AGU ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

RESEARCH ARTICLE
10.1029/2021WR031260

# Event and Catchment Controls of Heavy Tail Behavior of Floods

**Elena Macdonald[1], Bruno Merz[1,2], Björn Guse[1], Luzie Wietzke[1], Sophie Ullrich[1], Matthias Kemter[1,2,3], Bodo Ahrens[4], and Sergiy Vorogushyn[1]**

[1]Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences, Potsdam, Germany, [2]Institute for Environmental Sciences and Geography, University of Potsdam, Potsdam, Germany, [3]Potsdam Institute for Climate Impact Research, Potsdam, Germany, [4]Institute for Atmospheric and Environmental Sciences, Goethe University Frankfurt, Frankfurt am Main, Germany

**Key Points:**

- Differences between large and small floods are described by a novel slope indicator and are decisive for the emergence of heavy tails
- Heavy tail behavior is mainly controlled by the catchment response and event precipitation, and by flood seasonality and catchment area
- Pre-event moisture state within a catchment has no noticeable impact on tail heaviness

**Supporting Information:**

Supporting Information may be found in the online version of this article.

**Correspondence to:**

E. Macdonald,
elena.macdonald@gfz-potsdam.de

**Citation:**

Macdonald, E., Merz, B., Guse, B., Wietzke, L., Ullrich, S., Kemter, M., et al. (2022). Event and catchment controls of heavy tail behavior of floods. *Water Resources Research*, *58*, e2021WR031260. https://doi.org/10.1029/2021WR031260

Received 21 SEP 2021
Accepted 27 MAY 2022



**Abstract** In some catchments, the distribution of annual maximum streamflow shows heavy tail behavior, meaning the occurrence probability of extreme events is higher than if the upper tail decayed exponentially. Neglecting heavy tail behavior can lead to an underestimation of the likelihood of extreme floods and the associated risk. Partly contradictory results regarding the controls of heavy tail behavior exist in the literature and the knowledge is still very dispersed and limited. To better understand the drivers, we analyze the upper tail behavior and its controls for 480 catchments in Germany and Austria over a period of more than 50 years. The catchments span from quickly reacting mountain catchments to large lowland catchments, allowing for general conclusions. We compile a wide range of event and catchment characteristics and investigate their association with an indicator of the tail heaviness of flood distributions, namely the shape parameter of the GEV distribution. Following univariate analyses of these characteristics, along with an evaluation of different aggregations of event characteristics, multiple linear regression models, as well as random forests, are constructed. A novel slope indicator, which represents the relation between the return period of flood peaks and event characteristics, captures the controls of heavy tails best. Variables describing the catchment response are found to dominate the heavy tail behavior, followed by event precipitation, flood seasonality, and catchment size. The pre-event moisture state in a catchment has no relevant impact on the tail heaviness even though it does influence flood magnitudes.

**Plain Language Summary** For each river catchment, we can estimate how likely it is that floods of certain magnitudes occur in that river. This is called a probability distribution. In some rivers, the occurrence of extreme floods is more likely than in others – their probability distribution decays slower and has a so-called heavy tail. Here, we examine which factors lead to higher probabilities of extreme floods in some rivers compared to others. To this aim, we look at the annual maximum river flows of 480 rivers in Germany and Austria over a period of more than 50 years. As potential factors influencing the likelihood of extreme floods, we analyze characteristics describing the river catchments in general and characteristics describing specific flood events. Using modeling approaches, we find that how a catchment responds to heavy rainfall has the strongest effect on the probability of extremes. A catchment is more likely to experience extreme flooding, if the largest observed floods are characterized by (a) high runoff coefficients, meaning a large share of rainfall becomes direct streamflow and (b) short event time scales, meaning the floods are short but intense. The rainfall itself and the season in which floods usually occur also have an influence.

## 1. Introduction

Hydro-meteorological time series like annual maximum precipitation or streamflow may exhibit heavy-tailed distributions (Bernardara et al., 2008; Farquharson et al., 1992; Smith et al., 2018; Villarini et al., 2011). A statistical distribution is termed heavy-tailed if its tail decays slower than that of an exponential distribution, leading to a higher occurrence probability of extreme events (El Adlouni et al., 2008; Papalexiou et al., 2013). Several classes of heavy tail distributions are distinguished which characterize the degree of tail heaviness (El Adlouni et al., 2008; Wietzke et al., 2020). The Generalized Pareto (GP) and Generalized Extreme Value (GEV) distributions with positive shape parameters are two widely used heavy tail distributions for modeling precipitation and streamflow series, respectively.

The reliable estimation of heavy tail behavior of floods is crucial, for example, for the robust design of flood protection measures or insurance appraisal. However, estimating the upper tail behavior from observational time series of typical length, for instance, 30–50 years, is associated with large uncertainty (Papalexiou & Koutsoyiannis, 2013). Understanding the controls of heavy tails is thus essential for sound estimation of extreme floods and their exceedance probabilities. A better understanding and characterization of the propensity of a given catchment to generate heavy tail flood behavior would help to avoid surprises, and harmful consequences and reduce risks (Merz et al., 2015).

In various science domains, different indices have been proposed to characterize tail heaviness (Wietzke et al., 2020). In hydro-meteorological studies, heavy tail behavior is most often characterized by the shape parameter ($\xi$) of the GEV distribution (Morrison & Smith, 2002). Other frequently used indices are the shape parameter of the GP distribution (Coles, 2001), the skewness (McCuen & Smith, 2008), and the Upper Tail Ratio (UTR) (Lu et al., 2017; Smith et al., 2018). UTR is defined as the ratio between the flood of record and the 10-year return period flood (Villarini & Smith, 2010). While GEV and GP distributions with shape parameters larger than zero are heavy-tailed (El Adlouni et al., 2008), skewness and UTR are not directly related or linked to the formal definition of tail heaviness in relation to the asymptotic behavior of the exponential distribution (Smith et al., 2018; Wietzke et al., 2020).

Distributions of annual maximum series (AMS) of streamflow are shaped by the complex interplay of meteorological and hydrological processes. A number of studies have analyzed factors that control the tail behavior of flood distributions, though many studies do not explicitly use heavy tail indicators, describe the relative propensity to heavy tails, for example, by analyzing skewness or the coefficient of variation of flood series.

Villarini and Smith (2010) analyzed the GEV shape parameter of flood distributions at 572 gauges in the eastern US, where more than half of the gauges exhibited positive values. They found a dominant control of extreme rainfalls caused by landfalling hurricanes on flood tail heaviness. Particularly in the Appalachian Mountains, very high $\xi$ of 0.3 and above were detected for the majority of basins. Furthermore, a weak decrease of $\xi$, and thus tail heaviness, with catchment area increase was observed (Villarini & Smith, 2010). Smith et al. (2018) came to a similar conclusion by studying more than 8000 US catchments: extreme floods in many catchments characterized by a very high UTR are generated by tropical cyclones and heavy summer thunderstorms. The effect of flood-causing precipitation on flood tails was also observed by Thorarinsdottir et al. (2018) for 203 basins in Norway. Catchments with high rain, but low snowmelt contribution to flood flows showed mainly positive $\xi$ in the regional GEV model, whereas catchments with snowmelt-dominated floods often exhibited upper-bounded distributions ($\xi < 0$). A minor negative influence of the catchment area and the ratio of catchment area to catchment length on the GEV shape parameter was detected. Light flood tails in snow-dominated catchments were also observed in the Mediterranean and Middle East regions (Bernardara et al., 2008; Farquharson et al., 1992). Gaume (2006) suggested that the heavy tail behavior of floods can be inherited from the heavy tail behavior of precipitation. Specifically, he argued that for very high return periods, 500 years and above, the tail of annual maximum precipitation distributions for durations in the range of the catchment concentration time controls the asymptotic behavior of the annual maximum flow distribution. For lower return periods of precipitation, catchment controls seem to exert a strong influence on flood distributions. Gottschalk and Weingartner (1998) found a much stronger variability of flood distributions for 17 catchments in Switzerland compared to the variability of rainfall volumes scaled by the event duration. They attributed it to the variability of runoff coefficients. Similarly, McCuen and Smith (2008) analyzed AMS for 28 catchments in the US and suggested that catchment and river channel storage dominated flood skewness. While the rainfall skewness was fairly constant, flood skewness showed a pronounced variability. These studies suggest that the impact of the tail behavior of rainfall distributions on the upper tail of flood distributions can be strongly modulated by the flood generation processes within catchments.

Nonlinear catchment response to rainfall has been proposed as another cause of the heavy tail behavior of floods. The exceedance of the storage capacity over a significant share of the catchment area was shown to affect the skewness of flood distributions in 10 catchments of Southern Italy (Gioia et al., 2008). When subsurface storage thresholds are exceeded, much higher flood runoff is generated. This can lead to an inflection point or step-change in the flood frequency curve thus increasing the propensity for heavy tails, as shown by Rogger et al. (2012) for two small basins in Austria based on modeling experiments and detailed field surveys. Similarly, Basso et al. (2015) suggested that the nonlinearity of the storage-discharge relationship expressed by a high

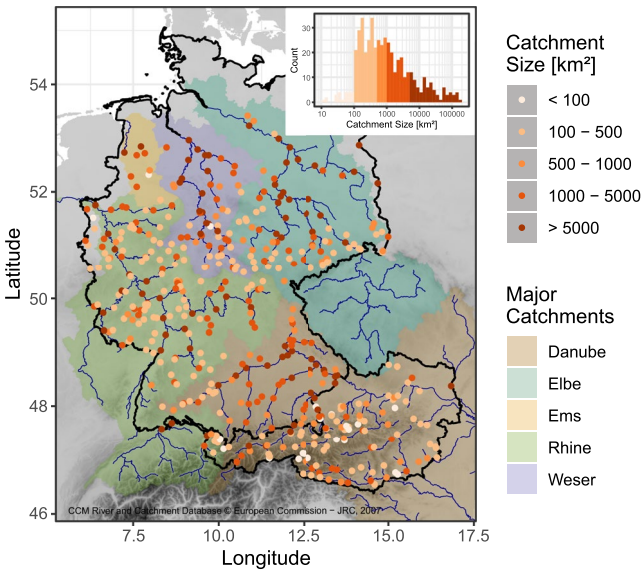


**Figure 1.** Locations of the 480 stream gauges used for analysis. The gauges are colored according to the catchment size. Catchment size distribution is shown in the inset histogram. Depicted river networks are from Vogt et al. (2007).

recession exponent is responsible for the occurrence of high daily streamflow, and Basso et al. (2016) demonstrated that this nonlinearity contributes to the heavy tail behavior of seasonal flood maxima. The nonlinearity of catchment response can be characterized by the variability of runoff coefficients across flood events. Merz and Blöschl (2009) found that a high skewness of flood flow distributions was related to strong variations in runoff coefficients in several Austrian basins. This high variation was particularly observed in dry catchments, whereas in wet basins, the runoff coefficients tended to be always high. In a similar vein, Guo et al. (2014) found higher coefficients of variation for flood flows in arid basins compared to wetter ones in the United States. The control of catchment aridity on runoff coefficients was also observed by Norbiato et al. (2009) for 14 Alpine basins. The mean runoff coefficient showed a strong positive correlation with mean annual precipitation.

An approach to characterize the complex interplay between meteorological and hydrological processes during flood generation is flood typology (Merz & Blöschl, 2003; Tarasova et al., 2019). The analysis of flood types can thus be viewed as another perspective on the integrated effect of individual control mechanisms on the heavy tail behavior of floods. Merz and Blöschl (2003) demonstrated a fundamentally different form of flood frequency curves, convex versus concave, for two catchments dominated by different flood types: long-rain and short-rain versus rain-on-snow, respectively. Fischer and Schumann (2020) identified different hydrograph shapes characterized by peak and volume for different flood types in catchments in Central Germany. The tail of mixed flood peak distributions was typically dominated by floods with low volumes but high peaks produced by intense thunderstorms.

The reviewed literature demonstrates the multitude of potential controls on flood tail behavior employing data-based, analytical, and modeling approaches. However, none of the previous studies realized a comprehensive multivariate analysis of a large number of factors to identify the controls of heavy flood tails and quantify their relative importance. To date, mainly univariate analyses or studies with a very limited number of factors were carried out. The analysis by Thorarinsdottir et al. (2018) for Norway can be considered an exception, though the catchment attributes used as flood frequency model covariates included primarily physiographic quantities, not directly reflecting flood generation mechanisms. Furthermore, only a few studies covered large hydro-climatic and physiographic regions with several hundred catchments (Merz & Blöschl, 2009; Smith et al., 2018; Thorarinsdottir et al., 2018; Villarini & Smith, 2010) to allow generalized conclusions on heavy tail controls. Finally, many studies did not specifically target flood tail indicators but considered the entire flood frequency curve (e.g., Rogger et al., 2012), flood skewness (McCuen & Smith, 2008; Merz & Blöschl, 2009), or the coefficient of variation of flood flows (Guo et al., 2014).

In this paper, we present a multivariate analysis of an extensive set of event and catchment characteristics with the aim of exploring the causes of heavy tails of flood distributions. Event characteristics include triggering precipitation event characteristics, antecedent catchment state as well as catchment response characteristics. Our data exploration approach first applies univariate analyses to selected potential heavy tail predictors, followed by linear and nonlinear multivariate modeling using multiple linear regression and random forests, respectively. We analyze a comprehensive dataset of event and catchment characteristics for 480 German and Austrian basins that cover a large range of flood generating processes – from quick reacting mountains to large lowland catchments. This allows us to draw general conclusions on the heavy tail behavior of floods.

## 2. Data

### 2.1. Streamflow Time Series

We use observed mean daily streamflow at gauging stations across Germany and Austria (Figure 1). Data are provided by the Hydrometric Services of the German Federal States and the Austrian Hydrometric Service. Out

of 639 gauges (492 in Germany and 147 in Austria), gauges with less than 50 years of observations, starting from the hydrological year 1951 (beginning on 1 November 1950), are excluded owing to large uncertainty in estimating the upper tail behavior. Furthermore, several gauges, for which peculiar estimates led to in-depth screening and a strong anthropogenic influence was identified (e.g., reservoir operation), are excluded. This results in overall 480 time series. Among these gauges, 100 gauges have small data gaps, with a mean of 1.7% missing data of the total series length of a respective gauge. We fill these gaps by the regression to the highest correlated daily streamflow series without gaps in the respective period. We extract annual maximum series (AMS) for hydrological years, defined as the period from November to the following October, to be subsequently used for the heavy tail analysis.

### 2.2. Catchments

For all Austrian gauges, the contributing catchments are provided by the central hydrographic office of Austria HZB. The remaining catchments in Germany and adjacent countries are delineated based on the 25 m resolution EU-DEM v1.1. The resulting catchment areas range over five orders of magnitude between 11 and 159,378 $km^2$ (Figure 1).

