HESS-2024-159: Constructing a geography of heavy-tailed flood distributions: insights from common streamflow dynamics

We thank the Editor and the Reviewer for providing comments. We have carefully incorporated their suggestions in the revised version of the manuscript and answered each point below.

The Editor's and Reviewers' comments are in black font with gray shading while our replies are in blue font. Text in the original manuscript is reported in red, with Lo referring to the line number in the previous version of the manuscript. Text in the tracked-revised version is presented in dark-blue, with Lr referring to the line number in the revised manuscript.

======

Response to the Editor

Following Referee's comments, I invite the authors to comply with the requests listed therein.

We have carefully addressed each of the comments and provided detailed responses. The manuscript has been revised accordingly to incorporate the requested changes.

Furthermore, I suggest the authors to provide detailed comments on the effectiveness of estimating heavy tail distribution starting from a few years of observations.

We thank the editor for this comment, which makes us understand that some parts of the previous text (e.g., Sec. 3.1 and Sec. 4.1) were not clear and gives us the chance to clarify them.

The possibility of reliably assessing the tail behavior (heavy versus non-heavy) from a few years of daily streamflow observations has been demonstrated in a previous publication by Wang HESS 2023. There we showed that a description of key hydrological processes occurring in watersheds (namely, the stochastic occurrence of precipitation events, filtering by soil moisture in the root zone, water storage in the catchment and its subsequent release) indicate that the asymptotic behavior of the distribution of daily streamflows, ordinary peak flows, and floods (maximum flows in a time interval) is linked to the median value of the hydrograph recession exponent in the watershed. We also demonstrated that this simple indicator reliably pinpoint heavy-tailed flow distributions identified in a dataset of river basins in Germany. The same approach is also capable of predicting river basins where extreme floods are more likely to occur (which is a proxy for heavy-tailed flood behavior), as shown by Basso et al. Nature Geoscience 2023.

Unlike conventional approaches relying on statistical fitting of maxima or peaks-over-threshold (which often yield only a few data points over short periods), the identified link between ordinary streamflow dynamics (which embodies the underlying storage-discharge dynamics) and tail behaviors enables leveraging the wealth of information contained in daily streamflow records. Since these are many also in short data series, the approach is tantamount to increasing the sample size. In this the approach resembles recent metastatistical methods (e.g., Marani and Ignaccolo, 2015; Marra et al., 2023), which also utilize (although in different ways) ordinary hydrologic events to estimate extremes. Differently from metastatistical methods, though, the approach we use derive the indicator of tail behavior from a mechanistic description of key process of runoff generation.

Several studies showed that hydrograph recession attributes can be robustly estimated from relatively short series of daily streamflow (e.g., Chen and Krajewski, 2016; Biswal and Marani, 2010; Dralle et al., 2017; Tashie et al., 2019). Nonetheless, a lower data limit exists, under which it is not possible to obtain a reliable estimation of the hydrograph recession exponent. This was found to be 5 years for a set of rivers in Germany (Wang HESS 2023).

We have enhanced the clarity of all these points of the revised manuscript as below. Once again thank you for the comment.

L_r 67-72: "We demonstrated that this simple indicator reliably pinpoint heavy-tailed flow distributions identified in a dataset of river basins in Germany. Unlike conventional approaches relying on statistical fitting of maxima or peaks-over-threshold (which often yield only a few data points over short periods), the identified link between ordinary streamflow dynamics (which embodies the underlying storage-discharge dynamics) and tail behaviors enables leveraging the wealth of information contained in daily streamflow records."

L_r 685-687: "The same approach is also capable of predicting river basins where extreme floods are more likely to occur (which is a proxy for heavy-tailed flood behavior), as shown by Basso et al. (2023)."

L_r 689-693: “In fact, the recession exponent reflects catchment nonlinearity, a robust driver of heavy-tailed flood behavior (Fiorentino et al., 2007; Struthers and Sivapalan, 2007; Gioia et al., 2008; Rogger et al., 2012; Basso et al., 2015; Merz et al., 2022; Basso et al., 2023; Wang et al., 2023), and it can be robustly estimated from relatively short series of daily streamflow (e.g., Chen and Krajewski, 2016; Biswal and Marani, 2010; Dralle et al., 2017; Tashie et al., 2019).”

