Reliable estimation of high floods: a method to select the most suitable ordinary distribution in the Metastatistical Extreme Value framework

Sumra Mushtaq[1], Arianna Miniussi[1], Ralf Merz[1], and Stefano Basso[1]

[1] Department of Catchment Hydrology, Helmholtz Centre for Environmental Research - UFZ, Halle (Saale), Germany.

Corresponding author: Sumra Mushtaq (sumra.mushtaq@ufz.de)

Abstract

Recent advances in the study of extreme values, namely the Metastatistical Extreme Value (MEV) framework, showed good performances for the estimation of extremes in several fields. Here we adopt MEV for flood frequency analysis and leverage its intrinsic property of allowing for the choice of the distribution which best describes ordinary peaks to improve flood estimation. To this end, we develop a non-parametric approach to select ex ante the most suitable distribution of ordinary peaks between Gamma and Log-Normal. The method relies on the tail ratio, which we define as the ratio between the empirical 99th and 95th percentile of the ordinary peaks, and is tested by using daily streamflow time series from 182 gauges in Germany. Based on the value of the tail ratio index, we choose either the Gamma or the Log-Normal distributions to represent the ordinary peaks in each gauge. The approach correctly identifies the most suitable distribution of ordinary peaks in a large majority of the analyzed basins, and is robust to changes of the considered dataset. The preliminary selection of the ordinary distribution based on the tail ratio index improves the estimation of frequent and rare floods with respect to MEV applied with a single distribution not tailored on the specific statistical properties of the ordinary peaks. Finally, by comparing the developed methodology with the standard Generalized Extreme Value (GEV) distribution, we show that we are able to reduce the estimation uncertainty of high flood quantiles.

Keywords: flood frequency estimation, extreme events, flood hazard, Metastatistical Extreme Value distribution, tail properties

1. Introduction

Floods repeatedly cause severe property damages and loss of human lives across the world (UNISDR 2015). More than half a million deaths occurred due to these events and 2.8 billion people were affected globally between 1980 and 2009 (Doocy et al., 2013). Focusing on a more local scale, Germany has experienced several intense flood events in recent decades. In hydrological terms, the June 2013 flood can be regarded as the most severe flood over the last 60 years (Merz et al., 2014), with high flood peaks in the Elbe catchment in eastern Germany, the Danube in the south, the Weser, and the Rhine in the western part of Germany. Most recently, in July 2021, heavy precipitation and severe thunderstorms induced flooding in western Germany, affecting more than 40,000 people and with at least 184 fatalities (Fekete and Sandholz 2021). From an economical viewpoint, whereas the August 2002 floods caused several billion euros losses (Thieken et al., 2005), floods in 2021 caused financial losses for more than 30 billion euros ('DW News', 2021). A reliable assessment of natural hazards at the origin of such disasters is crucial to improve preparedness of societies. In particular, accurate estimation of flood magnitude and frequency is key for the design and management of engineering and hydraulic structures, which are essential tools to reduce damages and thus protect a wide range of social and economic activities (Eash, 1997; Rosbjerg, 2013; Macdonald et al., 2006).

Flood frequency analysis is usually performed, according to the classical extreme value theory, by fitting a distribution on either annual maxima (e.g., Cunnane, 1973; Villarini and Smith, 2010; Seckin et al., 2011) or a few values over a high threshold (i.e., the Peak Over Threshold approach; Davison and Smith, 1990). Commonly used distributions are Gumbel, Log-Normal, Pearson Type III, Gamma, Log-Pearson and GEV (Bobée et al., 1993, Morrison and Smith, 2002; Haktinar et al., 2013). The possibility given by these approaches of inferring the upper tail of a distribution without knowing the whole stochastic underlying process is appealing and was necessary in time periods when only annual maxima were recorded. However, these approaches discard most of the information contained in the bulk of runoff events (Tarasova et al., 2020), thus typically providing estimations affected by large uncertainty (Miniussi et al., 2020). The limitations of traditional extreme value approaches have been highlighted by several studies (e.g., Klemeš, 1974; Iliopoulou and Koutsoyiannis, 2019; Lombardo et al., 2019). Among others, Hu et al. (2020) found that the estimation of the shape parameter of the GEV distribution (which chiefly controls its tail, i.e., the part of the distribution which describes rare events) is very sensitive to the sampling of flow data and highly affected by the length of the available time series. This problem can lead to huge uncertainty for the estimation of low-frequency quantiles, thus prohibiting the GEV to be a competitive candidate for the calculation of at-site flood statistics when only short observational series are available (Renard et al., 2013), as it occurs in large parts of the world (Müller and Thompson, 2016).

Recently, developments in extreme value statistics and novel theoretical approaches to appraise extreme discharges have proposed to estimate maxima by leveraging the information content of the full distribution of events (Marani and Ignaccolo, 2015; Basso et al., 2016), thus making more effective use of the available data. In particular, the Metastatistical Extreme Value (MEV) distribution (Marani and Ignaccolo, 2015; Zorzetto et al., 2016) postulates that extremes emerge from the set of ordinary peaks (i.e., all the independent values, not annual maxima or peak over threshold values), which are characterized by means of suitably selected probability distributions. The flexibility of MEV for the choice of the underlying distribution of ordinary events expressly allows for selecting the statistical model which better represents the upper tail of the distribution, a pivotal feature for achieving a correct characterization of extreme events.

The importance of the upper tail behavior of a distribution for the description of extremes is further highlighted by the number of past and recent studies dealing with the subject (e.g., Smith et al., 2018; Wietzke et al., 2020). To investigate the tail properties, previous studies employed different graphical methods, including the generalized Hill ratio plot, log-log plot (also called tail probability plot), and the mean excess function (Resnick 2007; El Adlouni et al., 2008). These graphical methods, being based on visual interpretation, lack objectivity and cannot be easily implemented when analyzing large datasets. These drawbacks call for the development of robust and objective methods (Cook et al., 2004). Recent attempts towards this direction have been done. For example, Nerantzaki and Papalexiou (2019) developed an automated procedure for the application of the mean excess function to investigate the tail properties of precipitation at the global scale. Wietzke et al. (2020) made a detailed inter-comparison of different upper tail indicators with the aim of indicating the most appropriate method. They considered the shape parameter of the GEV distribution, the Gini Index (Eliazar and Sokolova, 2010), the obesity index (Cooke and Nieboer, 2011) and the upper tail ratio (Smith et al., 2018). Their findings eventually suggest that the GEV shape parameter is the most suitable indicator for distinguishing between different tail behaviors. This indicator has however the drawback of being parametric, i.e., it requires the prior assumption of a distribution and it's fitting to the sample of maxima, which entails notable uncertainties (Basso et al., 2021).

In this work, we leverage the flexibility of MEV in seeking for the optimal distribution of ordinary events, and develop a methodology that enables selection ex ante of the best fitting ordinary peaks distribution. Related benefits for the estimation of flood quantiles are then evaluated. We thus intend to provide indications on how to suitably apply the MEV framework for the prediction of hydrologic extremes. The article is organized as follows: in Section 2, we describe the data and the methodology used in this study; Section 3 displays and discusses the results, which are finally summarized in Section 4.

2. Materials and Methods

2.1 Study Area and Data

We analyzed daily streamflow records from 182 gauges in Germany (Tarasova et al., 2018) which have time series longer than 30 years and less than 10% missing observations in each year. We utilized hydrological years, spanning from October to September. The catchment sizes range from 30 to 23,842 km^2 (median value: 581 km^2) and the lengths of the time series are comprised between 37 and 64 years (median value: 61 years). Catchments in the dataset cover the whole of Germany and its diverse climatic and physiographic conditions (Figure 1a). Germany is influenced by its specific position between continental climate in the east and maritime climate in the west of the country. The northwestern region of the country is dominated by circulation patterns linked to mid-latitude cyclone rainfall that can cause river flooding. Rainfall varies from North to South and reach its maximum in the Alpine Forelands and Southern Scarplands, with annual rainfall more than 2000 mm (Schädler et al., 2012) in the Alps. Annual rainfall amounts decrease from west to east (Figure 1b). Catchments in Germany exhibit different flood regimes that dominate during particular seasons: the central and western parts are dominated by winter floods, the north and east areas experience spring and summer floods, and the southern part of Germany is dominated by summer floods (Beurton and Thieken, 2009). The study area includes small to medium sized catchments which react faster than large catchments to heavy precipitation and are thus characterized by a more intense flood hazard (Schädler et al., 2012).