### 2.3. Hydrometeorological Time Series

Daily gridded precipitation and potential evapotranspiration from the E-OBS dataset (version 12) with 0.25° resolution (Haylock et al., 2008) are used to derive catchment-averaged series of precipitation and potential evapotranspiration. Catchment-averaged daily soil moisture series are computed by the conceptual, raster-based hydrological model mHM (Samaniego et al., 2010), set up and calibrated for all German and Austrian basins. We use relative soil moisture between wilting point and field capacity for the entire soil column parameterized in mHM. For catchment-averaged daily series of snowmelt and for the identification of flood generation processes, reanalysis data 4DAS (Primo et al., 2019), which is downscaled from the ERA20 C reanalysis to a spatial resolution of 0.11°, is used. The variables precipitation, soil moisture, soil pore space, snowmelt, convective available potential energy (CAPE), and convective inhibition (CIN) are aggregated to daily totals, as used by Kemter et al. (2020).

## 3. Methods

To analyze the heavy tail behavior of flood flows and identify their controls, we select the GEV shape parameter as a heavy tail indicator and compile a suitable set of predictor variables. The GEV shape parameter is most frequently used in hydro-meteorological studies to characterize heavy tail behavior (e.g., Morrison & Smith, 2002; Thorarinsdottir et al., 2018; Villarini & Smith, 2010). It assumes that the GEV is a suitable distribution of the annual maxima. This assumption is widely accepted as the asymptotic distribution of independent block maxima values converges to the GEV (Fisher & Tippett, 1928). It can, however, not be proven. Furthermore, Wietzke et al. (2020) showed that, compared to other upper tail indicators, the GEV shape parameter is well suited for analyzing different sites with variable record lengths and statistical moments, as done here. We explore the relationships between the heavy tail indicator and the predictor variables using univariate and multivariate models. The selected predictors describing flood events and catchment characteristics are summarized in Table 1 and explained in the subsequent sections. Figure 2 gives a schematic overview of all methodological steps involved as part of the analysis.

From here on, the following terminology is used: We refer to groups of variables describing coherent aspects as *event characteristics* (e.g., event precipitation) and *catchment characteristics* (e.g., catchment wetness). Each characteristic is described by one or more *variables* (e.g., precipitation volume, event duration, maximum precipitation intensity for event precipitation; and mean annual precipitation and aridity index for catchment wetness). The variables describing event characteristics are further aggregated to catchment-specific *indicators* (see Section 3.2.1 for details), while this is not necessary for variables describing catchment characteristics. Once the indicators and variables are analyzed against the GEV shape parameter of flood flows, we refer to them as *(potential) predictors*.

**Table 1**
*Potential Predictors for Heavy Tail Behavior of Flood Flows*

| Characteristic | Variable | Indicators[a] | Unit | Data source | Description | Potential predictors |
|---|---|---|---|---|---|---|
| | | | Event characteristics | | | |
| Event precipitation | Precipitation volume (*Pvol*) | Yes | mm | E-OBS | Catchment average event precipitation sum associated with AMS flow event | *Pvol_CV*, *Pvol_shape*, *Pvol_rho*, *Pvol_UTD*, *Pvol_slope*[b] |
| | Event duration (*Pdur*) | yes | days | E-OBS | Duration of precipitation event | *Pdur_CV*, *Pdur_shape*, *Pdur_rho*, *Pdur_UTD*, *Pdur_slope* |
| | Maximum precipitation intensity (*Pmax*) | yes | mm/day | E-OBS | Maximum daily precipitation intensity during precipitation event | *Pmax_CV*, *Pmax_shape*, *Pmax_rho*, *Pmax_UTD*, *Pmax_slope* |
| Antecedent catchment state | Flow at event start (*Qbegin*) | yes | $m^3/s$ | Gauge observation | Observed streamflow at the first day of flood event corresponding to AMS peak | *Qbegin_CV*, *Qbegin_shape*, *Qbegin_rho*, *Qbegin_UTD*, *Qbegin_slope* |
| | Soil moisture at event start (*SM*) | yes | - | mHM simulation | Catchment average relative soil moisture in relation to wilting point and field capacity on the first day of flood event | *SM_CV*, *SM_shape*, *SM_rho*, *SM_UTD*, *SM_slope* |
| | Precipitation before event start (*P10 d*ays) | yes | mm | E-OBS | Catchment average precipitation sum of 10 days prior to flood event | *P10 d_CV*, *P10 d_shape*, *P10 d_rho*, *P10 d_UTD*, *P10 d_slope* |
| Event catchment response | Runoff coefficient (*RC*) | yes | - | Gauge, E-OBS, 4DAS reanalysis | Runoff coefficient for flood event determined as a relation between direct runoff and event liquid precipitation plus snowmelt | *RC_CV*, *RC_shape*, *RC_rho*, *RC_UTD*, *RC_slope* |
| | Event time scale (*ETS*) | yes | day | Gauge observation | Ratio between direct runoff volume and direct peak flow of flood event | *ETS_CV*, *ETS_shape*, *ETS_rho*, *ETS_UTD*, *ETS_slope* |
| Event timing | Flood seasonality (*FS*) | No | - | Gauge observation | Mean date of flood peak occurrence | *FS_x*, *FS_y* |
| | Event unseasonality (*EUnS*) | yes | - | Gauge observation | Difference between flood peak occurrence date and mean date of flood peak occurrence | *EUnS_CV*, *EUnS_shape*, *EUnS_rho*, *EUnS_UTD*, *EUnS_slope* |
| Event types | Event types of top 5 (*Type5*) | No | - | 4DAS reanalysis | Types of the five biggest flood events compared to the types of all other AMS flood events | *Type5_pvalue* |
| | Event type shares (*Type_share*) | no | - | 4DAS reanalysis | Ratio of the frequency of one event type to the total number of AMS flood events | *Snow_share*, *Rain_share*, *SM_share* |
| | | | Catchment characteristics | | | |
| Catchment area | Catchment area (*Size*) | No | $km^2$ | EU-DEM, HZB (2009) | Area of catchments contributing to selected gauges | *Size* |
| Catchment wetness | Mean annual precipitation (*MAP*) | No | mm | E-OBS | Catchment average precipitation sum for delineated catchments | *MAP* |
| | Aridity index (*AI*) | No | - | E-OBS | Index based on the ratio of monthly precipitation and monthly potential evapotranspiration | *AI* |

**Table 1**
*Continued*

| Characteristic | Variable | Indicators[a] | Unit | Data source | Description | Potential predictors |
|---|---|---|---|---|---|---|
| Tail heaviness of rainfall in the flood season | Shape parameter of the maximum precipitation in the flood season (*MP_shape*) | No | - | E-OBS | Maximum d-day precipitation in a timeframe around the mean date of flood occurrence, with d being the catchment response time | *MP_shape* |
| Nonlinearity of catchment response | Flashiness index (*FI*) | No | - | Gauge observation | Ratio of absolute day-to-day fluctuations of streamflow relative to total flow in a year | *FI* |
| | *Q10/Q50* | No | - | Gauge observation | Ratio of daily 10th percentile streamflow to the median daily flow | *Q10/Q50* |
| Synchronicity of precipitation and catchment state | Phase correlation between P and SM (*P_SM_cor*) | No | - | E-OBS, mHM simulation | Spearman rank correlation of catchment average precipitation and soil moisture per month | *Pmean_SM_cor*, *Phigh_SM_cor* |
| | Phase correlation between P and Q (*P_Q_cor*) | No | - | Gauge observation, E-OBS | Spearman rank correlation of catchment average precipitation and streamflow per month | *Pmean_Q_cor*, *Phigh_Q_cor* |

[a]This column specifies for which variables the aggregated indicators as described in Section 3.2.1 are estimated ("yes") and for which this not the case ("no"). [b]The acronyms *CV*, *shape*, *rho*, *UTD*, and *slope* refer to the following indicators, which were derived for event characteristics: coefficient of variation, GEV shape parameter, Spearman rank correlation coefficient, upper tail dependence coefficient, and slope against flood return periods. Details are given in Section 3.2.1.

### 3.1. AMS Flood Events

To analyze the link between flood event characteristics and heavy tail behavior, we identify flood event hydrographs corresponding to the AMS peaks using an automated procedure applied by Guse et al. (2020) and further modified here to be applicable to a wide range of catchments with areas varying across five orders of magnitude. Each AMS flood hydrograph is characterized by a peak, a start, and an end point. The start and end points are located between an AMS peak and earlier and later independent peaks. First, the independent peaks are identified by backward and forward search starting at the AMS peak, and fulfill the following criteria: (a) the lowest streamflow between two independent peaks is below 70% of the smaller peak and below 20% of the AMS peak, (b) the smaller peak is greater than 20% of the AMS peak, and (c) the time lag (DT) between two peaks is at least several days. These criteria were empirically derived by Bacchi et al. (1992) and LAWA (2018) to avoid the identification of small peaks as independent flood events. Larger DT can be expected for larger catchments with longer response times and flatter recession curves. In extension to Guse et al. (2020), we relate DT to the catchment size and use DT = 7 days for catchments larger than 10.000 km$^2$, DT = 5 days for catchments between 10.000 and 1.000 km$^2$, and DT = 3 days for catchments with an area below 1.000 km$^2$.

To identify the start and end point of the hydrographs, we apply a gradient-based method proposed by Klein (2009). Flow gradients between two consecutive days are computed and threshold values at the 90th, 91st, and 92nd percentiles are empirically identified for the three catchment area classes. Gradients in larger catchments show higher variability than in smaller ones. If gradients below the respective threshold value are detected for 5, 4, or 3 consecutive days for the three catchment area classes, this corresponds to the flattening of the hydrograph, and a start or end point is taken as the last or first point within this time window. In case no start or end point is detected by the gradient search within a time window of 20, 15, or 10 days for the three area classes, the day with the lowest flow within this time window is taken. All thresholds are determined empirically by trial-and-error followed by the visual inspection of hydrograph separation and plausibility check for several gauges in the three area classes.

### 3.2. Event Characteristics

All the event characteristics as well as the catchment characteristics that are compiled as potential heavy tail predictors are selected based on an extensive literature review (see Merz et al., 2022, and references therein)

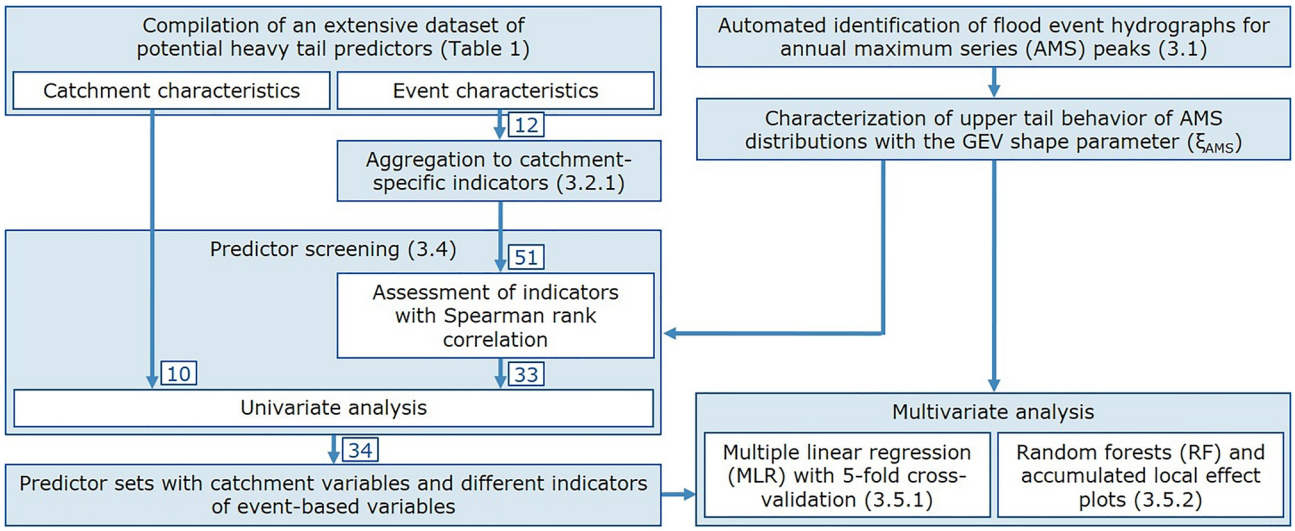


**Figure 2.** Schematic overview of the workflow. Numbers in parenthesis refer to the sections with detailed descriptions of the respective steps. Numbers in boxes next to arrows indicate how many variables/potential predictors are passed to the next step.

combined with expert opinion. In the initial set of potential predictors, we include characteristics that have extensively been discussed before as relevant for heavy tails of flood peaks (e.g., precipitation, catchment response), but also characteristics that have only been mentioned in this context and which an influence has not yet been tested (e.g., seasonality). For some of the characteristics, the associated literature is discussed in detail in the introduction, while for the remaining the reasons for including them are presented subsequently in the respective paragraphs. We do not claim that the set of potential predictors analyzed in this study is exhaustive. For example, predictors on the spatial variability of some variables (see Tarasova et al., 2020) could be added. We do believe though that the most relevant processes are covered with our set of characteristics.

### 3.2.1. Event Precipitation

For each identified flood event, the event precipitation is determined by a backward search starting from the streamflow peak date and including all consecutive wet days till the last wet day with precipitation exceeding 0.1 mm/day. For each event precipitation, *total volume (Pvol)*, *duration (Pdur)*, and *maximum intensity (Pmax)* are computed as event precipitation variables.

### 3.2.2. Antecedent Catchment State

We use three variables to describe the catchment state prior to the onset of the flood event. *Streamflow at the beginning* (*Qbegin*) of a flood event is widely applied to describe the catchment wetness state (e.g., Ettrick et al., 1987; Merz et al., 2018). We also use *catchment average soil moisture (SM)* at the beginning of a flood event, standardized with the monthly average and standard deviation of *SM* (Zink et al., 2017). Merz and Blöschl (2009) identified a strong control of pre-event soil moisture on runoff coefficients in Austrian catchments. A similar result is found by Norbiato et al. (2009) for a few alpine catchments. The third variable is antecedent precipitation, defined here as the *rainfall amount of 10 days* (*P10 d*) preceding a flood event. Antecedent precipitation has also been widely used as a proxy for catchment wetness (Schröter et al., 2015; Tarasova et al., 2018).

### 3.2.3. Event Catchment Response

The response of catchments to precipitation is described using two variables representing runoff generation and runoff concentration, respectively. The *runoff coefficient* (*RC*) is estimated as the ratio of direct runoff during a flood event and the sum of rainfall and snowmelt associated with the event. The concentration of runoff in a catchment is estimated using the *event time scale* (*ETS*). *ETS* is the ratio of direct runoff volume during a flood event and direct peak discharge (Gaál et al., 2012). A short *ETS* indicates a slim but pointy hydrograph and fast runoff concentration, while a long *ETS* corresponds to a wider hydrograph indicating a slow runoff concentration (Tarasova et al., 2018).