In particular, authors should clearly and concisely highlight the approximations and limitations of the methodolgy they outlined. They should e.g., clearly state that they assume that severe floods are caused by the same drivers and smaller floods.

Our approach is grounded in the assumption that the generation of runoff and floods in catchments results from the interaction between precipitation, infiltration, soil moisture dynamics, temporary retaining of water in the catchment storage and its final release as discharge. The methodology is applicable to all cases where the drivers of floods can be conceptualized in this way. This method may instead not be suitable for regions where the primary drivers differ significantly, such as:

- Regions where substantial accumulation of water in the form of snow occurs, where floods are primarily driven by melting processes rather than precipitation (notice, however, that Basso NatGeo 2023 shows the method to provide reliable results also under these conditions).
- Situations in which the role of soil moisture dynamics and the catchment water storage is bypassed, e.g., when very intense precipitation, impervious surfaces and drainage infrastructures primarily dictate flood responses.

These scenarios fall outside the scope of our framework and represent limitations of the methodology.
We now indicate these limitations at lines XXX of the revised version of the manuscript.

L_r 768-772: “We acknowledge that our methodology assumes flood generation arises from the interplay of precipitation, infiltration, soil moisture dynamics, temporary storage, and discharge. It is less applicable where these processes are bypassed, such as in regions dominated by snowmelt or where soil moisture and catchment storage play minimal roles (e.g., in cases of intense precipitation where impervious surfaces and drainage systems predominantly control flood responses). These represent inherent limitations of the framework.”

We hope this explanation clarifies the assumptions, applicability, and limitations of our methodology while highlighting its theoretical and practical strengths.

Overall, I still find the manuscript quite lenghty and difficult to read. Even though, from the first review round, the manuscript has been significantly improved substantially, I invite the authors to shorten it and improve its readability.

We have further refined and shortened the manuscript to improve its readability. Specifically, we have made adjustments based on Reviewer 1's first comment. A brief summary of the adjustments is provided below:

1. In the Introduction section, we have removed L_o 36-44, L_o 65-70, and consolidated parts from the Discussion section (L_o 661-664, L_o 690-694), refining them further.

2. In the Results section, we have removed L_o 566-573 and condensed the summary of L_o 482-492.

3. In the Discussion section, we have not only merged the paragraphs mentioned in the first point into the Introduction but also removed L_o 658-661, 664-666 and further condensed the lengthy paragraph L_o 698-719.

Further details can be found in our response to Reviewer 1's first comment.

Response to the Reviewer 1

This is the second-round review for this manuscript. The authors did a good job in responding to my previous concerns. The manuscript has been improved substantially. However, there remain certain issues to be further resolved before its publication.

Thank you for your thoughtful review and comments. Below, we have provided detailed responses to each point and made the necessary revisions to the manuscript accordingly.

Specific comments:

1) The manuscript remains a little bit lengthy and reading it through is not quite an enjoyable experience. Some of the texts in the discussion section can be concatenated into introduction. Some paragraphs in introduction can be discarded (Line 36-44). The same thing for the discussion section. In the Results section, please also be concise. Summarizing paragraphs such as Lines 482 onwards and Lines 566 onwards can be removed. Please carefully reconstruct these texts.

We appreciate the reviewer's valuable suggestions, and we have refined, consolidated, and restructured the manuscript as suggested. Specifically, the following adjustments have been made:

1. L_o 36-44 have been removed as suggested:

~~L_o 36-44: "To organize current knowledge on the drivers and underlying mechanisms of heavy-tailed flood distributions, Merz et al. (2022) conducted an extensive review of current studies and summarized their findings into nine hypotheses. Notably, they pointed out that while one might intuitively assume that heavy-tailed flood distributions are inherited from heavy-tailed rainfall distributions, the evidence does not always support this hypothesis. For instance, a study by McCuen and Smith (2008) revealed that cases with skewed rainfall distributions, implying longer and heavier tails, do not necessarily translate into skewed flood distributions. This finding is supported by similar results from Sharma et al. (2018), who discovered that although there has been a significant increase in rainfall extremes, a corresponding increase in flood extremes is not observed. Indeed, Gaume (2006) pointed out that the asymptotic behavior of flood distributions is primarily controlled by rainfall distributions only for situations with very large return periods."~~

