Water Resources Research

REVIEW ARTICLE

10.1029/2021WR030506

Key Points:

- Heavy tail behavior of flood peak distributions may lead to surprise and high flood damage
- We propose nine hypotheses on the mechanisms causing heavy tails in flood peak distributions
- We review to which extent the current knowledge supports or contradicts these hypotheses

Correspondence to:

B. Merz,
bmerz@gfz-potsdam.de

Citation:

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*, *58*, e2021WR030506. https://doi.org/10.1029/2021WR030506

Received 29 MAY 2021
Accepted 5 JUN 2022

Author Contributions:

Conceptualization: Bruno Merz, Günter Blöschl
Visualization: Elena Macdonald
Writing – original draft: Bruno Merz, Stefano Basso, Svenja Fischer, David Lun, Ralf Merz, Björn Guse, Alberto Viglione, Sergiy Vorogushyn
Writing – review & editing: Bruno Merz, Stefano Basso, Svenja Fischer, David Lun, Günter Blöschl, Ralf Merz, Björn Guse, Alberto Viglione, Sergiy Vorogushyn, Elena Macdonald, Luzie Wietzke, Andreas Schumann

Understanding Heavy Tails of Flood Peak Distributions

Bruno Merz[1,2], Stefano Basso[3], Svenja Fischer[4], David Lun[5], Günter Blöschl[5], Ralf Merz[3], Björn Guse[1], Alberto Viglione[6], Sergiy Vorogushyn[1], Elena Macdonald[1], Luzie Wietzke[1], and Andreas Schumann[4]

[1]GFZ German Research Centre for Geosciences, Potsdam, Germany, [2]Institute for Environmental Sciences and Geography, University of Potsdam, Potsdam, Germany, [3]UFZ Helmholtz Centre for Environmental Research, Potsdam, Germany, [4]Department of Civil and Environmental Engineering, Ruhr University Bochum, Bochum, Germany, [5]Institute of Hydraulic Engineering and Water Resources Management, Technical University Vienna, Vienna, Austria, [6]Department of Environment, Land and Infrastructure Engineering, Politecnico di Torino, Torino, Italy

Abstract Statistical distributions of flood peak discharge often show heavy tail behavior, that is, extreme floods are more likely to occur than would be predicted by commonly used distributions that have exponential asymptotic behavior. This heavy tail behavior may surprise flood managers and citizens, as human intuition tends to expect light tail behavior, and the heaviness of the tails is very difficult to predict, which may lead to unnecessarily high flood damage. Despite its high importance, the literature on the heavy tail behavior of flood distributions is rather fragmented. In this review, we provide a coherent overview of the processes causing heavy flood tails and the implications for science and practice. Specifically, we propose nine hypotheses on the mechanisms causing heavy tails in flood peak distributions related to processes in the atmosphere, the catchment, and the river system. We then discuss to which extent the current knowledge supports or contradicts these hypotheses. We also discuss the statistical conditions for the emergence of heavy tail behavior based on derived distribution theory and relate them to the hypotheses and flood generation mechanisms. We review the degree to which the heaviness of the tails can be predicted from process knowledge and data. Finally, we recommend further research toward testing the hypotheses and improving the prediction of heavy tails.

Plain Language Summary Statistical distributions are used to estimate the probability of flood peaks, which in turn is needed for risk management and the design of flood protection. Flood peak distributions often show heavy tail behavior, that is, extreme floods are more likely to occur than would be predicted by commonly used distributions that have exponential asymptotic (light tailed behavior). This heavy tail behavior may surprise flood managers and citizens, as human intuition tends to expect light tail behavior. In this review, we summarize the knowledge about the causes of heavy flood tails. To this end, we discuss the flood generation processes in the atmosphere, catchment, and river system, that tend to generate heavy-tailed flood peak distributions.

1. Introduction

Floods often come as a surprise. Examples of extreme floods that have occurred unexpectedly and have led to disastrous socio-economic consequences abound in the literature (Merz et al., 2015). Figure 1 shows one example time series with such a surprising flood. The 2002 flood peak of the River Kamp, Austria, was about three times larger than the highest flood in the 100-year observational period before and has indeed caused enormous damage triggering desperate emergency measures in the region (Blöschl et al., 2006). From a statistical perspective, the occurrence of such an event is very unlikely if the extreme value behavior conforms to an asymptotically exponential (light-tailed) distribution. However, if the underlying probability distribution has a heavy tail, its occurrence is less unlikely. A heavy upper tail implies that the extreme values are more likely to occur than would be predicted by distributions with exponential asymptotic behavior, such as Exponential, Gamma, and Gumbel distributions (El Adlouni et al., 2008). Because human intuition tends to expect light tail behavior, processes that show heavy tail behavior often lead to surprise (Taleb, 2007).

Heavy-tailed behavior of flood peak distributions is of the highest relevance for flood design and risk management. Neglecting heavy tail behavior, if it exists, results in underestimating the probability of occurrence of extremes. This underestimation may result in biased flood management measures, such as underestimated dike

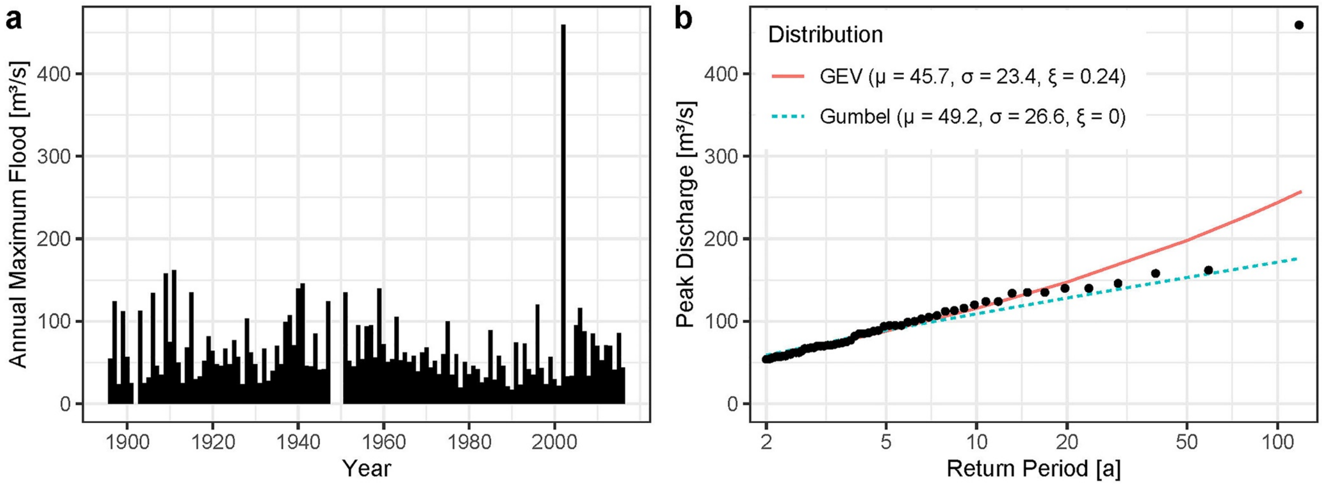

Figure 1. Times series of annual maximum streamflow (a) and flood frequency curve (b) of River Kamp, Austria. The flood peak in 2002 was roughly three times larger than the flood of record in the preceding 100 years. The fitted Generalized Extreme Value distribution shows heavy-tailed behavior. In addition, the light-tailed Gumbel distribution is fitted to the sample.

heights or inadequate insurance cover. These in turn will lead to much higher flood risks (damage times its probability) than flood management decisions based on unbiased probability estimates.

Unfortunately, the processes causing heavy tail behavior in flood peak distributions are not well understood and the literature on the subject is dispersed. There is some mechanistic understanding of the generation of heavy tail behavior of other geophysical phenomena, such as ocean rogue waves, wind gusts, and extreme precipitation (Böttcher et al., 2007; Toffoli et al., 2019; Wilson & Toumi, 2005), pointing to the non-linear interaction of component processes, but for the case of river floods the findings are unclear. In this review, we summarize in a coherent way the current knowledge of the processes that generate heavy tails in flood peak distributions.

We include atmospheric, catchment, and river system processes, such as rainfall mechanisms, runoff generation processes, and the construction of river embankments. We do not consider singularities, such as glacier lake outbreak floods or floods induced by massive landslides, which are caused by completely different mechanisms than the remaining floods in that catchment. An example of such an event is the Vajont disaster in 1963, where a landslide into a reservoir caused a flood wave that overtopped the dam and led to almost 2,000 fatalities (Delle Rose, 2012). The probability and magnitude of such singularities, or unrepeatable events, cannot be estimated by studying the other floods in the catchment. Their estimation requires assembling evidence about the relevant influencing factors and projecting them through a causal model (Hall & Anderson, 2002), applying, for instance, methods developed in the field of probabilistic risk analysis (Paté-Cornell, 2012). In this review, we propose nine hypotheses on generating mechanisms and discuss to which extent the current knowledge allows to support or falsify each hypothesis. We discuss the statistical conditions that may generate heavy tail behavior and relate them to the flood generation mechanisms. We explore the interplay of component processes and assess to which extent this information may guide the estimation of upper tail behavior. Component processes are defined as the elements of an aggregation. Examples of component processes are different flood types within a catchment that jointly constitute the sample of observed flood events or the aggregation of rainfall and runoff generation leading to flood peaks. Finally, we recommend future research on testing these hypotheses and improving the inference of upper tail behavior from data and process understanding.

2. Identifying Heavy Tails

2.1. Defining Heavy-Tailed Distributions

Our interest lies in the right tail of distributions, which characterizes, loosely speaking, the largest events and whether a distribution can be considered heavy-tailed or not. The most general definition (as it includes a wider class of distributions than other definitions) is that distribution with a distribution function F of a real-valued random variable X is heavy-tailed if and only if

10.1029/2021WR030506

Table 1
Heavy-Tailed Distributions and Their Characterization

Heavy-tailed	$E\left(e^{\lambda X}\right) = \infty \, \forall \lambda > 0$
Long-tailed	$\frac{\overline{F}(x+y)}{\overline{F}(x)} \rightarrow 1 \; for\; x \rightarrow \infty \, \forall y > 0$
Sub-exponential	$\lim_{x\rightarrow\infty} \frac{P(X_1+\cdots X_n > x)}{P(\max(X_1,\cdots,X_n) > x)} = 1$
Regularly varying (Asymptotic Pareto)	$\lim_{t\rightarrow\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-\alpha} \; \alpha > 0$
Exact Pareto tail	$F_P(x) = 1 - \left(\frac{u}{x}\right)^{\alpha} \quad \alpha > 0,\; x > u$
Stable (α-stable)	Pareto tail with $\alpha < 2$

Note. For a real-valued random variable X with distribution function F, we denote $\overline{F} = P(X > x)$. An overview of the different classes of tail heaviness for the most common distributions in hydrology is given in the study by El Adlouni et al. (2008).

$$E\left(e^{\lambda X}\right) = \int e^{\lambda X}\, dF(x) = \infty \; for\; all\; \lambda \in \mathbb{R}^{>0} \tag{1}$$

Equation 1 also refers to the moment generating function of X. If the expected value does not exist for all $\lambda > 0$, the random variable X is considered heavy-tailed. If any moments of X are not finite (do not exist), the moment generating function is not finite (does not exist) for all $\lambda > 0$ and X is heavy-tailed. An equivalent expression to Equation 1, for the random variables relevant in hydrology, is: a random variable X is heavy-tailed if the tail of its distribution function, $\overline{F}(x) = 1 - F(x)$, is a heavy-tailed function, meaning

$$\lim_{x\rightarrow\infty} \overline{F}(x)e^{\lambda x} = \infty \; for\; all\; \lambda > 0 \tag{2}$$

Equation 2 states that a random variable is heavy-tailed if the tail of the distribution function decays more slowly than any exponentially decreasing function, and therefore the behavior of the product, as x grows, is dominated by the exponential increase of $e^{\lambda x}$ in Equation 2.

Examples of heavy-tailed distributions are the Pareto distribution and the lognormal distribution. Subclasses of this definition are long-tailed distributions, sub-exponential distributions, asymptotic Pareto distributions (also called regularly varying), distributions with an exact Pareto tail, and (α-)stable distributions. All these definitions are nested with increasingly heavier tails, meaning that a long-tailed distribution is always heavy-tailed, a sub-exponential distribution is always long-tailed, and so on, but not vice versa. An overview of the definitions is given in Table 1.

Heavy-tailed distributions appear in flood statistics most commonly for two kinds of time series of flood peaks: annual maximum series (AMS) and Peak-over-Threshold series (POT). AMS consists of the largest flood peak per year. If the events are assumed to be independent, the asymptotic distribution converges to a Generalized Extreme Value (GEV) distribution with a distribution function

$$F(x) = exp\left(-\left(1 + \xi\, \frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right), \tag{3}$$

where $\xi \in \mathbb{R}^{\backslash 0}$ is the shape parameter, $\sigma > 0$ is the scale parameter, and $\mu \in \mathbb{R}$ is the location parameter (Fisher & Tippett, 1928). The constraint $1 + \xi\,(x - \mu)/\sigma > 0$ has to hold. The special case of $\xi = 0$ with

$$F(x) = exp\left(-exp\left(-\frac{x-\mu}{\sigma}\right)\right) \tag{4}$$

Corresponds to the light-tailed Gumbel distribution. If $\xi > 0$, the distribution represents the extreme value distribution of maxima of type II, also known as Fréchet distribution (Coles, 2001), and is heavy-tailed. Similar results also exist for short-range dependent data (Leadbetter & Rootzen, 1988).

POT series consists of all events above a threshold, which can be chosen hydrologically or statistically (Lang et al., 1999). Again, assuming independence of events and invoking the Pickands-Balkeema-de Haan Theorem (Balkema & de Haan, 1974; Pickands, 1975), the limit distribution for a given threshold μ is given by the Generalized Pareto Distribution (GPD):

$$F_\mu(x) = 1 - \left(1 + \frac{\kappa(x-\mu)}{\sigma}\right)^{-1/\kappa} \tag{5}$$

For $x > \mu$ if $\kappa > 0$ and $\mu < x < \mu - \sigma/\kappa$ for $\kappa < 0$, where $\kappa \in \mathbb{R}^{\backslash 0}$ is the shape parameter, $\sigma > 0$ is the scale parameter and $\mu \in \mathbb{R}$ is the location parameter. The distribution is heavy-tailed for $\kappa > 0$. Again, there exists a special case for $\kappa = 0$ with

$$F(x) = 1 - exp\left(-\frac{x-\mu}{\sigma}\right) \quad for\ x > \mu \tag{6}$$

Which can be interpreted as a shifted exponential distribution.

2.2. Quantifying the Upper Tail Behavior

There are different measures for quantifying the tail behavior of flood peak distributions, with the most widespread measure being either the shape parameter of the GEV distribution for AMS or the shape parameter of the GPD for POT (Coles, 2001). The shape parameters are directly related to the tail index $\alpha = 1/\xi$ and $\alpha = 1/\kappa$ for AMS and POT, respectively (The tail index refers to α in Table 1, rows 4 and 5). The GEV with a positive shape parameter belongs to the class of regularly varying distributions, while the Generalized Pareto distribution $\kappa > 0$ possesses an exact Pareto tail. The Gumbel and the exponential distribution are light-tailed distributions. For other distributions than the GPD and the GEV, such as the lognormal distribution, the GEV shape parameter is not an adequate indicator of the upper tail behavior. The skewness is also frequently used to characterize the behavior of the upper tail of flood peak distributions (McCuen & Smith, 2008). Although a skewed distribution is not necessarily heavy-tailed, skewness is often preferred as no assumption on the underlying distribution, besides the existence of the third moment, is required for the consistent estimation of skewness, as compared to more direct indices of tail heaviness. For the specific case of GEV, skewness and the shape parameter are directly related. Other quantitative metrics, such as the Upper Tail Ratio (Smith et al., 2018) or the Gini Index (Davidson, 2012) are less often used (see Wietzke et al. [2020] for a discussion on scalar upper tail indices). Further, graphical methods allow estimating the upper tail behavior from data, for example, by mean excess plots or the generalized Hill ratio plot (Embrechts et al., 1997), but more subjective choices must be made. In general, it is not a priori clear which estimator to use and a suitable choice depends on the possible values of α as well as the assumptions made for the distribution function (Embrechts et al., 1997).

There are various estimators for the different measures of upper tail behavior, but generally, very long time series are needed to reliably estimate these measures from observations (Papalexiou & Koutsoyiannis, 2013; Wietzke et al., 2020). Given the typical time series lengths in hydrology, light-tailed distributions can seem heavy-tailed and vice versa. When estimating an upper tail parameter, the sampling behavior depends on the actual underlying tail behavior, possible misspecifications of the model, the size of the sample, and the estimator itself. For example, the method of moments shows a strong underestimation for heavy-tailed distributions for a sample size smaller than 50, hence with this method the heavy-tailed behavior of flood time series of typical length might not be recognized, as pointed out by Koutsoyiannis (2004). Analyses of flood data of numerous catchments show patterns of the upper tail behavior of flood peak distributions that tend to be erratic or only weakly consistent in space (Bernardara et al., 2008; Merz & Blöschl, 2009). This behavior has been explained by the large influence of sampling uncertainty and the effect of single extreme floods, as the tail heaviness can be overestimated if a short time series contains an extreme event (Merz & Blöschl, 2009). Robust estimation techniques can reduce the impact of single extreme events on the estimation of the tail for small samples. Fischer and Schumann (2016) propose POT approaches based on monthly maxima and demonstrate their higher statistical robustness (Huber, 2004) against occasional extreme events compared to the approach using annual maxima series.

3. Statistical Perspectives on the Generation of Heavy Tails

Heavy tails of flood frequency curves can result from heavy tail characteristics of the component processes or may emerge from the non-linear superposition of light-tailed processes (Blöschl & Zehe, 2005). If we represent the component processes as random variables, we can investigate how different mathematical operations on these component processes influence the upper tail behavior of the resulting flood frequency curve. We organize the operations most relevant to flood generation into three groups: (S1) Arithmetic combination of random variables, (S2) Mixture of distribution functions, and (S3) Transformations of random variables (Figure 2). For all of these operations, two approaches are relevant from a statistical point of view: Asymptotic approximations, referring to limit theorems, and finite sample aggregations of random variables. Both approaches are discussed separately and further details can be found in Appendix A. In Section 3.4 we embed these statistical perspectives in the context of flood frequency analysis to discuss how they can contribute to better understanding the flood tail behavior.

10.1029/2021WR030506

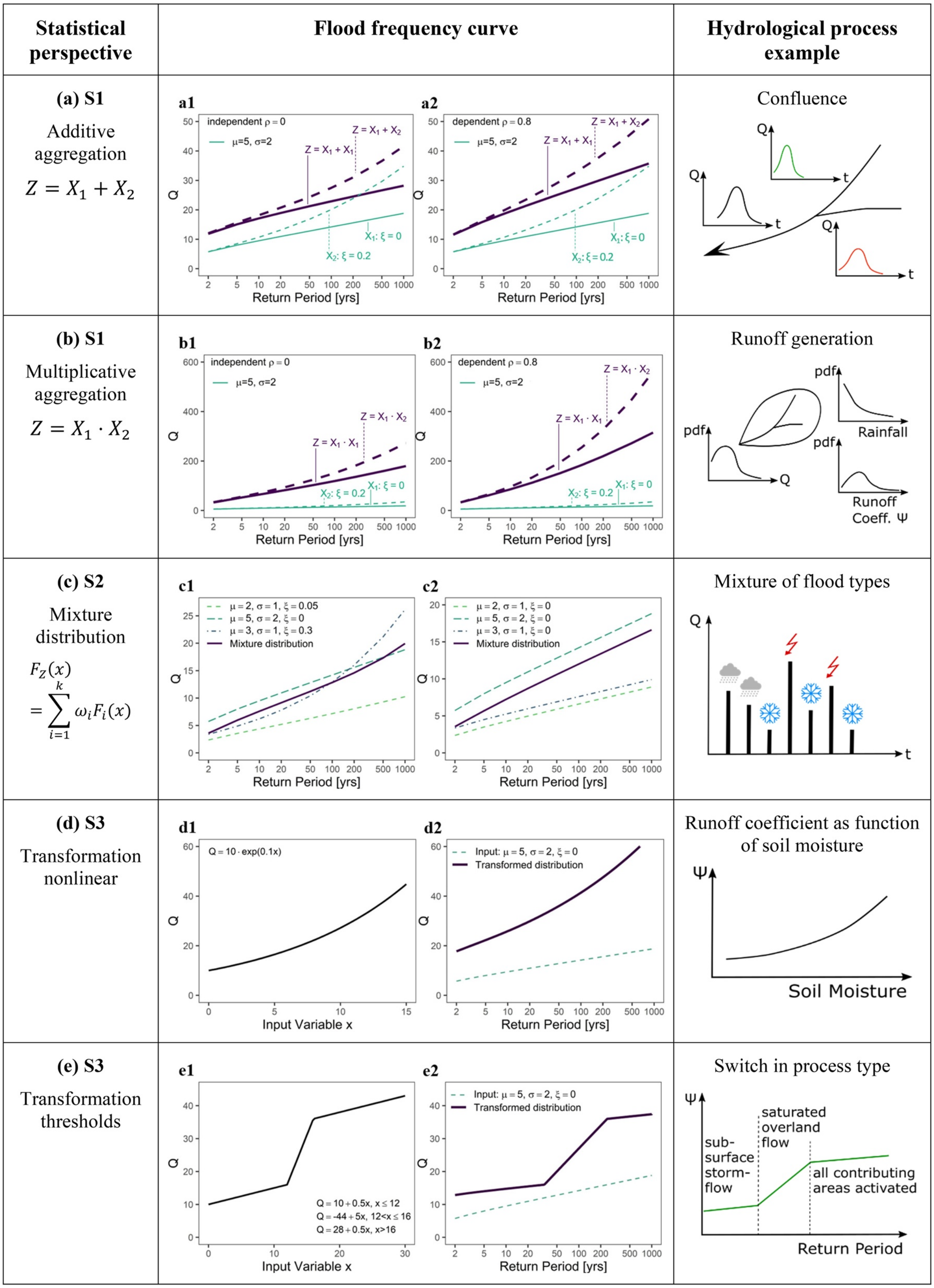

Figure 2.