#### 3.2.4. Event Timing

Evidence of an association between the timing of flood events and their magnitude in temperate regions are contradictory. While Smith et al. (2018) and Petrow et al. (2007) observed the occurrence of large floods in distinct seasons in the US and Germany, respectively, Merz and Blöschl (2009) found no dependency of flood moments on the event timing for Austrian catchments. The effect of seasonality on tail heaviness has not been analyzed so far to our best knowledge. Therefore, we assess *flood seasonality* (*FS*) as a potential predictor for the upper tail indicator based on the average date of flood occurrence. Following the approach by Burn (1997), we use directional statistics to describe occurrence dates of flood peaks in polar coordinates and estimate the mean date (MD) per catchment. MD differs substantially between the first of January (MD = $2\pi/365$) and 31st of December (MD = $2\pi$) even though the seasonality is almost identical, and therefore we decompose MD into an x-component (cos(MD); *FS_x*) and a y-component (sin(MD); *FS_y*). Furthermore, we introduce the deviation from MD as another variable of event timing. Smith et al. (2018) found for the US that "record floods" differed strongly in their seasonality from the rest of the AMS flood peaks. For each event, the absolute difference, bounded above at $\pi$, between the date of flood occurrence and MD is estimated and termed *event unseasonality* (*EUnS*).

#### 3.2.5. Event Types

For all flood events, an event type based on the dominant flood generation processes is derived. We consider five flood event types based on the triggering factors: convective precipitation, stratiform rainfall, soil moisture excess, snowmelt, and rain-on-snow. The classification is based on a simple decision tree developed by Kemter et al. (2020). An event is classified as generated by snowmelt if snowmelt is greater than rainfall, and as rain-on-snow, if rainfall is greater, but at least two-thirds of the catchment are covered with snow. If this is not the case and convective conditions prevail in at least 25% of the catchment, the event is classified as being generated by convective rainfall. All flood events not meeting the above criteria are assigned to be caused either by soil moisture excess if the mean soil water content exceeded 70% before the event, or by stratiform rainfall if this was not the case. We use Fisher's exact test (Fisher, 1934) to quantify how the occurrence of flood types differs between the five largest events and the remaining AMS floods. The null hypothesis is that the relative proportion of the flood types is the same in the two groups. The smaller the p-value (*Type5_pvalue*), the higher the probability that the five largest events are caused by otherwise rare processes. Furthermore, we group the five event types into snow-dominated (snowmelt and rain-on-snow), rain-dominated (convective precipitation and stratiform rainfall), and soil moisture excess, and estimate the share of event types per catchment as indicators (*Snow_share*, *Rain_share*, *SM_share*).

#### 3.2.6. Indicators of Event Characteristics

The variables describing the characteristics of flood events are first derived separately for each AMS event. In the next step, an integral value per gauge, termed *indicator*, is computed for each variable. For all event characteristics except flood seasonality and event types, five indicators are calculated and assessed: *CV*, *shape*, *rho*, *UTD*, and *slope* (details below). These are based on three different hypotheses regarding the association between a variable and the tail heaviness of flood distributions: (a) a larger variability of the variable favors heavier flood tails; (b) a heavier tail of the variable favors heavier flood tails; (c) a close association between (the upper tail of) the variable and the flood magnitude favors heavier flood tails. The derivation of indicators for flood seasonality and event types is explained in the respective paragraphs above.

The coefficient of variation (*CV*) is used to represent the first hypothesis. The second hypothesis is addressed by fitting a GEV distribution to the variables and adopting its shape parameter (*shape*) as an indicator. For the third hypothesis, three indicators are assessed, namely the Spearman rank correlation coefficient (*rho*) between the variable and AMS, the upper tail dependence coefficient (*UTD*) between the variable and AMS, and a *slope* value describing the relation of the variable with the return period of respective flood peaks. While the correlation coefficient is based on the entire time series, the *UTD* coefficient considers only the correlation in the upper tails (here: top decile) and can be understood as the probability of one margin exceeding a threshold under the condition that the other margin exceeds a threshold (Frahm et al., 2005). Here, the nonparametric estimator of Schmid and Schmidt (2007) is used (R package *copula*; Hofert et al. [2020]). The *slope* indicator is a novel coefficient proposed here, which takes the entire time series into account, but is dominated by the values associated with high flood return periods. This indicator is inspired by the results of Merz and Blöschl (2008), who reported

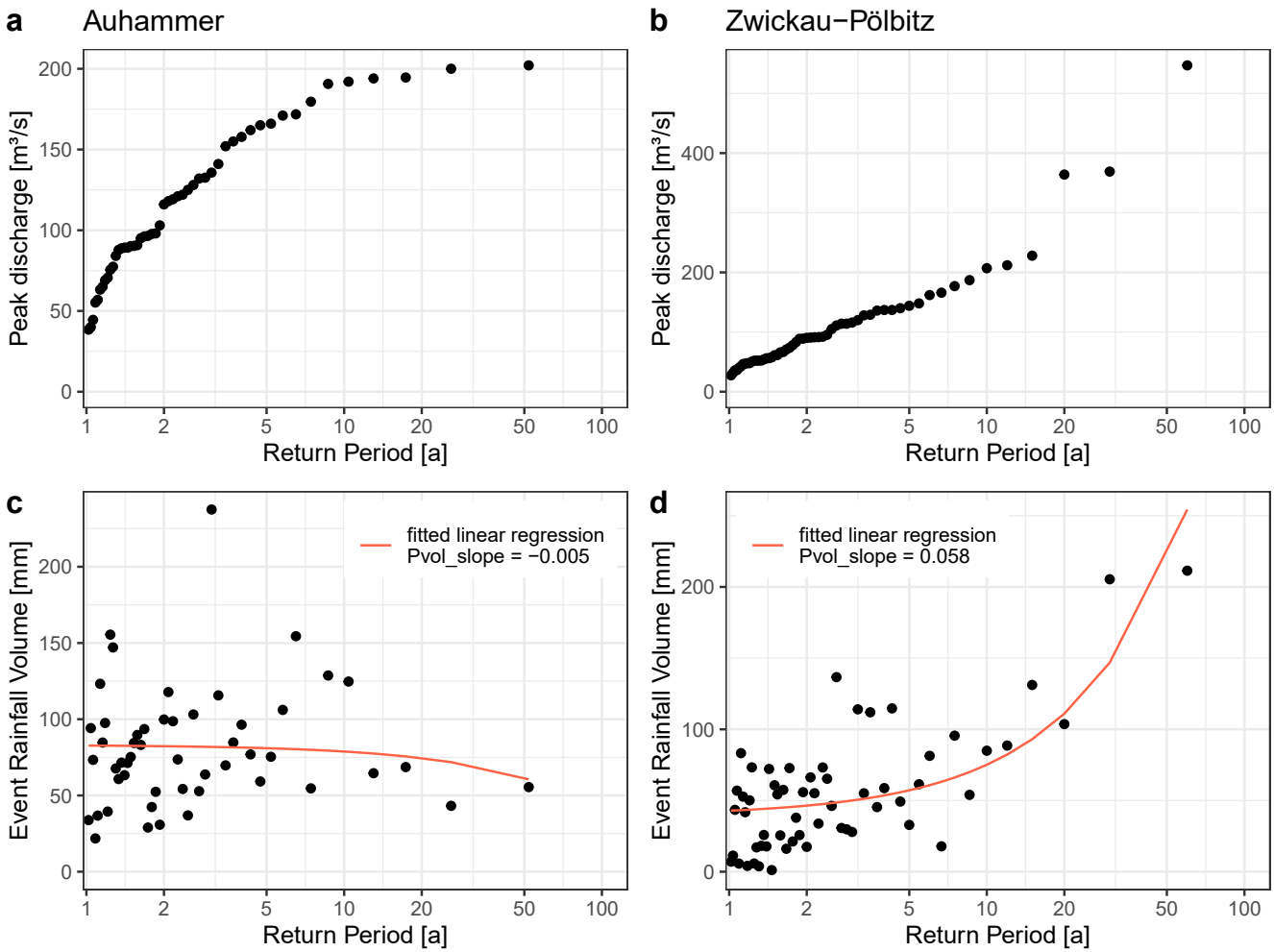


**Figure 3.** Two examples of the association of a variable, here event rainfall volume, with the return period of the associated flood events. In the Auhammer catchment (a), (c), the rainfall volumes associated with the largest flood events are slightly below average, resulting in a slope value close to 0; in the Zwickau-Pölbitz catchment (b), (d), the highest rainfall volumes are associated with the largest flood events, resulting in a slope value of 0.058. Note the logarithmic scale on the *x*-axes.

different associations of event runoff coefficients with the return periods of respective flood peaks. The *slope* describes whether the largest flood events in a catchment are also the events with the highest (or lowest) values of a variable (Figure 3). The variable is plotted against the return periods of the respective flood event peaks and a linear model is fitted by the least-square procedure. The slope of the linear model, normalized with the mean of the variable, is taken as the indicator.

### 3.3. Catchment Characteristics

#### 3.3.1. Catchment Area

Spatial aggregation can average out the effect of nonlinear behavior in flood generating processes taking place at small scales. This can lead to lighter flood tails in larger catchments. A weak decrease in tail heaviness with *catchment area* (*Size*) was found in some studies (e.g., Villarini & Smith, 2010), while others found no dependency (e.g., Smith et al., 2018).

#### 3.3.2. Catchment Wetness

For each basin, we derive *mean annual precipitation* (*MAP*) based on the regionalized E-OBS precipitation dataset. Further, we compute *average aridity index* (*AI*) following Knoben et al. (2018). *AI* is calculated based on the

Thornthwaite's Moisture Index ($MI_{month}$) according to mean monthly observations of precipitation ($P_{month}$) and potential evapotranspiration ($PET_{month}$):

$$\mathrm{MI}_{\mathrm{month}} = \begin{cases} 1 - \dfrac{PET_{month}}{P_{month}}, & P_{month} > PET_{month} \\ 0, & P_{month} = PET_{month} \\ \dfrac{P_{month}}{PET_{month}} - 1, & P_{month} < PET_{month} \end{cases}$$

The average aridity is then calculated as one-twelfth of the sum of monthly Thornthwaite's indices and can range from −1 (most arid conditions) to 1 (most humid conditions) (Knoben et al., 2018; Melsen & Guse, 2019).

### 3.3.3. Tail Heaviness of Rainfall in the Flood Season

Heavy tails of extreme precipitation can, under certain conditions, propagate to heavy tails of annual maximum floods (Gaume, 2006). Besides event precipitation, the tail heaviness of catchment maximum precipitation in the flood season (MP) is considered a potential predictor. Extreme precipitation events are regarded as relevant for the generation of flood peaks if they persisted over a timespan corresponding to the catchment response time d (Ganguli & Merz, 2019) and occurred within a time window around the mean date of flood occurrence MD. The width of the window is estimated according to Bertola et al. (2020) and depends on the concentration of occurrence dates around MD. We use the *GEV shape parameter of all MP* (*MP_shape*) in a catchment as a potential predictor.

### 3.3.4. Nonlinearity of Catchment Response

The overall, not event-specific response of a catchment to rainfall is an important description of its functioning. Catchments in which rapid changes in the daily streamflow are frequent might be prone to fast increases of streamflow after heavy precipitation which can then lead to extreme flood events. To quantify the rapidity and frequency of such streamflow changes, we use the Richards-Baker *flashiness index* (*FI*) (Baker et al., 2004). For each year $y$, the ratio of absolute day-to-day fluctuations of streamflow relative to the total flow in that year is estimated:

$$\mathrm{FI}_y = \frac{\sum_{i=1}^{n} |\mathrm{q}_i - \mathrm{q}_{i-1}|}{\sum_{i=1}^{n} \mathrm{q}_i}$$

with $q$ being the mean daily streamflow, $i$ indicating the day, and $n = 365$ (366). The annual values $FI_y$ are averaged to the mean flashiness index $FI$ which ranges between 0 (constant flow) and 2 (high flashiness). Another factor influencing the catchment response to rainfall is groundwater storage. As a variable describing the groundwater storage, the ratio of the 10th percentile of daily streamflow to the median daily flow, *Q10/Q50*, is taken (Norbiato et al., 2009). A high ratio of low to median flows corresponds to high subsurface water storage which can buffer heavy precipitation and cause lighter flood tails.

### 3.3.5. Synchronicity of Precipitation and Catchment State

If the seasonality of rainfall intensities and the seasonality of catchment wetness are in phase, this resonance can lead to higher flood magnitudes and can impact the shape of flood frequency curves (Sivapalan et al., 2005). To quantify whether rainfall and wetness seasonality are synchronized or out of phase, we estimate Spearman's rank correlation coefficient between monthly measures of catchment average precipitation and monthly measures of catchment wetness. The measures of precipitation are the monthly average of daily precipitation *Pmean*, and the number of times per month that the *d*-daily precipitation exceeds the 95th percentile *Phigh*. The measures of catchment wetness are the monthly average of daily soil moisture and the monthly average of daily streamflow. In total, four correlation coefficients are estimated (*Pmean_SM_cor*, *Phigh_SM_cor*, *Pmean_Q_cor*, *Phigh_Q_cor*).

### 3.4. Screening of Predictors

Including all variables and indicators, a total of 43 potential predictors are generated. Prior to applying multivariate models, a reduced expedient set of indicators and variables needs to be selected. We use different methods to distill the most promising predictors for explaining the variability of the upper tail indicator across our set of catchments.

First, the suitability of the five indicators used for event-based variables (i.e., *CV*, *shape*, *rho*, *UTD*, *slope*) is assessed using Spearman rank correlation against the shape parameter of AMS floods ($\xi_{AMS}$). The significance threshold for the correlation coefficients is set to 0.05, and the indicators looking most promising across all variables are selected for further analyses. In a second step, all potential predictors, both event and catchment characteristics, are subject to univariate analyses against $\xi_{AMS}$. We use again Spearman rank correlation along with a visual screening of scatterplots and locally estimated scatterplot smoothing (LOESS) curves. The aim is in particular to reduce the number of variables of which two or more describe similar aspects (e.g., *MAP* and *AI* for catchment wetness), and not to achieve a drastic dimensionality reduction of the variable space.