2. L_o 482-492 has been summarized as follows:

~~L_o 482-492: "To sum up this section, we have identified the conjunction of dry periods and higher temperatures as crucial meteorological factors significantly contributing to the dynamics of catchment storage, thereby influencing the nonlinearity of hydrological responses. We refer to catchment storage sensu Kirchner et al. (2009) and Botter et al. (2009), i.e., the varying amount of water contained in a catchment between dry and wet periods.. This capacity is dynamic and depends on various factors, such as soil moisture states, precipitation, and evapotranspiration (Merz and Blöschl, 2009; Zhou et al., 2022). These findings shed light on the interplay between catchments and meteorological conditions in the manifestation of heavy-tailed flood behavior. We acknowledge that these results are based on overarching climate conditions and do not encompass all climate types, and achieving an equal number of study sites across various climate regions might not always be feasible. We should be mindful of potential bias caused by sample sensitivity, particularly in regions with a limited number of cases (e.g., Csa, BSh, BWk in this study). Expanding the number of study sites in these climate regions could strengthen the current understanding."~~

L_r 513-518: "To sum up this section, we identify the conjunction of dry periods and higher temperatures as crucial factors contributing to the dynamics of catchment storage, as defined by Kirchner et al. (2009) and Botter et al. (2009), refers to the variable water volume in a catchment between dry and wet periods, shaped by factors like soil moisture, precipitation, and evapotranspiration (Merz and Blöschl, 2009; Zhou et al., 2022). Since achieving an equal distribution of study sites across climate types is challenging, we should remain mindful of potential bias due to sample sensitivity, particularly in regions with limited cases (e.g., Csa, BSh, and BWk)."

3. L_o 566-573 have been removed as suggested:

~~L_o 566-573: "In summary, while heavy-/nonheavy-tailed behavior is generally consistent across seasons, there is a certain probability for cases to exhibit seasonality. This seasonality of inferred heavy-tailed behavior shows a dynamic pattern of increasing during the growing season and decreasing during the dormant season. Regions with pronounced temperature variations across seasons, particularly with higher temperatures in summer, and characterized by relatively evenly distributed rainfall throughout the year tend to display such dynamics. This highlights the importance of both evapotranspiration and the temporal characteristics of rainfall in shaping flood tail behavior across seasons, aligning with previous studies (Guo et al., 2014 ; Basso et al., 2023)."~~

4. In addition to the suggestions provided by the reviewer, we have further removed and condensed the following sections to enhance the overall readability of the manuscript:

[Introduction]

~~L_o 65-70: “Importantly, it has shown its capacity to provide robust estimates for both short and long data records. This is mainly because it infers heavy-tailed behavior from common discharge dynamics, which allows for a more effective use of information contained in the data. For instance, when working with a 10-year data series, only 10 samples are available for annual maxima analysis, while a much larger number of of recession events (an average of 400 in Wang et al., 2023) can possibly be used for estimating hydrograph recession exponents and inferring from this the flood tail behavior.”~~

[Combine descriptions from the Discussion into the Introduction]

~~L_o 661-664: “Climate conditions have been found shaping the catchment geomorphology (Wu et al., 2023) and river network dynamics (Ward et al., 2020) which contribute to the degree of catchment response nonlinearity (Biswal and Marani, 2010). Meanwhile, the changes in flood generation processes can significantly affect the frequency of large floods (Tarasova et al., 2023), potentially altering flood tail behavior.”~~

L_r 50-52: “For instance, climate conditions have been found shaping the catchment geomorphology (Wu et al., 2023) and river network dynamics (Ward et al., 2020) which contribute to the degree of catchment response nonlinearity (Biswal and Marani, 2010).”