3.1. Arithmetic Combination of Random Variables (S1)

The flood peak at a confluence of two tributaries can be represented as an additive aggregation of the floods of the tributaries if they occur at the same time (Guse et al., 2020). Spatial aggregation of flood producing point precipitation can also be a relevant additive aggregation process:

$$Z = X_1 + X_2$$

Here, X_1 and X_2 are component processes that are additively aggregated to the variable Z. Assuming statistical independence, additive aggregations generally do not create heavy tails, if X_1 and X_2 are light-tailed, but usually propagate them, if at least one component is heavy-tailed. The most relevant cases in the context of floods are summarized in Table A1. For the sum of dependent random variables, the tail heaviness of the aggregated random variable depends on the heaviness of the marginal tails and the dependence of the tails (Albrecher et al., 2006; Kortschak & Albrecher, 2009). For the additive aggregation of many random variables, the aggregation via the mean may remove heavy tails, as a consequence of a variation on the classical Central Limit Theorem (Billingsley, 1995; Nair et al., 2017). However, heavy tails are preserved in the limiting distribution of mean-aggregated processes, if the condition of finite variance of the components is not met, resulting in α-stable distribution (Table 1, for details, see Nair et al. [2017]).

In contrast to additive processes, multiplicative processes can lead to heavy-tailed outcomes, even if the components are light-tailed. Here, X_1 and X_2 are component processes that are multiplicatively aggregated to the variable Z:

$$Z = X_1 \cdot X_2$$

This behavior is relevant, for example, for runoff generation where the flood peak is often formulated as the product of a runoff coefficient and a representative rainfall that can both be considered as random variables (e.g., Gaume, 2006; Gottschalk & Weingartner, 1998; Sivapalan et al., 2005; Viglione et al., 2009).

The conditions for the creation of a heavy-tailed distribution from the multiplicative aggregation of light-tailed random variables are complex, even for the independent case. For example, the product of two exponentially distributed random variables is heavy-tailed, the product of two normally distributed random variables is light-tailed, while the product of three normally distributed random variables is heavy-tailed (Foss et al., 2009). Rojas-Nandayapa and Xie (2018) suggested sufficient conditions for the product of any two random variables to be heavy-tailed. If at least one of the components is heavy-tailed, this property is propagated to the resulting random variable Z. Table A2 summarizes the most relevant cases for the product of random variables for different classes of heavy-tailed distributions. For dependent random variables, the dependence structure of the components is relevant for the tail behavior of the aggregated random variable (for several special cases see Ranjbar et al., 2013). For the multiplicative combination of many random variables the lognormal distribution, which is heavy-tailed, arises naturally (Nair et al., 2017).

Figures 2a and 2b illustrate the differences in terms of tail behavior between the sums and the products of random variables. The sum of two random variables with light tails (Gumbel distributed, $\xi = 0$) shows a light-tailed behavior, while for the sum of a Gumbel and a heavy-tailed GEV variable ($\xi = 0.2$), the heavy tail of the GEV is propagated. In contrast, if two random variables are multiplied, the resulting distribution shows an upper heavy tail, even for the case where both variables have a Gumbel distribution.

Figure 2. Illustration of the statistical perspectives S1-S3 and examples of related hydrological processes. (a) Sum and (b) product of two random variables for the independent and dependent case. Each subplot shows two variants: combination of two component distributions with $\xi = 0$ (solid) and combination of two component distributions with $\xi = 0$ (solid) and $\xi = 0.2$ (dashed). Although the component distributions are the same, the upper tail behavior of the resulting distributions varies clearly between the subplots. For instance, the product of two dependent random variables **(b2)** shows a much heavier tail than the sum of the same random variables **(a2)**. For the dependent case, a Gaussian copula with dependence parameter $\rho = 0.8$ was used. (c) Effects of mixing distributions on the tail behavior. The three component distributions (dashed) are combined with equal weights to a mixed distribution (solid). **(c1)** The tail behavior of the mixed distribution tends to follow the most dominant tail of the components. **(c2)** Mixing three Gumbel distributions (with light upper tails) preserves the tail behavior. (d) and (e) Effects of non-linear transformations on the upper tail behavior. **(d1)** and **(e1)** show the non-linear transformation, **(d2)** and **(e2)** show the input and distributions of the transformed random variables.

3.2. Mixture of Distribution Functions (S2)

While in the statistical perspective S1 we considered the arithmetic combination of random variables, here component processes contribute randomly, according to assigned probabilities, to the outcome of the aggregated process, resulting in a mixture distribution. For many catchments, it has been shown that floods are generated by different (atmospheric and catchment) processes and that the population of flood events in observed time series is a mixture of flood types (Tarasova et al., 2019). These can be represented by mixture distributions in terms of flood process types (Fischer, 2018), flood seasons (Strupczewski et al., 2012; Waylen & Woo, 1982), and precipitation types (Cavanaugh et al., 2015).

The distribution function of the mixture distribution Z is the weighted sum of the distribution functions F_i of the components X_i

$$F_Z(x) = \sum_{i=1}^{k} \omega_i F_i(x)$$

where ω_i are the weights with $\omega_i \geq 0$ and $\sum_{i=1}^{k} \omega_i = 1$.

The mixture of a finite number of light-tailed distributions can only result in a light-tailed mixture distribution. If all components are heavy-tailed, the heavy tail is generally propagated to the aggregated mixture distribution (Foss et al., 2009). The assumption of heavy-tailedness is a weaker assumption than the non-existence of moments. If the assumption of the non-existence of the moments of one of the components holds, the resulting mixture distribution is heavy-tailed. In empirical hydrological studies it has been pointed out that the tail index of a mixture distribution seems to be inherited from the component with the most pronounced tail (e.g., Carreau et al., 2009; Cavanaugh et al., 2015). Figure 2 (c) illustrates these effects with two examples. In Figure 2 (c1) three different tails of the components result in a heavy-tailed mixture distribution, while in Figure 2 (c2) the mixture of three light-tailed random variables that are Gumbel distributed results in a light-tailed mixture distribution.

3.3. Transformations of Random Variable (S3)

The transformation of rainfall to surface runoff in the case of infiltration excess overland flow can be considered a non-linear transformation, where runoff is zero as long as rainfall intensity is below infiltration capacity, but becomes the difference between rainfall and infiltration capacity when the infiltration capacity is exceeded. Another example of a transformation is applying the maximum transformation to continuous streamflow data to obtain annual maximum streamflow. In both cases the underlying concept is a transformation g of random variables X and its effect on the propagation of heavy tails to the transformed variable Z:

$$Z = g(X)$$

While linear transformations do not change the tail behavior, non-linear transformations can affect the tail of the transformed variable, depending on the transformation itself, its parameters as well as the tail characteristics of the input variable. Results for selected transformations that are most relevant for hydrological purposes are given in Appendix. The exponential transformation illustrated in Figure 2 (d) gives a heavy tail for a light-tailed input variable, while the simple thresholding in Figure 2 (e) preserves the light tail of the input variable. An example of exponential transformation is runoff generation where the runoff coefficient depends in a non-linear way on the soil moisture. Thresholding behavior can occur when process types switch, for instance, when subsurface stormflow occurs during small events and saturated overland flow during large events.

In flood frequency analysis the maximum operation is usually applied to daily flows to obtain yearly maxima. The heavy tails of components are preserved under the maximum operation (Foss et al., 2011; Mikosch, 1999), meaning that heavy tails of daily streamflow are propagated to the flood peaks. For the maxima of many random variables, the GEV arises as to the limiting distribution for most relevant hydrological variables, which motivates its frequent use in flood frequency analysis. More specifically, the aggregated maximum distribution is heavy-tailed, if the components lie in the domain of attraction of Frechet, resulting in a positive shape parameter for the GEV (Fisher & Tippett, 1928).

3.4. Statistical Perspectives and the Context of Flood Frequency Analysis

When discussing the hypotheses, we also consider whether the flood generation mechanisms can be related to the statistical conditions that can generate heavy tail behavior. There are, however, differences in the statistical and hydrological perspectives on heavy tails.

Statistically speaking, the property of heavy tails is an asymptotic property of a distribution. It is hard to infer such property from quantiles with finite return periods only. Large sample approximations, that is, an asymptotic theory suggesting a GEV distribution, may not always be accurate for small sample application due to a slow rate of convergence, and pre-asymptotic results may be a better approximation (Fisher & Tippett, 1928). In hydrological practice, we typically focus on return periods in the range of 50–200 years and in rare cases up to higher return periods. The hydrological practice thus investigates heavy tail properties for finite return periods and pre-asymptotic behavior. Further, we neither know the true distributions of the components nor the true operation/transformation. In flood frequency analysis using extreme value statistics, we assume one or several distributions, fit them to the observed data and estimate the upper tail behavior from the fitted distributions.

Wilson and Toumi (2005) discussed the difference between asymptotic and pre-asymptotic behavior for heavy precipitation events. Based on physical considerations, they represented daily precipitation as the product of three independent Gaussian random variables, that is advected mass, specific humidity, and precipitation efficiency. According to statistical theory, the tail of the distribution of this product has a stretched exponential form and is thus heavy tailed, but at the same time is formally in the domain of attraction of Gumbel ($\xi = 0$). Block maxima of such random variables would be heavy tailed as well for finite block sizes but converge to a Gumbel distribution in the limit. They argued that the often observed heavy tail behavior may be explained by the slow convergence to the $\xi = 0$ limit, that is, finite block sizes. These differences in the statistical (asymptotic) and hydrological (pre-asymptotic) perspectives are illustrated by an example for the statistical perspectives S2 (Mixture of distribution functions) and S3 (Transformation of random variables), respectively. Let's assume that, in the given catchment, floods are caused by extratropical rainfall systems. However, there is a small probability that the catchment is hit by a landfalling tropical cyclone. In the latter case, the much higher rainfall volumes will generate much higher flood peaks compared to the other floods. For the sake of the argument, we assume that the distributions of both flood processes are light-tailed. In the context of the hydrological perspective, we fit a preselected distribution to the observed flood peaks. We easily find heavy-tailed behavior for the mixture distribution when we disregard the heterogeneity of a sample (Figure 3, upper panel), although the true mixture distribution is light-tailed (statistical perspective, see Section 3.2). Repeating this example with bounded distributions also yielded some realizations with heavy tails (not shown), although mixtures of bounded GEV distributions cannot be heavy-tailed in a statistical sense.

A similar argument can be made for the transformation of random variables, for instance, in the case of threshold behavior. Let's assume that the light-tailed input, for example, event rainfall volume, is transformed into much higher flood peaks once a certain threshold is exceeded. As we do not know the transformation, we fit a preselected distribution to the observed flood peaks. Again, hydrological practice can easily find heavy-tailed behavior (Figure 3, lower panel), although the true behavior is light-tailed (statistical perspective, see Section 3.3).

Other sources of misestimation of the upper tail behavior are temporal changes in time series of flood peaks or flood-related variables. In such cases, the upper tail behavior of the flood peak distribution can vary in time, as shown for flood-rich and flood-poor periods by Lun et al. (2020). When dealing with non-stationary time series, the aggregation of non-identically distributed random variables may lead to a misclassification of the underlying model and to a falsely estimated tail behavior. Ignoring non-stationarity, a model is chosen that best fits the non-homogeneous observations. For example, a GEV might provide a good fit for a sample of non-stationary Gumbel-distributed variables. Figure 4 illustrates this effect. A step change in the mean behavior is falsely interpreted as heavy-tailed behavior when pre- and post-change time periods are not separated in the flood frequency analysis (Figure 4, upper panel). Multidecadal variation can also be falsely interpreted as heavy-tailed behavior (Figure 4, lower panel). Drawing blocks of 30 years out of a time series with a substantial multidecadal variation leads to a wide range of shape parameters depending on the specific block that is drawn. In such cases, a non-stationary distribution may be fitted to the data, where the parameters are estimated as functions over time (Delgado et al., 2010).

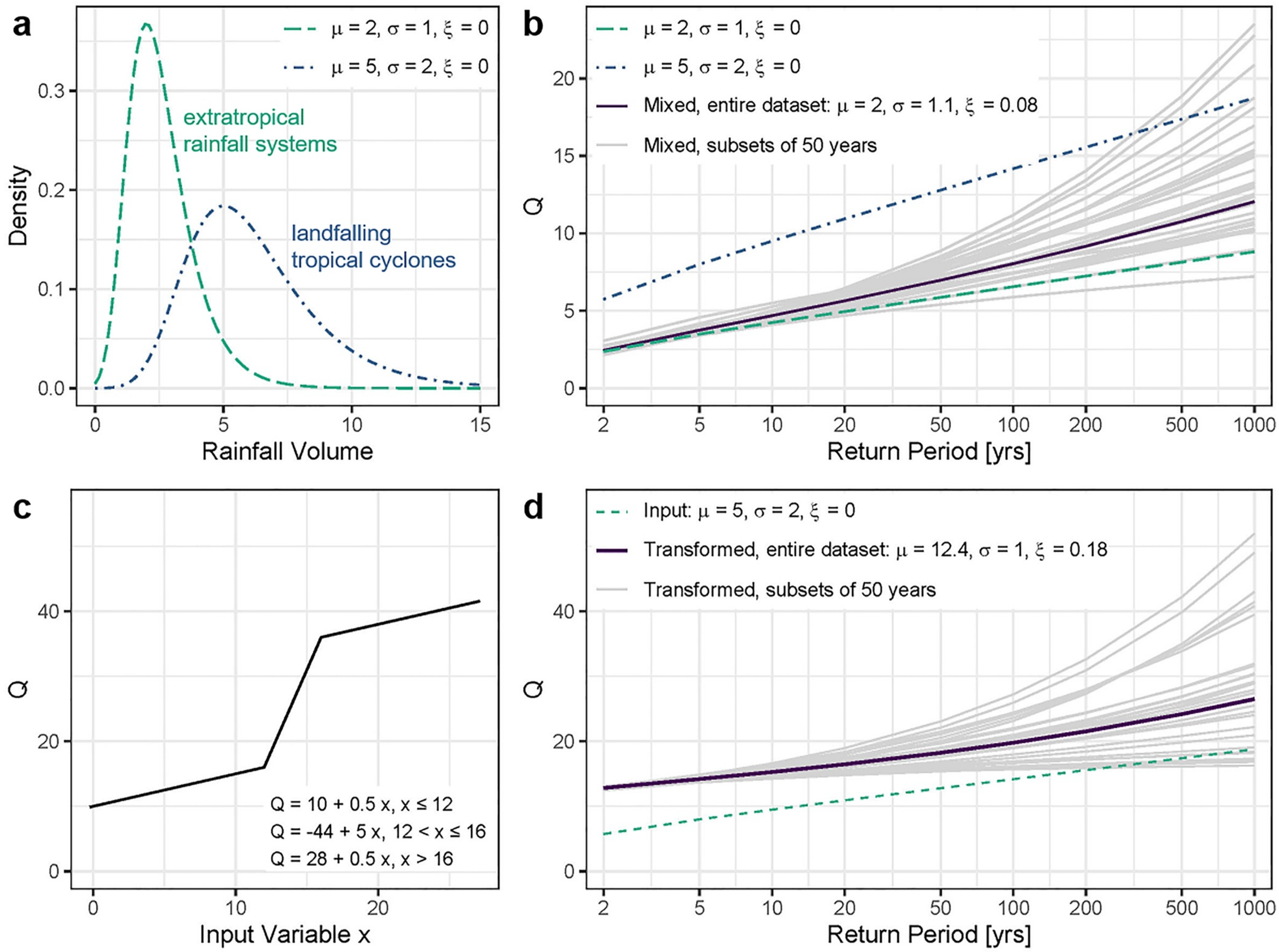

Figure 3. Top: Example for statistical perspective S2: Mixture of floods caused by extratropical rainfall systems and landfalling tropical cyclones. (a) Mixture of two Gumbel distributions (similar to Figure 2c) whereas 5% of the events are drawn from the distribution yielding higher rainfall volume. (b) The aggregated distribution is estimated by fitting a GEV to 100,000 values (denoted entire dataset) and to 30 random samples of size 50 (denoted subsets of 50 years). **Bottom**: Example for statistical perspective S3: (c) Non-linear transformation of rainfall in runoff (similar to Figure 2e), but (d) fitting of GEV to the transformed variable for 100,000 values (denoted entire dataset) and for 30 samples of size 50 drawn randomly from the transformed variable (denoted subsets of 50 years).

A final caveat is related to the typical distributions, such as the GEV, that are used in flood frequency analysis and the statistical perspectives S1-S3. Process-based simulations suggest that the flood peak distributions may have more complex shapes, for instance, an S-shape (Rogger, Pirkl, et al., 2012). Using field data in two Austrian catchments they found a sharp increase in the slope of the flood frequency curve when the storage capacity in parts of the catchment was exhausted and a decrease in the slope when only a few additional areas got saturated with increasing return periods. Along similar lines, Guse et al. (2010) and Fernandez et al. (2010) proposed an S-shaped flood frequency curve based on the argument that for any catchment under stationary conditions there should exist a maximum flood value that cannot be exceeded due to physical grounds. These studies indicate that the widely used distributions may not represent the true upper tail behavior.

These differences in statistical perspectives (asymptotic behavior) and hydrological practice (pre-asymptotic behavior) should be considered when estimating the upper tail behavior of flood peak distributions. First, one should be aware of the pre-asymptotic behavior which is considered an upper tail in hydrology. It is important to reflect on the range of return periods for which one can make sound statements and discuss whether processes may emerge with increasing return periods that are not at work for less extreme floods. Further, as we typically do neither know the true distribution functions and their temporal variation nor the underlying transformations, the statistical perspectives can give hints, but not the certainty, about the true upper tail behavior.

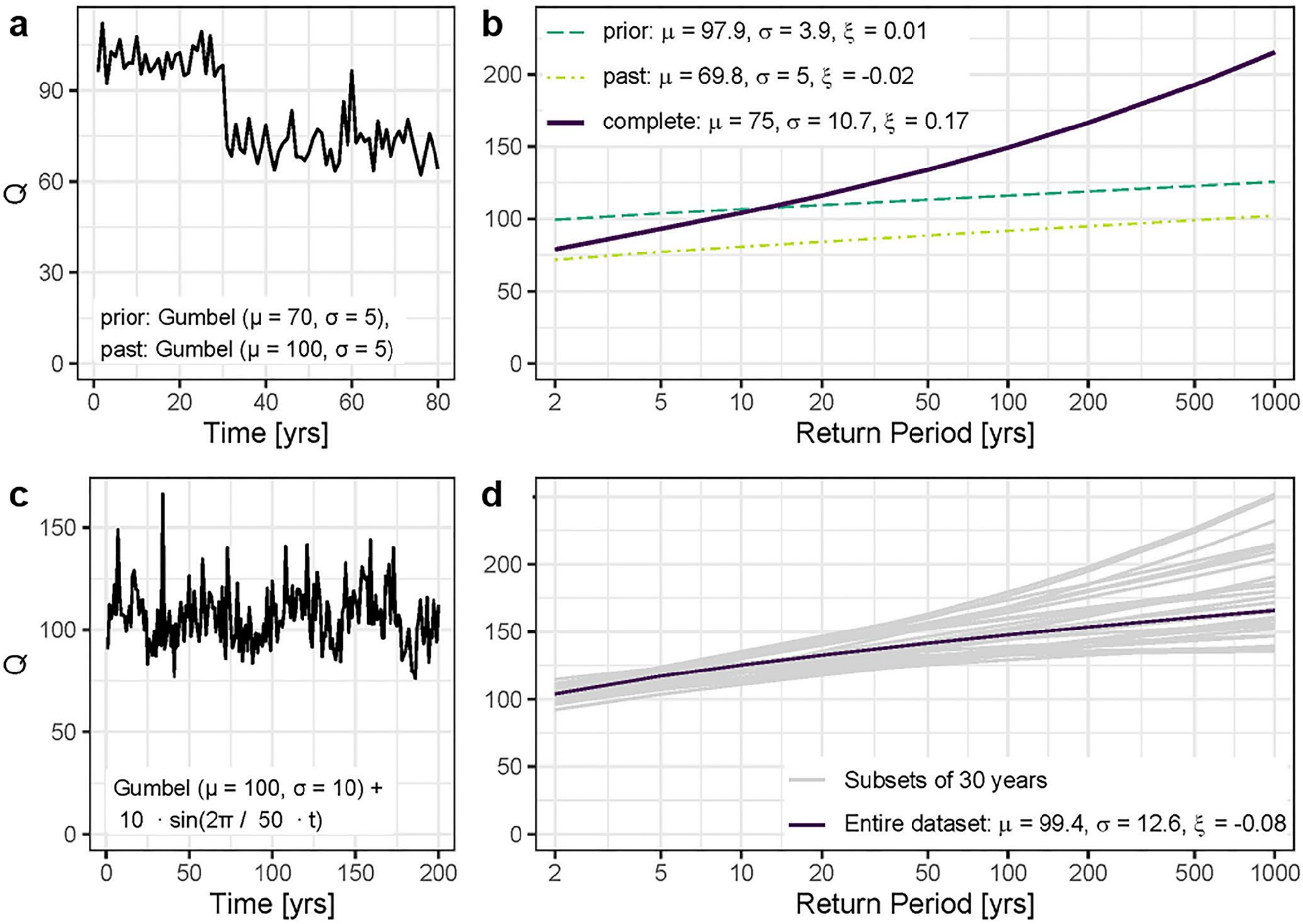

Figure 4. Effects of temporal changes on the upper tail behavior. Upper panel: (a) Time series (80 values generated) with a step change in the mean value. (b) The distributions for the periods prior to and after the step change show light-tailed behavior, but aggregating both periods and falsely fitting a GEV leads to heavy-tailed behavior. Lower panel: (c) Cyclic fluctuations (100,000 values generated but only 200 values shown) generated by the superposition of a sine wave with period of 50 years and random, Gumbel distributed values. (d) 30 block samples of length 30 are drawn randomly and illustrated (light gray). A large range of upper tail behavior can be obtained depending on the specific 30-year period drawn.

