We appreciate the feedback and acknowledge that our work, in line with previous studies (Papalexiou and Koutsoyiannis, 2013), provides support for the stability of the ξ-index when the data length is adequately long. However, it is important to note that reduced variability of the ξ-index in analyses performed with longer data lengths does not necessarily denote better identification of heavy-tailed behavior by the ξ-index than the a-index. The width (whiskers) of the boxplots for different indices (i.e., the variability mentioned by the reviewer) are not on the same scale (because each index has its own scale), making them incomparable. The estimated values of each index can thus be only compared within the same index.

In light of this, Figure 3 shows that the a-index exhibits consistent categorization into heavy/nonheavy-tailed flood behavior across all data lengths. The variability of the ξ-index notably diminishes with longer data series, as stressed by the reviewer. However, the categorization provided by the ξ-index becomes more inconsistent with diminishing length of the available data To elucidate this matter further, we introduce an additional subfigure (Figure 3d) illustrating the consistency of identified tail behavior against identification based on the entire data record. This clarifies that the differences in consistency between the two indices (ξ-index and a-index) are primarily observed in analyses with data length shorter than 10 years in this study. For data series longer than 10 years, both indices exhibit similar and increasing consistency, with the ξ-index slightly outperforming the a-index.

We apologize for any ambiguity stemming from the previous figure presentation. We improved the revised manuscript by:

1. Introducing Figure 3d and the accompanying caption:

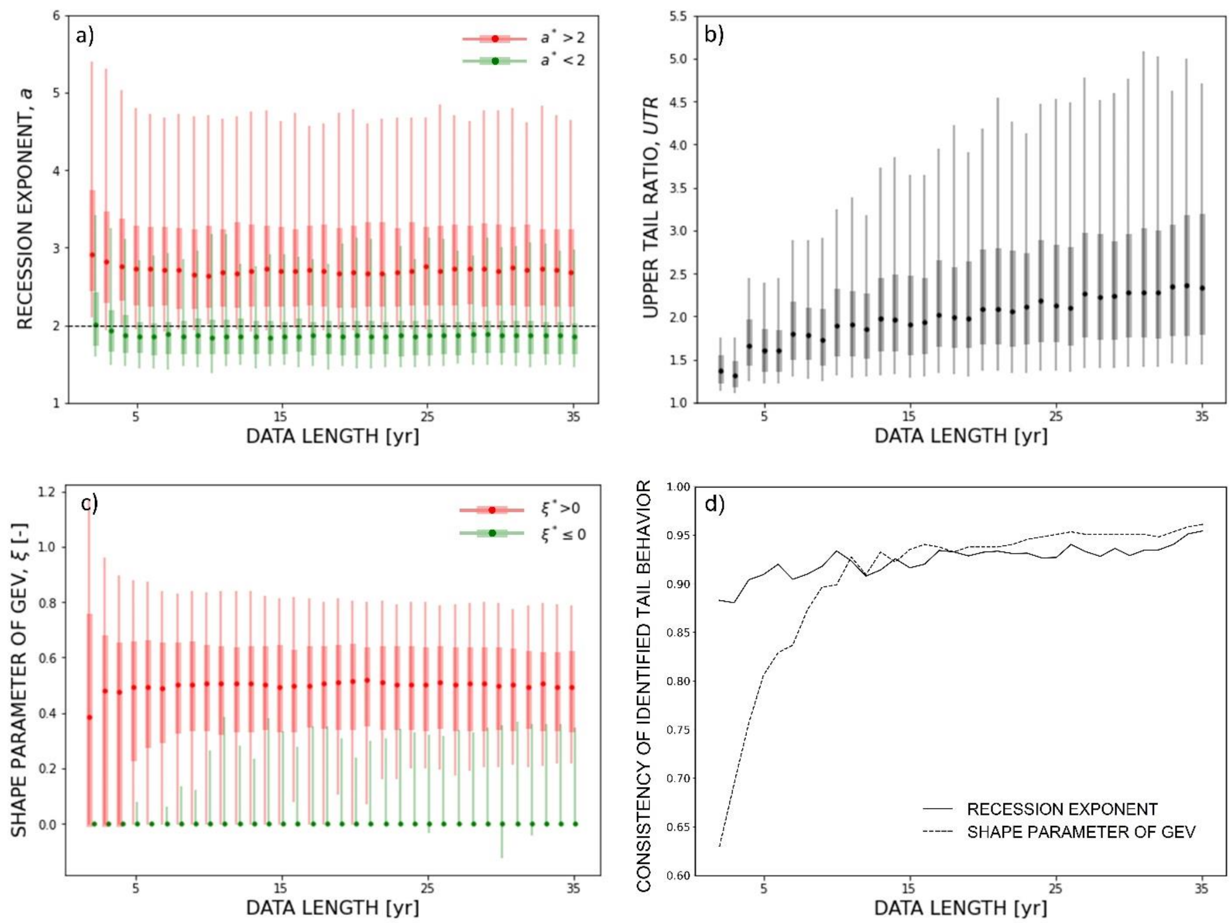

Figure 3. Stability of the categorization of case studies into heavy/nonheavy-tailed flood behavior for decreasing data lengths. Estimates of three different indices of tail behavior as a function of data length. (a) Hydrograph recession exponent a (i.e., the proposed index of this study). Two frequently 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 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.*

2. Enhancing our prior explanation in the Results Section by revising $L_0$251-$L_0$256 to *L_r289-L_r303*:

$L_0$251-$L_0$256: "Fig. 3c depicts the categorization of tail behavior using GEV shape parameter estimates. The results show that to achieve stable and consistent categorization of heavy-tailed behavior, a minimum test data length of 5 years is recommended. Underestimation of heavy-tailed behavior occurs with shorter data lengths. Although the median values of ξ range from 0.39 to 0.52 for the heavy-tailed cases and is equal to 0 for the nonheavy-tailed cases, the coefficient of variation shows some variation across the test data length, ranging from 0.37 to 1.03 for heavy-tailed cases. The coefficient of variation is not applicable for nonheavy-tailed cases due to their zero mean values."

L_r289-L_r303: "Fig. 3c illustrates the categorization of tail behavior using GEV shape parameter estimates. The results indicate that ξ estimates 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 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 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."

3. Adding a comparison between the a-index and ξ-index in the Discussion Section:

L_r382-L_r389: "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."

4) It is my understanding that the whole (not peaks-over-threshold and maxima) streamflow timeseries are fitted to equation (1) from where the a-index is estimated. If this is true, please discuss the method used to fit this distribution to data. For example, is it through the method of moments, where an n-number of moments is estimated from data to estimate the n-number of parameters of the distribution or through the method of curve-fitting, where one estimates the parameters of the distribution that result in the smaller small-square-error between the theoretical and empirical distribution?

We regret the misunderstanding made clear by this comment. At the same time, we are also glad that it gives us the chance to clarify our approach. We would like to stress that we do not fit Equation (1) to the empirical distribution of observed streamflow. Consequently, no method is used for fitting the distribution to the data and determine its parameters. Equation (1) serves as a physically-based framework that helps us understand the role of *a* in determining the tail behavior. The value of *a* is not determined through a fitting process. Rather, it is estimated from the analysis of hydrograph recessions, following a widely applied approach (see, e.g., Biswal and Marani, 2010; Jachens et al. 2020) as stated in the original manuscript at $L_0$115-$L_0$116. In particular, a power law is used to represent hydrograph recessions of a single event *i*,

$$\frac{dq}{dt} = -K_i \cdot q^{a_i}$$

Where *q* denotes the streamflow, *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.

We apologize for any confusion caused by our previous statement and have enhanced our description of *a* estimation in the Section of data and parameter estimation:

$L_0$115-$L_0$116: "We estimated the hydrograph recession exponent a for each case study as the median value of the exponents of power law functions fitted to $dq/dt{-}q$ pairs of individual hydrograph recession (Jachens et al., 2020; Biswal, 2021)."

L_r142-L_r151: "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."