### 3.5. Multivariate Analysis

Based on the preselected sets of potential predictors, we apply linear and nonlinear multivariate models, namely multiple linear regression (MLR) and random forests (RF), respectively. MLR has the advantage of a simple estimation procedure along with easily interpretable results (Molnar, 2021). Its main disadvantage is that MLR models can only include linear relationships between predictors and the predictand. In contrast, RF works well if nonlinear dependencies or interactions between predictors are present, but are less efficient in modeling linear relations (Molnar, 2021). Furthermore, RF has high predictive power, is robust to the inclusion of noisy or highly correlated predictors, can handle big data efficiently, and are fast in comparison to other machine learning methods (Tyralis et al., 2019). Originally, RF are difficult to interpret, but the RF variant used here counteracts this limitation and has been described as especially useful for interpretation and causal inference (Hothorn et al., 2006; Tyralis et al., 2019). Both MLR and RF have frequently been used and also compared to one another in hydrological studies. Conclusions on the superiority of one model over the other depend highly on the context and the (non-)linear dependencies present in the analyzed data. For example, Booker and Snelder (2012) found a poor performance of MLR compared to RF in the context of flow duration curves, and linked this to high-order interactions and complex non-linear relations. Ibarra-Berastegi et al. (2011) on the other hand, concluded that using either model resulted in equal performance when predicting precipitation and surface moisture fluxes. Therefore, we apply both multivariate models. In the models, we use predictor sets with different indicators of the event variables (e.g., *shape* and *slope*) and sets with just one indicator per variable (e.g., only *slope*).

#### 3.5.1. Linear Multivariate Analysis

We consider all possible combinations of 3 to 10 predictors for constructing multiple linear regression models. We limit the range of combinations to reduce the computational time, and because we assume the best model to lie within this range and that fewer or more predictors would result in a too simplified or too complex model, respectively. To avoid overfitting during the predictor selection (Lever et al., 2016), we use 5-fold cross-validation. The data is split randomly into five subsets and each subset once serves as test data to evaluate the performance of an MLR model trained on all remaining data (Hastie et al., 2009). The best model within each cross-validation fold is selected based on the lowest Bayesian Information Criteria (BIC). To analyze the degree of predictor multicollinearity, we compute the variance inflation factor (VIF) for each of the best predictor combinations per subset (Hirsch et al., 1992). For each predictor combination, VIF is related to the coefficient of determination ($R^2$) for an ordinary least square regression fitted to all predictors except one taken as a predictand. Predictor combinations with VIF >10 show a high degree of multicollinearity and are excluded from the analysis (Kutner et al., 2004; Montgomery et al., 2001). From the resulting five predictor sets, the one with the highest $R^2$ based on the test data is selected. An MLR model with these predictors is fitted to the entire dataset, using again five-fold cross-validation to get a more conservative estimation of the model performance. The above-described procedure is run for each of the sets of potential predictors (single/several indicators per variable) and the resulting models are evaluated based on the explained variance. Finally, the relative importance of the selected predictors is estimated (R package *relaimpo*; Grömping [2006]) and their interaction is analyzed for the best performing models.

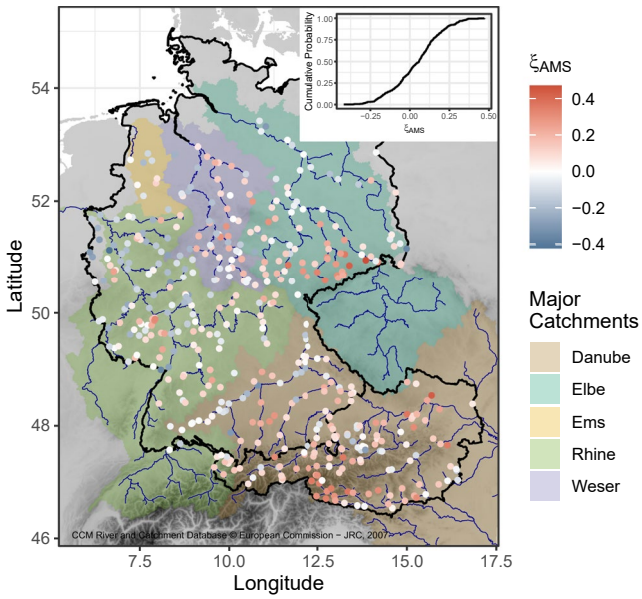


**Figure 4.** Spatial distribution of the generalized extreme value shape parameter of annual maximum series flood flows ($\xi_{AMS}$). Gauge locations are plotted and colored based on $\xi_{AMS}$. Depicted river networks are from Vogt et al. (2007).

#### 3.5.2. Nonlinear Multivariate Analysis

Random forests are generated for each of the sets of potential predictors (single/several indicators per variable). In a random forest, an ensemble of regression trees is combined, each of which is built based on recursive partitioning of the predictor space into regions with similar values of the response variable (Breimann, 2001). Each tree is constructed using a bootstrap sample from the entire dataset, and the remaining data are called out-of-bag (OOB) data (Breimann, 2001). We first select predictors from the predictor set using recursive feature elimination (RFE) with mean squared error (MSE) as an accuracy measure and based on 100 trees (R package *moreparty*; Robette [2020]). With the selected predictors, a random forest is generated using 500 conditional inference trees as base learners (R package *party*; Hothorn et al. [2021]). In contrast to regular regression trees, conditional inference trees are not biased toward variables with many potential split points (Hothorn et al., 2006). The performance of the random forests is evaluated based on the OOB predictions which result in more conservative accuracy measures (Strobl et al., 2009), similar to the cross-validation for MLR models. For the best-performing model, the influence of the selected predictors on the prediction is estimated with accumulated local effects (ALE) plots (Apley & Zhu, 2020; Robette, 2020). ALE plots have a similar purpose to partial dependence plots but are unbiased even in the presence of collinear input variables (Molnar, 2021). They estimate how the prediction of a model changes over small intervals of each predictor (local effects). The effects are then accumulated across all intervals and centered in a way that the mean effect is zero (for more details see Apley & Zhu [2020]; Molnar [2021]). The interaction between selected predictors is finally analyzed using scatterplots.

## 4. Results

### 4.1. Spatial Distribution of GEV Shape Parameters of Flood Flows

The GEV shape parameter of AMS flood flows ($\xi_{AMS}$) ranges between −0.418 and 0.471 with a median of 0.0577 for the 480 gauges. High values are found in the south and north-east of Austria, as well as in Saxony in Germany (Figure 4). Gauges in the west of Germany have mainly light-tailed distributions. In the south-west, center, and north-east of Germany slightly negative $\xi_{AMS}$ and values close to zero are estimated.

### 4.2. Correlation Analysis of Event-Based Indicators

The five indicators, that is, *CV*, *shape*, *rho*, *UTD*, and *slope*, applied to event-based variables are evaluated using Spearman rank correlation against $\xi_{AMS}$. Significant correlations are found for all indicators for at least four of the eight variables (Figure 5). No indicator clearly outperformed all others at this stage, so we keep one indicator for each of the three hypotheses stated in Section 3.2.1 for the following analyses. Out of the three indicators based on the third hypothesis, that is, a closer association between (the upper tail of) the variable and the flood magnitude favors heavier flood tails, the newly developed *slope* indicator has the highest number of significant correlations as well as the highest absolute correlation coefficients. Furthermore, the *UTD* is found to be suitable only for positive correlations between variables and the flood magnitude as it cannot capture dependence of the lower tail of a variable on the upper tail of the flood distribution. Hence, we use *CV*, *shape*, and *slope* for further analyses.

### 4.3. Univariate Analysis of Predictors

Univariate analyses are conducted for the remaining three indicators of all event characteristics as well as for all catchment characteristics – a total of 43 potential predictors. Only a selection is shown and discussed here, the remainder can be found in the supplement (Figure S1 in Supporting Information S1).

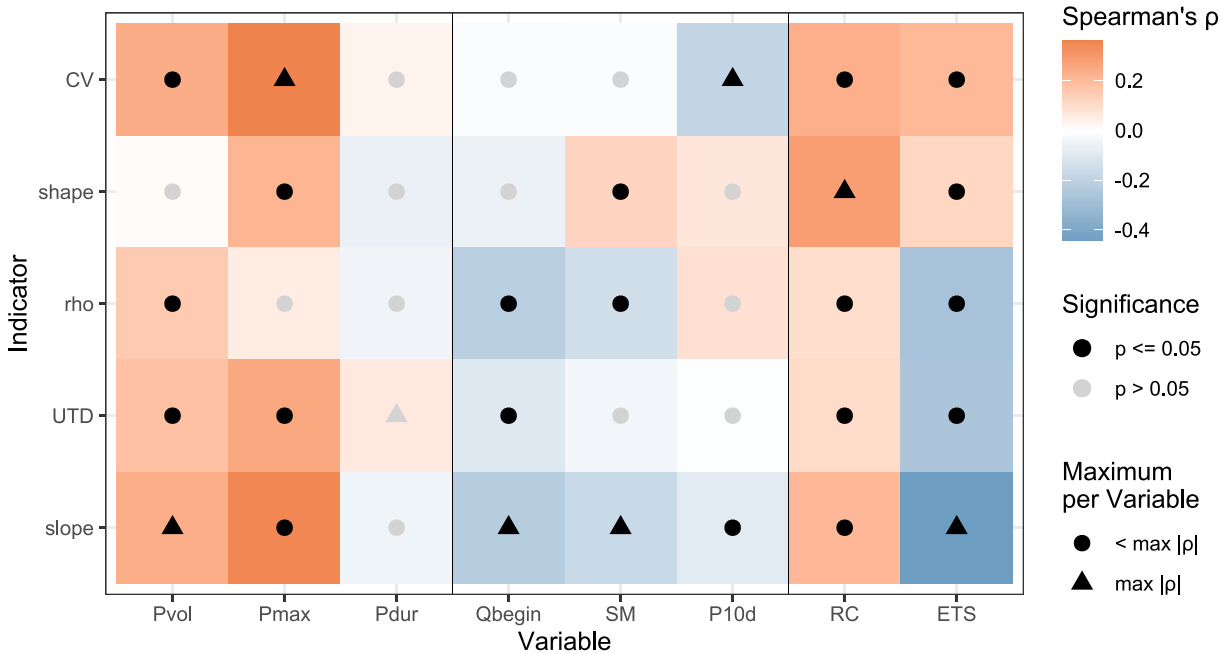


**Figure 5.** Spearman rank correlation of indicators of event-based variables against the generalized extreme value shape parameter of annual maximum series flood flows. Event-based variables are ordered along the horizontal matrix dimension and respective indicators computed for each variable are ordered along the vertical dimension. For each variable, the indicator with the highest absolute correlation coefficient is marked with a triangle. *Pvol* – precipitation volume, *Pmax* – maximum precipitation intensity, *Pdur* – precipitation duration, *Qbegin* – flow at event start, *SM* – soil moisture at event start, *P10 d* – precipitation before event start, *RC* – runoff coefficient, *ETS* – event time scale.

Indicators of the volume (*Pvol*) and maximum intensity (*Pmax*) of the event precipitation are found to be significantly correlated to $\xi_{AMS}$ while no significant correlations are found for the duration (*Pdur*) of the precipitation event (Figures 6a–6c). The slope indicators of *Pvol* and *Pmax* show positive correlations to $\xi_{AMS}$ meaning that higher slopes are associated with heavier tails of the AMS flood flows. High slopes in turn mean that, at the respective gauges, the largest floods are strongly associated with the highest volumes or intensities of precipitation, respectively. In general, mainly positive slopes are observed for those two variables, indicating that the return period of floods tends to increase with increasing precipitation volume and intensity across almost all gauges.

To reduce the number of potential predictors for the multivariate approaches, *Pdur* is excluded from further analyses at this stage. Similarly, other variables, of which two or more describe the same event or catchment characteristic and show similar behavior in the univariate analyses, are excluded, namely *SM* (antecedent catchment state), *MAP* (catchment wetness), and *P_SM_cor* (Synchronicity of precipitation and catchment state).

For the antecedent catchment state, the slope of *Qbegin* shows the strongest correlation to $\xi_{AMS}$ (Figure 6d). As indicated by the positive slope values, we find that at the majority of the gauges the return period of AMS flood flows tends to increase with a wetter catchment state. Nevertheless, the influence of the catchment state on the upper tail of flood peak distributions is less pronounced than that of other event characteristics.

The catchment response is characterized by the runoff coefficient (*RC*) and the event time scale (*ETS*). The slopes of *RC* are mainly positive and show a significant positive correlation with $\xi_{AMS}$, meaning that gauges, where the largest floods are associated with high *RC*, tend to have heavy-tailed flood distributions (Figure 6e). In contrast, the slopes of *ETS* are largely negative and are also negatively correlated with $\xi_{AMS}$. If the largest flood events in a catchment have a short *ETS* and are influenced by fast runoff generation processes, this favors heavy tail behavior.

The univariate analysis of the flood seasonality shows lighter tails for catchments dominated by winter floods and heavier tails for those dominated by summer floods (Figures 6g–6h). While a significant negative correlation is found for the x-component of the mean occurrence date ($FS\_x \approx -1 \triangleq$ Jun/Jul; $FS\_x \approx 1 \triangleq$ Dec/Jan) with $\xi_{AMS}$, this is not the case for the y-component ($FS\_y \approx -1 \triangleq$ Sep/Oct; $FS\_y \approx 1 \triangleq$ Mar/Apr).