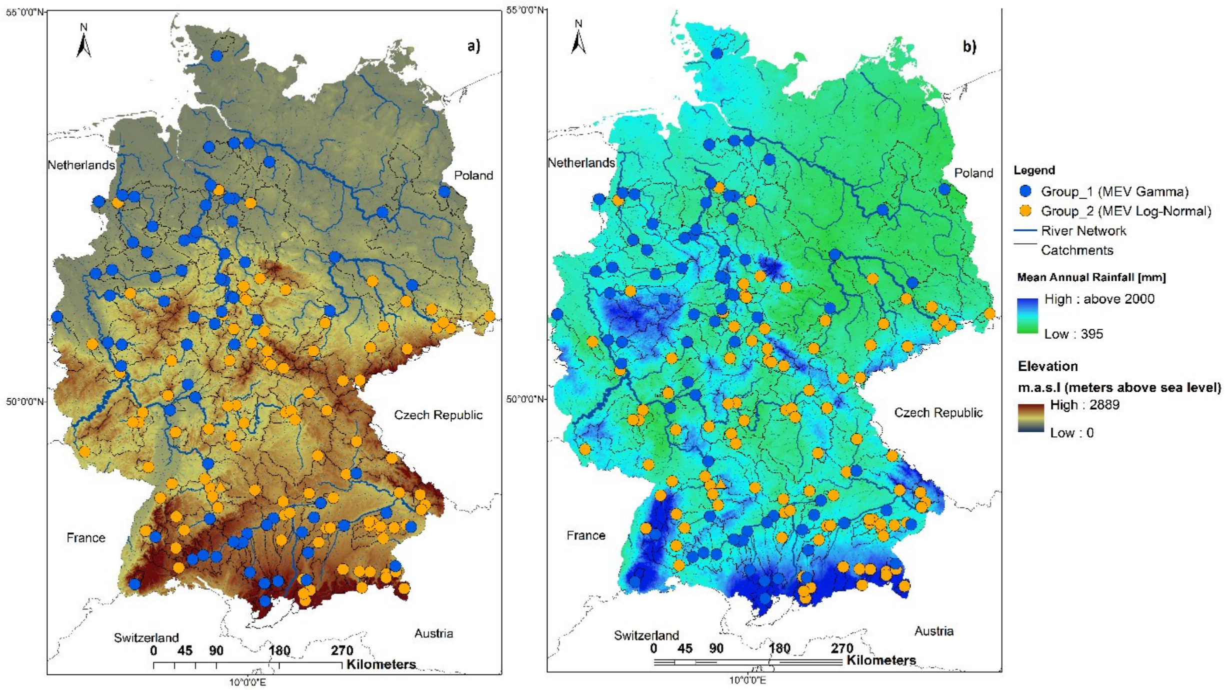

Figure 1 –Locations within Germany of the 182 streamflow gauges analyzed in the present work (blue and orange dots). The two colors refer to the subdivision of catchments into two groups for which either Gamma or Log-Normal are the most suitable distributions of ordinary peaks, which is proposed in this study according to the methods explained in Section 2.3 and later discussed in Section 3. The map background indicates the river network and (a) the country elevation (m a.s.l.) and (b) the mean annual rainfall (mm) in Germany.

2.2. The Metastatistical Extreme Value (MEV) distribution

2.2.1. Theoretical framework

In the present work we adopt a recently developed extreme value approach, the Metastatistical Extreme Value (MEV) distribution (Marani and Ignaccolo, 2015; Zorzetto et al., 2016; Marra et al., 2018) for the study of peak flows. This method postulates that extremes emerge from ordinary events (i.e., all the independent values, not annual maxima or peak over threshold values) and treats as random variables both the parameters of the ordinary distributions and the number of event occurrences in each year. This approach therefore leverages all the information content of the bulk of the events distribution, i.e., not only of its tail as done in traditional extreme value distribution methods (Gnedenko, 1943; Coles, 2001; the reader is referred to Serinaldi et al. (2020) for a systematic explanation of the underlying theory and the differences between traditional extreme value theory and MEV). Moreover, the MEV framework allows for flexibility in the choice of the distribution of ordinary values, by enabling the selection of the distribution that describes them best (further details on this point are provided in the following sections).

The discrete formulation of the MEV cumulative distribution function, expressed as an average over the M observational years, is (Zorzetto et al., 2016):

$$\zeta(x) = \frac{1}{M}\sum_{j=1}^{M}[F(x;\boldsymbol{\theta})]^{n_j} \tag{1}$$

where $F(x;\ \theta)$ is the ordinary distribution of peak flows, $\boldsymbol{\theta}$ is a vector of parameters of this distribution (notice that in the above formulation we adopt time invariant parameters of the ordinary distribution, as the subscript j does not appear in $\boldsymbol{\theta}$) and n_j is the number of events within each year (here considered variable across years).

Theoretically, MEV can account for the inter-annual variability of the ordinary distribution as well (for a detailed discussion on the topic we refer to Miniussi and Marani, 2020). However, as the number of independent ordinary events (peaks) available in observed streamflow series is typically limited, the advantages (robust parameter estimation and uncertainty reduction) of considering the whole sample for fitting the parameters of the ordinary distribution overcome those obtained from the use of yearly varying parameters (Miniussi et al., 2020).

In order to identify a set of independent ordinary events from daily streamflow records, which are typically highly correlated, we use the decorrelation procedure recommended by the guidelines of the US Water Resources Council (USWRC, 1976). This is a well-established method which has been adopted in many studies in different regions of the world (see, e.g., Lang et al. (1999) and Miniussi et al. (2020)). At first, we identify one peak within each time block of length $\mathrm{T} = 5\ \mathrm{days} + \log(\mathrm{A})$, where A is the basin area in square miles and the value of T is rounded off to the nearest integer number. This would result for German catchment sizes of one event peak each 6^{th} to 9^{th} day. The second step consists in checking that the magnitude of the minimum flow between two consecutive peaks is lower than 75% of the lowest one, to enable discarding secondary peaks which occur during recession events. If this requirement is not fulfilled, the smallest peak in the pair is eliminated. The set of all peak discharge values obtained from this procedure constitutes the sample of ordinary events, on which the ordinary distribution ($F(x;\boldsymbol{\theta})$ in Eq. (1)) is then fitted. In our dataset, the average number of flood peaks per year resulting from this selection process ranges from 8 to 24 (median value: 17) and their mean inter-arrival time is 23 days, which is much larger than the required minimum lag time of 6 to 9 days for event independence (USWRC, 1976).

2.2.2. Ordinary distributions

Once the ordinary peaks are selected, we must choose a distribution to describe their statistical properties. Miniussi et al. (2020) identified the Gamma distribution as the best performing distribution to model flood peaks in the conterminous United States. Yet, they also suggested that the statistical properties of the ordinary peaks must be investigated before choosing the distribution that describes them best. As mentioned, MEV allows for handling diverse ordinary distributions. Thus, we leverage this flexibility of MEV and examine as potential candidates for the set of German river basins the two parameters Gamma and Log-Normal distributions. Both these distributions are widely used for flood frequency analysis (Stedinger, 1980; Chen et al., 2004) while exhibiting different characteristics especially for what concerns their tail properties, being Gamma a light-tailed distribution (El Adlouni et al., 2008) and the Log-normal akin to a heavy-tailed one (Koutsoyiannis, 2020). Although Gamma and Log-Normal are used in this study, as in a preliminary scrutiny they were found to be the most suitable distributions for the considered dataset, the proposed approach applies to any other choice for the distributions of ordinary events.