4. Prevalence of Heavy Tails and Hypotheses on Their Causes

4.1. Prevalence of Heavy Tails in Flood-Related Data

Analyses of observed flood time series often suggest the presence of heavy tails, either due to pre-asymptotic results or asymptotic behavior. For example, Farquharson et al. (1992) found an average GEV shape parameter of 0.40 for 162 catchments in various arid and semi-arid regions around the world, indicating strong heavy-tailed behavior. For the AMS time series of 813 catchments from Austria, Italy, and Slovakia, the averaged values of the L-coefficient of skewness were found to be generally larger than Gumbel's fixed L-coefficient of skewness, indicating heavy tail behavior (Salinas et al., 2014). Bernardara et al. (2008) estimated positive GEV shape parameters for 60% of the 173 catchments analyzed in southeastern France. The GEV shape parameters of the AMS time series for 572 stations in the eastern US were found to be generally positive (Villarini & Smith, 2010). For about 32% of the stations, the shape parameter was larger than 0.2, and for about 9% it was greater than 0.33 suggesting very heavy tails. In the Appalachians, Morrison and Smith (2002) estimated GEV shape values larger than 0.5 in 28% of 104 catchments examined. Molnar et al. (2006) found that heavy-tailed power law distributions were a better fit than exponential distributions for daily discharges greater than 20% of the flood of record for 159 rivers across the US. Finally, Smith et al. (2018) concluded from the analysis of several thousand flood time series across the US that the sample properties of the Upper Tail Ratio (ratio of the flood of record and the magnitude of the 10-year flood) were most consistent with GEV distributions with positive shape parameter, implying heavy-tailed distributions. These results are broadly consistent with the earlier conclusion of Martins and Stedinger (2000) that hydrologic experience indicates that a likely range for annual flood data is $0 \leq \xi < 0.3$, that is, that flood peak distributions tend to have heavy tails.

Hydrological practice often applies the concept of PMP (Probable Maximum Precipitation) and PMF (Probable Maximum Flood) for designing sensitive infrastructure, such as hydro-power dams. The assumption of a maximum value implies a bounded distribution. Hence, there is a discrepancy between widespread observations of heavy flood tails and the PMF concept. However, the PMP/PMF concept has been criticized as logically inconsistent and delusive (e.g., Koutsoyiannis & Papalexiou, 2017; Salas et al., 2014). For instance, PMP estimates are based on combinations of observed maxima of selected drivers of precipitation, whereas one assumes the existence of deterministic upper limits, but determines these limits statistically – an approach that is logically inconsistent (Koutsoyiannis & Papalexiou, 2017). This problem is aggravated for PMF as the upper limits of the processes that are combined within flood events are not known. Given these inconsistencies, and the fact that there is no consensus on how to estimate the PMF (Felder & Weingartner, 2017), the PMP/PMF concept is not a strong argument against the notion of heavy flood tails.

Heavy-tail behavior seems to be not only widespread in flood peak data, but also in other variables related to flooding (Katz et al., 2002). Annual maxima of more than 15,000 daily rainfall time series around the globe showed heavy tails in 60% of the stations (Papalexiou et al., 2013). The spatial variability of the GEV shape parameter was normally distributed with a positive mean (0.114) and a standard deviation of 0.045 (Papalexiou & Koutsoyiannis, 2013). Cavanaugh et al. (2015) found that most locations of a global daily data set with more than 22,000 weather stations showed heavy tails in annual maximum precipitation. For more than 4,000 stations across the United States, Papalexiou et al. (2018) concluded that hourly extreme precipitation had a heavy (sub-exponential) tail, much heavier than exponential or Gamma tails. Heavy tails have also been found to occur rather frequently in loss data for floods and other natural hazards (Cooke & Nieboer, 2011).

4.2. Hypotheses on the Hydrological Causes of Heavy Tails

We screened the literature to develop hypotheses on the causes of heavy-tailed flood peak distributions. We included studies that investigated the upper tail behavior of flood peak distributions via data-based and simulation-based approaches. Data-based studies, several of them mentioned in Section 4.1, typically estimate an indicator of upper tail behavior, such as the shape of the GEV or the skewness, for a large set of catchments, and attempt to explain the variation of this indicator between catchments by catchment characteristics. Simulation-based studies estimate the flood peak distribution, including the upper tail behavior, via a hydrological model from rainfall characteristics (derived flood frequency analysis; Eagleson, 1972). Under simplifying assumptions on the rainfall and runoff generation, flood peak distributions can be derived analytically (e.g., Basso et al., 2016; Sivapalan et al., 2005; Viglione et al., 2009). Coupling a stochastic weather generator with a rainfall-runoff model allows investigating more complex settings (e.g., Beven, 1987; Struthers & Sivapalan, 2007). Some of these simulation-based studies have specifically investigated, for instance, how threshold processes in the runoff generation affect the upper tail of the flood frequency curve (Rogger et al., 2013).

We extracted from these data-based and simulation-based studies any hints about the mechanisms that may cause heavy tails in flood peak distributions. Some papers directly proposed hypotheses about the emergence of heavy tails, in other cases we have formulated the hypotheses based on the findings reported in the papers. We developed nine hypotheses and associated them with the compartments atmosphere, catchment, and river network (Table 2). This association is based on the main mechanisms that are assumed to cause heavy-tailed flood distributions.

5. Atmosphere

5.1. Heavy Flood Tails Are Inherited From Heavy Rainfall Tails

Given the prevalence of heavy tails in rainfall distributions (e.g., Cavanaugh et al., 2015), either due to pre-asymptotic results or asymptotic behavior (see Section 3.4), and the high relevance of rainfall characteristics for flood peak distributions, the hypothesis that heavy flood tails are inherited from the rainfall distribution seems obvious. Hence, we review the question of whether heavy (light) tails of rainfall distributions lead to heavy (light) tails of flood peak distributions.

Table 2
Hypotheses on the Hydrological Causes of Heavy Tails in Flood Peak Distributions

No.	Source of upper-tail behavior	Hypothesis	Hypothesis derived from …
Atmosphere			
H1	Rainfall	Heavy flood tails are inherited from heavy rainfall tails.	Gaume (2006)
H2	Characteristic flood generation process	The characteristic flood generation process shapes the upper flood tail (e.g., rain vs. snowmelt).	Bernardara et al. (2008)
H3	Mixture of flood types	Mixture of flood types generates heavy flood tails.	Villarini and Smith (2010)
Catchment			
H4	Runoff generation	Non-linear response to precipitation causes heavy flood tails.	Gioia et al. (2008), Rogger et al. (2013)
H5	Water balance	Drier catchments have heavier flood tails due to interaction of water balance processes.	Farquharson et al. (1992), Berghuijs et al. (2014)
H6	Catchment size	Smaller catchments have heavier flood tails due to less pronounced spatial aggregation effects.	Villarini and Smith (2010)
River system			
H7	Reservoirs	Construction of reservoirs increases tail heaviness.	Maheshwari et al. (1995), Assani et al. (2006)
H8	Confluences	Confluences lead to downstream heavy-tail behavior.	Vorogushyn and Merz (2013)
H9	River embankments	Dikes increase the tail heaviness up to certain point (dike failure).	Apel et al. (2009)

Note. The hypotheses are assigned to the compartments atmosphere, catchment and river network.

There is not much data-based evidence supporting this hypothesis. McCuen and Smith (2008) compared the skew of annual maximum rainfall and annual maximum streamflow for 28 streamflow gauges on the US east coast. They obtained almost identical skew values for rainfall but a very large variation in flood skew values, that is, very different upper flood tails resulting from rather homogeneous rainfall tail behavior. They argued that this larger variation in flood skew was caused by catchment processes, in particular those related to catchment and channel storage, which transferred the upper tail behavior of rainfall into the upper tail behavior of flood peaks. Only for the extreme case, the catchment will be saturated, acting much like an impervious surface, and the runoff characteristics will follow the rainfall characteristics. Otherwise, catchment storage will determine how the flood distribution is related to the rainfall distribution.

The finding of McCuen and Smith (2008) that similar upper tail behavior of rainfall can lead to different upper tail behavior of flood peaks resonates well with the results of Gottschalk and Weingartner (1998). They estimated flood peak distributions for 17 small Swiss catchments, assuming that they can be derived from the product of the rainfall volume, scaled with respect to its duration, and the runoff coefficient. Their analyses demonstrated that identical distributions of rainfall volume gave rise to the very different behavior of the peak flow distribution depending on the distributions of runoff coefficients.

Gaume (2006) investigated the relationship between the upper tail of the rainfall and flood distributions by derived flood frequency analyses. He concluded that his analysis "… confirms and extends the results of previous works, that is, the shape of the flood peak distribution is asymptotically controlled by the rainfall statistical properties, given limited and reasonable assumptions concerning the rainfall-runoff process…". He suggested that for very large return period floods, for example, beyond 500 years, the distribution of the maximum mean rainfall intensity over a duration in the order of the time of concentration of the catchment should be considered as the possible flood peak asymptotic distribution, as schematically illustrated in Figure 5. The suggestion that the flood tail heaviness follows the tail heaviness of precipitation for large return periods is also the basic assumption of the GRADEX method which is widely used in practice, particularly in France (Naghettini et al., 2012). The return period beyond which the flood distribution follows the rainfall distribution is typically set to much lower values, for instance, to 10–20 years for relatively impermeable watersheds and up to 50 years for watersheds with high infiltration capacity (Naghettini et al., 2012). Such low values are in contrast to McCuen and Smith (2008) and Gaume (2006) who suggested that the tail behavior of the flood distribution was inherited from the rainfall distribution in the extreme case only, that is, for very large return periods.

10.1029/2021WR030506

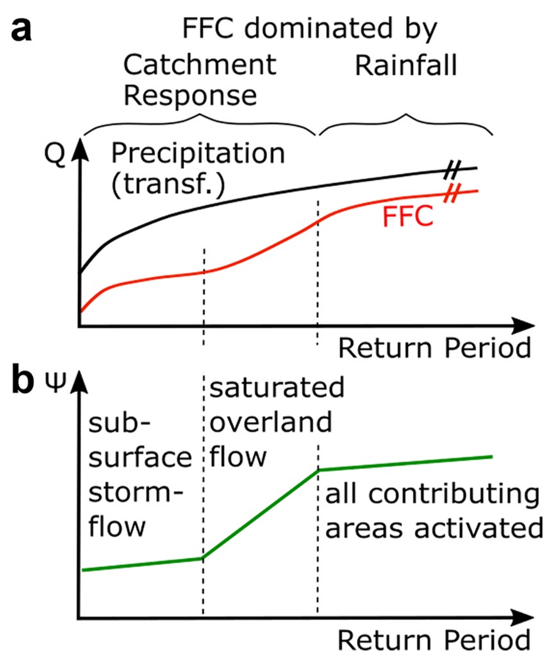

Figure 5. (a) Schematic frequency curves for flood peaks and for precipitation and (b) runoff coefficient as a function of return period (logarithmic axes). Only for very large return periods, the flood frequency curve is controlled by the rainfall statistical properties and both frequency curves run in parallel. For lower return periods the catchment response dominates the behavior of the flood frequency curve. The catchment shows a nonlinear response caused by a switch between subsurface stormflow and saturated overland flow and a threshold when all areas contributing to event runoff are active.

As the rainfall is combined (statistical perspective S1) or transformed (S3) to flood peaks, the tail behavior of the flood time series depends on the type of combinations or transformations and on the tail behavior of rainfall and catchment response. A large variety of tail behavior (in the typical range of return periods of interest) is possible, in particular, if threshold processes in the rainfall-runoff process occur (e.g., Viglione et al., 2009; Rogger, Kohl, et al., 2012, 2013). As the flood generation process is often well described as the product of random variables representing the rainfall and catchment response, an increase in tail heaviness when moving from rainfall to flood peaks should not be surprising (see also Section 6.1, hypothesis H4). In the extreme case when the catchment is saturated, rainfall will be transformed in a linear way to runoff, and the upper tail of flood peaks will follow the upper tail of precipitation. The return period beyond which the upper-tail flood behavior converges to the upper-tail rainfall behavior depends on the flood generation processes and is expected to vary from catchment to catchment.

In summary, our review suggests that the hypothesis 'Heavy flood tails are inherited from heavy rainfall tails' is little plausible as the runoff generation processes strongly modulate the upper tail behavior of streamflow. This statement is, however, limited to the range of return periods where catchment processes, such as catchment and river network storage or snow cover and snowmelt, exert a substantial influence on flood magnitude. For very high return periods, the catchment response loses its influence and the flood tail tends to be dominated by the tail of the rainfall distribution. A highly relevant question is beyond which return period the flood peak distribution is determined by the rainfall distribution. We recommend systematic studies to understand how this threshold varies between catchments and how it is related to climate and catchment characteristics.

5.2. The Characteristic Flood Generation Process Shapes the Upper Flood Tail

Catchments show different flood generation processes (Merz et al., 2020). In arid regions floods tend to be generated by heavy precipitation and infiltration of excess overland flow, whereas in humid regions subsurface stormflow and saturation excess overland flow dominate flood generation (Blöschl et al., 2017; Farquharson et al., 1992). In high elevation and high latitude catchments, snowmelt plays a dominant role. Here, we review the question of whether the characteristic flood generation process in a catchment leaves a fingerprint on the flood frequency curve with effects on its upper tail.

Merz and Blöschl (2003) compared two catchments, characterized by different flood types, in Austria. The flood frequency curve steepens with an increasing return period for the Fahrafeld catchment whose majority of floods, including the largest events, are long-rain floods. Flood peaks tended to increase with increasing rainfall. This relationship gets progressively steeper, reflecting the non-linearity of runoff generation with increasing event rainfall depth. In contrast, the frequency curve flattens out at large return periods for the Obermühl catchment for which most events, including the largest ones, are rain-on-snow floods. They suggested that the different flood generation processes caused the difference in the upper tail behavior. As the meltwater release is limited by the available energy, one would expect the tail of the distribution of the rain-on-snow-dominated catchments to be flatter than that of the rainfall-dominated catchments.

This suggestion is supported by regional studies. For more than 200 large catchments in Norway, the shape parameter of the GEV distribution is mainly explained by the average fraction of rain during the flood event (Thorarinsdottir et al., 2018). In regions where snowmelt dominates the event water input, lower, and often negative, shape parameters are found, suggesting an upper threshold for these spring floods caused by snowmelt. Similarly, Bernardara et al. (2008) found that the hydrological regime was the best predictor for explaining the shape parameter of the 173 flood time series of the Rhone-Mediterranean region in southeast France. About 97% of the catchments belonging to the Mediterranean regime showed heavy-tail behavior, whereas catchments of the continental regime and snowmelt-dominated catchments showed substantially lower fractions (44% and 47%, respectively). For a global data set of semiarid and arid regions, Farquharson et al. (1992) noted less heavy flood tails in Iran and Jordan. They attributed this behavior to the dominant flood generation processes, arguing that floods in Iran are largely generated from snowmelt and that the flood regime in Jordan is damped by groundwater contributions. Data analyses of 321 catchments across the US by Berghuijs et al. (2014) suggest a link between the regional growth curve (i.e., the normalized flood frequency curve) and the seasonal water balance. Growth curves of snow-dominated catchments and of other clusters tend to show light tail behavior, whereas arid clusters exhibit growth curves with heavy tails.

Hence, there is some data-based evidence supporting hypothesis H2 that the dominant flood generation process determines the upper tail behavior. Regions, where floods are caused by snowmelt, tend to have lighter tails, and regions with stronger non-linearity in flood generation tend to show more pronounced tail heaviness.

5.3. Mixture of Flood Event Types Generates Heavy Tails

In many environments, the basic assumptions of extreme value statistics, that is, that all floods are realizations of the same distribution, are little plausible. Hirschboeck (1988) introduced the concept of flood-hydroclimatology and suggested that unusually large floods may be related to specific large-scale atmospheric circulation anomalies (Hirschboeck, 1988). In these and similar cases, the flood peak time series may be more meaningfully represented as a mixture of different flood types. Here, we evaluate the hypothesis that such mixtures tend to produce heavy-tailed flood distributions (H3). Although H2 (Section 5.2) and H3 are both related to flood types, they are formulated as separate hypotheses. H2 evaluates the evidence that the dominant flood generation process in a region determines the upper tail, whereas H3 investigates the effect of mixtures of flood types on the upper tail.

Floods have been classified into event types from a hydroclimatic (large-scale circulation patterns and atmospheric state), hydrological (catchment-scale precipitation patterns and antecedent catchment state), and hydrograph-based perspective (Tarasova et al., 2019). Using the hydroclimatic perspective, Petrow et al. (2007) showed for the Mulde catchment in Germany that although the majority of floods were caused by westerly atmospheric flow, extreme floods were triggered by a specific atmospheric situation, Vb weather pattern, a slowly moving low-pressure field over the Gulf of Genoa, which can transport large amounts of moisture. Barth et al. (2017) found that annual maximum streamflow time series in the western United States often contained events generated from distinctly different hydrometeorological mechanisms. Smith et al. (2018) analyzed the Upper Tail Ratio for more than 8,000 US gauging stations. They found that often the flood-generating mechanism of strange floods, that is, floods that led to high values of Upper Tail Ratio, was rare and contrasted with the common flood-generating mechanism in these catchments. Warm season thunderstorms and tropical cyclones were identified as major flood types for record floods, whereas the broader population of annual floods was dominated by snowmelt floods and winter/spring storm systems. There is a clear contrast between the seasonality of record floods and the bulk of floods. The distribution of record floods has a maximum around mid-June and a secondary maximum around the beginning of September, corresponding to landfalling tropical cyclones on the east coast.

Using the hydrological perspective, Merz and Blöschl (2003) stratified a large set of catchments and flood events in Austria into five flood types. They found significant changes in the relative frequency of the flood types with magnitude. Large floods were frequently caused by short-rain events but rarely caused by rain-on-snow events and almost never by snowmelt events. The unit peak discharge values varied between different flood types, with rain-on-snow and snowmelt floods showing lower values. Mixed populations of flood types, that is, high winter-storm rainfall-driven floods and the more typical, smaller spring snowmelt floods, were found at high-elevation stations in the Sierra Nevada (Gotvald et al., 2012). Similarly, Tarasova et al. (2020) analyzed the variation of runoff generation event types from small events to large floods in 172 German catchments and found coherent spatial patterns of this variation. Using the hydrograph-based perspective, Fischer and Schumann (2020) identified three flood types with different hydrograph characteristics for catchments in the Harz region in Germany. Floods with high volumes and small peaks are mainly affected by snowmelt, whereas floods with high peaks and small volumes are often caused by intense summer thunderstorms. The latter type typically dominates the upper tail of the flood peak distribution.

The findings of these studies suggest that, in many catchments, large floods are associated with flood types that are different from those of small floods. They do however not investigate explicitly how changes in process types with flood magnitude affect the upper flood tail. A direct association between the mixture of flood types and heavy flood tails has been proposed by Morrison and Smith (2002) for Appalachian catchments in the eastern United States. They found a regional differentiation, where the central region showed strong heavy-tail behavior. They attributed this behavior to the mixture of three flood types, that is, organized systems of thunderstorms, tropical storms, and extratropical cyclones. Similarly, Villarini and Smith (2010) analyzed the upper tail behavior of annual maximum streamflow from 572 stations in the eastern United States, where floods can be triggered by different storm types. They compared the shape parameter estimated for the full record against the estimate after removing the floods caused by tropical cyclones. Removing tropical cyclone floods significantly lowers the shape parameter, suggesting that anomalously heavy flood tails are linked to landfalling tropical cyclones.

In summary, mixtures of flood types can generate heavy tail behavior due to two effects. First, the mixture may contain a heavy-tailed distribution whose upper tail behavior then tends to dominate the upper tail behavior of the mixture distribution. For instance, in catchments with snow-related floods and rainfall-driven floods, the heavier tail behavior of rainfall-driven floods tends to dominate the mixture behavior. Second, there may be a different process that occurs very rarely but generate much higher flood peaks. An example is a landfalling hurricane, hitting a catchment on the east coast of the United States (Villarini & Smith, 2010). In this case, the mixture distribution can be heavy-tailed from the hydrological perspective, even when the distributions of both regular floods and hurricane-related floods are light-tailed (see Figure 3).