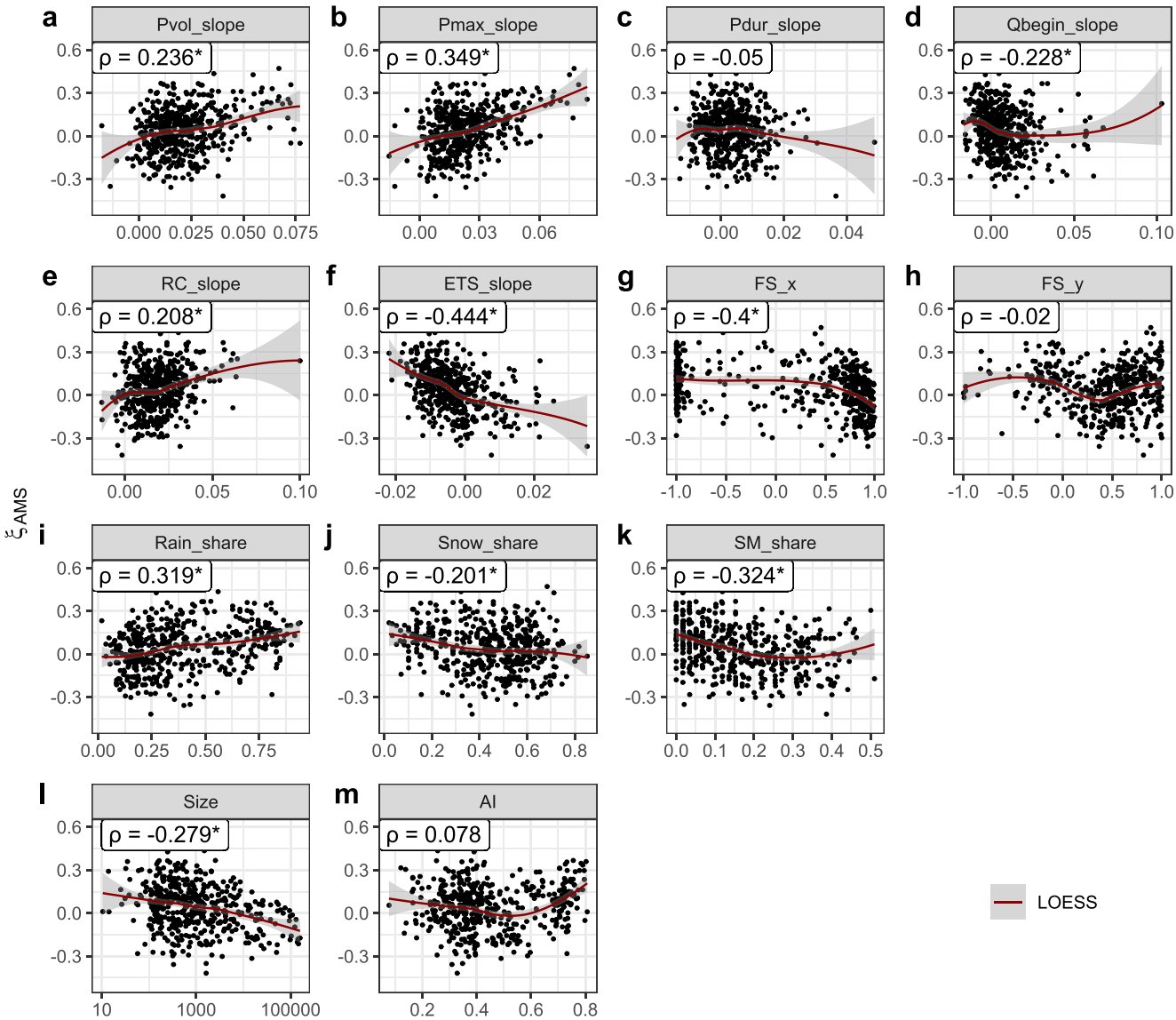


**Figure 6.** Univariate analysis of potential predictors against the generalized extreme value shape parameter of annual maximum series flood flows ($\xi_{AMS}$). Spearman rank correlations (ρ) which are significant at the 0.05 level are marked with an asterisk. Locally estimated scatterplot smoothing (LOESS) regression and the corresponding 95% confidence bound are indicated by red line and gray shading (note that the LOESS regression has large uncertainty in areas with scarce data; in these areas, the relation between the shape parameter and the predictors is very sensitive to outliers). *Pvol* – precipitation volume, *Pmax* – maximum precipitation intensity, *Pdur* – precipitation duration, *Qbegin* – flow at event start, *RC* – runoff coefficient, *ETS* – event time scale, *slope* – slope indicator of event variables, *FS* – flood seasonality (with x and y component), *Rain/Snow/SM_share* – share of the events classified as rain/snow/soil moisture induced, *Size* – catchment area, *AI* – aridity index.

We find that catchments dominated by snow-related events tend to have lighter tails than catchments with a high share of rain-induced flood events (Figures 6i–6k). This resonates well with the analysis of seasonality. The share of rain-induced AMS events in a catchment (*Rain_share*) shows a significant positive correlation with $\xi_{AMS}$, while the share of snow-related events (*Snow_share*), as well as the share of events induced by soil moisture excess (*SM_share*), show significant negative correlations with $\xi_{AMS}$.

With regards to catchment characteristics, the catchment area is found to be significantly correlated with $\xi_{AMS}$, with heavier tails for small catchments and lighter tails with increasing catchment size (Figure 6l). In contrast to that, no clear association between the average aridity index characterizing catchment wetness and $\xi_{AMS}$ is detected (Figure 6m).

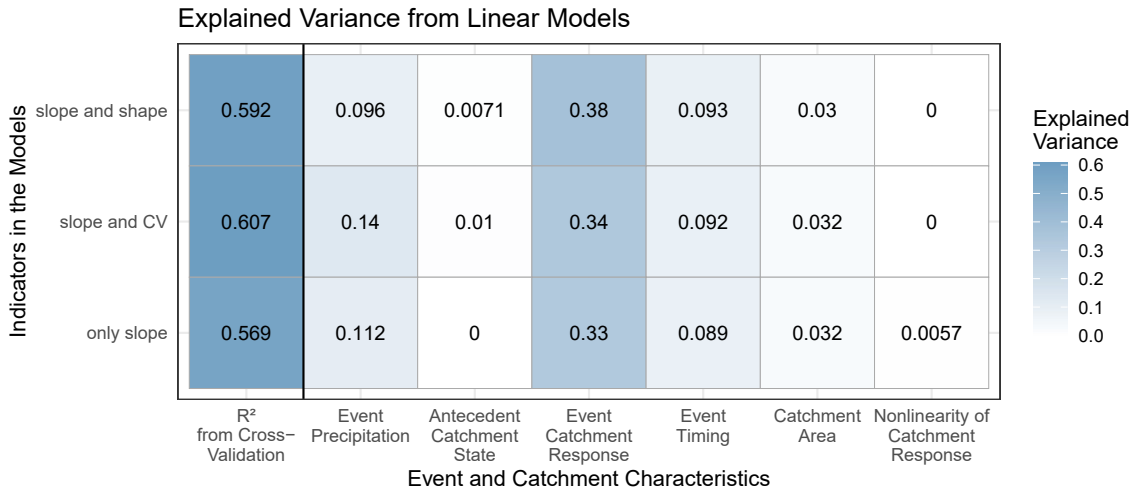


**Figure 7.** Explained variance by the event and catchment characteristics of which predictors are selected in the three best performing linear regression models. The denoted values show the total explained variance by all predictors describing the respective characteristic.

### 4.4. Multiple Linear Regression

We construct MLR models for different sets of potential predictors (single/several indicators per variable). In each set, the MLR model with the lowest BIC has between six and nine predictors. The coefficient of determination ($R^2$) based on cross validation ranges between 0.235 and 0.607 for these models. Distinctively higher values of $R^2$ are reached by the MLR models containing the novel *slope* indicator in the set of predictors compared to the ones without it ($R^2 \geq 0.569$ vs. $R^2 \leq 0.386$). Results for the latter group of models are therefore discarded.

The three models with *slope* indicators (i.e., *slope* and *shape*, *slope* and *CV*, only *slope*) differ slightly in the selected predictors but are very similar in the event and catchment characteristics which are found to be relevant. In all three models, predictors representing event catchment response, event precipitation, and event timing have the highest explained variance (Figure 7). In general, event characteristics are found to be of greater importance for the heavy tail behavior of flood flows than catchment characteristics. The only catchment characteristic, which is selected as a predictor in each of the three best performing models, is the catchment area. Event characteristics describing the antecedent catchment state and the dominant flood generation process have little or no impact on the tail heaviness of AMS flood distributions according to these models.

The predictors which have the highest relative importance in each of the three models are *ETS_slope* and *RC_slope*, which both characterize the event catchment response (Figure 8). *ETS_slope* has a negative model coefficient meaning that the predictand, that is, $\xi_{AMS}$, increases with decreasing *ETS_slope*. Catchments, where large floods have shorter *ETS* compared to smaller floods, show heavy tail behavior, as also indicated in the univariate analysis. In contrast, $\xi_{AMS}$ increases with increasing *RC_slope* in all three models. Catchments, where the runoff coefficient tends to increase with the return period of flood events, show heavier tails in their flood distribution. In terms of explained variance, the two indicators of the event catchment response are followed by a predictor representing the event precipitation. In the model with *slope* and *CV* indicators, this is *Pmax_slope*, in the two others, it is *Pvol_slope*. The model coefficient is positive in all three cases. If the largest floods in a catchment are strongly associated with very high precipitation volume or intensity, it can lead to heavy tail behavior.

The event timing is found to be of high relative importance in all three models. The x-component of the flood seasonality is selected as a predictor in all cases with a negative model coefficient. Low values of *FS_x*, that is, a mean date of flood occurrence in summer, lead to high $\xi_{AMS}$, while tails are lighter for catchments with high *FS_x* corresponding to dominance of winter floods. Two of the models also include *EUnS_slope*, which characterizes the deviation from the mean seasonality, as a predictor, but with low explained variance. A last predictor, which is selected in all three models, is the size of the catchment. Small catchment sizes show heavier flood tails, as indicated by the negative model coefficient. The remaining predictors differ between the three models, but have only a very low explained variance and are not further discussed.

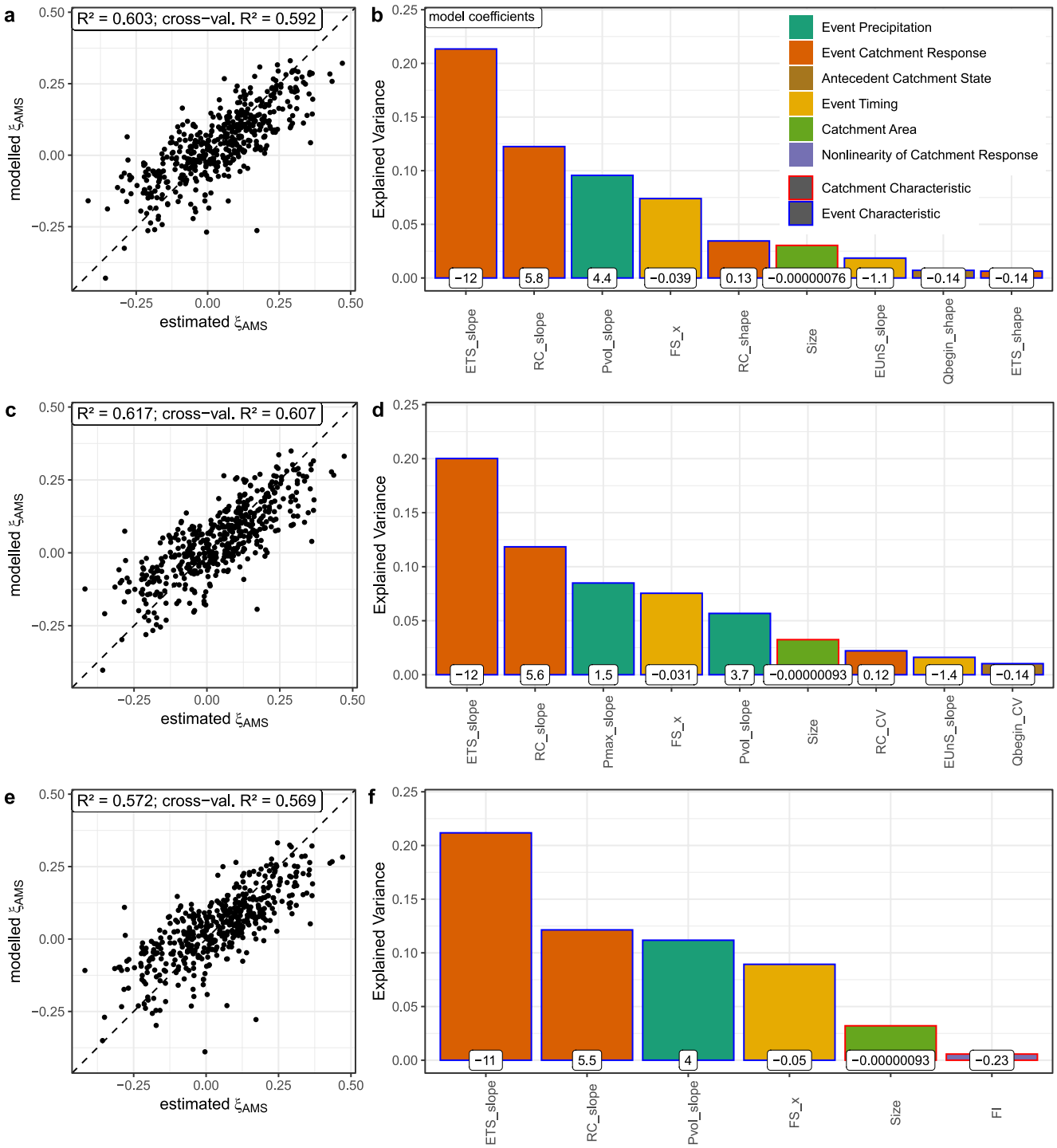


**Figure 8.** Results from the multiple linear regression models based on predictor sets with *slope* and *shape* (a), (b), *slope* and *CV* (c), (d) and only *slope* (e), (f) indicators for event-based variables. **a**, **c**, and **e** show modeled generalized extreme value shape parameters of annual maximum series flood flows ($\xi_{AMS}$) against $\xi_{AMS}$ estimated from time series. **b**, **d**, and **f** show the predictors selected in the best performing models along with their model coefficients and their relative importance for the respective model output. *Pvol* – precipitation volume, *Pmax* – maximum precipitation intensity, *Qbegin* – flow at event start, *RC* – runoff coefficient, *ETS* – event time scale, *EUnS* – event unseasonality, *slope* – slope indicator of event variables, *shape* – GEV shape parameter of event variables, *CV* – coefficient of variation of event variables, *FS_x* – x-component of flood seasonality, *Size* – catchment area, *FI* – flashiness index.