The cumulative distribution function of the Gamma distribution is (Davis, 1959; Lancaster, 1966; Hosking, 1990):

$$F(x;\ \alpha, \beta) = \frac{\beta^{-\alpha} \int_0^x t^{\alpha-1}\, e^{\left(-\frac{t}{\beta}\right)} dt}{\Gamma(\alpha)} \qquad (2)$$

where α and β are respectively the shape and scale parameters, and Γ is the gamma function.

The cumulative distribution function of the Log-Normal distribution is instead (McAlister, 1879; Bobée et al., 1993):

$$F(x; \mu, \sigma) = \frac{\Phi(logx - \mu)}{\sigma} \qquad (3)$$

where μ and σ are respectively the scale and shape parameters of the Log-Normal distribution, and Φ is the standard normal distribution function.

We estimate parameters of the Gamma and Log-Normal distributions by fitting them on the sample of ordinary peaks (selected as explained in section 2.2) by means of L-Moments (Greenwood et al., 1979; Hosking, 1990). The choice of L-moments is motivated by their lower sensitivity to outliers and to limited sizes of the calibration samples when compared to Maximum Likelihood (Hosking and Wallis, 1997).

2.3. Tail ratio method and sensitivity analysis

We propose and apply in this study a methodology that enables the choice of the best ordinary distribution to be used in Eq. (1) ex ante, based on the statistical properties of the ordinary peaks. This non-parametric approach avoids the need to evaluate every time the goodness-of-fit of different distributions on randomly extracted samples of ordinary peaks. Rather, it allows for appraising the behavior of the upper tail of the ordinary peaks (a crucial step for the selection of the best ordinary distribution) and thus choose a priori, between Gamma and Log-Normal, the most suitable distribution to characterize them.

The method relies on the calculation of the tail ratio index, which is defined as the ratio between the empirical 99th and 95th percentiles of ordinary peaks, similarly to the procedure suggested by Nerantzaki and Papalexiou (2019) to evaluate the tail behavior of daily precipitation. In the present study, we tested different ratios (99th to 90th, as in Nerantzaki and Papalexiou (2019), 99th to 95th and intermediate ratios) with similar results. In the end, we found that in the case of flood peaks considering the ratio between the 99th and 95th percentile allows to describe best the upper tail of the distribution in our dataset. The method is inspired by the upper tail ratio of Smith et al. (2018), although pursuing a different aim and with clear advantages when compared to it. The upper tail ratio is defined as the ratio between the maximum event in the record and the empirical 10-years return level. It is computed from annual maxima and is therefore sensitive to the observational sample size. Conversely, our proposed method (the tail ratio index) is calculated accounting for the whole distribution of ordinary peaks. First, its definition is consistent with the rationale underlying the MEV framework, which considers extremes emerging from ordinary events. Second, the whole set of ordinary peaks (a median number of 1184 in the considered catchments) is used in the tail ratio index instead of annual maxima, which results in more robust calculations.

Therefore, we believe that the proposed method is transferrable to other areas, with potential limitations in catchments characterized by drier conditions, in which the low number of ordinary peaks might impair the robustness of its computation.

The steps involved in the establishment and evaluation of the proposed tail ratio method are the following: 1) we select the best ordinary distribution for each station evaluating the goodness-of-fit of Gamma and Log-Normal distributions on the complete series of ordinary peaks, by means of the skill score metric (Murphy and Winkler, 1992; Hashino et al., 2006); 2) we calculate the tail ratio index of each series and seek a threshold to distinguish relatively lighter and heavier tailed distributions. We choose to this purpose the threshold of the tail ratio index that allows maximizing the number of catchments for which the optimal (i.e., the one with the highest skill score value) distribution is selected; 3) in order to evaluate the robustness of the threshold identified on the whole German dataset and the possibility to apply it to different datasets (i.e., subsamples of the original one), we randomly select for 100 times 50% of the basins and calculate again the threshold value obtained from each resampling; 4) we check the capability of this threshold to correctly distinguish between catchments for which either the Gamma or the Log-Normal distributions are optimal (as in step 2), by using the remaining 50% of the basins in the dataset as test catchments.

2.4. Cross-validation and benchmarking

In order to evaluate the predicting capabilities of the tested flood estimation approaches, we perform a Monte-Carlo experiment through which we extract (1000 times) from the available data record 10 random years of observations, which we use as calibration sample. The remaining data are each time used as validation sample. By means of this procedure, any potential systematic variability in the time series is disregarded (Vogel et al., 1993; Haktanir et al., 2013; Zorzetto et al., 2016) and the estimation uncertainty can be evaluated.

The length of the calibration sample was chosen as to have enough ordinary events to enable a robust estimation of the ordinary distributions parameters and at the same time to allow for the assessment of the estimation uncertainty of the different approaches. In fact, being the minimum length of the analyzed streamflow records equals to 30 years, with a calibration sample longer than 10 (e.g., 20) years we could not evaluate the estimation accuracy in a predictive fashion, as no observed maximum would be associated with an empirical return period greater than 20 years in the validation sample. Moreover, we stress that we aim at quantifying errors with respect to the empirical values, as the performance (e.g., in terms of accuracy improvement) of an extreme value approach cannot be evaluated when quantiles with return periods higher than the empirical ones are taken into account. Furthermore, assessing the accuracy of extreme value methods constrained on short calibration samples is pivotal for, e.g., data scarce areas, where only short time series are available.

We also benchmark MEV estimations with the traditional GEV approach, which constitutes the suggested distribution for flood frequency analysis in Germany (Petrow et al., 2007). GEV parameters are estimated by fitting the distribution through L-moments (Greenwood et al., 1979; Hosking, 1990) on yearly maxima in the calibration sample. In order to evaluate the estimation accuracy of both MEV and GEV, we use the non-dimensional error (Zorzetto et al., 2016), which quantifies the performance of different models. For each random extraction in the Monte-Carlo experiment, we calculate the non-dimensional error $\epsilon = \frac{x_{est} - x_{obs}}{x_{obs}}$ between estimated and observed maxima in the validation sample which have a return period longer than the calibration sample size. We finally pool together the non-dimensional error values obtained for different ranges of the ratio T/S between return period T and calibration sample size S.

3. Results and Discussion

We first apply the MEV approach by using either a Gamma or a Log-Normal distribution for each of the 182 German gauges. We then evaluate the capabilities of MEV-Gamma (Figure 2a) and MEV-Log-Normal (Figure 2b) to estimate observed maxima in the validation sample. We find that high quantiles are frequently underestimated by MEV-Gamma (see blue dots below the 45 degree line in Figure 2a). Although this issue is addressed by using MEV-Log-Normal, overestimation prevails in the latter case.

The contrasting behavior of the two distributions suggests that the ordinary peaks are in some catchments characterized by lighter tails, for which the Gamma is a suitable distribution and the Log-Normal has a too heavy tail, and vice versa in other basins. These results are consistent with previous findings highlighting that flood records may exhibit both light and heavy-tail behaviors (Katz et al., 2002; Bernardara et al., 2008). An example of this instance is provided in Figure 3. The blue lines in Figure 3a show that using a Gamma distribution to fit the empirical distribution of ordinary peaks (grey dots) would lead to an under-estimation of the exceedance probability of the largest flows. This results into a significant underestimation of the largest annual maxima as shown in Figure 3c. The underestimation issues of the Gamma distribution that we highlight here are in line with previous findings (Papalexiou et al., 2013). The flood frequency curve is instead correctly estimated by using MEV-Log-Normal (Figure 3d) because the underlying Log-Normal distribution is capable to correctly represent the empirical exceedance probability of the largest ordinary flows, as represented in Figure 3b.