~~L_o 690-694: “This study focuses a binary distinction between heavy and non-heavy-tailed distributions, rather than assessing the degree of heaviness, for two key reasons. First, identifying heavy-tailed distributions is inherently challenging. Second, the identification itself holds significant hydrological importance, regardless of the degree of heaviness. In fact, the presence of a heavy tail alone can serve as a critical warning of a relatively high probability of extreme events, making it a crucial issue also in studies using other indices (e.g., Macdonald et al., 2022).”~~

L_r 93-97: “This study emphasizes distinguishing between heavy and non-heavy-tailed distributions rather than quantifying tail heaviness. Identifying heavy-tailed distributions is inherently challenging, yet it is hydrologically significant. In fact, the presence of a heavy tail alone can serve as a critical warning of a relatively high probability of extreme events, making it a crucial issue also in studies using other indices (e.g., Macdonald et al., 2022).”

[Discussion]

~~L_o 658-666: “Regions with relatively uniform hydroclimatic conditions (the Atlantic Europe and Northern Europe) tend to exhibit a single/dominant propensity of flood tail behavior. Conversely, in regions characterized by diverse conditions (the continental United States), inferred flood tail behavior presents a balance between heavy- and nonheavy-tailed cases in terms of frequency and distribution. … Our findings in Figure 3e exemplify how different flood generation processes, influenced by the interplay of varied hydrometeorological and terrain conditions, result in opposite flood tail propensities.”~~

~~L_o 698-719: “In this study, we also found that the relationship between flood tail behavior and the expansion of catchment scales can be explained by changes in catchment nonlinearity, which are influenced by distinct flood generation processes. Previous studies have presented diverse perspectives on the relationship between flood tail behavior and catchment scales. While some studies have suggested that smaller catchments tend to exhibit heavier tails (e.g., Meigh et al., 1997; Pallard et al., 2009), others have noted a similar trend but with only a weak correlation (Merz and Blöschl, 2009; Villarini and Smith, 2010). Meanwhile, some studies have found no significant relationship between these two variables (Morrison and Smith, 2002; Smith et al., 2018). These studies have explored this topic without reaching a consensus, and many conclusions lack sufficient evidence and a clear understanding. In contrast, our findings (Figure 8) distinctly differentiate between various patterns by considering region classifications based on distinct dominant flood generation processes, thereby providing a mechanistic understanding. As a catchment expands, it encompasses more diverse terrain, which in turn facilitates a wider range of altitudes and flood types. In regions where tail behavior is primarily influenced by evapotranspiration dynamics (Figure 8c), the presence of diverse altitudes tends to moderate the effect of higher temperatures, reducing the influence of high evapotranspiration on the emergence of heavy tails. In regions where tail behavior is primarily controlled by snowmelt (Figure 8d) (mainly composed of catchments in Norway in this study), it has been shown that larger catchments are more likely to encompass a mix of flood types, including snowmelt-driven and rainfall-driven floods (Vormoor et al., 2016). Merz et al. (2022) suggested that heavier-tailed behavior in rainfall-driven floods tends to dominate in such mixed conditions. Our findings support this hypothesis by demonstrating an increase in tail heaviness as catchment area enlarges. In regions where heavy tails are pronounced due to the strong nonlinearity resulting from the interplay of uneven rainfall and high evapotranspiration, there is no significant relationship between catchment nonlinearity and catchment area (Figure 8b). This lack of relationship may be because the expansion of the catchment area does not appear to significantly enhance or reduce this interplay.”~~

L_r 739-746: “In this study, we found that the relationship between flood tail behavior and catchment scale can be explained by changes in catchment nonlinearity, influenced by distinct flood generation processes. Previous studies have suggested varied relationships between flood tail behavior and catchment scale, with some indicating smaller catchments exhibit heavier tails (Meigh et al., 1997; Pallard et al., 2009), while others report weak correlations (Merz and Blöschl, 2009; Villarini and Smith, 2010) or no significant relationship (Morrison and Smith, 2002; Smith et al., 2018). In contrast, our findings (Figure 8) clarify these patterns by considering region classifications based on dominant flood generation processes, which determine whether the nonlinearity of hydrological response increases or decreases when catchments expand."

2) Please justify using a constant threshold of 5 days to calculate the recession exponents across different basins.