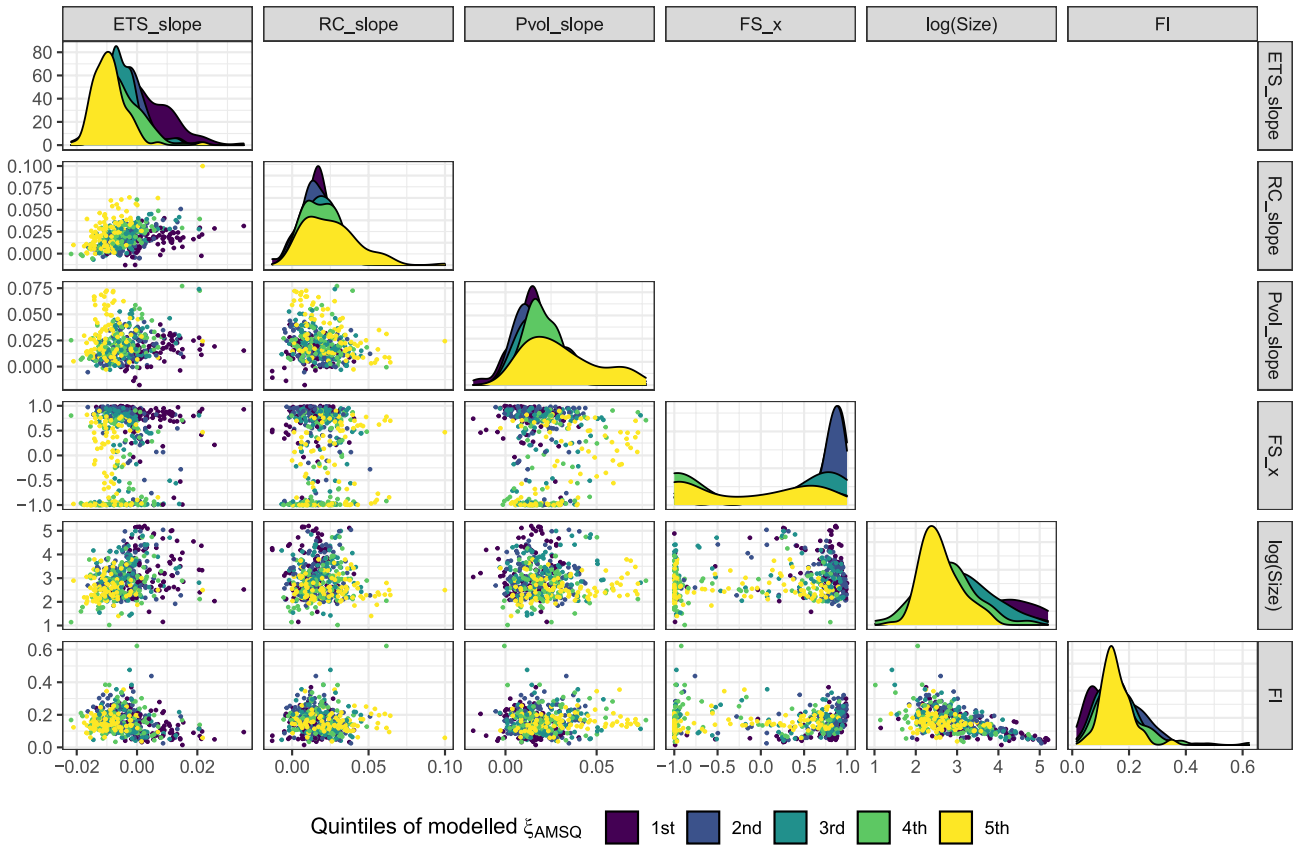


**Figure 9.** Interaction between the predictors selected in one of the best-performing multiple linear regression models. Lighter colors correspond to higher values of the predictand, that is, the generalized maximum value shape parameter of annual maximum series flood flows ($\xi_{AMS}$). Note that the axes differ between subplots: The *x*-axis (*y*-axis) of each subplot reports the values of the predictor stated at the top (right). High (low) values of *FS_x* correspond to a mean date of flood occurrence in winter (summer). *Pvol* – precipitation volume, *RC* – runoff coefficient, *ETS* – event time scale, *slope* – slope indicator of event variables, *FS_x* – x-component of flood seasonality, *Size* – catchment area, *FI* – flashiness index.

Since the three MLR models agree on the majority of predictors which are found to be most important and are also very similar in terms of overall explained variance, we select only one of them for further analysis. The model with only *slope* indicators is less complex than the other two (6 instead of 9 predictors), while having only a slightly lower $R^2$. So, we further focus on this model to analyze the interaction between selected predictors presented in Figure 9. In this plot matrix, the diagonal plots show the histograms of predictors and scatter plots below the diagonal show pairwise relationships between predictors. Histograms are segregated according to five quantile ranges of the predicted $\xi_{AMS}$, that is, the first range corresponds to the lower 20% of $\xi_{AMS}$ values and the fifth range to the upper 20%. These ranges are colored accordingly and the same color code is used in the scatter plots to pinpoint pairwise interactions of predictors and resulting $\xi_{AMS}$.

If *ETS_slope* is low, $\xi_{AMS}$ tends to be high, almost independently of the remaining predictors (Figure 9). Similarly, a high *RC_slope* leads to high $\xi_{AMS}$ even in combination with low *Pvol_slope* values or in winter-flood-dominated catchments where flood distributions would usually have lighter tails. The positive model coefficient of *Pvol_slope* suggests that $\xi_{AMS}$ increases with *Pvol_slope*. If, however, a high value of *Pvol_slope* coincides with a moderate to the high value of *ETS_slope* or a flood seasonality with mainly winter floods, the flood distribution tail is not as heavy. As seen before, catchments with mainly summer floods ($FS\_x \approx -1$) tend to have high $\xi_{AMS}$. Catchments with mainly winter floods have lighter tails, but $\xi_{AMS}$ can be high if *ETS_slope* is low and/or the catchment size is small. In small catchments, $\xi_{AMS}$ tends to be high. This is not the case though for the smallest catchments (<100 $km^2$) and if *ETS_slope* is high or winter floods are dominating. *FI* has only a weak impact on $\xi_{AMS}$ as can be seen in Figure 8. The negative model coefficient indicates that a low *FI* leads to a high $\xi_{AMS}$, but for the lowest values $\xi_{AMS}$ is low.

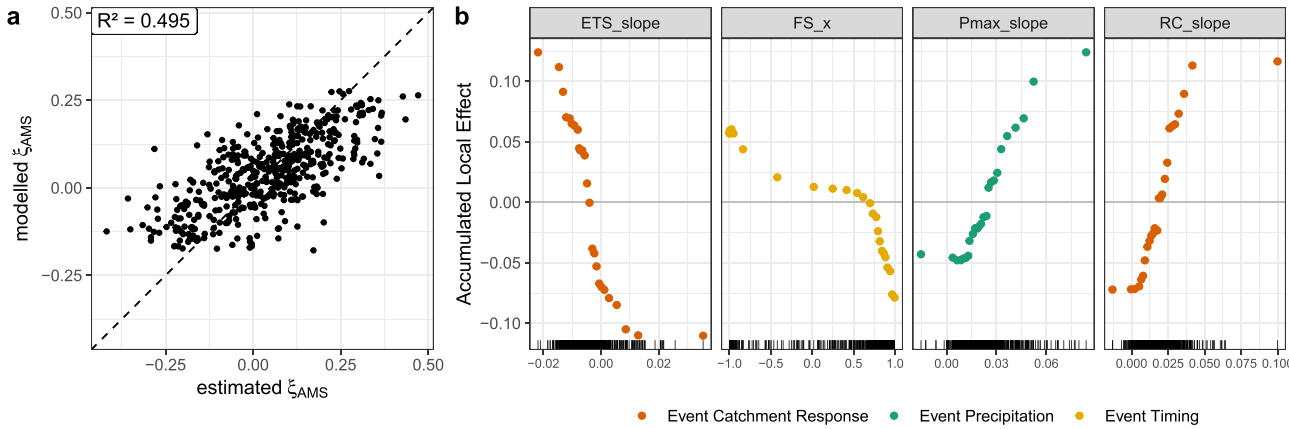


**Figure 10.** Results from the best-performing random forest model. (a) Modeled generalized extreme value shape parameters of annual maximum series flood flows ($\xi_{AMS}$) against $\xi_{AMS}$ estimated from time series and (b) Accumulated local effects (ALE) for the four selected predictors. ALE plots show how over the range of a predictor the model outcome differs from the mean prediction, in units of the predictand. Here, ranges were set to encompass 20 catchments each, leading to the uneven spacing of points. High (low) values of *FS_x* correspond to a mean date of flood occurrence in winter (summer). *Pmax* – maximum precipitation intensity, *RC* – runoff coefficient, *ETS* – event time scale, *slope* – slope indicator of event variables, *FS_x* – x-component of flood seasonality.

### 4.5. Random Forest

As for the MLR models, we start the nonlinear multivariate analysis with different sets of potential predictors (single/several indicators per variable). In all cases, the automatic predictor selection results in three to five predictors, which we then use to generate random forests. The $R^2$ values based on OOB data range from 0.274 to 0.495, with values above 0.49 only for models including *slope* indicators. In fact, in all three cases with *slope* indicators in the potential predictor sets (i.e., *slope* and *shape*, *slope* and *CV*, and only *slope*), the exact same four predictors are selected. So even if *shape* and *CV* indicators are in the set, they are not selected during the predictor selection, resulting in the same random forest model for the three sets. The following results focus on this best-performing model.

The four selected predictors are *ETS_slope*, *RC_slope*, *Pmax_slope,* and *FS_x* (Figure 10), characterizing the event catchment response, event precipitation, and event timing. This conforms with the event characteristics which are found to be most important in the MLR model. Predictors of the antecedent catchment state and of the flood types are not selected, as also none of the catchment characteristics.

In the random forest model, low values of *ETS_slope* lead to high values of $\xi_{AMS}$ (Figure 10), as also in the MLR model. If the largest flood events in a catchment are associated with very short *ETS* so that *ETS_slope* is for example, around −0.02, the model predicts $\xi_{AMS}$ to be higher by roughly 0.12 compared to the average prediction across all catchments (Figure 10b). Similarly, the model output is clearly above the average predicted $\xi_{AMS}$ for high values of *RC_slope* or *Pmax_slope*. High values of *FS_x* on the contrary lead to comparatively low values of $\xi_{AMS}$. This reflects the effect of flood seasonality, for example, for winter floods lighter tails can be expected and the opposite for summer floods. The effects of the four predictors on $\xi_{AMS}$ in the random forest model align with the results from univariate and linear multivariate analyses. No distinct nonlinear relationships between $\xi_{AMS}$ and the predictors *ETS_slope*, *RC_slope,* and *Pmax_slope* are found. For *FS_x*, the ALE plot shows a steep decrease toward negative effects for catchments dominated by winter floods, while the increase toward a positive effect for summer floods is less pronounced.

The interactions between the predictors selected in the random forest model are very similar to the ones already presented for the MLR model (Figure S2 in Supporting Information S1). The only additional predictor is *Pmax_slope*, but it behaves almost identical in relation to other predictors as *Pvol_slope*, which is included in the MLR model.

## 5. Discussion

Various event and catchment variables that can potentially determine the tail heaviness of flood flows are derived. For event-based variables, indicators are developed which are expected to pinpoint the effect of variable properties on tail heaviness. Besides more conventional measures, such as the coefficient of variation or the shape parameter of a fitted GEV distribution, the slope of a linear model relating the variable with the return period of the respective flood events is proposed. It is found (a) has the highest Spearman rank correlation against $\xi_{AMS}$ for most variables and (b) to improve model performance in both linear and nonlinear multivariate models. Most or all selected event-based predictors are *slope* indicators in the best-performing linear and nonlinear models, respectively.

The *slope* indicator captures well how a variable is associated with flood magnitude, especially for high return periods. For some variables, this association is found to have a crucial effect on the tail heaviness of flood flows. The upper tail behavior of flood distributions is per definition determined by the few largest events. In contrast to other examined indicators, the *slope* indicator is also dominated by the characteristics of those few large events and allows us to quantify if the highest (or lowest) values of a variable are associated with the largest flood events in a catchment. We evaluate the novel *slope* indicator to be of high value for the analysis of heavy-tailed flood flows.

In this study, we apply a multiple linear regression and nonlinear random forest model to analyze the controls of flood $\xi_{AMS.}$ The linear model is found to slightly outperform the nonlinear model. The best-performing linear regression, which is based on the predictor set with only *slope* indicators, explains 57% of the variance of $\xi_{AMS}$. With the random forest model generated for the same predictor set, 50% of the variance can be explained. In the examination of the results from the random forest approach, no distinct nonlinear relationships between predictors and predictand are observed for three of the four selected predictors. In contrast, the fourth one – flood seasonality – shows a nonlinearity with a stronger effect on $\xi_{AMS}$ for winter-flood-dominated catchments. The difference in the explained variance between the linear and the nonlinear model is relatively small and might partly be caused by differences in the estimation of the model performance (5-fold cross validation for linear vs. out-of-bag error for nonlinear approach). While the random forest model allows insights into nonlinear effects of predictors, the slightly better performing linear regression has the advantage of being easier to interpret and still capturing important influences on $\xi_{AMS}$.

Our analysis of flood tail behavior based on event and catchment characteristics suggests unequivocally that the event time scale and runoff coefficient along with event precipitation characteristics such as maximum precipitation intensity or precipitation volume are the dominant controls of flood tail heaviness. The importance of these factors is given by this very order. All the best models identify the slopes of *ETS* and *RC* relationships to flood return periods to be the most powerful predictors of flood tail heaviness. In the linear models, these two factors together explain more than half of the total explained variance. Catchments with large negative *ETS* slopes are associated with heavier tails. In other words, heavier tails prevail in catchments, where larger flood peaks are associated with relatively small direct runoff volumes compared to smaller floods. So, if catchments show a flashier response for larger events compared to smaller ones, this favors heavy-tailed flood distributions. *ETS*, originally proposed by Gaál et al. (2012), is controlled by the precipitation timescale and the time of runoff concentration. Precipitation timescale represents a characteristic duration of the rainfall event triggering an AMS flood (Viglione, Chirico, Woods, et al., 2010). Thus, both precipitation characteristics and runoff generation processes influence *ETS*. For the bulk of rainfall-runoff events, not specifically focused on AMS events, Tarasova et al. (2018) found a strong correlation (Spearman's $R^2 \geq 0.4$) between *ETS* and precipitation volume in many German catchments, suggesting a considerable influence of rainfall characteristics on *ETS*. In our flood event set consisting of annual maximum floods, this correlation vanishes ($R^2$ is primarily below 0.15, not shown). This result suggests that for AMS floods, *ETS* becomes rather controlled by runoff concentration than by precipitation. A negative *ETS* slope, that is, small *ETS* for large floods, indicates an efficient runoff concentration in a catchment and rapid transport of large runoff volumes in short times. This leads to peak hydrographs. Such behavior occurs in catchments where the prevailing runoff generation mode shifts from subsurface storm flow for small floods toward overland surface flow due to infiltration or saturation excess. For instance, Viglione, Chirico, Komma, et al. (2010) showed a large flood driven by long-rain rainfall in the Kamp region, Austria, to have a fast and concentrated response in contrast to small short-rain floods. For the latter, higher average hillslope travel times for rainfall excess and higher spatial average variance in hillslope routing times were detected. Furthermore,

connectivity of runoff contributing areas increases with flood magnitude. This process can additionally increase the efficiency of water transport in the landscape and lead to a rapid increase in runoff (Tromp-van Meerveld & McDonnell, 2006; Viglione, Chirico, Komma, et al., 2010; Western et al., 2001).

In a similar vein, a positive runoff coefficient slope, that is, large *RC* for large floods characterizes the propensity of a catchment to generate non-proportional large direct runoff volumes for large compared to small precipitation and snowmelt events. For the bulk of rainfall-runoff events, that is, a set dominated by small events, Merz et al. (2006) identified rainfall depth and antecedent catchment state as major controls of the temporal distribution of *RC*s within 337 Austrian catchments. Our multivariate analyses suggest that catchments with steeper *RC* slopes tend to show heavier flood tails. It is not a prerequisite for the emergence of heavy flood tails to have high *RC*s across all flood events, as is often the case in wetter catchments (Merz et al., 2006). Rather, the gradient of *RC*s between small and large floods is decisive for the upper tail behavior. The occurrence of high *RC*s for large floods can indicate the presence of threshold behavior related to saturation of runoff contributing to catchment areas and their connectivity, which leads to step changes in flood frequency curves (Rogger et al., 2012). Indeed, flood frequency curves for some of our analyzed catchments could be interpreted to have a step change following a visual inspection. Overall, our result suggests that the flood tail heaviness is mainly controlled by the runoff generation processes. These processes modulate the effect of precipitation input characterized by the slope of event precipitation intensity or volume. These two precipitation-related indicators follow both *ETS_slope* and *RC_slope* in terms of explained variance in the linear models.