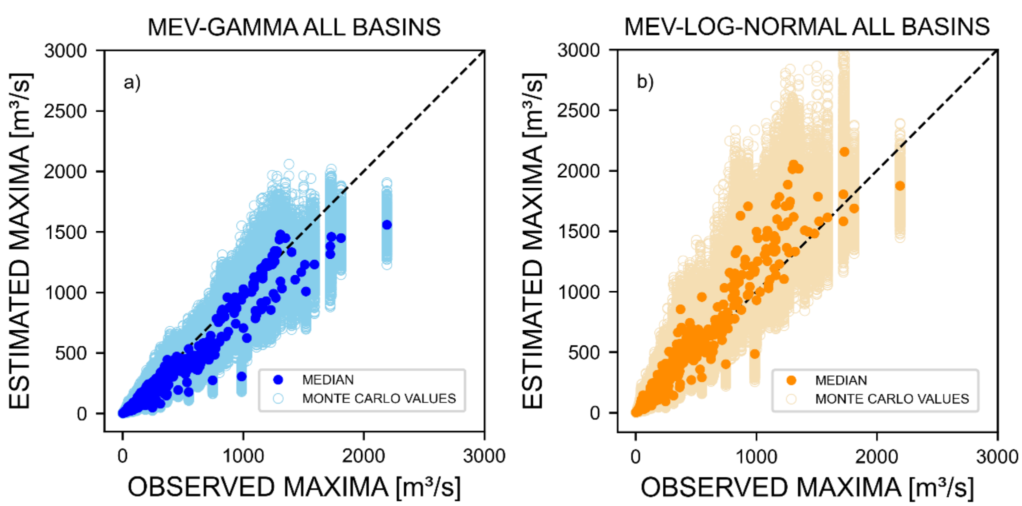

Figure 2 - Quantile-quantile plots of streamflow maxima with return period longer than the length of the calibration sample (10 years) for MEV-Gamma (blue dots in panel a) and MEV-Log-Normal (orange dots in panel b). Observed annual maxima and

return periods are empirically estimated by using the validation samples. Light and dark colors indicate respectively the 1000 resamples of the Monte-Carlo approach and their median values, for all basins.

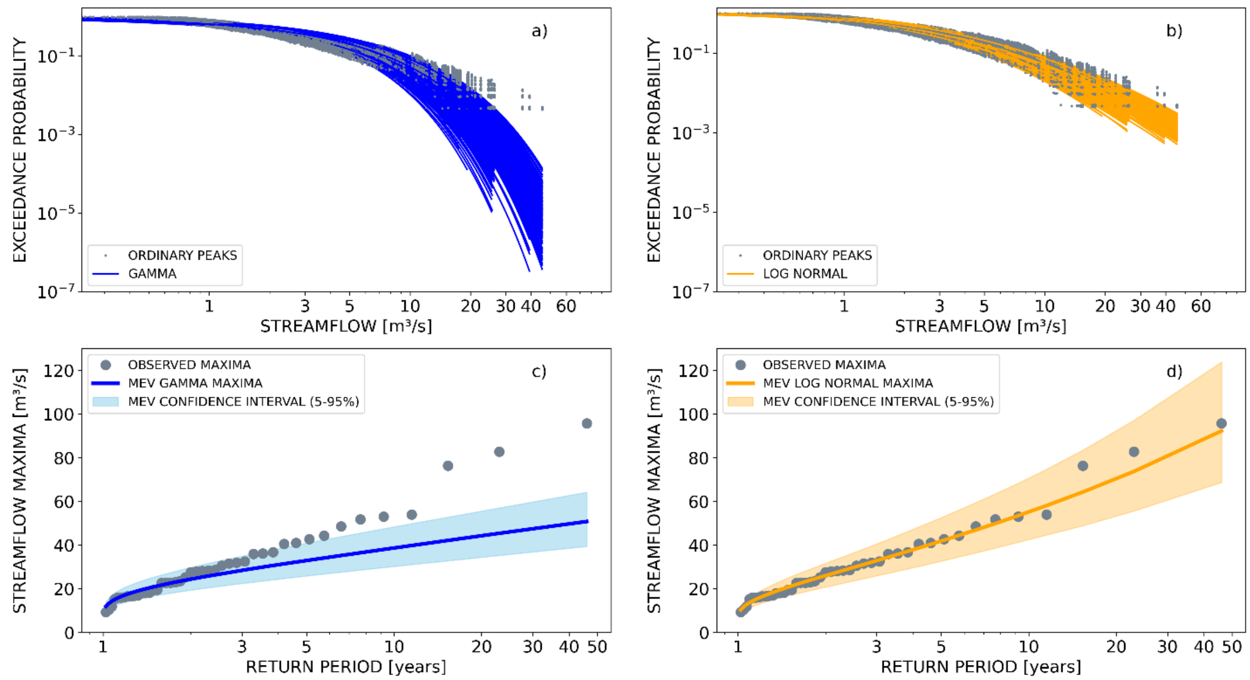

Figure 3 - (a-b) Exceedance probability of ordinary peaks plotted as a function of streamflow values (in double logarithmic scale) for an exemplary case study indicated with a triangle in Figure 1 (the Neckar River at Rockenau, Area = 70 km^2). Gamma (panel a, blue solid lines) and Log-Normal (panel b, orange solid lines) probability distributions are fitted on observed ordinary peaks obtained from 1000 resamples of 10 years long calibration samples (grey dots). (c-d) Resulting flood frequency curves for the same catchment: grey dots represent the median values of the 1000 validation samples obtained from the cross-validation procedure, whose empirical frequencies (and thus return periods, defined as the inverse of the exceedance probability) have been estimated by means of Weibull plotting position; blue (panel c) and orange (panel d) lines show the median MEV estimates for the corresponding quantiles by respectively using Gamma and Log-Normal distributions to fit the ordinary peaks. Matching shaded areas indicate 90% confidence intervals (5th-95th percentiles).

This explanatory case highlights how the choice of a suboptimal distribution of ordinary peaks largely affects the estimation of maxima, as also signaled by previous studies (Cunnane, 1985; Haddad and Rahman, 2011; Hu et al., 2020) which indicated the importance of selection of distribution for the estimation of high quantiles. In fact, when MEV-Gamma is applied, maxima are generally underestimated to a more or less large extent, while the application of the Log-Normal distribution allows for an accurate appraisal of high flood magnitudes. Hence the need of understanding, ideally a priori, what is the best distribution that should be used to describe the tail behavior of ordinary peaks an issue highlighted in past studies (El Adlouni et al., 2008).

To reach this goal, we calculate the tail ratio index introduced in Section 2.3 to evaluate the relative tail behavior of the distribution of ordinary peaks ex ante and independently from the MEV procedure.

Figure 4 visualizes the results of the procedure detailed in Section 2.3. The blue line shows the fraction of cases for which the tail ratio index of the catchments in which the Gamma distribution is identified as the best fitting one is lower than the corresponding value of the tail ratio index reported on the x-axis. For increasing values of the tail ratio index we are able to correctly match the best ordinary distribution for an increasing fraction of cases for the Gamma distribution. A symmetric and opposite comment can be done in the case of catchments where the Log-Normal distribution is the most suitable one (orange line in Figure 4). Ordinary peaks for which the Log-Normal distribution is the best fit tend indeed to have higher values of tail ratio index, and consequently the lower the threshold the most likely it is to correctly classify them. It is important to note that the fraction of cases represented by the y-axis in Figure 4 refers to the number of catchments for which either distribution is the best, not to the whole dataset.