Reviewer #2:

In this article, the author investigated a new method for representing the probability distribution function of flood with a heavy tail. A new method is to fit a power raw function to recession parts of hydrographs and then the derived power parameter *a* is utilized for the probability distribution function of flood as the reviewer's understanding. The result showed a high capability of power parameter a derived from recession parts of hydrographs to represent the power parameter of the probability distribution function for streamflow data including flood data. The advantage of this bland new approach is to utilize recession parts of hydrographs, which means usage of restricted data but numerous numbers of data. Instead of using all the data, this approach enables us to handle a huge amount of data to reduce uncertainty of flood estimation. However, another uncertainty, which is the representativeness of the power parameter from the recession hydrograph to the probability distribution function of the flood, rises. Moreover, it must be well discussed whether a power parameter from a recession hydrograph could be considered as a power parameter of a probability distribution function.

We thank the Reviewer for the positive feedback and suggestions, which we have addressed point by point below.

1) Please discuss how the power parameter which is fitted to hydrograph recession can be applied to the power parameter of the distribution function of flood.

Thank you for your comment. In this paper, we present mathematical reasons of why the hydrograph recession exponent may serve as an indicator of heavy-tailed flood behavior, and support them by means of empirical data. The mathematical derivations are based on the description of key hydrological processes leading to runoff generation embedded in the physically-based extreme value distribution (PHEV) framework. In this context, the tail behaviors of daily flow, ordinary peak flow and flow maxima distributions are solely determined by a power law function (i.e., heavy-tailed behavior) when the hydrograph recession exponent is above two. We discussed the case of daily flow distributions in the main text ($L_0$60-$L_0$93) and emphasized that the same critical value of the recession exponent can also be applied to identify heavy-tailed behavior for ordinary peak flows and flood distributions, with detailed explanations originally provided in Appendix A (originally noted in $L_0$94-$L_0$98). Additionally, we highlighted that the observations support the theoretical findings for all three cases proposed in this work, i.e., streamflow distributions, ordinary peak distributions, and flood distributions, which are discussed in figures 1 and 2 in the main text.

To address the reviewer's comment, we have integrated the mathematical explanations regarding ordinary peak flows and flood distributions into the main text to enhance their visibility:

L_r91-L_r95: "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):

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

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

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

L_r117-L_r127: "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 $q \longrightarrow +\infty$ 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 $C_3 = \lambda\tau C_2$)

$$\lim_{q\to+\infty} p_j(q) = \lim_{q\to+\infty}\left\{C_2\cdot \overbrace{\underbrace{q^{1-a}}_{\mapsto 0}}^{\mapsto 0}\overbrace{\underbrace{\left(e^{\frac{-1}{\alpha K(2-a)}\cdot q^{2-a}}\right)}_{\mapsto e^0=1}}^{\mapsto 0}\right\}, \qquad \begin{matrix}\text{for } 1<a<2\\[2em] \text{for } a>2\end{matrix} \tag{5}$$

$$\lim_{q\to+\infty} p_M(q) = \lim_{q\to+\infty}\left\{C_3\cdot \overbrace{\underbrace{q^{1-a}}_{\mapsto 0}}^{\mapsto 0}\overbrace{\underbrace{\left(e^{\frac{-1}{\alpha K(2-a)}\cdot q^{2-a}}\right)}_{\mapsto e^0=1}}^{\mapsto 0}\right\}, \qquad \begin{matrix}\text{for } 1<a<2\\[2em] \text{for } a>2\end{matrix} \tag{6}$$

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."

This theoretical justification of why the recession exponent is an index of heavy-tailed flood behavior is now presented in Section 2 (*L_r71-L_r127*) . Furthermore, Figures 1 and 2 in the Results Section offer data validation supporting this representation (*L_r176-L_r253*).

We have also better discussed the justification of the hydrograph recession exponent representing heavy-tailed flood behavior by emphasizing the nonlinearity of catchment response as a plausible control of heavy-tailed behavior. The discussion in $L_0$302-$L_0$305 is now enhanced as *L_r350-L_r354*:

$L_0$302-$L_0$305: "Based on a mechanistic description of hydrological dynamics and validation with observations from a large dataset, our findings directly demonstrate that heavy-tailed flood behaviors emerge as a result of the nonlinearity of the catchment hydrologic response, which can thus be used as a metric to assess tail behaviour of flood distributions from common streamflow dynamics."

L_r350-L_r354: "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."