The analysis of interactions between predictor variables indicates the dominant role of *ETS_slope* and *RC_slope* for the modeled $\xi_{AMS}$ values. Their impact on tail heaviness is not affected or modulated by other predictors. In contrast, the impact of several variables is dampened by the event catchment response if the indicators of *RC* and in particular *ETS* are particularly low and high, respectively.

A strong association of large floods with high precipitation volumes, that is, high values of *Pvol_slope*, generally leads to heavy-tailed flood distributions. If, however, the large floods are also associated with medium to long time scales of the catchment response, tail heaviness is reduced. This resonates well with findings from McCuen and Smith (2008), who argued that rainfall distribution properties are not directly passed on to flood distribution properties, but are modulated by catchment processes such as catchment and channel storage. Long *ETS* indicates slow runoff concentration and the contribution of slow runoff components, which can dampen the immediate effect of heavy precipitation.

The flood seasonality and catchment area are also found to exert an impact on tail heaviness but are less important than the event catchment response and event precipitation. Catchments that are dominated by summer floods show a higher propensity to heavy-tailed flood distributions than those with mainly winter floods. The random forest approach shows that this relationship is slightly nonlinear. The seasonality of floods is sometimes used as a surrogate for different flood-generating processes (e.g., Villarini & Smith, 2010). Here, we use event types to address dominant flood generation mechanisms, but the respective predictors are not selected in any of the best-performing models. Presumably, the selected flood seasonality component *FS_x* allows sufficient discrimination of flood types with respect to tail heaviness in Germany and Austria. Catchments with a high share of event types that predominantly occur in winter, that is, snow-dominated floods, tend to have lighter tails than those with a high share of rain-dominated floods, which are typical in summer.

The catchment area has a negative effect on the tail heaviness of flood distributions, with lighter tails for larger catchments, in our study region encompassing Germany and Austria. There have been previous studies finding this effect (e.g., Merz & Blöschl, 2009; Villarini & Smith, 2010), but also others have not detected such an association (e.g., Smith et al., 2018). A possible explanation for the influence of the catchment area is the spatial aggregation of processes in larger catchments. While local, convective rainfalls of high intensity can lead to relatively large floods in small catchments, their effect may be averaged out in larger catchments (Merz & Blöschl, 2009). Similarly, nonlinear behavior in the runoff generation at small scales, for example, threshold processes, can be averaged out with increasing catchment size, as catchment properties and relevant processes become more heterogeneous (Rogger et al., 2012). This spatial aggregation of precipitation and runoff generation can result in lighter tails for larger catchments.

The effect of catchment size on the tail heaviness of flood flows is superimposed by the effect of *ETS*, in particular for high values of *ETS_slope*. Small catchments have heavy-tailed distributions as long as the large floods

have shorter *ETS* than the smaller floods. If, however, the largest floods are dominated by slow processes of the catchment response, that is, the value of *ETS_slope* is high, even small catchments have light-tailed flood distributions. Long *ETS* is for example, common in catchments where lakes prolong the runoff concentration (Tarasova et al., 2018). Here, temporal smoothing is taking place, similar to the spatial aggregation discussed above for large catchments. The effect on flood distributions is the same: lighter tails in catchments with stronger aggregation, be it spatial or temporal.

Indicators of average catchment soil moisture at the event start as simulated by a hydrological model and show weak correlations with $\xi_{AMS}$. Slightly higher correlations are detected for *Qbegin* indicators. *Qbegin* is an integral variable characterizing not only the catchment moisture state at the onset of the flood event but is also affected by connectivity within a catchment. *Qbegin* slopes are mainly positive, though a significant portion of detected slopes is negative (~33%). Most of these negative slopes have very small absolute values (abs(*Qbegin_slope*) < 0.005) and are likely to result from uncertainties in the detection of the starting point of a hydrograph for small events. Though the univariate analysis indicates some association of *Qbegin_slope* with $\xi_{AMS}$, our multivariate models suggest no explanatory power of this indicator for the upper tail behavior of floods. At the first glance, it may sound counterintuitive that the initial catchment state does not play a role in the heavy tail behavior of floods, while its role in flood generation, in general, is extensively documented (e.g., Merz & Plate, 1997; Zehe & Blöschl, 2004). We believe this lies in the nature of heavy tails which are determined by a few largest events. Higher catchment wetness increases the propensity for flooding compared to dryer catchment states given the same precipitation input. The effect of antecedent catchment state and, in particular, the spatial variability of soil moisture is pronounced for medium-sized, but not large events, as detected in a small German catchment (Merz & Plate, 1997). This resonates with findings by Grillakis et al. (2016) for catchments in Austria and Crete, who suggested that the effect of pre-event soil moisture on the peak flow of flash floods depends on the precipitation event magnitude. Also in the large Elbe catchment in Germany, Nied et al. (2017) detected a decreasing role of the pre-event catchment state compared to event-triggering weather patterns for the occurrence of larger floods with return periods above 10 years compared to smaller floods ($T \geq 2$ years). Thus, there is evidence for consistent sensitivity of floods to pre-event catchment states for small and large catchments and this sensitivity is controlled by the magnitude of meteorological drivers. The occurrence of high moisture states appears to be not sufficient for the emergence of heavy tails. In catchments that tend to be wet, variability of pre-event soil moisture or streamflow at the event start is relatively low, so other factors control the occurrence of extremes. In catchments that tend to be dry, the variability in pre-event conditions is large, but still not sufficient to explain the occurrence of heavy tails. Additional factors, such as intensive or high volume precipitation and/or unusual catchment response characterized by the *ETS_slope* and *RC_slope* indicators discussed above, are required for the generation of extraordinary events.

There are some limitations to the underlying data. Using catchment-averaged variables can mask very heterogeneous conditions, especially for large catchments, and thus introduce some uncertainty. For example, a local, high-intensity rainfall may be averaged out in large catchments (Merz & Blöschl, 2009). For small catchments, on the other hand, a higher resolution of the gridded precipitation data would be beneficial, as it would better represent local conditions. Similarly, a higher temporal resolution of the data would also be beneficial for small catchments, as the mean daily flow can underestimate flood peaks of a short duration. High-resolution data is sparse though and going below a daily scale would distinctively reduce the number of gauges with sufficiently long time series in our study region.

The negative shape parameters estimated for some catchments should be interpreted with caution. They correspond to distributions that are bounded from above. Upper limits of streamflow are also assumed by the concept of probable maximum precipitation and probable maximum flood, which is used in hydrological practice for designing sensitive infrastructure. However, this concept has been critized as flawed (Koutsoyiannis & Papalexiou, 2017; Salas et al., 2014). Further, for a global analysis of extreme daily rainfall, Papalexiou and Koutsoyiannis (2013) reported a decreasing percentage of negative shape parameters with increasing record length, suggesting exclusively positive values for records of sufficient length. The availability of streamflow records of appropriate length is limited though. Therefore, it might be worth to further exploring alternative distributions, as recently demonstrated by Zaghloul et al. (2020).

The linear and nonlinear multivariate analyses provide useful insights regarding the processes that are most relevant for the emergence of heavy tails in flood peak distributions. Understanding the controls of heavy tails can

be beneficial for flood risk management, especially in situations with limited funds or otherwise restricted possibilities. Unfortunately, using the derived MLR or RF as predictive models for catchments with no or only short records of streamflow data is not practical. This is because the predictors which were found to have the highest explanatory power, that is, *ETS_slope* and *RC_slope*, are based on streamflow time series themselves. Nevertheless, knowing the influence that differences in *ETS* and *RC* between small and large floods have on heavy tail behavior, is already an important step. It strongly suggests further studying these variables in more detail to get a better understanding of the factors that lead to distinct differences in the runoff generation for different flood magnitudes. This in turn can help in identifying catchments that are prone to these factors and that are also prone to exhibiting heavy-tailed flood distributions. Once catchments with a higher probability of the occurrence of extreme floods are detected, these can be prioritized in flood risk management practices.

## 6. Conclusions

We conducted a multivariate analysis of an extensive set of event and catchment characteristics with the aim of exploring the controls of heavy tail behavior of flood distributions. Tail heaviness is described by the shape parameter of the Generalized Extreme Value distribution fitted to the annual maximum flow series. The analyzed event characteristics include variables related to the antecedent catchment state, flood-generating event precipitation, event catchment response, and timing and types of flood events. For these event-based variables, different indicators were estimated and we expected to pinpoint the association between a variable and the tail heaviness of flood distributions. Among those indicators, the slope indicator was found to be the most powerful. It captures

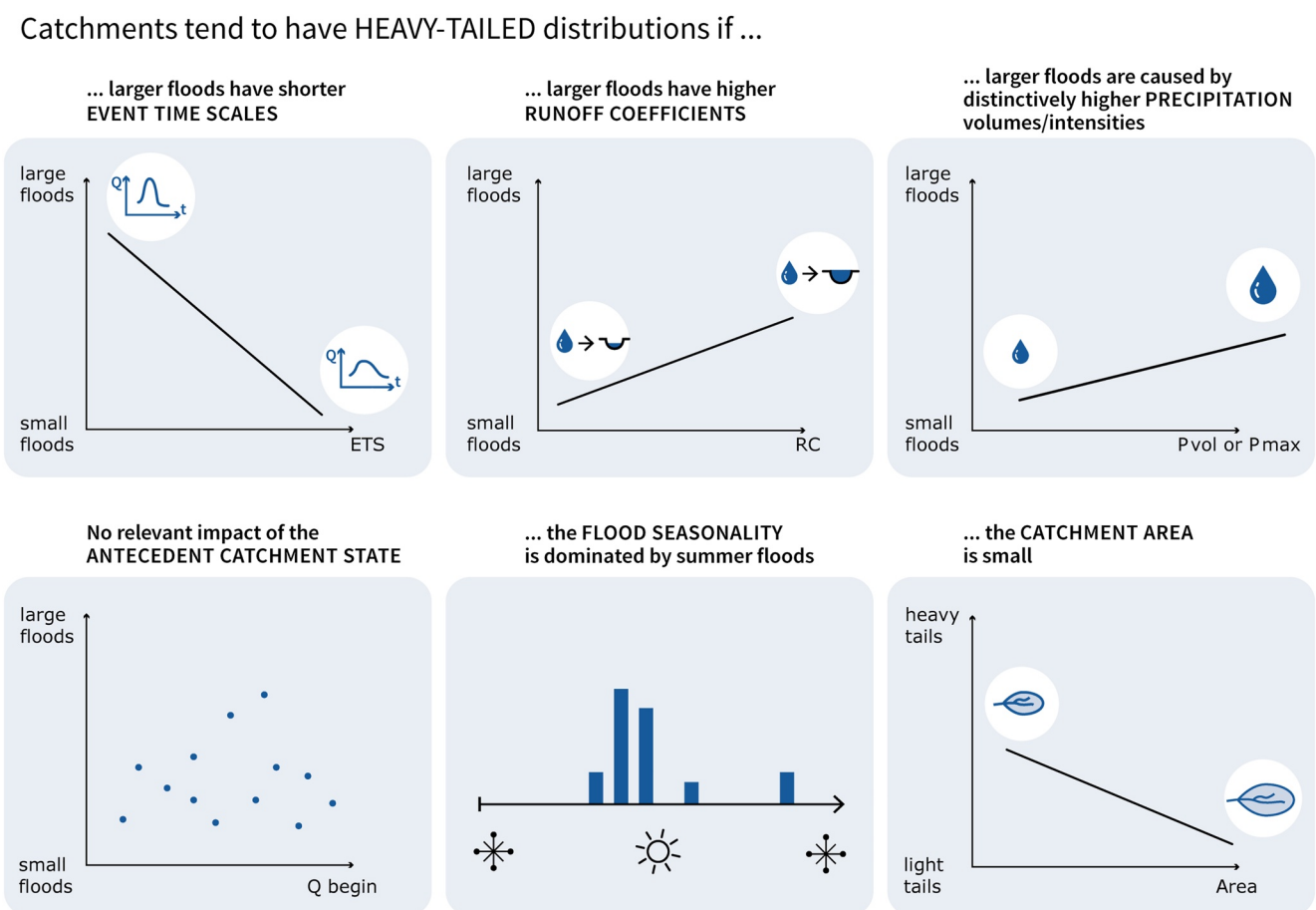


**Figure 11.** Summary of the findings from the multivariate analysis of potential controls of heavy tail behavior of flood peak distributions. The results are based on data from 480 gauges across Germany and Austria, which range from quickly reacting mountain catchments to large lowland catchments. *ETS* – event time scale, *RC* – runoff coefficient, *Pvol* – precipitation volume, *Pmax* – maximum precipitation intensity, *Qbegin* – flow at event start.

how a variable is associated with the flood return period, that is, whether the largest flood events at a gauge also have the highest or lowest values of a respective variable. For some variables, this association was found to have a crucial effect on the tail heaviness of flood flows.

Univariate as well as linear and nonlinear multivariate analyses were conducted to scrutinize potential heavy tail predictors. In both the multiple linear regression and the random forest approach, variables describing the event catchment response were identified as the dominant controls of heavy tail behavior, followed by event precipitation characteristics. The main controls of heavy-tailed distributions are summarized in Figure 11. Catchments, where the largest floods have smaller ratios of direct runoff volume to peak flow than smaller floods, tend to have heavy-tailed distributions. Similarly, associations of the largest events with the highest runoff coefficients or with the highest flood-event causing precipitation volumes result in heavier tails. The flood seasonality and catchment area also exert an impact on tail heaviness, but to a lesser extent. Heavier tails were found for smaller catchments and for those dominated by summer floods. Pre-event catchment moisture state is not decisive for the emergence of heavy tails in flood series. In summary, our analysis shows that catchment response is decisive for the emergence of heavy tails, but not the pre-event catchment state. The strong role of catchment response suggests that in catchments with heavy tail behavior, extreme floods are associated with faster, rainfall-driven runoff generation, as opposed to a slower runoff generation for smaller floods. Particularly, small catchments can reach threshold states under severe precipitation events, where flood generation processes shift, causing extreme floods. A better understanding of the factors that lead to an unusual catchment response for large compared to smaller floods promises high benefits for flood design and risk assessment.

## Data Availability Statement

Data on the characteristics of catchments and of flood events that were used for deriving potential heavy tail predictors in the study are available at GFZ Data Services via https://doi.org/10.5880/GFZ.4.4.2021.004 (Macdonald et al., 2021).

**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)" (project number 278017089) is gratefully acknowledged. MK was funded by the DFG Research Training Group "Natural Hazards and Risks in a Changing World" (NatRisk Change GRK 2043). We thank Simon Michael Papalexiou and two anonymous reviewers for their thorough and constructive reviews. Open access funding enabled and organized by Projekt DEAL.