6. Catchment

6.1. Non-Linear Response to Precipitation Causes Heavy Flood Tails

Here, we hypothesize that the non-linear response of catchments to precipitation, including threshold processes of runoff generation, causes heavy tails of flood peak distributions even when precipitation is light-tailed.

The link between the upper flood tail and the non-linearity of runoff response has mainly been investigated by simulation studies, several of them using hypothetical catchments. Based on a derived flood frequency model, Gioia et al. (2008) explained the highly skewed flood distributions observed in 10 catchments in Southern Italy by threshold mechanisms. They suggested that, whereas ordinary floods were caused by rainfall events exceeding a threshold infiltration rate in a small source area, extremes occurred when a threshold storage value was exceeded in a large portion of the catchment. In a similar vein, Rogger, Pirkl, et al. (2012) related thresholds in catchment response to catchment storage capacity in two Austrian alpine catchments. They used detailed field surveys of hydro-geologic storage capacity and surface runoff generation to specify the parameters of a distributed continuous runoff model and simulate soil saturation patterns. Their model results suggest that a sudden increase in the slope of the flood frequency curve is caused by the exceedance of the storage capacity, which generates surface runoff in large parts of the catchments (schematically illustrated in Figure 5). They also noted that this may occur more easily if the storage capacity is uniformly distributed within the catchments.

The partial area concept, that is, that only a fraction of the catchment area contributes to event runoff and that this fraction varies in time, is a plausible explanation for heavy flood tails. In this concept, a small contributing area generates ordinary floods. As the contributing area expands, the magnitude of runoff events increases, possibly leading to a steepening in the flood frequency curve, as demonstrated by field (e.g., Rogger, Pirkl, et al., 2012) and simulation studies (e.g., Fiorentino et al., 2007). Along these lines, the controls of some physically-based parameters underlying the partial area runoff generation on the skewness of the flood peak distribution were investigated by Gioia et al. (2012). Based on a derived flood frequency curve which accounts for two threshold mechanisms associated with ordinary (i.e., exceedance of a constant infiltration rate in a small area) and extreme (i.e., exceedance of a storage threshold over a large portion of the basin) events, they showed that light-tailed rainfall distributions can be transformed into heavy-tailed flood distributions by the catchment response.

Using a stochastic rainfall model coupled with a deterministic rainfall-runoff scheme, Kusumastuti et al. (2007) and Struthers and Sivapalan (2007) suggested that a change in the dominant runoff generating mechanism manifests itself as an inflection point in the flood frequency curve. The return period at which this transition occurs tends to increase with increasing catchment storage capacity (e.g., deeper soil) and increasing catchment aridity (Kusumastuti et al., 2007; Struthers & Sivapalan, 2007; Rogger, Kohl, et al., 2012; Rogger, Pirkl, et al., 2012;

Rogger et al., 2013). The spatial heterogeneity of rainfall and runoff generation also plays an important role. Struthers and Sivapalan (2007) showed that varying soil depths within the catchment can mask the effect of storage thresholds. In addition, the magnitude of the step change decreases with enhanced temporal variability of the amount of water stored in the catchment prior to rainfall events, and for heterogeneous spatial distributions of the storage deficit. It instead grows with the increasing size of the variably saturated region (Rogger et al., 2013). Hence, the appearance of a step change is related to the fraction of areas where fast runoff is generated at the same time.

Using a simplified conceptualization of rainfall-runoff processes, Basso et al. (2015) explained the emergence of heavy tails of flow distributions by highly non-linear storage-discharge relationships. In a follow-up study, Basso et al. (2016) also showed that this non-linearity contributes to heavier tails in the distributions of seasonal flood maxima. They suggested that the non-linear storage-discharge relationship could arise from the hydraulics of surface flow draining a hillslope (Guerin et al., 2019; Rupp & Selker, 2006), the expansion of the stream network contributing to flow (Biswal & Marani, 2010; Mutzner et al., 2013), and the spatial heterogeneity of hydraulic properties between hillslopes (Harman et al., 2009). The latter explanation challenges the idea that spatial heterogeneity thins streamflow tails due to smoothing of threshold effects, as proposed by Struthers and Sivapalan (2007) and Rogger, Pirkl, et al. (2012).

The superposition of precipitation and non-linear catchment response can be linked to the statistical perspectives S1 and S3. In the first case, flood runoff is expressed as a multiplicative process of rainfall and catchment response. In the second case, the actual water storage in a catchment is a random variable that plays a key role in determining the catchment runoff. In the case of non-linear catchment response, streamflow results from a non-linear (e.g., power law) transformation of this random variable, which can produce heavy-tailed flow distributions. In both cases, that is, assuming S1 (multiplicative processes) or S3 (non-linear transformation), we may expect heavy-tailed flood behavior caused by the catchment response to rainfall.

In summary, the available studies provide clear evidence that the catchment response plays an important role in the emergence of heavy-tailed flood peak distributions. However, there are contrasting answers to the question of whether spatial variability in runoff generation enhances or dampens the tail heaviness, and several studies focus on step changes in the flood frequency curve and not directly on the upper tail behavior.

6.2. Drier Catchments Have Heavier Flood Tails Due To Interaction of Water Balance Processes

Several studies analyzing observed flood time series have found that flood frequency curves of catchments subject to drier climatic conditions have heavier tails than wetter catchments. Farquharson et al. (1992) combined flood frequency curves of 162 catchments around the world into regional curves and noted that the tails in the arid regions such as South Africa or Saudi-Arabia were much heavier than the tail of the humid region of Great Britain. A follow-up study consisting of many catchments in sub-tropical and tropical regions (Meigh et al., 1997), as well as a recent study comparing semi-arid and Mediterranean catchments (Metzger et al., 2020), support the conclusion that drier regions are associated with heavier tails. Merz and Blöschl (2009) found the skewness of annual maximum floods in 459 Austrian catchments to be negatively correlated with mean annual precipitation and positively correlated with evapotranspiration. Molnar et al. (2006) found heavy-tailed power law distributions to better fit daily discharges greater than 20% of the flood of record than exponential distributions for 159 catchments across the US and noted that the parameter of the power law decreases with increasing aridity suggesting increasing heavier tails. A link between the shape of the flood frequency curve and the annual water balance has also been identified by Guo et al. (2014) for 266 catchments across the US. By pooling flood growth curves according to aridity, they showed that the pooled growth curves were distinct for each aridity class and that the tail heaviness increased with aridity. Berghuijs et al. (2014) supported this conclusion by grouping 321 catchments across the continental U.S. into clusters with similar seasonal water balance behavior and finding heavier flood tails for the arid clusters. Studying about 8,000 catchments in the US, Smith et al. (2018) found that the Upper Tail Ratio was somewhat larger (median ≈ 2.1) in the regions with less than 650 mm annual precipitation compared to more humid regions (median ≈ 1.5).

Although data-based studies have consistently found that drier catchments tend to have heavier flood tails, the mechanisms responsible for enhanced tail heaviness are less clear. Guo et al. (2014) suggested that the higher non-linearity in the rainfall-runoff transformation in drier catchments caused heavier flood tails compared to wet

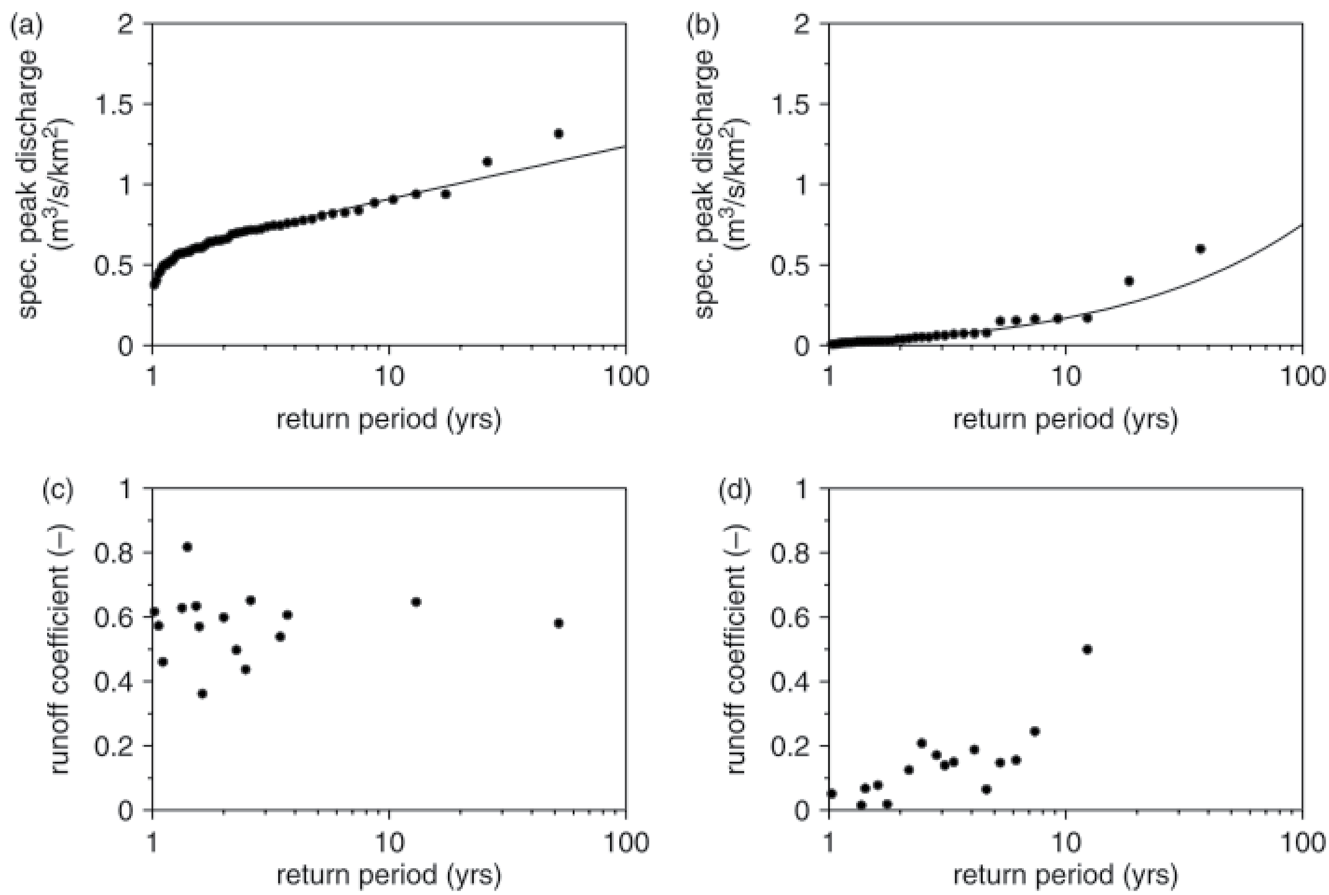

Figure 6. Flood frequency curves (a), (b) and runoff coefficients for annual maximum streamflow (c), (d) in two Austrian catchments. The drier catchment (right) displays a larger variation of runoff coefficients and a heavier upper tail of the flood peak distribution (Taken from Merz & Blöschl, 2009).

catchments. This suggestion is derived from their observation that the upper tails of the regional growth curves vary more between different aridity classes for annual maximum streamflow compared to annual maximum rainfall. Similarly, Merz and Blöschl (2009) suggested that runoff generation was responsible for the higher skewness of annual maximum floods in drier catchments in Austria. While in wet regions, runoff coefficients were typically large and did not increase much with flood magnitudes, they were smaller in dry catchments but increased substantially with flood magnitudes thus resulting in large flood skewness. Figure 6 illustrates that the sharp increase in runoff coefficients in the drier catchment (760 mm annual rainfall) is aligned with an upward curvature of the flood frequency curve on a semi-logarithmic plot while the consistently high runoff coefficients in the wetter catchment (1,800 mm annual rainfall) are aligned with a downward curvature of the flood frequency curve. As flood peaks can be described as the product of two random variables, that is, runoff coefficient and a representative rainfall (e.g., Gottschalk & Weingartner, 1998), statistical perspective S1 suggests that flood peaks in drier catchments tend to show heavier tails as their runoff coefficient distributions tend to be more skewed.

The data-based conclusion of Merz and Blöschl (2009) and Guo et al. (2014) is supported via simulation by Viglione et al. (2009). They found that the ratio between the return periods of maximum floods and rainfall depended on the average wetness of the catchment. In a dry system, where large runoff coefficients rarely occur, a single event with a high runoff coefficient can produce a flood with a return period that is hundreds of times larger than the one of the corresponding rainfall but low runoff coefficient. By contrast, the return period of floods never exceeds a few times that of the corresponding rainfall in a wet system, where runoff coefficients are always high.

Another contribution, besides runoff generation, to heavier flood tails in drier areas may result from heavier-tailed rainfall distributions in these areas. When analyzing hourly rainfall extremes of over 4,000 stations across the

United States, Papalexiou et al. (2018) found clear spatial patterns of the tail indices. These patterns show similarities with the main Köppen-Geiger climate classes with heavier tails in regions classified as arid.

Besides heavier rainfall tails and more non-linear runoff generation in drier catchments, also the interaction of rainfall, evapotranspiration, and runoff generation contributes to heavier flood tails. Focusing on this interaction, a mechanistic explanation of the higher skewness of streamflow (which translates into higher skewness of annual maximum peaks; statistical perspective S3) observed in drier catchments is provided by Botter (2010). Considering the stochastic character of the catchment water balance, he showed that the skewness is inversely proportional to the streamflow-producing rainfall frequency, which is lower in drier catchments due to more erratic rainfall regimes and higher evapotranspiration rates (and hence a higher threshold beyond which rainfall produces floods streamflow). Similarly, the physically-based theoretical derivation of flood frequency curves proposed by Basso et al. (2016) suggests that the tail of the seasonal flood distributions is controlled by the ratio (called persistency index) between the frequency of streamflow-producing rainfall (i.e., rainfall frequency suitably reduced by the effect of evapotranspiration) and the average catchment response time. Catchments with lower persistency index (named erratic regimes) display heavier tails. Dry climates (i.e., reduced runoff frequency caused by small rainfall frequency or high evapotranspiration) favor erratic regimes. This conclusion resonates with statistical perspective S3. In summary, data-based and mechanistic modeling studies conclude that drier catchments tend to show heavier flood tails due to heavier rainfall tails, more pronounced non-linear runoff generation, and the more erratic interaction between rainfall, evapotranspiration, and streamflow.

6.3. Smaller Catchments Have Heavier Flood Tails Due To Less Pronounced Spatial Aggregation Effects

Flood frequency curves may differ between catchments of different sizes for a number of reasons. For instance, in small catchments local, high intensities rainfall bursts of convective origin can be the main cause of flooding, while in large catchments such effects may be averaged out while other processes, such as flood routing, may become more important (Merz & Blöschl, 2009; Rosbjerg et al., 2013).

The data-based evidence for the hypothesis that smaller catchments have heavier flood tails is mixed. Meigh et al. (1997) stratified regional flood frequency curves in tropical and sub-tropical countries according to catchments size which suggested heavier tails in smaller catchments. For instance, for the Philippines, the GEV shape parameter increased steadily from 0.07 to 0.29 when going from the group of largest catchments (>2,500 km^2) to the smallest catchments (<25 km^2). Villarini and Smith (2010) found a decreasing shape parameter with catchment size (decrease by 0.07 per order of magnitude) of the annual maximum flood series in 572 basins of the eastern US. St. George and Mudelsee (2019) detected the 10 largest ratios of the flood of record to the second largest flood in 2,790 US catchments in rather small catchments; seven out of 10 catchments have areas smaller than 1,000 km^2. Merz and Blöschl (2009) found a weak (Spearman's correlation $r = -0.17$) negative correlation between skewness of maximum annual flood records and catchment area in 459 Austrian catchments. For 813 catchments with areas from 4.6 to 131,488 km^2 in Austria, Italy, while Slovakia, Salinas et al. (2014) reported a decrease of L-skewness with increasing catchment size for drier and medium wet catchments with mean annual precipitation up to 860 and 1,420 mm/yr, respectively, while for wetter catchments they observed no dependence. Other studies have not detected such an association. Morrison and Smith (2002) and Northrop (2004) found that the GEV shape parameter did not depend on the catchment area for the Appalachian Mountains and for 1,000 catchments in the UK, respectively. Similarly, the analysis of more than 5,500 flood series across the US, with approved stationarity according to the Mann-Kendall test, did not show a catchment size effect of the shape parameter (Smith et al., 2018). Finally, a non-monotonic relation between the upper tail behavior and spatial scale has been found by Pallard et al. (2009). They investigated the relationship between the drainage density, which tends to be inversely proportional to catchment area (Murphey et al., 1977), and the skewness of the annual maximum flood series. Simulation studies and data from 44 catchments in the Po River Basin suggest a U-shape scaling, with high skewness for small drainage densities and increasing skewness with increasing density beyond the local minimum.

These studies which analyze how the tail behavior of observed flood time series changes with spatial scale do neither provide a clear statement on this hypothesis nor do they provide a clear process explanation. However, they point to the possibility that the less heavy tails in larger catchments, identified in some of these studies, are a consequence of spatial aggregation. This effect can work via aggregation of precipitation and aggregation of runoff generation.

Overeem et al. (2010) investigated changes in the upper tail behavior of rainfall with increasing area using weather radar data in the Netherlands. The regionally estimated GEV shape parameter decreased with increasing rainfall area, from 0.17 for 6 km^2 to 0.07 for 1,700 km^2. They suggested that this decrease is a consequence of the decreasing spatial dependence with increasing scale. Similarly, Skaugen et al. (1996) argued that the distribution of extreme areal precipitation results from the sum of positively skewed and spatially correlated point precipitation within the area. For perfect spatial correlation, the distributions of the areal and point precipitation are identical, whereas in the absence of spatial correlation and for sufficiently fast decaying spatial correlation the spatial rainfall will converge to a Gaussian distribution because of the central limit theorem, assuming the variance exists (statistical perspective S1). In reality, rainfall lies between these two limits and areal rainfall characteristics depend on the spatial dependence of rainfall (Skaugen et al., 1996). Dyrrdal et al. (2016) found decreasing GEV shape parameters of areal precipitation with increasing area in the southwest of Norway (from around −0.05 at 1 km^2 to around −0.15 at 14,000 km^2). For the southeast of Norway, however, the behavior was more complex with an increase in shape parameters from around 0.1 at 1 km^2 to around 0.25 at 6,000 km^2, followed by a decrease to 0.20 at 14,000 km^2. This more complex behavior and the substantial variability in the relation of the shape parameter of the areal precipitation distribution and the aggregation area suggest that this relation depends on many factors including prevailing precipitation types and orographic enhancement of precipitation.

Spatial aggregation of runoff generation may also contribute to lighter flood tails in larger catchments. Heavy flood tails in the form of step changes in the flood frequency curve have been attributed to runoff generation threshold processes (see Section 6.1). Rogger, Pirkl, et al. (2012) argued that the catchment scale plays a role in the occurrence of such step changes for catchments where direct runoff is generated by storage excess. As the catchment size increases, the storage tends to become spatially more heterogeneous. Simultaneous saturation of a large part of the catchment is, therefore, less likely than in a small catchment. This argument can be extended to other flood generation processes with non-linear behavior. Distinct non-linear behavior at small scales tends to be averaged out at larger scales when increasingly different areas and processes are involved in the runoff generation.

This hypothesis can be linked to the statistical perspective S1, as the streamflow in large catchments can be interpreted as the spatial aggregation of rainfall and runoff generated in many small subareas of the catchment. The mean of a large number of random variables tends toward the normal distribution if the variables are uncorrelated or the decay of the correlation is sufficiently fast.

In summary, this hypothesis is only weakly supported by observed data. Spatial aggregation of precipitation and runoff generation may contribute to less heavy flood tails in larger catchments. However, the emergence or the shifting dominance of processes with increasing scale, such as precipitation types and runoff generation mechanisms, may inhibit a clear change of upper tail behavior in the catchment area.

7. River Network

7.1. Construction of Reservoirs Increases Tail Heaviness

Reservoirs regulating river flows may differently affect small and large floods and thus change the shape of the flood peak probability distribution (Volpi et al., 2018). The impact of reservoirs on floods is strongly related to their function. All reservoirs have a limited storage capacity. This capacity in relation to the volume of the inflowing flood determines the retention effect. The reduction of the peak depends also on the flood management strategy, the shape of the flood hydrograph, the options to consider flood forecasts, and the technical characteristics of the reservoir. The majority of reservoirs affect small floods strongly but do not reduce substantially the peak of extreme floods (Figure 7a). As a consequence, the construction of reservoirs may transform the flood frequency curve toward a heavier tail compared to the pre-reservoir situation.

Only a few data-based studies are available to evaluate this hypothesis. Assani et al. (2006) found higher skewness values for the annual maximum flows of 60 rivers regulated for hydropower production compared to 88 pristine catchments in Quebec (Canada). Similarly, Botter et al. (2010) found that damming of rivers in the Italian Alps increased the skewness of streamflow as compared to the pre-dam period. Mei et al. (2017) analyzed data from 38 rivers across the United States for which extensive records of pre- and post-damming annual flood peaks exist. The probability distributions of post-damming flood peaks exhibit heavier tails compared to the corresponding natural discharges.