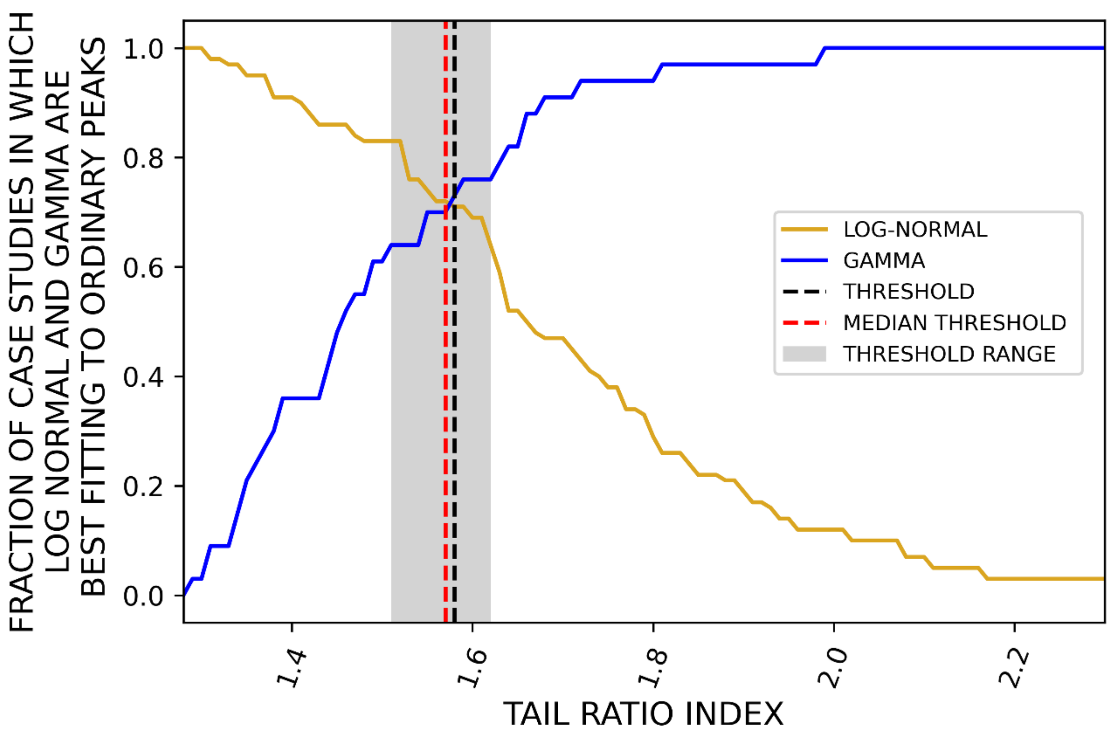

Figure 4- Identification of the threshold of the tail ratio index optimizing the number of catchments for which the most suitable distribution of ordinary peaks is selected ex ante based on the value of the tail ratio index itself. The most suitable distribution of ordinary peaks is identified on the basis of the skill score of the fitted ordinary distributions. The blue line (for the Gamma distribution) and the orange line (in the case of the Log-Normal distribution) show the fraction of cases for which the Gamma or Log-Normal is the most suitable distribution. The crossing point of these two curves represents the tail ratio index value that optimizes the correct assignment of catchments to two groups for which Gamma and Log-Normal are respectively the most suitable distributions of ordinary peaks (black dashed line, named threshold in the following).To evaluate the robustness of the identified threshold and its uncertainty when applied to different sets of catchments, we recalculate it by randomly selecting 50% of the catchments from the whole dataset . We repeat the random selection 100 times and each time validate the threshold capability to correctly identify the most suitable distribution for the remaining 50% of catchments. The resulting range of thresholds over 100 random extractions and their median value are respectively displayed with a grey shaded area and a red dashed line.

The crossing point of these two curves, identified with a black dashed line in Figure 4, represents the threshold on the tail ratio index which allows for maximizing the number of catchments where the correct identification of the best fitting distribution of ordinary peaks can be obtained ex ante by means of the tail ratio index. The identified threshold, equals to 1.58 for the whole set of German basins, allows for correctly identifying 74% of the catchments in each distribution group. Approximately 30% of the catchments for which the correct distribution of ordinary peaks is not identified based on the tail ratio index (i.e., 14 out of 48) display values of the tail ratio index close to the threshold (within the shaded region of Figure 4). For the remaining cases, the distribution of ordinary peaks indicates the possibility of a heavier tail than the one exhibited by most of the flood frequency curves generated in the Monte Carlo procedure.

In order to check the robustness of the selected threshold, with the aim to apply it to different subsets of river basins, we follow the procedure detailed in Section 2.3.

The thresholds obtained with this procedure span a narrow range between 1.51 and 1.62 (grey shaded area in Figure 4), with a median value of 1.57 (red dashed line in Figure 4). The identified threshold exhibits weak dependence on the subset of selected basins, proving its capability to correctly distinguish ex ante between basins where either the Gamma or the Log-Normal distributions best fit the ordinary peaks. These results show that the identified threshold is robust, appropriate for German catchments and can thus be applied to different datasets other than the one used in this study.

Following up on the previous analyses, we hence split the whole set of case studies into two groups. Group 1 (76 basins; blue dots in Figure 1) includes catchments with a tail ratio index lower than 1.58, for which we apply the MEV-Gamma, whereas Group 2 (106 basins; yellow dots in Figure 1) comprises catchments with values of the tail ratio index greater than 1.58, for which the MEV-Log-Normal is used. The performance of the MEV approach after this a priori selection of the most suitable distribution of ordinary peaks by means of the tail ratio index is illustrated in Figure 5a,b. Quantile-Quantile (QQ) plots in Figure 5 show an improvement of the performance of MEV when the distribution of ordinary peaks is chosen by considering the tail features of the underlying data versus neglecting them (i.e., the performance increases between Figure 2a,b and Figure 5a,b). These results agree to those of Papalexiou et al. (2013), who emphasized the importance of the upper part (i.e., the tail) of the probability distribution. In particular, the under and overestimation issues resulting from the application of MEV-Gamma and MEV-Log-Normal to the whole set of German case studies (Figure 2) largely decrease compared to the two classified groups. We observe that errors reduce of a median value of 57% and 58% across all quantiles and 56% and 40% for high quantiles (i.e., with ratio between the return period T and the calibration sample size S greater than 3) when these distributions are separately applied to Group 1 and 2. This indicates that the preliminary selection of catchments (exhibiting relatively lighter and heavier tails) based on the tail ratio index is effective and allows for improving the estimation of floods in the MEV framework.