Using a constant threshold of 5 days means that we include all cases with recession durations longer than 5 days, which accounts for the majority of cases as supported by numerous studies (e.g., Shaw and Riha, 2012; Thomas et al., 2015; Dralle et al., 2017; Tashie et al., 2020). This criterion excludes cases with very short recession durations (≤4 days), which typically represent catchments with very small drainage areas or extremely low permeability (e.g., urbanized regions). Notably, Thomas et al. (2015), who specifically studied hydrograph recession in high-population-density areas using 45 small to meso-scale catchments (ranging from 5.15 to 903.91 km², with a median size of 84.95 km²), set a minimum and constant threshold of 10 days for recession duration across basins—twice as strict as our criterion.

Furthermore, Chen and Krajewski (2016) (Section 3.2) demonstrated that analyzing recession exponents based on shorter durations (<6 days) could introduce significant biases. Given that our dataset contains no catchments smaller than 4 km² (median size: 1240 km²) and focuses on relatively natural catchments, we emphasize that this threshold is both well-supported by existing literature and necessary to avoid bias in subsequent analyses.

We regret the lack of clarity in our original description and appreciate the reviewer's comment, which has provided us with the opportunity to further clarify this aspect in the revised manuscript, as detailed below.

~~L_o 203-207: "To reduce noise from short events (Ye et al., 2014) and ensure sufficient sample sizes (i.e., a sufficient number of analyzed recessions) to obtain representative values of recession parameters (Shaw, 2016), it is common practice to set a minimum recession duration. Following previous studies, we did not vary this duration across basins and set it equal to 5 days (Biswal and Marani 2010; Shaw and Riha, 2012; Dralle et al., 2017; Jachens et al., 2020; Tashie et al., 2020a)."~~
L_r 217-223: "To reduce noise from short events (Ye et al., 2014; Chen and Krajewski, 2016) and ensure sufficient sample sizes (i.e., a sufficient number of analyzed recessions) to obtain representative values of recession parameters (Shaw, 2016), it is common practice to set a minimum recession duration. Based on the catchment sizes in our dataset, we selected 5 days as the minimum threshold for recession duration (i.e., analyzing all cases with recession durations longer than 5 days), a choice well-supported by existing literature (e.g., Biswal and Marani 2010; Shaw and Riha, 2012; Thomas et al., 2015; Chen and Krajewski, 2016; Dralle et al., 2017; Jachens et al., 2020; Tashie et al., 2020a)."

3) Line 258-261, the authors mention that the river gauge location is used when the boundary data is missing. How many are these, considering that the diversity in physiographic attributes within the basin is important in determining tail heaviness.

Among the four main study regions—Germany, the US, the UK, and Norway—all catchments in the first two regions have complete boundary information. However, 33 out of 82 catchments in the UK and 69 out of 82 catchments in Norway lack boundary data. While we acknowledge the potential bias introduced by this limitation, the relatively low climate variability (based on Köppen climate classification) in both the UK and Norway (see Figure 3) suggests that the impact of this bias is likely minimal. To address this, we have rewritten L_o 256-261 to enhance clarity.

~~L_o 256-261: "To determine the dominant hydroclimatic characteristic of each catchment, we overlay the Köppen climate map (Beck et al., 2018) and a derived potential evapotranspiration map (Zomer and Trabucco, 2022) with the river gauge and catchment boundary data. For the former, the most prevalent climate within the catchment (determined by overlapping areas within the boundary, or by the river gauge location if the boundary data is absent) is assigned as the representative feature. For the latter, we compute the catchment average value (or determined by the river gauge location if the boundary data is absent)."~~
L_r 279-287: "To determine the dominant hydroclimatic characteristics of each catchment, we overlaid the Köppen climate map (Beck et al., 2018) and a derived potential evapotranspiration map (Zomer and Trabucco, 2022) with river gauge and catchment boundary data. For the climate map, the most prevalent climate type within each catchment boundary was assigned as the representative feature. For potential evapotranspiration, the catchment average value was calculated. Of the 575 catchments in our dataset, 473 have boundary information. For the remaining 102 catchments lacking boundary data, representative features were determined based on the river gauge location. While we acknowledge the potential bias introduced by this limitation, it is worth noting that all 102 catchments are located in either the UK or Northern Europe. Due to the relatively low climate variability in these regions, the impact of this bias is expected to be minimal."