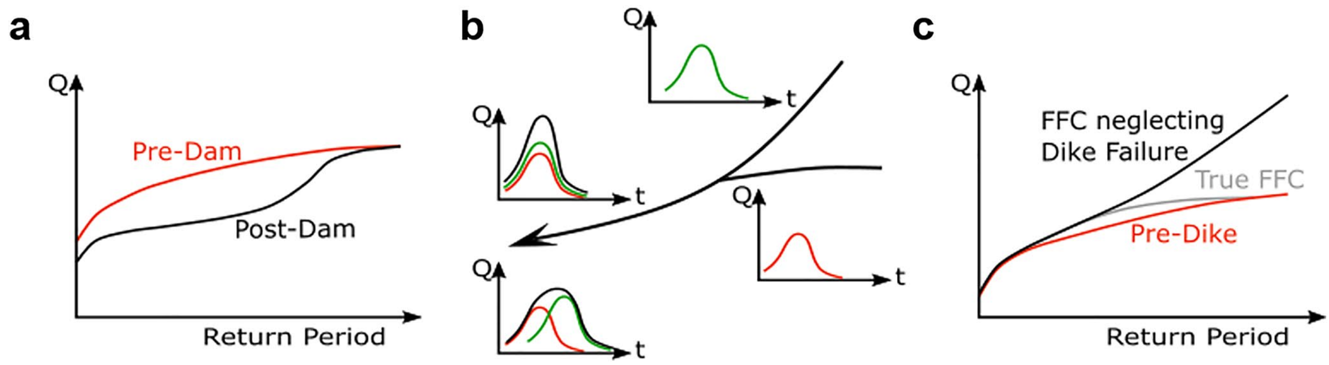

Figure 7. Schematic of the river network hypotheses. (a) H7: Construction of reservoirs tends to reduce small floods strongly but not the peak of extreme floods, leading to heavier flood tails for the period after dam construction. (b) H8: The relative timing of the two upstream peaks controls the superposition at the river confluence. Coincidence of the two upstream flood peaks for large floods but not for smaller ones will cause heavier tails downstream. (c) H9: Construction of dikes along rivers tends to increase flood peaks downstream. When dike breaches occur during extreme floods, and large water volumes are abstracted from the main channel, the flood peaks are capped, reflecting the original conditions prior to dike construction. Flood frequency curves sketched with a logarithmic axis for return period.

Maheshwari et al. (1995) analyzed the effects of the operation of weirs, diversions, and reservoirs on the flood regime of the Murray river in Australia based on a hydrological model. Their results suggest heavier tails of the distributions of annual maximum flows for the regulated cases as compared to the natural conditions, resulting from a decrease in peaks lower than the 2-year flood whereas peaks larger than the 20-year flood are little affected. By coupling a stochastic rainfall generator, a rainfall-runoff and a reservoir routing model for a hypothetical reservoir at the outlet of a small catchment, Ayalew et al. (2013) found that the flood frequency curves of unregulated and regulated flows converged for low probability events (i.e., the distribution of regulated flood peaks is more skewed as the high probability floods are smaller in the regulated case), and that a break in the regulated curve occurred when different release structures (i.e., sluice gate and spillway) were activated. Although tails of flood peak distributions are not explicitly mentioned in the study by Wang et al. (2017), their material suggests that heavier tails might result from reservoirs with intermediate values of the ratio between storage capacity and mean annual streamflow volume.

The dissimilar attenuation for different return periods is caused by the non-linear relation between outflow and volume of water stored in the reservoir. The retention storage and the reservoir's spillways, which might be activated when the retention storage approaches saturation, impose a non-linear transformation of the random variable incoming flow (through the water level in the reservoir) resulting in different attenuation of small and large floods which enhances flood tail heaviness (statistical perspective S3).

The position of the reservoir within the catchment also plays an important role in the enhancement of flood tail heaviness by reservoir construction (Volpi et al., 2018). Ayalew et al. (2015) showed that the interplay between the spatial configuration (i.e., single vs. multiple reservoirs placed in series or parallel), the storage and release capacity of the reservoirs, and their location in the drainage network controls the divergence between natural and regulated flood frequency curves and the return period at which the slope of the latter breaks, which in turn affects the tail behavior. Ayalew et al. (2017) showed that 133 small dams built in the Soap Creek watershed in Iowa, USA produced heavier tails compared to natural conditions, due to varying degrees of attenuation of flows with different exceedance probability. The effect fades moving downstream as the ratio between overall storage capacity and drainage area decreases.

7.2. Confluences Lead to Downstream Heavy-Tail Behavior

Flood hydrographs often show substantial changes at confluences. When the upstream flood waves arrive at the same time at the confluence, the downstream peak will be the sum of the two, while temporal decoupling leads to a weaker superposition (Blöschl et al., 2013; Geertsema et al., 2018; Vorogushyn & Merz, 2013). For instance, the devastating floods in 1988 and 1998 in Bangladesh were characterized by the synchronous arrival of the peaks of the Brahmaputra and Ganges Rivers (Mirza, 2002). The flood peaks downstream of a confluence can be interpreted as the superposition of the peaks of the two upstream river branches. Hence, this hypothesis is linked to the statistical perspective S1, that is, the sum of two random variables, which propagates heavy-tail behavior depending on the correlation. The relative timing of the two upstream peaks plays a major role as it controls the superposition of the peaks of the tributaries. If the temporal coincidence of floods from the upstream rivers does not depend on the flood magnitude, one would not expect a systematic effect on the upper tail of the flood frequency curve downstream of the confluence, while a coincidence for large floods but not for smaller ones will cause heavier tails downstream (Figure 7b).

There are a number of approaches to estimating the flood frequency distribution at river confluences, for instance using copulas (Bender et al., 2016; Wang et al., 2009). However, only a few studies investigate the mechanisms that can lead to flood coincidence at river confluences via simulation approaches (Pattison et al., 2014; Seo & Schmidt, 2013; Skublics et al., 2014) and past event analysis (Geertsema et al., 2018; Guse et al., 2020). The main factors that control whether flood waves arrive at the same time at the confluence are the space-time distribution of event precipitation and the lateral flow in the catchment and river network. These factors, in turn, depend on the spatial correlation of event precipitation and antecedent catchment wetness, the spatial organization of the confluence catchments, and the distribution of flow path lengths and flow velocities, as illustrated by Seo and Schmidt (2013) using a stylized model. For rivers with long flood wave durations, for example, larger lowland rivers, the exact timing of the arrival of discharge peaks is of smaller relevance, and the magnitude and duration of the flood waves play a more important role in enhanced flood peaks at confluences (Geertsema et al., 2018).

Guse et al. (2020) analyzed flood wave superposition for 37 river confluences in Germany by comparing the event characteristics of three flood time series, that is, at the downstream station and at the two upstream stations. For most confluences, the tributary peak typically arrived earlier than the main river peak. It is conceivable that specific atmospheric situations lead to a delayed arrival time of the tributary peak and/or an earlier arrival time of the main river peak, and that these situations occur differently for small and large floods. However, at most confluences, Guse et al. (2020) did not find systematic differences in the relative timing of the upstream peaks between floods of different magnitudes, so flood wave superposition was not found to cause heavy-tail behavior downstream in this dataset. At a few confluences, there is potential for a high impact of flood wave superposition. In the case of the Inn/Danube and Mosel/Rhine confluences, floods in the two upstream sub-basins are generated in areas far apart from each other and possibly by different storms or snowmelt events. If both regions are impacted by high precipitation or snowmelt, a flood coincidence may occur, possibly leading to a heavy flood tail downstream.

In summary, the few available studies and the related statistical perspective S1 do not suggest that flood wave superposition at river confluences is a relevant aspect for enhancing the upper-tail behavior of flood peak distributions downstream.

7.3. Dikes Increase the Tail Heaviness up to Certain Point (Dike Failure)

When rivers are embanked by dikes cutting off floodplain retention areas, flood hydrographs are affected by decreasing river cross section area, increasing water levels, and flood wave celerity. This, in turn, increases flood discharges downstream. When dike breaches occur during extreme floods, and large water volumes are abstracted from the main channel and retained in the floodplain storage, the flood peaks are capped, reflecting the original conditions prior to dike construction. Hence, the construction of dikes tends to transform the flood frequency curve toward heavier tails. This effect occurs, however, only up to a certain point, namely the point when dikes fail (Figure 7c). This point will often be the return period of overtopping, although other failure mechanisms can lead to breaches at lower return periods (Vorogushyn et al., 2009).

Observation-based and model-based extrapolation of the flood frequency curve rarely includes dike breaches and their system effects on downstream locations (Vorogushyn et al., 2018). This negligence can lead to wrong estimates of tail heaviness, that is, estimating heavier tails than in reality, for the range of return periods when dikes fail.

Several studies have shown that river flows downstream are influenced by upstream dike failures (Apel et al., 2004; de Bruijn et al., 2014; Curran et al., 2019; Vorogushyn et al., 2012). The effect of dike failures on the downstream flood frequency curve was formally studied by Apel et al. (2009) by simulating the propagation of flood waves along the Lower Rhine. For extreme floods, the model simulates substantial retention effects due to dike breaches, significantly decreasing flood peaks downstream of breach locations. The downstream flood frequency curves derived from their simulation model and an extreme value statistical approach differ substantially beyond return periods of around 1,000 years. The GEV shape parameter decreases from 0.24 for the statistical approach which neglects dike breaches to 0.17 for the simulation approach which includes breaches. The magnitude of this mechanism depends on several factors, such as the ratio of flood wave volume above the threshold to the retention capacity of the dike hinterland. Also, the time point of dike failure is decisive whether the peak flow is capped or the hinterland is filled prior to the arrival of the peak flow. In the latter case the effect on downstream upper tail behavior is limited (Skublics et al., 2014).

8. Synthesis

8.1. Assessment of the Nine Hypotheses

Table 3 summarizes the key findings of our review. We also assign a degree of plausibility and evidence to each hypothesis, visualized in Figure 8. This assignment is to some extent subjective, but we provide a justification for each assignment in Table 3. We define the degree of plausibility to represent the consistency of the hypothesis with known process reasoning. Specifically, we assign a hypothesis:

- High plausibility, if process knowledge highly supports the hypothesis, meaning that there is a clear mechanistic understanding of the emergence of heavy flood tail behavior.
- Medium plausibility, if process knowledge tends to support the hypothesis, but does not provide unambiguous mechanistic explanations of the emergence of heavy tail behavior.
- Low plausibility, if there is no clear mechanistic understanding supporting this hypothesis, or if process reasoning opposes the hypothesis.

The degree of evidence represents the number and types of studies (based on observations, simulation, and statistical perspectives) and the level of their agreement. Specifically, we assign a hypothesis:

- Robust evidence, if there are several studies of relevance for this hypothesis which include, as a whole, all three types (observations, simulations, statistical perspective), and which all agree on the hypothesis.
- Medium evidence, if neither robust nor limited evidence applies.
- Limited evidence, if there are very few studies available or if the available studies do not agree.

We assign high plausibility to the five hypotheses 'Mixture of flood types generate heavy flood tails' (H3), 'Non-linear response to precipitation causes heavy flood tails' (H4), 'Drier catchments have heavier flood tails due to interaction of water balance processes' (H5), 'Construction of reservoirs increases tail heaviness' (H7), and 'Dikes increase the tail heaviness up to a certain point (dike failure)' (H9). This does not mean that those catchments, where we find these potential sources of heavy flood tails, have necessarily heavy-tailed flood peak distributions. The hypotheses rather state conditions that are favorable for the emergence of heavy tails.

Our review assigns low plausibility to the hypotheses 'Heavy flood tails are inherited from heavy rainfall tails' (H1), 'Smaller catchments have heavier flood tails due to less pronounced spatial aggregation effects' (H6), and 'Confluences lead to downstream heavy tail behavior' (H8). Concerning H6 and H8, the low plausibility is a consequence of the fact that we could not identify clear mechanistic explanations, although some hints and possible explanations are found in the literature.

From the low plausibility of H1, based on robust evidence, one shall not conclude that the tail behavior of rainfall can be neglected. Rather it has to be seen as a kind of pre-disposition. If the flood-producing rainfall is heavy-tailed, then the upper flood tail tends to be heavy-tailed. However, the catchment and river network processes are able to strongly modulate the upper tail behavior, and they can easily transform light-tailed rainfall distributions into heavy-tailed flood peak distributions. Further, we reiterate here that our assignment of low plausibility for H1 is limited to the range of return periods where catchment response exerts a substantial influence on flood peaks. This range is typically the focus of flood design and risk management. However, in the extreme case, for instance, when the catchment is saturated, the catchment response loses influence and the flood behavior is dictated by the precipitation characteristics.

Table 3
Summary of Key Findings for the Nine Hypotheses, Including a Justification for the Assigned Degree of Plausibility and Degree of Evidence

Hypothesis	Key findings	Plausibility	Evidence
Atmosphere			
H1 – Heavy flood tails are inherited from heavy rainfall tails.	Although tail heaviness of precipitation matters, it is often not the dominant factor for the flood tail, as catchment and river network processes strongly modulate the rainfall control. For given rainfall characteristics, a wide range of flood tail behavior is possible in the range of return periods that are of typical interest to flood risk management. For very high return periods, catchment response loses its influence and the flood tail tends to be dominated by the tail of the rainfall distribution.	***Low*** Clear mechanistic understanding: runoff generation strongly influences upper tail behavior of floods.	***Robust*** Several relevant studies; including observations, simulations, and statistical perspective; all studies agree.
H2 – The characteristic flood generation process shapes the upper flood tail (e.g., rain vs. snowmelt).	Catchments, where flood generation is dominated by snowmelt, tend to show lighter tails compared to catchments with rainfall-driven floods. Climatic regions whose flood generation processes are characterized by stronger non-linearity tend to show heavier flood tails.	***Medium*** Studies do not provide clear process explanations, but rather hint to possible mechanisms.	***Medium*** Several relevant studies; based on observations only; all agree that characteristic flood generation process may affect upper tail.
H3 – Mixture of flood types generates heavy flood tails.	Can work through 2 effects. (1) If one of the component distributions is heavy-tailed, then the mixture distribution tends to be heavy-tailed. (2) There is a different process that occurs very rarely, but generates much higher peaks. In this case, the distribution of the superposition can be heavy-tailed, even when distributions of both regular floods and rare floods are light-tailed.	***High*** Clear understanding how mixtures can generate heavy tails.	***Limited*** Hardly any studies that explicitly investigate the relation between the mixture of flood types and flood tails.
Catchment			
H4 – Non-linear response to precipitation causes heavy flood tails.	Catchment response plays an important role in emergence of heavy flood tails. Non-linearity in runoff generation, either as non-linear, gradually increasing, or as threshold response, can enhance tail heaviness. Threshold mechanisms in runoff generation cause inflection points and step changes in flood frequency curve. Light-tailed precipitation can be transformed by non-linear catchment response into heavy-tailed flood time series.	***High*** Several mechanisms identified how non-linearity of catchment response causes heavy tails.	***Robust*** Several relevant studies; including observations, simulations, statistical perspective; all agree on statement that non-linearity enhances upper tail heaviness.

Table 3
Continued

Hypothesis	Key findings	Plausibility	Evidence
H5 – Drier catchments have heavier flood tails due to interaction of water balance processes.	Supported by data-based studies at global scale (arid vs. humid catchments) and regional scale (drier vs. wetter catchments). Suggested to result from heavier rainfall tails and from interplay of climatic conditions and catchment response. Dry catchments have heavier tails as precipitation falls on wide range of soil moisture conditions determined by intense evapotranspiration and longer interevent periods, which in turn leads to highly variable runoff coefficients. Rainfall in wet catchments occurs on a narrow range of soil moisture conditions, reflected in more constant runoff coefficients.	***High*** Several mechanisms are suggested.	***Robust*** Several relevant studies; including observations, simulations and statistical perspective; all studies agree.
H6 – Smaller catchments have heavier flood tails due to less pronounced spatial aggregation effects.	Only weakly supported by observed data. Spatial aggregation of precipitation and runoff generation may contribute to less heavy flood tails in larger catchments. Shifting dominance of processes with increasing scale may impede a clear change of upper tail behavior with catchment area.	***Low*** No clear mechanistic explanation available.	***Limited*** Several relevant studies; almost all based on observations; no agreement.
River system			
H7 – Construction of reservoirs increases tail heaviness.	Majority of studies suggest heavier flood tails in regulated rivers. The actual effect of a reservoir depends on its main purpose, operation rules, location within the catchment, and its volume compared to the flow volume contributed by the upstream catchment.	***High*** Clear mechanistic understanding.	***Medium*** Several relevant studies; including observations, simulations, statistical perspective; mostly agreement.
H8 – Confluences lead to downstream heavy-tail behavior.	Available studies and statistical perspective do not conclude that flood wave superposition at river confluences enhances tail heaviness of flood peak distributions downstream. This would only be the case if the arrival of flood waves from the upstream branches would coincide for large floods but not for small floods.	***Low*** No clear mechanistic understanding provided why and under which conditions confluences should enhance tail heaviness.	***Limited*** Very few studies available.
H9 – Dikes increase the tail heaviness up to certain point (dike failure).	As dike breaching is threshold process, it only affects events the far upper tail of flood peak distribution. Neglecting breaches, as typically done, can lead to overestimation of tail heaviness, as breaches can restore the original conditions prior to dike construction.	***High*** Clear mechanistic understanding.	***Limited*** Very few studies are available.

Note. The plausibility of each hypothesis is assessed as low, medium and high, and the degree of evidence is characterized as limited, medium and robust. Plausibility represents the consistency of a hypothesis with process reasoning. Degree of evidence represents the amount, types and agreement of studies.

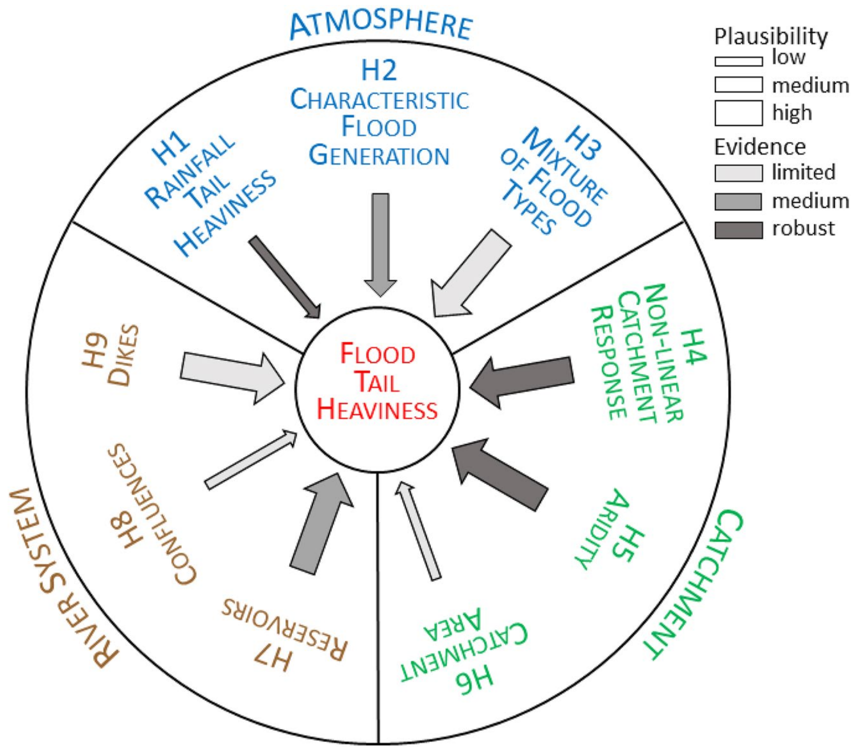

Figure 8. Degree of plausibility and degree of evidence supporting the nine hypotheses. Plausibility is assessed as low, medium, and high, based on the consistency of each hypothesis with process reasoning. Degree of evidence is characterized as limited, medium, and robust, based on the amount, types, and agreement of studies. Justifications for each classification are given in Table 3.

The degree of evidence varies from limited to robust. It is particularly limited to the hypotheses assigned to the river system. Obviously, in terms of heavy flood tails, the flood literature has focused more strongly on atmospheric and catchment processes compared to river processes, including river training.

Figure 9 shows the relations between the nine hypotheses. For four of them, that is, the three hypotheses assigned to the river network (H7-H9) and hypothesis H4 on the role of the non-linearity of runoff generation, the factors are directly linked to the enhancement of flood tail heaviness. The latter plays a particularly important role as a modulator of the effects of the hypotheses linked to the atmosphere and catchment (H1-H3, H5, H6). 'Rainfall tail heaviness' (H1) is directly linked to H4, as together they represent the rainfall-runoff processes.

Both (H1, H4) are affected by the remaining factors. Hence, the effects of 'Characteristic flood generation' (H2) and 'Mixture of flood types' (H3), but also 'Aridity' (H5) and 'Catchment area' (H6) are considered as higher order effects. For instance, the dominant flood generation process (H2) can affect the tail heaviness of rainfall (H1) and catchment response (H4). Regions where floods are caused by heavy, often convective, rainfall tend to have heavier rainfall tails and more non-linear catchment response compared to regions with snow-dominated flood regimes (Bernardara et al., 2008; Smith et al., 2019). Catchments, where floods are generated by a mixture of flood types (H3), tend to show heavier flood tails. This can work through mixtures of meteorological processes (H3 affects H1) or through mixtures of catchment processes (H3 affects H4). An example of the first case is the mixture of floods caused by extratropical rainfall systems and landfalling tropical cyclones (Villarini & Smith, 2010). An example of the latter case is a mixture of subsurface stormflow for long-rain winter floods and infiltration excess runoff for high-intensity summer floods. Similarly, 'Aridity' (H5) and 'Catchment area' (H6) can influence 'Rainfall tail heaviness' (H1) and 'Non-linear catchment response' (H2). Drier regions tend to show more non-linear runoff generation (H5 affects H4), which in combination with more erratic rainfall regimes and probably heavier rainfall tails (H5 affects H1) tends to increase the tail heaviness of floods. Larger catchments tend to show less non-linear runoff generation (H6 affects H4) and lighter rainfall tails (H6 affects H1).