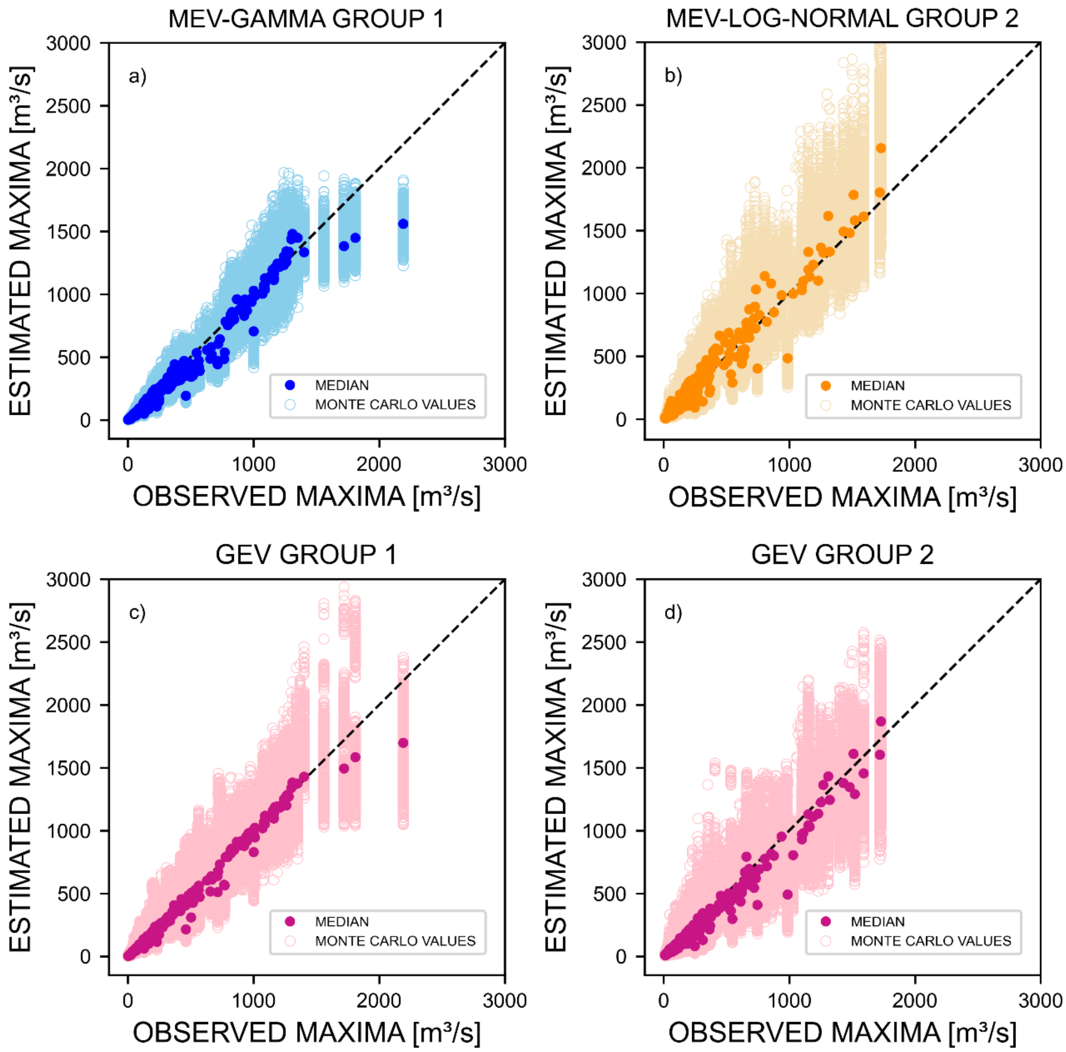

Figure 5 – Quantile-quantile plots of flood magnitudes estimated for two subsets of river basins that have been separated ex ante based on the tail ratio index of the ordinary peaks. When the tail ratio index is lower than 1.58 we apply the MEV-Gamma (panel a, blue and light-blue dots), whereas when the tail ratio index is greater than this threshold we use the MEV-Log Nor8mal (panel b, orange and light orange dots). Results are compared with GEV estimates in the two groups (panels c and d, dark pink and light pink dots). The median results among all the Monte-Carlo realizations for quantiles corresponding to return periods longer than the calibration samples (10 years) are shown in darker color dots, while all the resampling values (1000 for each catchment) are shown with lighter colors.

We also compare these results with estimates obtained by means of the traditional GEV distribution (Figure 5c,d). Despite acknowledging the limitations of fitting a three-parameter distribution on a relatively small sample, this situation is not uncommon in hydrological practice (e.g., Kobierska et al., 2018). The comparison between Figure 5a,b and Figure 5c,d shows that the preliminary selection of the distribution of ordinary peaks in the MEV approach enables improved MEV versus GEV estimates, especially in Group 1. Here, MEV-Gamma guarantees an average performance similar to GEV (dark blue and pink dots in Figure 5a,c) and a remarkable decrease of the uncertainty of the estimate (light blue and pink dots in Figure 5a,c). These results are supported by previous findings (Basso et al., 2021), who showed that MEV exhibits lower uncertainty than GEV. A similar result is obtained for Group 2, although in this case the uncertainty of MEV-Log-Normal is comparable to that of GEV, especially for the largest flood values. However, GEV is affected by larger uncertainty than MEV-Log-Normal in cases where the magnitude of streamflow maxima for a same return period is smaller, as highlighted by the following analysis.

Table 1. Median (Med) and maximum negative (i.e., underestimation, Max^-) and positive (i.e., overestimation, Max^+) values of non-dimensional error for all distributions and catchment groups and for different ranges of the ratio (T/S) between return period (T) and calibration sample size (S).

	MEV-Gamma						MEV-Log-Normal						GEV					
	All Basins			Group 1			All Basins			Group 2			Group 1			Group 2		
	Max^-	Max^+	Med	Max^-	Max^+	Med	Max^-	Max^+	Med	Max^-	Max^+	Med	Max^-	Max^+	Med	Max^-	Max^+	Med
1< T/S<=2	-0.81	0.90	-0.2	-0.61	0.66	-0.08	-0.78	2.28	0.08	-0.78	2.28	0.02	-0.6	1.39	-0.04	-0.78	3.12	-0.07
2< T/S<=3	-0.81	0.96	-0.23	-0.64	0.71	-0.1	-0.77	2.71	0.12	-0.77	2.71	0.06	-0.63	1.63	-0.05	-0.82	3.88	-0.09
3< T/S<=6	-0.90	1.17	-0.28	-0.73	0.86	-0.12	-0.82	3.93	0.15	-0.82	3.91	0.09	-0.78	2.12	-0.08	-0.91	5.86	-0.12

In order to provide a more complete overview of the estimation accuracy of the three extreme value distributions we complement the QQ-plots in Figure 5a-d with Figure 6, which shows the non-dimensional error between observed and estimated quantiles for all random extractions in the Monte Carlo procedure and all the gauges (Zorzetto et al., 2016). These results are also summarized in Table 1. In both cases, we pool the non-dimensional error values in different ranges of the ratio T/S between return period and calibration sample, each of which includes at least 10,000 values.

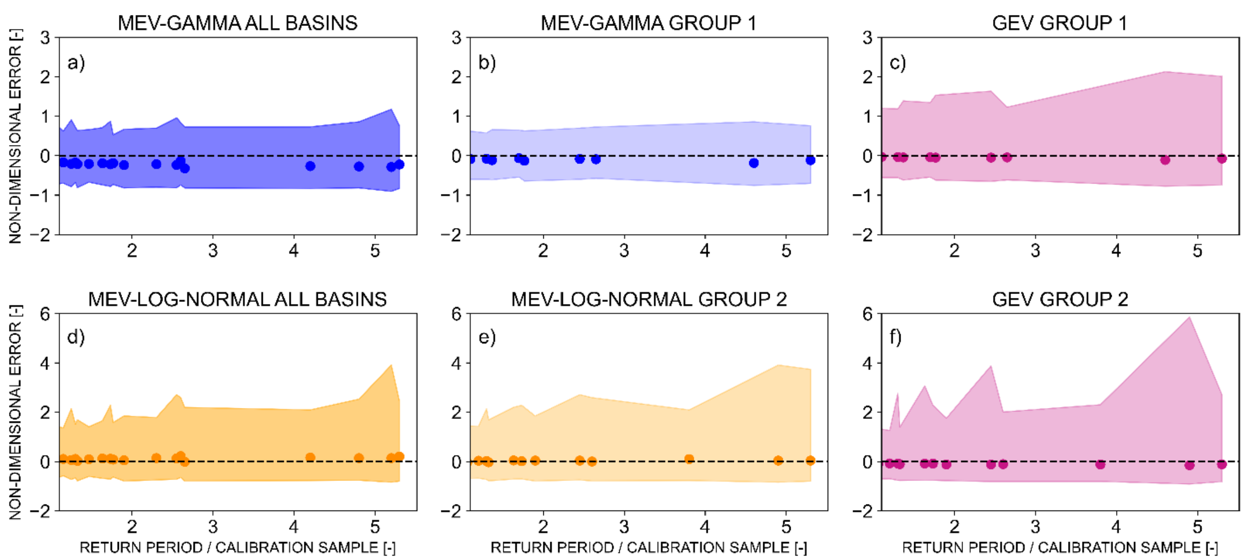

Figure 6 – Non-dimensional error between observations and quantiles of the analyzed statistical distributions computed in the Monte Carlo simulation procedure (1000 realizations), plotted as a function of different ranges of the ratio between the return period and the calibration sample size (T/S). Results for quantiles corresponding to return periods longer than the calibration sample size are displayed in order to evaluate the performance in a predictive way. The figures are plotted by pooling values of non-dimensional error from all stations which are included in specified ranges (bins) of T/S. Dots represent the median values of

T/S bins that include at least 10,000 values; shaded regions encompass the minimum and maximum values of each respective bin. Panels (a, d) show the non-dimensional error of MEV-Gamma and MEV-Log-Normal applied to the whole dataset regardless of the tail features of the distributions. Panels (b,e) show the non-dimensional error of MEV-Gamma and MEV-Log-Normal respectively applied to Group 1 and Group 2, and panels (c,f) show the non-dimensional error of GEV applied to the same groups.