4) Line 244-246, what exactly is the criteria for “extreme observations”. This needs to be clarified.

We have inserted a brief definition to improve clarity (see below), as suggested by the reviewer. The detailed framework employed to address this issue has already been outlined in the earlier description within Section 3.2.

~~L$_o$ 245-246: “In other words, only the most extreme observations are analyzed to determine whether the empirical distributions exhibit power-law behavior in their tails.”~~

L$_r$ 266-269: “In other words, only the most extreme observations—defined as those located within the identified tail of the empirical distribution, where the tail is determined based on the optimized lower boundary calculated using the framework proposed by Clauset et al. (2009)—are analyzed to assess whether the empirical distributions exhibit power-law behavior in their tails.”

5) There are quite few gauges within a couple of Koppen climate zones (e.g., BSh, BWk, and Csa). I believe this would introduce biases in the analysis. Could the authors test what if these zones are removed?

We thank the reviewer for pointing this out. As shown in Figure 7a, regions with a limited number of case studies (BSh, BWk, and Csa) do not pass the significance test, and we have indicated this lack of significance in the figure. Beyond Figure 7a, the other two related results—Figure 4 and Table 1—remain unaffected even if these three groups are excluded, and the conclusions drawn from them would not change.

In the previous version manuscript (L$_o$ 496-498), we acknowledged the limitation regarding the small number of case studies, and we have further clarified and addressed this point in the revised version. Please refer to the following sections for details.

~~L$_o$ 496-498: “We should be mindful of potential bias caused by sample sensitivity, particularly in regions with a limited number of cases (e.g., Csa, BSh, BWk in this study). Expanding the number of study sites in these climate regions could strengthen the current understanding.”~~

L$_r$ 517-520: “…we should remain mindful of potential bias due to sample sensitivity, particularly in regions with limited cases (e.g., Csa, BSh, and BWk). While excluding these groups does not affect the conclusions of Figure 4 and Table 1, increasing the number of study sites in these climates could enhance understanding.”

======

We sincerely appreciate the comments of both the Editor and the Reviewer, and we hope the responses provided above satisfactorily address the concerns.

References

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

Chen, B., & Krajewski, W. (2016). Analysing individual recession events: sensitivity of parameter determination to the calculation procedure. Hydrological Sciences Journal, 61(16), 2887–2901. https://doi.org/10.1080/02626667.2016.1170940

Dralle, D. N., Karst, N. J., Charalampous, K., Veenstra, A., & Thompson, S. E. (2017). Event-scale power law recession analysis: Quantifying methodological uncertainty. Hydrology and Earth System Sciences, 21(1), 65–81. https://doi.org/10.5194/hess-21-65-2017

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

Marra, F., Amponsah, W., & Papalexiou, S. M. (2023). Non-asymptotic Weibull tails explain the statistics of extreme daily precipitation. Advances in Water Resources, 173, 104388. https://doi.org/https://doi.org/10.1016/j.advwatres.2023.104388

Shaw, S. B., & Riha, S. J. (2012). Examining individual recession events instead of a data cloud: Using a modified interpretation of dQ/dt-Q streamflow recession in glaciated watersheds to better inform models of low flow. Journal of Hydrology, 434–435, 46–54. https://doi.org/10.1016/j.jhydrol.2012.02.034

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., Scaife, C. I., & Band, L. E. (2019). Transpiration and subsurface controls of streamflow recession characteristics. Hydrological Processes, 33(19), 2561–2575. https://doi.org/10.1002/hyp.13530

Thomas, B. F., Vogel, R. M., & Famiglietti, J. S. (2015). Objective hydrograph baseflow recession analysis. Journal of Hydrology, 525, 102–112. https://doi.org/10.1016/j.jhydrol.2015.03.028

Wang, H., Merz, R., Yang, S., & Basso, S. (2023). Inferring heavy tails of flood distributions through hydrograph recession. Hydrol. Earth Syst. Sci, 27(24), 4369–4384. https://doi.org/10.5194/hess-27-4369-2023