2) In the text, a parameter is mentioned as "a physically-based index". However, this index was derived by fitting the raw power function to the recession part of the hydrograph. In this sense, using the word "physically" may lead to confusion. Rather, "recession-derived" is suited, the reviewer thinks (for example, "a recession-based exponent").

Thank you for the suggestion. In response to the Reviewer's comment, we have improved the clarity of the writing by avoiding phrases like "physically-based index."

1. In response to the reviewer's comment, we have emphasized that the proposed index is derived from recession analysis and is expressed as an empirical exponent:

L_r142-L_r151: "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."

2. We have also implemented the following modifications:

$L_0$98-$L_0$99: "we propose the hydrograph recession exponent a as a physically-based index for identifying heavy-tailed flood behavior"

L_r125-L_r127: "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."

$L_0$102: "To test the proposed physically-based index of heavy-tailed flood behavior"

L_r129: "To test the proposed index of heavy-tailed flood behavior"

$L_0$239: "The physically-based index provides consistent categorization"

L_r277: "The proposed index provides consistent categorization"

$L_0$311-$L_0$312: "The physically-based index proposed in this study offers a solution to these limitations. Our proposed index…"

L_r360- L_r361: "The proposed index finds a solution to these limitations through a mathematical description of hydrological processes. Such an index…"

$L_0$318: "Overall, our physically-based index offers a promising solution"

L_r367: "Overall, our proposed index offers a promising solution"

3) In L147, the skewness of the histogram of KS statistics was mentioned. What does it mean in this study? More explanation is effective for the readers.

Thank you for the suggestions. In this study, we used the KS statistic to assess how well the power law describes the empirical data, indicating whether the data exhibits heavy-tailed behavior or not. A KS statistics skewed towards lower values, therefore, suggests a stronger presence of heavy-tailed behavior in the observations.

We found that this was indeed the case for instances with recession exponents above two. This finding aligns with the indications provided by the recession exponent, demonstrating a consistency between the suggested index and the observed heavy-tailed behavior in the data.

Furthermore, we observed significant differences in the skewness of the KS statistics between cases with recession exponents above two and those with recession exponents below two. Data from the former group are more likely to display heavy-tailed behavior. Conversely, data from the latter group (i.e., with recession exponents below two) are less likely to show heavy-tailed behavior.

These results further support the effectiveness of our proposed index in identifying heavy-tailed behavior in the data. In response to the Reviewer's comment, we have made improvements in our description to enhance clarity for readers:

$L_0$145-$L_0$146: “Then, we use the KS statistic κ to evaluate the reliability of the fitted power law distribution in describing the data ($\kappa \in [0,\infty]$, $\kappa=0$ denotes the highest reliability).”

L_r179-L_r182: “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.”

4) For figs. 1d-1f, why does the accuracy index become one when the threshold of KS statistics becomes larger? The large threshold of KS statistics means considering almost all cases. However, the index shown on the vertical axis is the probability of a>2, which is never all the cases.

We believe there is a misunderstanding regarding the accuracy calculation when considering the threshold of KS statistics. The accuracy does not become one when the threshold of KS statistics becomes larger, rather the opposite. This is indicated by the inverse x-axis in Figure 1. In fact, the accuracy tends to approach one when the threshold of KS statistics becomes smaller.

By lowering the threshold of KS statistics, we impose a stricter criterion for including cases in the calculation of the conditional probability of accuracy (described in L157-L159). As depicted in Figures 1d-1f, the accuracy approaches one when the threshold of KS statistics is set to a very small value. This indicates that only cases with very high reliability (i.e., cases with KS statistics below this very small value) are taken into account for the accuracy calculation. Among these cases, almost every one exhibits a recession exponent above two, as evident from the accuracy being close to one.

Upon recognizing the potential for misleading interpretations, we have enhanced the clarity of the previous description:

L_0160-L_0161: “Notice that the smaller the κ_r threshold, the more reliable is the description of the data through power law distributions.”

L_r196-L_r200: “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).”

We greatly appreciate these valuable comments and have incorporated all the suggested modifications into a revised version of the manuscript. Additionally, we have provided a marked version that highlights these changes. The newly referenced literature has been included in the revised version accordingly.

References

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., 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

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

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

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

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

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

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)

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

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

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

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

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

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

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

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

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

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

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

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