MEV-Gamma (Figure 6a) and MEV-Log-Normal (Figure 6d) respectively under and overestimates when applied to the whole dataset. Figure 6a once again highlights that MEV-Gamma, if applied without accounting for the statistical properties of the ordinary peaks, is affected by underestimation (notice that dark blue dots in Figure 6a tend to stay under the dashed black line, which represents the unbiased result with a relative error equal to zero). These results are in line with the findings of Papalexiou et al. (2013). When comparing the relative error of MEV-Gamma (Figure 6b) and GEV (Figure 6c) for the basins in group 1 (or, likewise, the relative error of MEV-Log-Normal (Figure 6e) and GEV (Figure 6f) for the basins in group 2), we highlight that despite similar performances of MEV and GEV in terms of median values, the latter is affected by much larger uncertainty. These results agree with the findings of, e.g., Odry and Arnaud (2017), who stressed the importance of a reduced uncertainty when estimating high flow values. Notably, in group 1 the overestimation magnitude of MEV-Gamma is the half of GEV across all T/S (Figure 6b-c). Differences between the relative performance of MEV Log-Normal and GEV are less pronounced in group 2, but substantially larger overestimation (i.e., non-dimensional error reaching a value of 6) can occur when employing GEV.

We finally present the comparison between the non-dimensional errors of MEV applied on the whole dataset (i.e., regardless of the observed tail of the distribution of ordinary peaks; Figure 6a,d) and of both MEV-Gamma (Figure 6b) and MEV-Log-Normal (Figure 6e) applied to their respective groups. In this way, we want to underline the benefits deriving from the flexibility of MEV in the choice of a suitable distribution of ordinary peaks when the catchments within a dataset exhibit different tail properties. Remarkably, we are able to significantly decrease both the underestimation by which MEV-Gamma is affected when blindly applied to the whole dataset (blue dots in Figure 6b are closer to the dashed line than in Figure 6a) and its uncertainty range (blue shaded area is narrower in Figure 6b than in Figure 6a), and to slightly reduce on average the overestimation issues of MEV-Log-Normal (Figure 6e). These results are highly relevant, as both underestimation and overestimation issues are drawbacks when estimating high return levels of hydrological variables: in the first case because of the underestimation of risks (Papalexiou et al., 2013), in the latter because of the enormous economic costs that can derive from applying design values which are larger than needed (Cho et al., 2004).

Finally, the subdivision into two groups of the set of German catchments by means of the tail ratio index, here proposed to select the most suitable distribution of ordinary events in the MEV approach, exhibits a clear geographical organization (Figure 1). Catchments for which a Gamma distribution (lighter tail) is the most suitable are primarily located in the Western part of Germany (Figure 1, blue dots), where winter floods triggered by precipitation on wet soils are dominant (Tarasova et al., 2020), with some catchments also gathered in the Alpine Forelands. Catchments for which a Log-Normal distribution (which has a relatively heavier tail) is the best choice are instead mostly located in the Eastern and Southern areas of the country (Figure 1, orange dots), regions affected by the occurrence of Vb-cyclones (Hofstätter et al., 2016) causing rare but intense rainfall on dry soils which produces large floods. The geographical clustering we identified recalls the spatial pattern recently showed by Tarasova et al. (2020) for the differences among processes triggering ordinary, frequent and upper tail floods in Germany. Our results suggest that their postulated outcomes in terms of distinct distributions of ordinary events is well-founded, and confirm the possibility to improve estimation of upper tail floods in this context by means of a Metastatistical Extreme Value approach. Although beyond the scope of this work, a quantitative investigation of the consistency between the identified spatial patterns and of the mix of hydrological processes giving rise to distributions of floods and ordinary peaks with relatively lighter (Gamma) and heavier (Log-Normal) tails in specific areas of Germany is a noteworthy research direction which is the subject of current work.

4. Conclusion:

In this study we adopted the Metastatistical Extreme Value (MEV) distribution for flood frequency analysis in a set of 182 catchments in Germany. Our goal was to optimize this novel framework based on the statistical properties of the ordinary peaks.

To this end, we developed a non-parametric approach to select ex ante the ordinary distribution of streamflow peaks, which allows for accurately estimating high flow quantiles. The proposed method (tail ratio index) makes a step forward in the evaluation of the tail behavior. In particular, it leverages the information content of ordinary events and avoids any graphical evaluation, hence allowing for the analysis of large datasets. It enables a binary classification of ordinary distributions characterized by lighter versus heavier tails. We identified a threshold that discriminates between these two categories and used it to choose ex ante if either a Gamma or a Log-Normal distribution is the most suitable for describing the ordinary events. Our approach correctly identifies the ordinary distribution in 74% of the basins in the dataset. We proved that the proposed value of the threshold is robust to random resampling of the catchments used to determine it, and can thus be reliably employed. We deem the proposed tail ratio index to be applicable to select the most suitable distribution of ordinary peaks in other regions of the world.

Selecting the distribution of ordinary peaks by means of this approach allows for reducing underestimation and overestimation issues compared to a blind (i.e., without investigating the tail features of the ordinary peaks) application of MEV to the whole dataset. Namely, we reduce MEV-Gamma underestimation and Log-Normal overestimation issues of a median 57% and 58% respectively. Finally, we benchmarked our results against the standard distribution used in flood frequency analysis, i.e., the Generalized Extreme Value (GEV) distribution. In 135 out of the 182 analyzed basins, the use of a tailored ordinary distribution in the MEV framework decreases the uncertainty and improves the estimation for return periods greater than the calibration sample. These are both relevant features to achieve a reliable estimation of extreme floods.

By proposing and verifying an easy-to-use method to leverage knowledge on tail properties, the study supports selection of the appropriate model to analyze ordinary peaks and floods in the MEV framework. It thus contributes to reduce under and overestimation of rarer floods, improving our capability to reliably assess the hazard posed by infrequent hydrological events.

Acknowledgement and Data:

This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project number 421396820 “Propensity of rivers to extreme floods: climate-landscape controls and early detection (PREDICTED)" and Research Group FOR 2416 “Space-Time Dynamics of Extreme Floods (SPATE)”. The financial support of the Helmholtz Centre for Environmental Research - UFZ is as well acknowledged. We thank the Bavarian State Office of Environment (LfU, https://www.gkd.bayern.de/de/fluesse/abfluss) and the Global Runoff Data Centre (GRDC) prepared by the Federal Institute for Hydrology (BfG, http://www.bafg.de/GRDC) for providing the discharge data for Germany, which are available in the linked repositories. We thank Larisa Tarasova for useful discussions and advices. The first author is also thankful to the Higher Education Commission of Pakistan (HEC) and the German Academic Exchange services (DAAD) for providing financial support to this study (DAAD- 91716446).