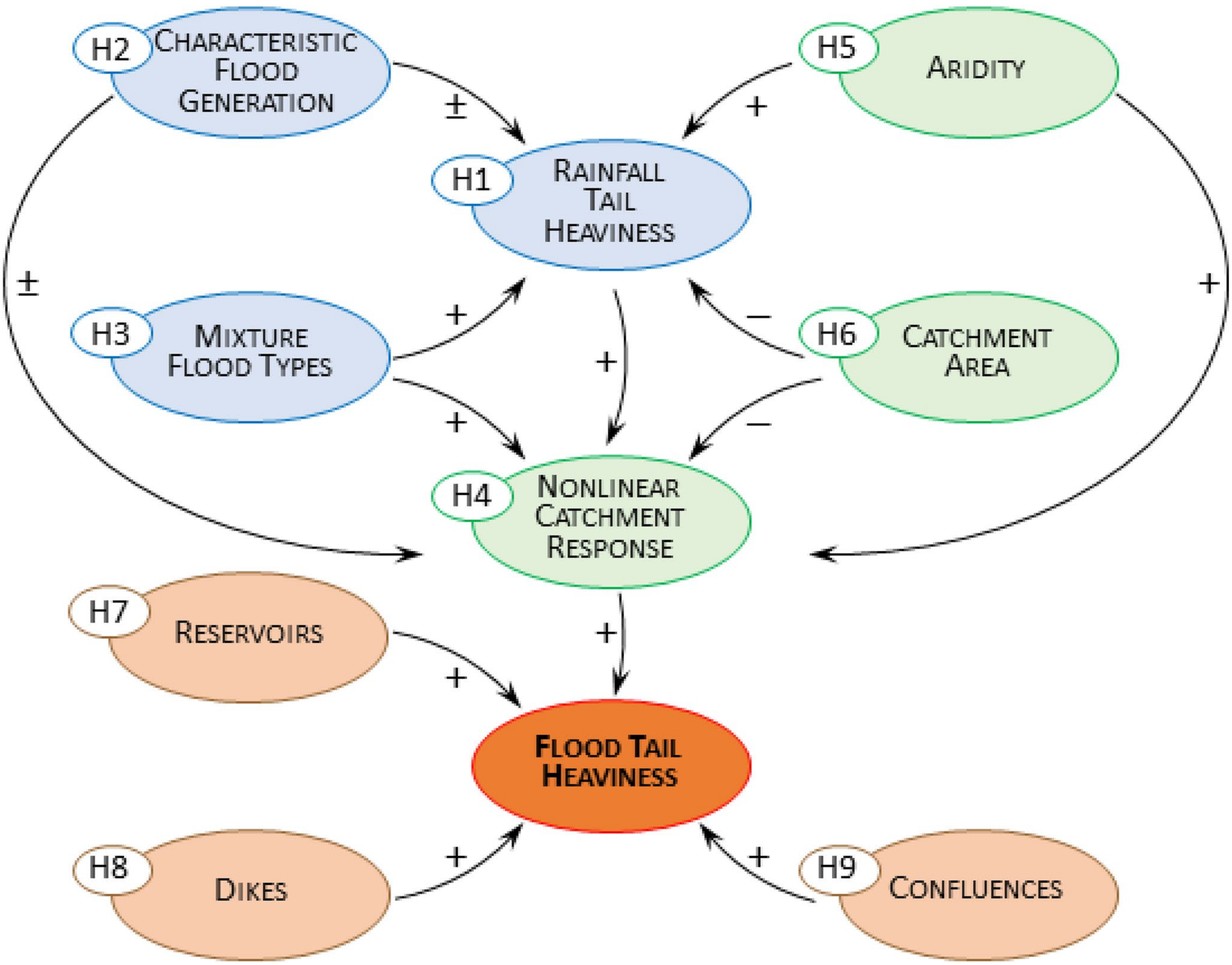

Figure 9. Relations between the nine hypotheses. + and − arrows represent positive/negative cause-effect relations. ± represents an effect that can work positively or negatively. The arrows should not be understood as equally substantiated. Colors denote the compartments atmosphere (blue), catchment (green) and river network (brown). Details are given in the respective hypotheses' sections.

The fact that several inter-related factors that apply simultaneously in a given catchment may explain, besides the uncertainty in estimating heavy tail behavior, why observation-based studies have been limited in understanding the generation of heavy flood tails.

8.2. Implications for Estimating the Flood Tail Behavior

In this section, we discuss the potential avenues of taking advantage of process knowledge for the estimation of tail heaviness. This task is required when estimates of flood discharges of large return periods are needed. Examples are the design of structural flood defense measures with protection standards in the order of 100 or 1,000 years, or flood risk assessments which comprise, in the ideal case, flood scenarios from frequent events to the worst-case scenario. Based on our review, we recommend four guiding questions when confronted with the challenge of estimating tail heaviness. The level of detail and the specific approaches to answer these questions will vary strongly between assessments based on the specific circumstances, such as the purpose of the assessment, data availability, available expertise, and resources.

1. Which processes lead to floods and affect flood peaks in the catchment under study? Do these processes favor heavy-tailed flood peak distributions?

Our review has identified a range of factors that favor heavy flood tails. We recommend identifying the flood generation processes aiming at understanding whether factors that favor heavy tails apply in the catchment under study. This is, in the first instance, the non-linearity of runoff generation (H4). Examples are a non-linear increase in the contributing area with increasing rainfall or a more rapid transfer of direct runoff during extreme rainfall events leading to much higher flood peaks compared to more frequent rainfall events. Runoff generation has to be analyzed in combination with the rainfall regime and the upper tail of the flood-triggering rainfall (H1), as more erratic rainfall and runoff behavior can signal heavier flood tails. We also recommend identifying the flood types, as the mixture of types can favor heavy-tailed behavior (H3), particularly if one type occurs in rare cases only. If the catchment is characterized by a single flood type, this specific type can also support the estimation of the upper tail behavior, as for instance, snow-related types tend to show lighter tails (H2). In the case of the presence of reservoirs (H7) and dikes (H9) in the catchment, their possible influence on flood peaks should be considered as well. This guiding question about the flood processes resonates with the call of Merz and Blöschl (2008) to enhance flood frequency analysis by causal information expansion.

2. Are large floods different from small floods? Might there occur processes in extreme cases which we do not see in the observations?

Heavy flood tails can emerge in two, principally different, ways: First, the processes that cause heavy tail behavior work across the entire spectrum of events, and small and large events show the same system dynamics. For instance, a highly non-linear catchment response enhances the tail heaviness of floods, and this non-linearity is already manifest in frequent rainfall-runoff events. In that case, the estimation of flood tail heaviness can use much larger sets of observations including small floods, and even ordinary streamflow values following the meta-statistical approach (Marani & Ignaccolo, 2015; Tarasova et al., 2020). Second, heavy flood tails are produced by the emergence of processes that do not occur during small events. Examples are threshold processes in runoff generation that occur only in rare cases, unseasonal floods, such as an unusually intensive rainfall in winter on the frozen ground (Wendi et al., 2019), or rare meteorological phenomena, such as landfalling hurricanes in the eastern United States (Villarini & Smith, 2010). Smith et al. (2018) coined the term 'strange floods' for events for which the flood-generating mechanism is rare in the given catchment and contrasts markedly with the common flood-generating mechanisms.

We recommend analyzing whether the flood processes, identified under guiding question 1, occur across the whole spectrum of events. To this end, flood processes and flood characteristics, such as flood-causing storm tracks, flood types, flood timing, runoff coefficients, or flood routing characteristics should be compared between small and large floods (e.g., Nakamura et al., 2013; Smith et al., 2018; Tarasova et al., 2020). Further, we recommend reflecting whether processes might occur in extreme cases which are not included in the observations. Neglecting such processes could lead to an underestimation of tail heaviness. Merz et al. (2015) discussed unexpected incidents in flood risk assessments and provide recommendations to better understand and reduce the potential for surprise. One approach is, for example, the development of downward counterfactuals. These are alternative realizations of past events with a worse outcome than in reality (Woo, 2019). In relation to H8, for example, one could investigate how past floods would have developed in case the flood waves of the upstream rivers would have arrived at the confluence at the same time.

3. How can this process knowledge be used to inform the estimation of the upper tail behavior?

Typically, flood discharges of large return periods and the upper tail behavior of flood distributions are estimated in three ways: at-site extreme value statistics based on observed flood peaks and extrapolation of the local flood frequency curve, regional approaches based on observed peaks, and derived flood frequency using simulation models. When using at-site flood frequency analysis, the answers to guiding questions 1 and 2 help to understand whether we can safely extrapolate from small floods to floods of large return periods and whether the fitted distribution shows plausible upper tail behavior. Merz and Blöschl (2008) provide examples of how hydrological reasoning and soft data, for example, landforms from maps, and hydrological activity from field trips, can support formal flood frequency analysis. If process reasoning suggests that processes can occur that are not or very rarely included in the observations, then one should consider enhancing the analysis by regional and simulation approaches. For instance, in the case of embanked rivers, the process of dike failure is typically not included in the observations. A simulation approach can assist in understanding how this process affects the upper tail.

When using regional flood frequency analysis, guiding questions 1 and 2 can help decide how to group catchments, for example, whether the shape of the flood frequency curve or its growth curve can be assumed as constant within a region. When the tail-influencing processes vary slowly in space, assuming a constant shape value throughout the region is a plausible strategy. This is particularly the case for climatic factors. An example is the distribution of flood types, for example, whether floods are only caused by midlatitude cyclones or also by tropical cyclones, is a regional phenomenon. Other processes can vary rapidly in space. An example is the non-linearity of the runoff generation, which is often determined by local characteristics. Rogger, Pirkl, et al. (2012) pointed to this difference in their discussion about step changes in flood frequency curves. When such step changes are triggered by climatic controls, then pooling catchments in a climatic homogeneous region seems valuable. When they are caused by local characteristics, pooling can mask the variability of the upper tail behavior and lead to the over-/underestimation of floods with large return periods.

When using simulation to estimate the flood frequency curve, guiding questions 1 and 2 can inform the model development. It should be secured that the processes that shape the upper tail are included in the model. For example, the incorporation of reservoirs and dikes and their potential failures into hydrological and hydrodynamic models can increase the robustness of tail estimation (Apel et al., 2009). Often knowledge and data about these processes are scarce, yet one should reflect on whether the model is a plausible representation of these processes. This requires going beyond the comparison of simulated and observed hydrographs via model performance measures. Model calibration and validation could focus on the replication of factors relevant to tail-heaviness. For instance, hydrological models could be tested to whether they adequately reproduce flood types and their frequencies, when their mixture is deemed to affect flood tails. Complex model structures and overparameterization should be avoided, as they add uncertainty without assuring the capability to mimic key elements to estimate flood tails, such as the catchment response and recession behaviors (Biswal & Singh, 2017). Simplified descriptions of catchment dynamics (Kirchner, 2009), whose parameters are easily constrained by means of daily flow series, could be used instead and their emerging tail properties studied. Well-established knowledge on the role played by physical watershed attributes, such as the dendritic structure of rivers (Biswal & Marani, 2010; Rinaldo et al., 2006) for the runoff response, which is commonly used to estimate event-based peak flows, could be as well leveraged to assess flood probabilities and their tails.

Irrespective of the selected approach one needs to carefully consider whether the flood processes or the underlying drivers can be assumed to be constant in time. Sources of time-varying flood peak distributions, such a climate-related flood-rich and flood-poor periods or construction of reservoirs, should be considered in the estimation of the upper tail behavior.

4. What could be the consequences of erroneous estimation of the flood tail heaviness?

Answering the three guiding questions above does not guarantee that the flood tail heaviness is estimated correctly. There will still be cases where the tail heaviness is strongly underestimated, that is, the probability of a large flood discharge occurring is much higher than estimated. The final guiding question thus attempts to understand the consequences of a possible underestimation of tail heaviness. They will vary strongly from case to case and depend on the exposure and vulnerability of the flood-prone areas and on the purpose of the assessment. An underestimation of the tail heaviness when designing the flood defense of a major infrastructure leads, for instance, to more severe consequences than a similar underestimation when designing a dike for agricultural areas. The answer to this question is thus essential information for developing risk management strategies (Merz et al., 2015).

8.3. Recommendations for Further Research

Based on our review, we propose the following research strands to better understand the generation of heavy-tailed flood distributions with the final aim of improved upper tail estimates.

1. Clarifying the hypotheses where clear mechanistic understanding is missing or evidence is limited

Our review assigns low and medium plausibility to four hypotheses (H1, H2, H6, H8) due to the lack of a clear mechanistic understanding. We recommend further clarifying these hypotheses with carefully designed studies to improve the mechanistic understanding of flood tail heaviness. Simple correlation analyses between flood tail indicators and catchment indicators representing potential drivers of tail heaviness do not suffice to unravel the underlying mechanisms, as several drivers work at the same time. The contributions of drivers that may be jointly responsible for heavy tail behavior (e.g., more likely heavy-tailed rainfall distributions in smaller catchments and, at the same time, higher non-linearity in runoff generation) should be assessed. Carefully designed studies should also help to enhance the degree of evidence for the hypotheses where it is limited (H3, H6, H8, H9). For instance, the effects of dikes and confluences, and the related role of the dendritic structure of river networks, on tail heaviness have not received much attention in the hydrological literature. Simulation experiments could be designed to systematically quantify the effects of dikes and their failures on tail heaviness and how these effects depend on the characteristics of the river-dike-floodplain system. For the river confluence problem, systematic simulation experiments could analyze the variability in catchment response time as a function of atmospheric conditions to clarify whether certain atmospheric constellations could produce temporal peak matching and enhance tail heaviness downstream of the confluence.

2. Clarifying the role of changes in system dynamics with increasing flood magnitude

The question of whether the dynamics of the flood generation change from small to large floods is central to understanding flood tail heaviness. We recommend rigorously investigating whether and how processes, from atmospheric through catchment to river network processes, change with flood magnitude. Large-sample hydrology, exploiting datasets of large sets of catchments, has the potential to derive robust conclusions and test hydrological hypotheses across a variety of regions (Addor et al., 2020; Andréassian et al., 2009). It is thus a prime candidate for addressing the question of whether large and small floods are different.

Research on flood event classification would also help to answer this question. Flood events often show complex space-time dynamics which challenges their classification. To date, there is no agreement about the ingredients of a good flood event classification and hardly any attempt to compare or validate the results of different classifications (Stein et al., 2020; Tarasova et al., 2019). More rigorous testing including uncertainty analysis and extending classification methods to include indicators of space-time dynamics of flood events are required (Tarasova et al., 2019). Improved flood type classification would help in understanding whether small and large floods belong to different flood types.

However, if the system dynamics are the same for small and large floods and even for daily streamflow oscillations, the flood tails could be derived from the tail behavior of daily flows. Besides process studies, the statistical literature offers novel concepts in this regard including metastatistical approaches for which only singular applications in hydrology exist (e.g., Miniussi, Marani, & Villarini, 2020). It has been shown that interactions of stochastic fluctuations occurring at different timescales (e.g., daily and interannual) lead to heavier tails in the distribution of the considered random variable (Porporato et al., 2006). The Metastatistical Extreme Value (MEV) distribution (Zorzetto et al., 2016) acknowledges the existence of these stochastic fluctuations in rainfall, streamflow, and the resulting floods and provides a rigorous approach to considering them for the estimation of extremes (Marra et al., 2018, 2019, 2020; Schellander et al., 2019; Miniussi & Marani, 2020; Miniussi, Marani, & Villarini, 2020, Miniussi, Villarini, & Marani, 2020; Zorzetto & Marani, 2019, 2020).

3. Clarifying the geography of heavy flood tails

Observation-based analyses of upper tail flood behavior show rather erratic spatial patterns. Although it is clear that these spatial patterns are disturbed by large sampling uncertainty when estimating the upper tail behavior, a question is whether geography of heavy flood tails can be established which is able to inform flood design and risk management. Such geography would map the upper tail behavior across regions and would identify the hotspots of heavy tail behavior.

A way forward would be to identify the dominant factors of flood generation within regions and catchments and to investigate their upper tail behavior and how they combine to generate extreme floods. For instance, Su and Smith (2021) investigated annual maximum values of precipitable water and of vertically-integrated water vapor flux across the conterminous US, arguing that these factors are key ingredients for extreme precipitation. They found spatial clusters and larger regions of heavy-tailed behavior for precipitable water and water vapor flux, respectively. These areas of heavy tail behavior could be linked to tropical cyclones and extratropical systems. Studying Upper Tail Ratios across China, Yang, Yang, and Smith (2021) found large upper tail ratios mainly distributed north of the Yangtze River, with a striking concentration in the middle reach of the Yellow River.

These record floods were caused by anomalous moisture transport, for example, zonal moisture transport associated with tropical cyclones, and/or synoptic configurations, for example, blocking, and interaction with complex terrain. The interaction between storm properties, for example, size, motion, convective intensity, and mountainous terrain has been identified as a decisive factor for 'hotspots' of extreme floods, for example, in Colorado (Smith et al., 2019) and the Blue Mountains of eastern Oregon (Smith et al., 2018).

4. Investigating the events that dominate the upper tail

An improved understanding of heavy flood tails would particularly benefit from in-depth investigations of outstanding flood events. For instance, Rössler et al. (2014) analyzed an extreme flood in Switzerland that was caused by sustained snowfall followed by an atmospheric river, associated with exceptional moisture transport to the catchment, rapid temperature increase, and high rainfall intensities. The resulting rain-on-snow flood showed specific characteristics and could only be simulated by extensive changes to the hydrological model that had been used successfully for flood forecasting prior to this event. Similarly, analyses of record floods and 'strange floods' have shown that the largest floods are often caused by specific interactions or space-time variabilities of atmospheric, catchment and river processes (e.g., Dettinger et al., 2004; Neimann et al., 2011; Smith et al., 2018, 2019; Yang et al., 2017). More systematic reconstructions and documentation of extreme floods, using the entire range of observational and simulation approaches, would contribute to unraveling the ingredients of the events that dominate the upper tail of flood peak distributions.

5. Clarifying the effects of temporal changes on upper flood tails

Detecting changes in flood time series and attributing them to the underlying drivers has developed into an active research topic. These studies often focus on climate change (e.g., Blöschl et al., 2017) and climate variability (e.g., Hodgkins et al., 2017), but there are also attempts to consider other drivers, such as reservoirs or land management (e.g., Yang, Yang, Villarini, et al., 2021), and to separate the effects of different drivers (e.g., Viglione et al., 2016). Almost all studies, however, do not distinguish between small and large floods and thus assume implicitly that small and large floods change in parallel. Approaches are required that allow understanding whether the upper tail of flood peak distributions is affected differently than the bulk of the floods. To this end, the recently developed 'Extreme Event Attribution', which attempts to quantify how anthropogenic climate change has affected the likelihood of specific extreme (flood) events (e.g., Kay et al., 2018), is highly promising. Another promising research line is the attempt to detect and attribute changes that differentiates between small and large floods. Bertola et al. (2020, 2021) showed that the 100-year flood changed differently than the 2-year flood in some European regions and that these changes can be attributed to different drivers. Similarly, recent studies have shown that small and large floods respond differently to changes in precipitation (Brunner et al., 2021; Wasko et al., 2019, 2021).

9. Conclusions

We proposed nine hypotheses on the mechanisms causing heavy tails in flood peak distributions. Based on our review, we draw the following conclusions:

1. 'Heavy flood tails are inherited from heavy rainfall tails' (H1): Although the tail heaviness of rainfall matters, H1 is hardly plausible, as the transformation of rainfall to flood peaks strongly modifies the tail. However, in the extreme case, for example, when the catchment gets saturated, the catchment response loses its influence and the flood tail tends to be dominated by the tail of the rainfall distribution.
2. 'The characteristic flood generation process shapes the upper flood tail' (H2): Data-based studies have consistently reported that regions dominated by different flood generation processes often show differences in flood tail behavior. Snowmelt-dominated flood regimes tend to show lighter tails compared to rainfall-driven flood regimes, and there is a tendency toward heavier flood tails in regions where flood generation is characterized by stronger non-linearity.
3. 'Mixture of flood types generates heavy flood tails' (H3): Mixing flood types propagates heavy tails of component distributions to the mixed distribution and is able to generate heavy tails. H3 is highly plausible, but there are hardly any studies on the relation between flood-type mixtures and tail heaviness.
4. 'Non-linear response to precipitation causes heavy flood tails' (H4): Highly plausible hypothesis, as there is a clear mechanistic understanding of how the non-linearity of runoff generation, in combination with the

10.1029/2021WR030506

rainfall regime, is able to produce heavy flood tails. For instance, non-linearity in catchment response, either as non-linear, gradually increasing, or as threshold response, can easily transform light-tailed precipitation into heavy-tailed flood time series.