## References

Apley, D. W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society: Series B*, *82*(4), 1059–1086. https://doi.org/10.1111/rssb.12377

Bacchi, B., Brath, A., & Kottegoda, N. (1992). Analysis of the relationships between flood peaks and flood volumes based on crossing properties of river flow processes. *Water Resources Research*, *28*, 2773–2782. https://doi.org/10.1029/92WR01135

Baker, D. B., Richards, R. P., Loftus, T. T., & Kramer, J. W. (2004). A new flashiness index: Characteristics and applications to Midwestern rivers and streams. *Journal of the American Water Resources Association*, *40*(2), 503–522. https://doi.org/10.1111/j.1752-1688.2004.tb01046.x

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

Bernardara, P., Scherzer, D., Sauquet, E., Tchiguirinskaia, I., & Lang, M. (2008). The flood probability distribution tail: How heavy is it? *Stochastic Environmental Research and Risk Assessment*, *22*(1), 107–122. https://doi.org/10.1007/s00477-006-0101-2

Bertola, M., Viglione, A., Vorogushyn, S., Lun, D., Merz, B., & Blöschl, G. (2020). Do small and large floods have the same drivers of change? A regional attribution analysis in Europe. *Hydrology and Earth System Sciences*, *August*, 1–26. https://doi.org/10.5194/hess-2020-396

Booker, D. J., & Snelder, T. H. (2012). Comparing methods for estimating flow duration curves at ungauged sites. *Journal of Hydrology*, *434*(435), 78–94. https://doi.org/10.1016/j.jhydrol.2012.02.031

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Burn, D. (1997). Catchment similarity for regional flood frequency analysis using seasonality measures. *Journal of Hydrology*, *202*, 212–230. https://doi.org/10.1016/S0022-1694(97)00068-1

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

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

Ettrick, T. M., Mawdlsey, J. A., & Metcalfe, A. V. (1987). The influence of antecedent 806 catchment conditions on seasonal flood risk. *Water Resources Research*, *23*, 481–488. https://doi.org/10.1029/WR023i003p00481

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

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. (1934). *Statistical methods for research workers* (5th ed.). Oliver & Boyd.

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

Frahm, G., Junker, M., & Schmidt, R. (2005). Estimating the tail dependence coefficient: Properties and pitfalls. *Insurance: Mathematics and Economics*, *37*(1), 80–100. https://doi.org/10.1016/j.insmatheco.2005.05.008

Gaál, L., Szolgay, J., Kohnová, S., Parajka, J., Merz, R., Viglione, A., & Blöschl, G. (2012). Flood timescales: Understanding the interplay of climate and catchment processes through comparative hydrology. *Water Resources Research*, *48*(4), W04511. https://doi.org/10.1029/2011WR011509

Ganguli, P., & Merz, B. (2019). Extreme coastal water levels exacerbate fluvial flood hazards in Northwestern Europe. *Scientific Reports*, *9*(1), 1–14. https://doi.org/10.1038/s41598-019-49822-6

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

Gioia, A., Iacobellis, V., Manfreda, S., & Fiorentino, M. (2008). Runoff thresholds in derived flood frequency distributions. *Hydrology and Earth System Sciences*, *12*(6), 1295–1307. https://doi.org/10.5194/hess-12-1295-2008

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

Grillakis, M. G., Koutroulis, A. G., Komma, J., Tsanis, I. K., Wagner, W., & Blöschl, G. (2016). Initial soil moisture effects on flash flood generation – A comparison between basins of contrasting hydro-climatic conditions. *Journal of Hydrology*, *541*, 206–217. https://doi.org/10.1016/j.jhydrol.2016.03.007

Grömping, U. (2006). Relative importance for linear regression in R: The package relaimpo. *Journal of Statistical Software*, *17*(1), 1–27. Retrieved from https://cran.r-project.org/package=relaimpo

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*, 937–953. https://doi.org/10.1002/2013WR014374

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*, 1633–1648. https://doi.org/10.5194/hess-24-1633-2020

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). Springer. Retrieved from http://www.springerlink.com/index/D7X7KX6772HQ2135.pdf

Haylock, M., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., & New, M. (2008). A European daily high-resolution gridded dataset of surface temperature, precipitation and sea-level pressure. *Journal of Geophysical Research*, *113*, D20119. https://doi.org/10.1029/2008JD010201

Hirsch, R. M., Helsel, D. R., Cohn, T. A., & Gilroy, E. J. (1992). Statistical analysis of hydrological data. In D. A. Maidment (Ed.), *Handbook of Hydrology* (pp. 17.1–17.55). McGraw-Hill. https://doi.org/10.1097/00004669-199201000-00016

Hofert, M., Kojadinovic, I., Maechler, M., & Yan, J. (2020). copula: Multivariate Dependence with Copulas. R package version 1.0-1. [Dataset]. Retrieved from https://CRAN.R-project.org/package=copula

Hothorn, T., Hornik, K., Strobl, C., & Zeileis, A. (2021). party: A Laboratory for Recursive Partytioning. R package version 1.3-7 [Dataset]. Retrieved from https://CRAN.R-project.org/package=party

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational & Graphical Statistics*, *15*(3), 651–674. https://doi.org/10.1198/106186006X133933

Ibarra-Berastegi, G., Saénz, J., Ezcurra, A., Elías, A., Diaz Argandoña, J., & Errasti, I. (2011). Downscaling of surface moisture flux and precipitation in the Ebro Valley (Spain) using analogues and analogues followed by random forests and multiple linear regression. *Hydrology and Earth System Sciences*, *15*(6), 1895–1907. https://doi.org/10.5194/hess-15-1895-2011

Kemter, M., Merz, B., Marwan, N., Vorogushyn, S., & Blöschl, G. (2020). Joint trends in flood magnitudes and spatial extents across Europe. *Geophysical Research Letters*, *46*, e2020GL087464. https://doi.org/10.1029/2020GL087464

Klein, B. (2009). Ermittlung von Ganglinien für die risikoorientierte Hochwasserbemessung von Talsperren (Doctoral dissertation), Doktorarbeit an der Ruhr-Universität Bochum, Fakultät fuer Bau- und Umweltingenieurwissenschaften, Lehrstuhl für Hydrologie, Wasserwirtschaft und Umwelttechnik.

Knoben, W., Woods, R., & Freer, J. (2018). A quantitative hydrological climate classification evaluated with independent streamflow data. *Water Resources Research*, *54*, 5088–5109. https://doi.org/10.1029/2018WR022913

Koutsoyiannis, D., & Papalexiou, S. M. (2017). *Extreme rainfall: Global perspective, Handbook of applied Hydrology*. In V. P. Singh (Ed.), (2nd ed., pp. 74.1–74.16). McGraw-Hill.

Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). *Applied linear regression models* (4th ed.). McGraw-Hill Irwin.

LAWA. (20*18). Leitfaden zur Hydrometrie des Bundes und der Länder – Pegelhandbuch, 5. Aufl., Bund/Länderarbeitsgemeinschaft Wasser (LAWA) and Ministerium für Umwelt, Klima und Energiewirtschaft Baden-Württemberg*.

Lever, J., Krzywinski, M., & Altman, N. (2016). Points of significance: Model selection and overfitting. *Nature Methods*, *13*(9), 703–704. https://doi.org/10.1038/nmeth.3968

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

Macdonald, E., Merz, B., Ullrich, S., Lüdtke, S., Schröter, K., Guse, B., et al. (2021). Flood event and catchment characteristics in Germany and Austria [Dataset]. GFZ Data Services. https://doi.org/10.5880/GFZ.4.4.2021.004

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)

Melsen, L. A., & Guse, B. (2019). Hydrological drought simulations: How climate and model structure control parameter sensitivity. *Water Resources Research*, *55*(12), 10527–10547. https://doi.org/10.1029/2019WR025230

Merz, B., Basso, S., Fischer, S., Lun, D., Blöschl, G., Merz, R., et al. (2022). *Understanding heavy tails of flood peak distributions*. Submitted to Water Resources Research. https://doi.org/10.1029/2021WR030506

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*, W08432. https://doi.org/10.1029/2007WR006744

Merz, R., & Blöschl, G. (2009). A regional analysis of event runoff coefficients with respect to climate and catchment characteristics in Austria. *Water Resources Research*, *45*, W01405. https://doi.org/10.1029/2008WR007163

Merz, R., Blöschl, G., & Parajka, J. (2006). Spatio-temporal variability of event runoff coefficients. *Journal of Hydrology*, *331*, 591–604. https://doi.org/10.1016/j.jhydrol.2006.06.008

Merz, B., Nguyen, D., Apel, H., Gerlitz, L., Schröter, K., Steirou, E.-S., & Vorogushyn, S. (2018). Spatial coherence of flood-rich and flood-poor periods across Germany. *Journal of Hydrology*, *559*, 813–826. https://doi.org/10.1016/j.jhydrol.2018.02.082

Merz, B., & Plate, E. J. (1997). An analysis of the effects of spatial variability of soil and soil moisture on runoff. *Water Resources Research*, *33*(12), 2909–2922. https://doi.org/10.1029/97WR02204

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

Molnar, C. (2021). *Interpretable machine learning - a Guide for Making Black Box models explainable*. Retrieved from https://christophm.github.io/interpretable-ml-book/

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2001). *Introduction to linear regression analysis* (3rd ed.). John Wiley and Sons, Inc.

Morrison, J. E., & Smith, J. A. (2002). Stochastic modeling of flood peaks using the generalized extreme value distribution. *Water Resources Research*, *38*(12), 41-1–41-2. https://doi.org/10.1029/2001wr000502

Nied, M., Schröter, K., Lüdtke, S., Nguyen, V. D., & Merz, B. (2017). What are the hydro-meteorological controls on flood characteristics? *Journal of Hydrology*, *545*, 310–326. https://doi.org/10.1016/j.jhydrol.2016.12.003

Norbiato, D., Borga, M., Merz, R., Blöschl, G., & Carton, A. (2009). Controls on event runoff coefficients in the eastern Italian Alps. *Journal of Hydrology*, *375*(3–4), 312–325. https://doi.org/10.1016/j.jhydrol.2009.06.044

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

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 Discussions*, *4*(2), 589–625. https://doi.org/10.5194/hessd-4-589-2007

Primo, C., Kelemen, F. D., Feldmann, H., Ahrens, B., & Ahrens, B. (2019). A regional atmosphere-ocean climate system model (CCLMv5.0clm7-NEMOv3.3-NEMOv3.6) over Europe including three marginal seas: On its stability and performance. *Geoscientific Model Development Discussions*, *12*(12), 5077–5095. https://doi.org/10.5194/gmd-12-5077-2019

Robette, N. (2020). Moreparty: A Toolbox for conditional inference random forests. R package version (0.2.0) [Dataset]. Retrieved from https://CRAN.R-project.org/package=moreparty

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

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.

Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, *46*, W05523. https://doi.org/10.1029/2008WR007327

Schmid, F., & Schmidt, R. (2007). Multivariate conditional versions of Spearman's rho and related measures of tail dependence. *Journal of Multivariate Analysis*, *98*(6), 1123–1140. https://doi.org/10.1016/j.jmva.2006.05.005

Schröter, K., Kunz, M., Elmer, F., Mühr, B., & Merz, B. (2015). What made the June 2013 flood in Germany an exceptional event? A hydro-meteorological evaluation. *Hydrology and Earth System Sciences*, *19*, 309–327. https://doi.org/10.5194/hess-19-309-2015

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*(6), 1–17. https://doi.org/10.1029/2004WR003439

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

Strobl, C., Malley, J., & Tutz, G. (2009). *An introduction to recursive partitioning: Rationale, Application and characteristics of classification and regression trees*. Bagging and Random Forests (Issue 55). Retrieved from https://epub.ub.uni-muenchen.de/10589/1/partitioning.pdf

Tarasova, L., Basso, S., Wendi, D., Viglione, A., Kumar, R., & Merz, R. (2020). A process-based framework to characterize and classify runoff events: The event typology of Germany. *Water Resources Research*, *56*(5). https://doi.org/10.1029/2019WR026951

Tarasova, L., Basso, S., Zink, M., & Merz, R. (2018). Exploring controls on rainfall-runoff events: 1. Time series-based event separation and temporal dynamics of event runoff response in Germany. *Water Resources Research*, *54*(10), 7711–7732. https://doi.org/10.1029/2018WR022587

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

Tromp-van Meerveld, H. J., & McDonnell, J. J. (2006). Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis. *Water Resources Research*, *42*, W02411. https://doi.org/10.1029/2004WR003800

Tyralis, H., Papacharalampous, G., & Langousis, A. (2019). A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water*, *11*(5), 910. https://doi.org/10.3390/w11050910

Viglione, A., Chirico, G. B., Komma, J., Woods, R., Borga, M., & Blöschl, G. (2010). Quantifying space-time dynamics of flood events. *Journal of Hydrology*, *394*, 213–229. https://doi.org/10.1016/j.jhydrol.2010.05.041

Viglione, A., Chirico, G. B., Woods, R., & Blöschl, G. (2010). Generalized synthesis of space-time variability in flood response: An analytical framework. *Journal of Hydrology*, *394*, 198–212. https://doi.org/10.1016/j.jhydrol.2010.05.047

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

Villarini, G., Smith, J. A., Baeck, M. L., Vitolo, R., Stephenson, D. B., & Krajewski, W. F. (2011). On the frequency of heavy rainfall for the Midwest of the United States. *Journal of Hydrology*, *400*(1–2), 103–120. https://doi.org/10.1016/j.jhydrol.2011.01.027

Vogt, J., Soille, P., De Jager, A., Rimaviciute, E., Mehl, W., Foisneau, S., et al. (2007). A pan-European river and catchment database. *EUR 22920 EN*. OPOCE. JRC40291. https://doi.org/10.2788/35907

Western, A. W., Blöschl, G., & Grayson, R. B. (2001). Toward capturing hydrologically significant connectivity in spatial patterns. *Water Resources Research*, *37*(1), 83–97. https://doi.org/10.1029/2000WR900241

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

Zaghloul, M., Papalexiou, S. M., Elshorbagy, A., & Coulibaly, P. (2020). Revisiting flood peak distributions: A pan-Canadian investigation. *Advances in Water Resources*, *145*, 103720. https://doi.org/10.1016/j.advwatres.2020.103720

Zehe, E., & Blöschl, G. (2004). Predictability of hydrologic response at the plot and catchment scales: Role of initial conditions. *Water Resources Research*, *40*, W10202. https://doi.org/10.1029/2003wr002869

Zink, M., Kumar, R., Cuntz, M., & Samaniego, L. (2017). A high-resolution dataset of water fluxes and states for Germany accounting for parametric uncertainty. *Hydrology and Earth System Sciences*, *21*(3), 1769–1790. https://doi.org/10.5194/hess-21-1769-2017