5. 'Drier catchments have heavier flood tails due to interaction of water balance processes' (H5): Data-based studies at the global and continental scale, and to a lesser extent at the regional scale, support H5. There are some mechanistic explanations of how the interplay of rainfall, evapotranspiration, and runoff generation yields heavier flood tails in drier catchments.
6. 'Smaller catchments have heavier flood tails due to less pronounced spatial aggregation effects' (H6): Spatial aggregation of precipitation and runoff generation may contribute to less heavy flood tails in larger catchments. However, H6 is little plausible as there is no convincing mechanistic understanding supporting it and only weak support by data. Rather, we assume that the emergence or shifting dominance of processes with increasing scale impedes a clear relation between catchment size and flood tail heaviness.
7. 'Construction of reservoirs increases tail heaviness' (H7): Highly plausible hypothesis given clear mechanistic understanding, that is, strong retention of small floods but small or no retention of large floods. Studies mostly agree that the construction of reservoirs tends to produce more skewed flood distributions, although there is a lack of studies that explicitly analyze the effects on the tail behavior. The actual effect of a reservoir depends on its purpose, the operation rules, its location within the catchment, and its volume compared to the flow volume contributed by the upstream catchment.
8. 'Confluences lead to downstream heavy-tail behavior' (H8): Little plausible hypothesis as there is no clear mechanistic understanding of why and under which conditions confluences enhance tail heaviness. There is also limited evidence for H8 due to a small number of studies on the relation between confluences and upper flood tails.
9. 'Dikes increase the tail heaviness up to a certain point (dike failure); (H9): Clear mechanistic understanding supports H9, although there are only a few studies on this topic. Neglecting dike breaches in flood frequency analysis, as typically done, can lead to wrong estimates of tail heaviness beyond the threshold where dikes fail.

Our discussion of statistical perspectives helps to understand how heavy flood tails emerge through the aggregation, mixture, or transformation of components. When confronted with the estimation of the tail behavior in a given catchment, however, we recommend identifying the processes to understand whether factors apply that favor heavy flood tails. To clarify the hypotheses where clear mechanistic understanding is missing or evidence is limited, we recommend carefully designed simulation and data exploration studies. Such studies should particularly investigate whether/how flood generation changes with increasing magnitude. Comparisons of flood generation processes between small and large floods should consider the entire spectrum of processes, from the atmospheric triggering mechanisms to catchment and river network processes.

Appendix A

A1. Arithmetic Combination of Random Variables (S1)

Table A1
Assumptions on Independent Random Variables X and Y, and Additional Constraints, in Order for Their Sum to Belong to a Specific Subclass of Heavy-Tailed Distributions

X	Y	Additional constraints	$X + Y$
Long-tailed	Long-tailed		Long-tailed
Long-tailed	$\overline{F}_Y(x) = o\left(\overline{F}_X(x)\right)$		Long-tailed
Sub-exponential	Sub-exponential	$\overline{F}_X$ and $\overline{F}_Y$ are weakly tail-equivalent[1]	Sub-exponential
Sub-exponential	$\overline{F}_Y(x) = o\left(\overline{F}_X(x)\right)$		Sub-exponential
Regularly varying	Regularly varying		Regularly varying
Stable (α-stable)	Stable (α-stable)		Stable (α-stable)

Note. These results can be generalized to the sum of n variables. For details see Foss et al. (2011) and Embrechts et al. (1997)
[1] Two distributions F and G without upper bounds on their support are called weakly tail-equivalent if there exist $c_1 > 0$ and $c_2 < \infty$ such that for any x > 0: $c_1 \le \frac{\overline{F}(x)}{\overline{G}(x)} \le c_2$.

Table A2
Assumptions on Random Variables X and Y, and Additional Constraints, in Order for Their Product to Belong to a Specific Subclass of Heavy-Tailed Distributions

X	Y	Additional constraints	$X \cdot Y$
Long-tailed	Long-tailed		Long-tailed
Long-tailed	Light-tailed		Long-tailed
Sub-exponential		See Proposition 3.1. and Theorem 3.1. in Su and Chen (2006)	Sub-exponential
Regularly varying with index α	Regularly varying with index α		Regularly varying with index α
Regularly varying with index α		$EY^{\alpha+\varepsilon} < \infty$ for some $\varepsilon > 0$	Regularly varying with index α

Note. All random variables are assumed to be non-negative and independent. In the case of sub-exponential distributions, sufficient conditions for the sub-exponentiality of the product are too long to conveniently fit in the table, but can be found in Su and Chen (2006). For details see Mikosch (1999) and Su and Chen (2006).

A2. Transformations of Random Variables (S3): Results for Selected Transformations That Are Relevant for Hydrological Purposes

Halliwell (2013) investigates the impact of power and exponential transformations on the tails of positive, real-valued random variables and provides sub-categories for light-tailed distributions that he calls light- and medium-tailed distributions. Distributions which do not have all moments finite remain heavy-tailed under power and exponential transformations. Light- and medium-tailed distributions can become heavy-tailed or light-tailed depending on the exponent and on the distribution. Similarly, using an exponential function for the transformation preserves heavy tails, turns medium-tailed variables into heavy-tailed ones and has to be investigated more closely for light-tailed variables. Logarithmic transformations turn light- and medium-tailed variables into light-tailed and have to be more closely investigated for heavy-tailed variables. Examples are the normal distribution which can be transformed to the Lognormal distribution, a heavy-tailed distribution, or the Pareto distribution which is obtained by applying the exponential transformation to the exponential distribution.

For the reciprocal transformation $T(x) = 1/x$ general statements are not available. However, if the density of X is greater than 0 in 0, $T(X)$ does not have finite mean and hence not existing moments (Lehmann & Shaffer, 1988). A classic example is the Cauchy distribution which arises as the reciprocal of a normally distributed random variable.

Very general transformations for the generation of heavy tails from any random variable, the so-called Lambert-W-transformations, are given by Goerg (2015). These transformations have a tail parameter, determining the tail behavior, and can also be used to transform heavy-tailed data into a light-tailed distribution. They are directly connected to the Tukey-transformations which can generate heavier tails for the transformed random variables (Fischer, 2010).

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

Data were not used, nor created for this research.

Acknowledgments
The financial support of the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) for the research group FOR 2416 "Space-Time Dynamics of Extreme Floods (SPATE)" and for the project PREDICTED (421,396,820) is gratefully acknowledged. We also acknowledge the support of the Austrian Science Funds (FWF) "SPATE" project (Grant No. I 3174). We thank James Smith and two anonymous reviewers for their thoughtful and thorough reviews. Open Access funding enabled and organized by Projekt DEAL.

References

Addor, N., Do, H., Alvarez-Garreton, C., Coxon, G., Fowler, K., & Mendoza, P. (2020). Large-sample hydrology: Recent progress, guidelines for new datasets and grand challenges. *Hydrological Sciences Journal*, *65*(5), 712–725. https://doi.org/10.1080/02626667.2019.1683182

Albrecher, H., Asmussen, S., & Kortschak, D. (2006). Tail asymptotics for the sum of two heavy-tailed dependent risks. *Extremes*, *9*(2), 107–130. https://doi.org/10.1007/s10687-006-0011-1

Andréassian, V., Perrin, C., Berthet, L., Le Moine, N., Lerat, J., Loumagne, C., et al. (2009). HESS opinions: Crash tests for a standardized evaluation of hydrological models. *Hydrology and Earth System Sciences*, *13*, 1757–1764. https://doi.org/10.5194/hess-13-1757-2009

Apel, H., Merz, B., & Thieken, A. H. (2009). Influence of dike breaches on flood frequency estimation. *Computers & Geosciences*, *35*(5), 907–923. https://doi.org/10.1016/j.cageo.2007.11.003

Apel, H., Thieken, A. H., Merz, B., & Blöschl, G. (2004). Flood risk assessment and associated uncertainty. *Natural Hazards and Earth System Sciences*, *4*(2), 295–308. https://doi.org/10.5194/nhess-4-295-2004

Assani, A. A., Stichelbout, É., Roy, A. G., & Petit, F. (2006). Comparison of impacts of dams on the annual maximum flow characteristics in three regulated hydrologic regimes in Québec (Canada). *Hydrological Processes*, *20*, 3485–3501. https://doi.org/10.1002/hyp.6150

Ayalew, T. B., Krajewski, W. F., & Mantilla, R. (2013). Exploring the effect of reservoir storage on peak discharge frequency. *Journal of Hydrologic Engineering*, *18*, 1697–1708. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000721

Ayalew, T. B., Krajewski, W. F., & Mantilla, R. (2015). Insights into expected changes in regulated flood frequencies due to the spatial configuration of flood retention ponds. *Journal of Hydrologic Engineering*, *20*, 1–10. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001173

Ayalew, T. B., Krajewski, W. F., Mantilla, R., & Wright, D. B. (2017). Effect of spatially distributed small dams on flood frequency: Insights from the Soap Creek watershed. *Journal of Hydrologic Engineering*, *22*. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001513

Balkema, A. A., & de Haan, L. (1974). Residual life time at great age. *Annals of Probability*, *2*(5), 792–804. https://doi.org/10.1214/aop/1176996548

Barth, N. A., Villarini, G., Nayak, M. A., & White, K. (2017). Mixed populations and annual flood frequency estimates in the Western United States: The role of atmospheric rivers. *Water Resources Research*, *53*(1), 257–269. https://doi.org/10.1002/2016wr019064

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

Bender, J., Wahl, T., Müller, A., & Jensen, J. (2016). A multivariate design framework for river confluences. *Hydrological Sciences Journal*, *61*(3), 471–482. https://doi.org/10.1080/02626667.2015.1052816

Berghuijs, W. R., Sivapalan, M., Woods, R. A., & Savenije, H. H. G. (2014). Patterns of similarity of seasonal water balances: A window into streamflow variability over a range of time scales. *Water Resources Research*, *50*(7), 5638–5661. https://doi.org/10.1002/2014WR015692

Bernardara, P., Schertzer, D., Eric, S., Tchiguirinskaia, I., & Lang, M. (2008). The flood probability distribution tail: How heavy is it? *Stochastic Environmental Research and Risk Assessment*, *22*, 107–122. https://doi.org/10.1007/s00477-006-0101-2

Bertola, M., Viglione, A., Lun, D., Hall, J., & Blöschl, G. (2020). Flood trends in Europe: Are changes in small and big floods different? *Hydrology and Earth System Sciences*, *24*(4), 1805–1822. https://doi.org/10.5194/hess-24-1805-2020

Bertola, M., Viglione, A., Vorogushyn, S., Lun, D., Merz, B., & Blöschl, G. (2021). Do small and large floods have the same drivers of change? A regional attribution analysis in Europe. *Hydrology and Earth System Sciences*, *25*(3), 1347–1364. https://doi.org/10.5194/hess-25-1347-2021

Beven, K. (1987). Towards the use of catchment geomorphology in flood frequency predictions. *Earth Surface Processes and Landforms*, *12*(1), 69–82. https://doi.org/10.1002/esp.3290120109

Billingsley, P. (1995). *Probability and measure* (3rd ed.). John Wiley & Sons (ISBN: 0-471-00710-2).

Biswal, B., & Marani, M. (2010). Geomorphological origin of recession curves. *Geophysical Research Letters*, *37*(24), l24403. https://doi.org/10.1029/2010GL045415

Biswal, B., & Singh, R. (2017). Incorporating channel network information in hydrologic response modelling: Development of a model and inter-model comparison. *Advances in Water Resources*, *100*, 168–182. https://doi.org/10.1016/j.advwatres.2016.12.015

Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A. P., Merz, B., Arheimer, B., et al. (2017). Changing climate shifts timing of European floods. *Science*, *357*(6351), 588–590. https://doi.org/10.1126/science.aan2506

Blöschl, G., Merz, R., & Reszler, C. (2006). Floods in Austria. In O. Vasiliev, P. van Gelder, E. Plate, & M. Bolgov (Eds.), *Extreme Hydrological Events: New Concepts for Security*, (Vol. 78, pp. 81–90). Springer. https://doi.org/10.1007/978-1-4020-5741-0_6

Blöschl, G., Nester, T., Komma, J., Parajka, J., & Perdigão, R. A. P. (2013). The June 2013 flood in the upper Danube basin, and comparisons with the 2002, 1954 and 1899 floods. *Hydrology and Earth System Sciences*, *17*(12), 5197–5212. https://doi.org/10.5194/hess-17-5197-2013

Blöschl, G., & Zehe, E. (2005). On hydrological predictability. *Hydrological Processes*, *19*(19), 3923–3929. https://doi.org/10.1002/hyp.6075

Böttcher, F., Barth, S., & Peinke, J. (2007). Small and large scale fluctuations in atmospheric wind speeds. *Stochastic Environmental Research and Risk Assessment*, *21*, 299–308. https://doi.org/10.1007/s00477-006-0065-2

Botter, G. (2010). Stochastic recession rates and the probabilistic structure of stream flows. *Water Resources Research*, *46*, W12527. https://doi.org/10.1029/2010WR009217

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*, 1–14. https://doi.org/10.1029/2009WR008523

Brunner, M. I., Swain, D. L., Wood, R. R., Willkofer, F., Done, J. M., Gilleland, E., & Ludwig, R. (2021). An extremeness threshold determines the regional response of floods to changes in rainfall extremes. *Communications Earth & Environment*, *2*(1), 173. https://doi.org/10.1038/s43247-021-00248-x

Carreau, J., Naveau, P., & Sauquet, E. (2009). A statistical rainfall-runoff mixture model with heavy-tailed components. *Water Resources Research*, *45*, W10437. https://doi.org/10.1029/2009WR007880

Cavanaugh, N. R., Gershunov, A., Panorska, A. K., & Kozubowski, T. J. (2015). The probability distribution of intense daily precipitation. *Geophysical Research Letters*, *42*(5), 1560–1567. https://doi.org/10.1002/2015GL063238

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer. https://doi.org/10.1007/978-1-4471-3675-0

Cooke, R. M., & Nieboer, D. (2011). Heavy-tailed distributions: Data, Diagnostics, and new developments. *Resources for the Future*, *11–19*, 64. RFF DP.

Curran, A., de Bruijn, K. M., Klerk, W. J., & Kok, M. (2019). Large scale flood hazard analysis by including defence failures on the Dutch river system. *Water*, *11*(8), 1732. https://doi.org/10.3390/w11081732

Davidson, R. (2012). Statistical inference in the presence of heavy tails. *The Econometrics Journal*, *15*, C31–C53. https://doi.org/10.1111/j.1368-423X.2010.00340.x

de Bruijn, K. M., Diermanse, F. L. M., & Beckers, J. V. L. (2014). An advanced method for flood risk analysis in river deltas, applied to societal flood fatality risk in The Netherlands. *Natural Hazards and Earth System Sciences*, *14*, 2767–2781. https://doi.org/10.5194/nhess-14-2767-2014

Delgado, J. M., Apel, H., & Merz, B. (2010). Flood trends and variability in the Mekong river. *Hydrology and Earth System Sciences*, *14*(3), 407–418. https://doi.org/10.5194/hess-14-407-2010

Delle Rose, M. (2012). Decision-making errors and socio-political disputes over the Vajont dam disaster. *Disaster Advancement*, *5*(3), 144–152.

Dettinger, M., Redmond, K., & Cayan, D. (2004). Winter orographic precipitation ratios in the Sierra Nevada – Large-scale atmospheric circulations and hydrologic consequences. *Journal of Hydrometeorology*, *5*(6), 1102–1116. https://doi.org/10.1175/jhm-390.1

Dyrrdal, A. V., Skaugen, T., Stordal, F., & Førland, E. J. (2016). Estimating extreme areal precipitation in Norway from a gridded dataset. *Hydrological Sciences Journal*, *61*(3), 483–494. https://doi.org/10.1080/02626667.2014.947289

Eagleson, P. S. (1972). Dynamics of flood frequency. *Water Resources Research*, *8*(4), 878–898. https://doi.org/10.1029/WR008i004p00878

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

10.1029/2021WR030506

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

Farquharson, F. A. K., Meigh, J. R., & Sutcliffe, J. V. (1992). Regional flood frequency analysis in arid and semi-arid areas. *Journal of Hydrology*, *138*(3), 487–501. https://doi.org/10.1016/0022-1694(92)90132-F

Felder, G., & Weingartner, R. (2017). Assessment of deterministic PMF modelling approaches. *Hydrological Sciences Journal*, *62*(10), 1591–1602. https://doi.org/10.1080/02626667.2017.1319065

Fernandez, W., Naghettini, M., & Loschi, R. (2010). A Bayesian approach for estimating extreme flood probabilities with upper bounded distribution functions. *Stochastic Environmental Risk Assessment*, *24*(8), 1127–1143. https://doi.org/10.1007/s0047-010-0365-4

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, M. (2010). Generalized Tukey-type distributions with application to financial and teletraffic data. *Statistical Papers*, *51*(1), 41. https://doi.org/10.1007/s00362-007-0114-z

Fischer, S. (2018). A seasonal mixed-POT model to estimate high flood quantiles from different event types and seasons. *Journal of Applied Statistics*, *45*(15), 2831–2847. https://doi.org/10.1080/02664763.2018.1441385

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

Fischer, S., & Schumann, A. (2020). Spatio-temporal consideration of the impact of flood event types on flood statistic. *Stochastic Environmental Research and Risk Assessment*, *34*(9), 1331–1351. https://doi.org/10.1007/s00477-019-01690-2

Fisher, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, *24*(02), 180–190. https://doi.org/10.1017/S0305004100015681

Foss, S., Korshunov, D., & Zachary, S. (2011). *An introduction to heavy-tailed and subexponential distributions* (Vol. 6, pp. 0090–6778). Springer.

Foss, S., Korshuov, D., & Zachary, S. (2009). Convolutions of long-tailed and subexponential distributions. *Journal of Applied Probability*, *46*(3), 756–767.

Gaume, E. (2006). On the asymptotic behavior of flood peak distributions. *Hydrology and Earth System Sciences*, *10*(2), 233–243. https://doi.org/10.5194/hess-10-233-2006

Geertsema, T. J., Teuling, A. J., Uijlenhoet, R., Torfs, P. J. J. F., & Hoitink, A. J. F. (2018). Anatomy of simultaneous flood peaks at a lowland confluence. *Hydrology and Earth System Sciences*, *22*(10), 5599–5613. https://doi.org/10.5194/hess-22-5599-2018

Gioia, A., Iacobellis, V., Manfreda, S., & Fiorentino, M. (2008). Runoff thresholds in derived flood frequency distributions. *Hydrology and Earth System Sciences*, *12*, 1295–1307. https://doi.org/10.5194/hess-12-1295-2008

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*(3), 937–951. https://doi.org/10.5194/hess-16-937-2012

Gottschalk, L., & Weingartner, R. (1998). Distribution of peak flow derived from a distribution of rainfall volume and runoff coefficient, and a unit hydrograph. *Journal of Hydrology*, *208*(3–4), 148–162. https://doi.org/10.1016/S0022-1694(98)00152-8

Gotvald, A. J., Barth, N. A., Veilleux, A. G., & Parrett, C. (2012). Methods for determining magnitude and frequency of floods in California, based on data through water year 2006. *U.S. Geological Survey: Scientific Investigation Report*, *2012–5113*, 38.

Guerin, A., Devauchelle, O., Robert, V., Kitou, T., Dessert, C., Quiquerez, A., et al. (2019). Stream-discharge surges generated by groundwater flow. *Geophysical Research Letters*, *46*, 7447–7455. https://doi.org/10.1029/2019GL082291

Guo, J., Li, H.-Y., Leung, L. R., Guo, S., Liu, P., & Sivapalan, M. (2014). Links between flood frequency and annual water balance behaviors: A basis for similarity and regionalization. *Water Resources Research*, *50*(2), 937–953. https://doi.org/10.1002/2013WR014374

Guse, B., Hofherr, T. H., & Merz, B. (2010). Introducing empirical and probabilistic regional envelope curves into a mixed bounded distribution function. *Hydrology and Earth System Sciences*, *14*(12), 2465–2478. https://doi.org/10.5194/hess-14-2465-2010

Guse, B., Merz, B., Wietzke, L., Ullrich, S., Viglione, A., & Vorogushyn, S. (2020). The role of flood wave superposition in the severity of large floods. *Hydrology and Earth System Sciences*, *24*(4), 1633–1648. https://doi.org/10.5194/hess-24-1633-2020

Hall, J., & Anderson, M. (2002). Handling uncertainty in extreme or unrepeatable hydrological processes - The need for an alternative paradigm. *Hydrological Processes*, *16*, 1867–1870.

Halliwell, L. J. (2013). (Vol. 2, pp. 1–27). Casualty Actuarial Society E-Forum. *Classifying the tails of loss distributions*.

Harman, C. J., Sivapalan, M., & Kumar, P. (2009). Power law catchment-scale recessions arising from heterogeneous linear small-scale dynamics. *Water Resources Research*, *45*(9), W09404. https://doi.org/10.1029/2008WR007392

Hirschboeck, K. K. (1988). *Flood hydroclimatology*. Wiley.

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.

Huber, P. J. (2004). *Robust statistics* (Vol. 523). John Wiley & Sons.

Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, *25*(8–12), 1287–1304. https://doi.org/10.1016/S0309-1708(02)00056-8

Kay, A. L., Booth, N., Lamb, R., Raven, E., Schaller, N., & Sparrow, S. (2018). Flood event attribution and damage estimation using national-scale grid-based modelling: Winter 2013/2014 in great Britain. *International Journal of Climatology*, *38*(14), 5205–5219.

Kirchner, J. W. (2009). Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward. *Water Resources Research*, *45*, W02429. https://doi.org/10.1029/2008WR006912

Kortschak, D., & Albrecher, H. (2009). Asymptotic results for the sum of dependent non-identically distributed random variables. *Methodology and Computing in Applied Probability*, *11*(3), 279–306. https://doi.org/10.1007/s11009-007-9053-3

Koutsoyiannis, D. (2004). Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation/Statistiques de valeurs extrêmes et estimation de précipitations extrêmes: I. Recherche théorique. *Hydrological Sciences Journal*, *49*(4), 590. https://doi.org/10.1623/hysj.49.4.575.54430

Koutsoyiannis, D., & Papalexiou, S. M. (2017). In V. P. Singh (Ed.), *Extreme rainfall: Global perspective, Handbook of applied hydrology* (2nd ed., pp. 741–7416). McGraw-Hill.

Kusumastuti, D. I., Struthers, I., Sivapalan, M., & Reynolds, D. A. (2007). Threshold effects in catchment storm response and the occurrence and magnitude of flood events: Implications for flood frequency. *Hydrology and Earth System Sciences*, *11*(4), 1515–1528. https://doi.org/10.5194/hess-11-1515-2007

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

Leadbetter, M. R., & Rootzen, H. (1988). Extremal theory for stochastic processes. *Annals of Probability*, *16*(2), 431–478.

Lehmann, E. L., & Shaffer, J. P. (1988). Inverted distributions. *The American Statistician*, *42*(3), 191–194. https://doi.org/10.1080/00031305.1988.10475560
Lun, D., Fischer, S., Viglione, A., & Blöschl, G. (2020). Detecting flood-rich and flood-poor periods in annual peak discharges across Europe. *Water Resources Research*, *56*, e2019WR026575. https://doi.org/10.1029/2019WR026575
Maheshwari, B. L., Walker, K. F., & McMahon, T. A. (1995). Effects of regulation on the flow regime of the river Murray, Australia. *Regulated Rivers: Research & Management*, *10*, 15–38. https://doi.org/10.1002/rrr.3450100103
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., Borga, M., & Morin, E. (2020). A unified framework for extreme subdaily precipitation frequency analyses based on ordinary events. *Geophysical Research Letters*, *47*. https://doi.org/10.1029/2020GL090209
Marra, F., Nikolopoulos, E., Anagnostou, E., & 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
Marra, F., Zoccatelli, D., Armon, M., & Morin, E. (2019). A simplified MEV formulation to model extremes emerging from multiple nonstationary underlying processes. *Advances in Water Resources*, *127*, 280–290. https://doi.org/10.1016/j.advwatres.2019.04.002
Martins, E. S., & Stedinger, J. R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, *36*(3), 737–744. https://doi.org/10.1029/1999WR900330
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)
Mei, X., Van Gelder, P. H. A. J. M., Dai, Z., & Tang, Z. (2017). Impact of dams on flood occurrence of selected rivers in the United States. *Frontiers of Earth Science*, *11*, 268–282. https://doi.org/10.1007/s11707-016-0592-1
Meigh, J. R., Farquharson, F. A. K., & Sutcliffe, J. V. (1997). A worldwide comparison of regional flood estimation methods and climate. *Hydrological Sciences Journal*, *42*(2), 225–244. https://doi.org/10.1080/02626669709492022
Merz, B., Vorogushyn, S., Lall, U., Viglione, A., & Blöschl, G. (2015). Charting unknown waters – On the role of surprise in flood risk assessment and management. *Water Resources Research*, *51*(8), 6399–6416. https://doi.org/10.1002/2015WR017464
Merz, R., & Blöschl, G. (2003). A process typology of regional floods. *Water Resources Research*, *39*(12), 1340. https://doi.org/10.1029/2002WR001952
Merz, R., & Blöschl, G. (2008). Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information. *Water Resources Research*, *44*(8), W08432. https://doi.org/10.1029/2007wr006744
Merz, R., & Blöschl, G. (2009). Process controls on the statistical flood moments - A data based analysis. *Hydrological Processes*, *23*(5), 675–696. https://doi.org/10.1002/hyp.7168
Merz, R., Tarasova, L., & Basso, S. (2020). The flood cooking book: Ingredients and regional flavors of floods across Germany. *Environmental Research Letters*, *15*, 114024. https://doi.org/10.1088/1748-9326/abb9dd
Metzger, A., Marra, F., Smith, J. A., & Morin, E. (2020). Flood frequency estimation and uncertainty in arid/semi-arid regions. *Journal of Hydrology*, *590*, 125254. https://doi.org/10.1016/j.jhydrol.2020.125254
Mikosch, T. (1999). *Regular variation, subexponentiality and their applications in probability theory*. (Vol. 99013). Eurandom.
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). https://doi.org/10.1029/2019WR026535
Miniussi, A., Marani, M., & Villarini, G. (2020). Metastatistical extreme value distribution applied to floods across the continental United States. *Advances in Water Resources*, *136*, 103498. https://doi.org/10.1016/j.advwatres.2019.103498
Miniussi, A., Villarini, G., & Marani, M. (2020). Analyses through the metastatistical extreme value distribution identify contributions of tropical cyclones to rainfall extremes in the eastern United States. *Geophysical Research Letters*, *47*(7). https://doi.org/10.1029/2020GL087238
Mirza, M. M. Q. (2002). Global warming and changes in the probability of occurrence of floods in Bangladesh and implications. *Global Environmental Change*, *12*(2), 127–138. https://doi.org/10.1016/S0959-3780(02)00002-X
Molnar, P., Anderson, R. S., Kier, G., & Rose, J. (2006). Relationships among probability distributions of stream discharges in floods, climate, bed load transport, and river incision. *Journal of Geophysical Research*, *111*(2). https://doi.org/10.1029/2005JF000310
Morrison, J. E., & Smith, J. A. (2002). Stochastic modeling of flood peaks using the generalized extreme value distribution. *Water Resources Research*, *38*(12), 411–4112. https://doi.org/10.1029/2001WR000502
Murphey, J. B., Wallace, D. E., & Lane, L. J. (1977). Geomorphic parameters predict hydrograph characteristics in the southwest. *Journal of the American Water Resources Association*, *13*, 25–37. https://doi.org/10.1111/j.1752-1688.1977.tb01987.x
Mutzner, R., Bertuzzo, E., Tarolli, P., Weijs, S. V., Nicotina, L., Ceola, S., et al. (2013). Geomorphic signatures on Brutsaert base flow recession analysis. *Water Resources Research*, *49*(9), 5462–5472. https://doi.org/10.1002/wrcr.20417
Naghettini, M., Gontijo, N. T., & Portela, M. M. (2012). Investigation on the properties of the relationship between rare and extreme rainfall and flood volumes, under some distributional restrictions. *Stochastic Environmental Research and Risk Assessment*, *26*(6), 859–872. https://doi.org/10.1007/s00477-011-0530-4
Nair, J., Wierman, A., & Zwart, B. (2017). The fundamentals of heavy tails: Properties, emergence, & estimation. Forthcoming book. Preprint. available at https://adamwierman.com/book/
Nakamura, J., Lall, U., Kushnir, Y., Robertson, A. W., & Seager, R. (2013). Dynamical structure of extreme floods in the U.S. Midwest and the United Kingdom. *Journal of Hydrometeorology*, *14*, 485–504. https://doi.org/10.1175/JHM-D-12-059.1
Neiman, P. J., Schick, L. J., Ralph, F. M., Hughes, M., & Wick, G. A. (2011). Flooding in Western Washington: The connection to atmospheric rivers. *Journal of Hydrometeorology*, *12*(6), 1337–1358.
Northrop, P. J. (2004). Likelihood-based approaches to flood frequency estimation. *Journal of Hydrology*, *292*(1), 96–113. https://doi.org/10.1016/j.jhydrol.2003.12.031
Overeem, A., Buishand, T. A., Holleman, I., & Uijlenhoet, R. (2010). Extreme value modeling of areal rainfall from weather radar. *Water Resources Research*, *46*(9). https://doi.org/10.1029/2009WR008517
Pallard, B., Castellarin, A., & Montanari, A. (2009). A look at the links between drainage density and flood statistics. *Hydrology and Earth System Sciences*, *13*, 1019–1029. https://doi.org/10.5194/hess-13-1019-2009
Papalexiou, S. M., AghaKouchak, A., & Foufoula-Georgiou, E. (2018). A diagnostic framework for understanding climatology of tails of hourly precipitation extremes in the United States. *Water Resources Research*, *54*(9), 6725–6738. https://doi.org/10.1029/2018WR022732
Papalexiou, S. M., & Koutsoyiannis, D. (2013). Battle of extreme value distributions: A global survey on extreme daily rainfall. *Water Resources Research*, *49*(1), 187–201. https://doi.org/10.1029/2012WR012557

Papalexiou, S. M., Koutsoyiannis, D., & Makropoulos, C. (2013). How extreme is extreme? An assessment of daily rainfall distribution tails. *Hydrology and Earth System Sciences*, *17*(2), 851–862. https://doi.org/10.5194/hess-17-851-2013

Paté-Cornell, E. (2012). On "Black Swans" and "perfect storms": Risk analysis and management when statistics are not Enough. *Risk Analysis*, *32*(11), 1823–1833.

Pattison, I., Lane, S. N., Hardy, R. J., & Reaney, S. M. (2014). The role of tributary relative timing and sequencing in controlling large floods. *Water Resources Research*, *50*(7), 5444–5458. https://doi.org/10.1002/2013WR014067

Petrow, T., Merz, B., Lindenschmidt, K.-E., & Thieken, A. H. (2007). Aspects of seasonality and flood generating circulation patterns in a mountainous catchment in south-eastern Germany. *Hydrology and Earth System Sciences*, *11*(4), 1455–1468. https://doi.org/10.5194/hessd-4-589-2007

Pickands, J., III. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, *3*(1), 119–131. https://doi.org/10.1214/aos/1176343003

Porporato, A., Vico, G., & Fay, P. A. (2006). Superstatistics of hydro-climatic fluctuations and interannual ecosystem productivity. *Geophysical Research Letters*, *33*, L15402. https://doi.org/10.1029/2006GL026412

Ranjbar, V., Amini, M., Geluk, J., & Bozorgnia, A. (2013). Asymptotic behavior of product of two heavy-tailed dependent random variables. *Acta Mathematica Sinica-English Ser.*, *29*(2), 355–364. https://doi.org/10.1007/s10114-012-0680-4

Rinaldo, A., Botter, G., Bertuzzo, E., Uccelli, A., Settin, T., & Marani, M. (2006). Transport at basin scales: 1. Theoretical framework. *Hydrology and Earth System Sciences*, *10*(1), 19–29. https://doi.org/10.5194/hess-10-19-2006

Rogger, M., Kohl, B., Pirkl, H., Viglione, A., Komma, J., Kirnbauer, R., et al. (2012). Runoff models and flood frequency statistics for design flood estimation in Austria - do they tell a consistent story? *Journal of Hydrology*, *456*(457), 30–43. https://doi.org/10.1016/j.jhydrol.2012.05.068

Rogger, M., Pirkl, H., Viglione, A., Komma, J., Kohl, B., Kirnbauer, R., et al. (2012). Step changes in the flood frequency curve: Process controls. *Water Resources Research*, *48*(5), 1–15. https://doi.org/10.1029/2011WR011187

Rogger, M., Viglione, A., Derx, J., & Blöschl, G. (2013). Quantifying effects of catchments storage thresholds on step changes in the flood frequency curve. *Water Resources Research*, *49*(10), 6946–6958. https://doi.org/10.1002/wrcr.20553

Rojas-Nandayapa, L., & Xie, W. (2018). Asymptotic tail behaviour of phase-type scale mixture distributions. *Annals of Actuarial Science*, *12*(2), 412–432. https://doi.org/10.1017/S1748499517000136

Rosbjerg, D., Blöschl, G., Burn, D. H., Castellarin, A., Croke, B., Di Baldassarre, G., et al. (2013). Prediction of floods in ungauged basins. In G. Blöschl, M. Sivapalan, T. Wagener, A. Viglione, & H. Savenije (Eds.), *Runoff prediction in Ungauged basins* (pp. 189–226). Cambridge University Press.

Rössler, O., Froidevaux, P., Börst, U., Rickli, R., Martius, O., & Weingartner, R. (2014). Retrospective analysis of a nonforecasted rain-on-snow flood in the Alps – A matter of model limitations or unpredictable nature? *Hydrology and Earth System Sciences*, *18*(6), 2265–2285.

Rupp, D. E., & Selker, J. S. (2006). On the use of the Boussinesq equation for interpreting recession hydrographs from sloping aquifers. *Water Resources Research*, *42*(12). https://doi.org/10.1029/2006WR005080

Salas, J. D., Gavilán, G., Salas, F. R., Julien, P. Y., & Abdullah, J. (2014). Uncertainty of the PMP and PMF. In S. Eslamian (Ed.), *Handbook of Engineering hydrology* (pp. 575–603). Taylor & Francis Group.

Salinas, J. L., Castellarin, A., Kohnová, S., & Kjeldsen, T. R. (2014). Regional parent flood frequency distributions in Europe – Part 2: Climate and scale controls. *Hydrology and Earth System Sciences*, *18*, 4391–4401. https://doi.org/10.5194/hess-18-4391-2014

Schellander, H., Lieb, A., & Hell, T. (2019). Error structure of metastatistical and generalized extreme value distributions for modeling extreme rainfall in Austria. *Earth and Space Science*, *6*(9), 1616–1632. https://doi.org/10.1029/2019EA000557

Seo, Y., & Schmidt, A. R. (2013). Network configuration and hydrograph sensitivity to storm kinematics. *Water Resources Research*, *49*(4), 1812–1827. https://doi.org/10.1002/wrcr.20115

Sivapalan, M., Blöschl, G., Merz, R., & Gutknecht, D. (2005). Linking flood frequency to long-term water balance: Incorporating effects of seasonality. *Water Resources Research*, *41*, W06012. https://doi.org/10.1029/2004WR003439

Skaugen, T., Creutin, J.-D., & Gottschalk, L. (1996). Reconstruction and frequency estimates of extreme daily areal precipitation. *Journal of Geophysical Research*, *101*(D21), 26287–26295. https://doi.org/10.1029/96JD01384

Skublics, D., Seibert, S. P., & Ehret, U. (2014). Abbildung der Hochwasserretention durch hydrologische und hydrodynamis-che Modelle unter unterschiedlichen Randbedingungen –Sensitivitätsanalyse am Donauabschnitt zwischen Neu-Ulmund Donauwürth. *Hydrol. Wasserbewirts.*, *58*, 178–189. https://doi.org/10.5675/HyWa_2014,3_2

Smith, J. A., Baeck, M. L., Yang, L., & Goodrich, D. C. (2019). The paroxysmal precipitation of the desert: Flash floods in the southwestern United States. *Water Resources Research*, *55*(12), 10218–10247.

Smith, J. A., Cox, A. A., Baeck, M. L., Yang, L., & Bates, P. (2018). Strange floods: The upper tail of flood peaks in the United States. *Water Resources Research*, *54*(9), 6510–6542. https://doi.org/10.1029/2018WR022539

Stein, L., Pianosi, F., & Woods, R. (2020). Event-based classification for global study of river flood generating processes. *Hydrological Processes*, *34*(7), 1514–1529.

St George, S., & Mudelsee, M. (2019). The weight of the flood-of-record in flood frequency analysis. *Journal of Flood Risk Management*, *12*, e12512. https://doi.org/10.1111/jfr3.12512

Strupczewski, W. G., Kochanek, K., Bogdanowicz, E., & Markiewicz, I. (2012). On seasonal approach to flood frequency modelling. Part I: Two-component distribution revisited. *Hydrological Processes*, *26*, 705–716. https://doi.org/10.1002/hyp.8179

Struthers, I., & Sivapalan, M. (2007). A conceptual investigation of process controls upon flood frequency: Role of thresholds. *Hydrology and Earth System Sciences*, *11*(4), 1405–1416. https://doi.org/10.5194/hess-11-1405-2007

Su, C., & Chen, Y. (2006). On the behavior of the product of independent random variables. *Science in China, Series A*, *49*(3), 342–359. https://doi.org/10.1007/s11425-006-0342-z

Su, Y., & Smith, J. A. (2021). An atmospheric water balance perspective on extreme rainfall potential for the contiguous US. *Water Resources Research*, *57*(4), e2020WR028387.

Taleb, N. N. (2007). *The Black swan: The impact of the highly Improbable* (p. 444). Random House.

Tarasova, L., Basso, S., & Merz, R. (2020). Transformation of generation processes from small runoff events to large floods. *Geophysical Research Letters*, *47*, e2020GL090547. https://doi.org/10.1029/2020GL090547

Tarasova, L., Merz, R., Kiss, A., Basso, S., Blöschl, G., Merz, B., et al. (2019). Causative classification of river flood events. *Water*, *6*(4), e1353. https://doi.org/10.1002/wat2.1353

Thorarinsdottir, T. L., Hellton, K. H., Steinbakk, G. H., Schlichting, L., & Engeland, K. (2018). Bayesian regional flood frequency analysis for large catchments. *Water Resources Research*, *54*(9), 6929–6947. https://doi.org/10.1029/2017wr022460

Toffoli, A., Ducrozet, G., Waseda, T., Onorato, M., Abdolahpour, M., & Nelli, F. (2019). *Ocean currents trigger rogue waves*. Eng. Conf. Paper presented at 29th Intern. Ocean and Polar https://onepetro.org/ISOPEIOPEC/proceedings-abstract/ISOPE19/All-ISOPE19/ISOPE-I-19-238/21164

Viglione, A., Merz, B., Dung, N. V., Parajka, J., Nester, T., & Blöschl, G. (2016). Attribution of regional flood changes based on scaling fingerprints. *Water Resources Research*, *52*(7), 322–5340.

Viglione, A., Merz, R., & Blöschl, G. (2009). On the role of the runoff coefficient in the mapping of rainfall to flood return periods. *Hydrology and Earth System Sciences*, *13*(5), 577–593. https://doi.org/10.5194/hess-13-577-2009

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

Volpi, E., Di Lazzaro, M., Bertola, M., Viglione, A., & Fiori, A. (2018). Reservoir effects on flood peak discharge at the catchment scale. *Water Resources Research*, *54*, 9623–9636. https://doi.org/10.1029/2018WR023866

Vorogushyn, S., Bates, P. D., de Bruijn, K., Castellarin, A., Kreibich, H., Priest, S., et al. (2018). Evolutionary leap in large-scale flood risk assessment needed. *Water*, *5*(2), e1266. https://doi.org/10.1002/wat2.1266

Vorogushyn, S., Lindenschmidt, K.-E., Kreibich, H., Apel, H., & Merz, B. (2012). Analysis of a detention basin impact on dike failure probabilities and flood risk for a channel-dike-floodplain system along the river Elbe, Germany. *Journal of Hydrology*, *436–437*, 120–131. https://doi.org/10.1016/j.jhydrol.2012.03.006

Vorogushyn, S., & Merz, B. (2013). Flood trends along the Rhine: The role of River training. *Hydrology and Earth System Sciences*, *17*, 3871–3884. https://doi.org/10.5194/hess-17-3871-2013

Vorogushyn, S., Merz, B., & Apel, H. (2009). Development of dike fragility curves for piping and micro-instability breach mechanisms. *Natural Hazards and Earth System Science*, *9*(4), 1383–1401. https://doi.org/10.5194/nhess-9-1383-2009

Wang, C., Chang, N.-B., & Yeh, G.-T. (2009). Copula-based flood frequency (COFF) analysis at the confluences of river systems. *Hydrological Processes*, *23*, 1471–1486. https://doi.org/10.1002/hyp.7273

Wang, W., Li, H. Y., Leung, L. R., Yigzaw, W., Zhao, J., Lu, H., et al. (2017). Nonlinear filtering effects of reservoirs on flood frequency curves at the regional scale. *Water Resources Research*, *53*, 8277–8292. https://doi.org/10.1002/2017WR020871

Wasko, C., & Nathan, R. (2019). Influence of changes in rainfall and soil moisture on trends in flooding. *Journal of Hydrology*, *575*, 432–441. https://doi.org/10.1016/j.jhydrol.2019.05.054

Wasko, C., Shao, Y., Vogel, E., Wilson, L., Wang, Q. J., Frost, A., & Donnelly, C. (2021). Understanding trends in hydrologic extremes across Australia. *Journal of Hydrology*, *593*, 125877. https://doi.org/10.1016/j.jhydrol.2020.125877

Waylen, P., & Woo, M.-K. (1982). Prediction of annual floods generated by mixed processes. *Water Resources Research*, *18*(4), 1283–1286. https://doi.org/10.1029/WR018i004p01283

Wendi, D., Merz, B., & Marwan, N. (2019). Assessing hydrograph similarity and rare runoff dynamics by cross Recurrence plots. *Water Resources Research*, *55*(6), 4704–4726. https://doi.org/10.1029/2018WR024111

Wietzke, L., 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). https://doi.org/10.1029/2005GL022465

Woo, G. (2019). Downward counterfactual search for extreme events. *Frontiers of Earth Science*, *7*(340). https://doi.org/10.3389/feart.2019.00340

Yang, L., Smith, J., Baeck, M. L., Morin, E., & Goodrich, D. C. (2017). Flash flooding in arid/semiarid regions: Dissecting the hydrometeorology and hydrology of the 19 August 2014 storm and flood hydroclimatology in Arizona. *Journal of Hydrometeorology*, *18*(12), 3103–3123.

Yang, L., Yang, Y., & Smith, J. (2021). The upper tail of flood peaks over China: Hydrology, hydrometeorology, and hydroclimatology. *Water Resources Research*, *57*(11), e2021WR030883.

Yang, L., Yang, Y., Villarini, G., Li, X., Hu, H., Wang, L., et al. (2021). Climate more important for Chinese flood changes than reservoirs and land use. *Geophysical Research Letters*, *48*(11), e2021GL093061.

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

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

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