References

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

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

Bernardara, P., Schertzer, 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), s. 107–122. https://doi:10.1007/s00477-006-0101-2

Beurton, S., & Thieken, A. H. (2009). Seasonality of floods in German. *Hydrological Sciences Journal,54* (1), 62-76. http://doi.org/10.1623/hysj.54.1.62

Bobée, B., Cavadias, G., Ashkar, F., Bernier, J., & Rasmussen, P. (1993). Towards a systematic approach to comparing distributions used in flood frequency analysis. *Journal of Hydrology*, *142*(1-4), 121-136. http://dx.doi.org/10.1016/0022-1694(93)90008-w

Chen, Y., Xu, S., Sha, Z., Gelder, P., & Gu, S. (2004). Study on L-moment estimations for log normal distribution with historical flood data. In: Y.B. Chen, K. Takara, I.D. Cluckie & F.H. De Smedt, eds. *GIS and RS in hydrology, water resources and environment*, Vol. **1**. Wallingford, U.K.: IAHS Press, pp. 107-113.

Cho, H.-K., Bowman, K. P., & North, G. R. (2004). A Comparison of Gamma and Lognormal Distributions for Characterizing Satellite Rain Rates from the Tropical Rainfall Measuring Mission. *Journal of Applied Meteorology*, *43*(11), 1586–1597. https://doi.org/10.1175/jam2165.1

Cook, B. I., Mann, M. E., D'Odorico, P., & Smith, T. M. (2004). Statistical simulation of the influence of the NAO on European winter surface temperatures: Applications to phenological modeling. *Journal of Geophysical Research: Atmospheres*, *109*(D16). https://doi:10.1029/2003jd004305

Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. Springer Series in Statistics, 1st ed. Springer, New York. https://doi.org/10.1007%2F978-1-4471-3675-0

Cooke, R. M. & Nieboer, D. (2011). Heavy-Tailed Distributions: Data, Diagnostics, and New Develop ments. *SSRN electronic journal*. https://doi:10.2139/ssrn.1811043

Cunnane, C. (1973). A particular comparison of annual maxima and partial duration series methods of flood frequency prediction. *Journal of Hydrology*, *18* (3–4), 257–271. https://doi.org/10.1016/0022-694(73)1694(73)90051-6

Cunnane, C. (1985). Factors affecting choice of distribution for flood series. *Hydrological sciences journal*, *30* (1), s. 25–36. https://doi:10.1080/02626668509490969

Davis, P.J. (1959). Leonhard Euler's Integral: A Historical Profile of the Gamma Function: In Memoriam: Milton Abramowitz. *The American Mathematical Monthly 66*, 849. http://doi.org/10.2307/2309786

Davison, A. C., & Smith, R. L. (1990). Models for Exceedances Over High Thresholds. *Journal of the Royal Statistical Society*: *Series B (Methodological)*, *52*(3), 393–425. http://doi.org/10.1111/j.2517-6161.1990.tb01796.x

Doocy, S., Daniels, A., Murray, S., & Kirsch, T. D. (2013). The Human Impact of Floods: a Historical Review of Events 1980-2009 and Systematic Literature Review. PLoS Currents. https://doi:10.1371/currents.dis.f4deb457904936b07c09daa98ee8171a

Eash, D. A. (1997). Effects of the 1993 flood on the determination of flood magnitude and frequency in Iowa. *Circular*. http://dx.doi.org/10.3133/cir1120k

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

Eliazar, I. I., & Sokolov, I. M. (2010). Gini characterization of extreme-value statistics. *Physica A: Statistical Mechanics and its Applications*, *389*(21), 4462-4472. https://doi.org /10.1016 /j.physa.2010.07.005

Fekete, A. & Sandholz, S. (2021). Here Comes the Flood, but Not Failure? Lessons to Learn after the Heavy Rain and Pluvial Floods in Germany 2021. *Water*, *13* (21), s. 3016.doi:10.3390/w13213016

Germany floods: Up to 30 billion euros needed for recovery fund (2021, August 9), DW News. Retrieved

from http://dw.com
Gnedenko, B. (1943). Sur La Distribution Limite Du Terme Maximum D'Une Serie 788 Aleatoire. *Annals of Mathematics*, *44* (3), 423-453. http://doi.org/10.2307/1968974
Greenwood, J. A., Landwehr, J. M., Matalas, N. C. and Wallis, J. R. (1979). Probability weighted Moments: definition and relation to parameters of several distributions expressable in inverse form. *Water Resources Research*, *15*, 1049-1054. https://doi.org/10.1029/WR015i005p01049
Haddad, K. & Rahman, A. (2011). Selection of the best fit flood frequency distribution and parameter estimation procedure: a case study for Tasmania in Australia. *Stochastic environmental research and risk assessment*, *25* (3), s. 415–428. https://doi:10.1007/s00477-010-0412-1
Haktanir, T., Cobaner, M., & Gorkemli, B. (2013). Assessment of Right-Tail Prediction Ability of Some Distributions by Monte Carlo Analyses. *Journal of Hydrologic Engineering, 18*(5), 499-517. https://doi.org/10.1061/(asce)he.1943-5584.0000620
Hashino, T., Bradley, A. A., & Schwartz, S. S. (2006). Evaluation of bias-correction methods for ensemble streamflow volume forecasts. *Copernicus GmbH*. https://doi.org/10.5194/hessd-3-561-2006
Hofstätter, M., Chimani, B., Lexer, A., & Blöschl, G. (2016). A new classification scheme of European cyclone tracks with relevance to precipitation. *Water Resources Research*, *52* (9), 7086-7104. https://doi.org/10.1002/2016WR019146
Hosking, J. R. M. (1990). L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society*: *Series B (Methodological)*, *52*(1), 105-124. https://doi.org/10.1111/j.2517-6161.1990.tb01775.x
Hosking, J. R. M & Wallis, J. R. (1997). Regional Frequency Analysis. Cambridge University Press. https://doi.org/10.1017%2Fcbo9780511529443
Hu, L., Nikolopoulos, E. I., Marra, F., & Anagnostou, E. N. (2020). Sensitivity of flood frequency analysis to data record, statistical model, and parameter estimation methods: An evaluation over the contiguous United States. *Journal of Flood Risk Management, 13*(1). https://doi.org/10.1111/jfr3.12580
Iliopoulou, T., and Koutsoyiannis, D., 2019. Revealing hidden persistence in maximum rainfall records. *Hydrological Sciences Journal*, *64* (14), 1673–1689.https://doi.org/10.1080/02626667.2019.1657578
Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources, 25*(8), 1287–1304. https://doi.org/10.1016/S0309-1708(02)00056-8
Klemes, V. (1974). Some problems in pure and applied stochastic hydrology. In *Proceedings of the symposium on statistical hydrology, miscellaneous publication* (No. 1275, pp. 2-15).
Kobierska, F., Engeland, K. & Thorarinsdottir, T. (2018). Evaluation of design flood estimates – a case study for Norway. *Hydrology research*, *49* (2), s. 450–465. https://doi:10.2166/nh.2017.068
Koutsoyiannis, D. (2020). Stochastics of Hydroclimatic Extremes—A Cool Look at Risk. National Technical University of Athens: Athens, Greece. http://hdl.handle.net/11419/6522
Lancaster, H. O. (1966) Forerunners of the Pearson chi squared. *Australian Journal of Statistics*, *8*, 117-126. https://doi.org/10.1111%2Fj.1467-842x.1966.tb00262.x
Lang, M., Ouarda, T.B.M.J., Bobée, B. (1999). Towards operational guidelines for over-threshold modeling. *Journal of Hydrology 225*(3-4), 103–11. https://doi.org/10.1016%2Fs0022-1694%2899%2900167-5
Lombardo, F., Napolitano, F., Russo, F., & Koutsoyiannis, D. (2019). On the Exact Distribution of Correlated Extremes in Hydrology. *Water Resources Research, 55*(12), 10405–10423. https://doi.org/10.1029/2019wr025547
Macdonald, N., Werritty, A., Black, A. R., & McEwen, L. J. (2006). Historical and pooled flood frequeny analysis for the River Tay at Perth, Scotland. *Area*, *38*(1), 34–46. https://doi.org/10.1111/j.1475-4762.2006.00673.x
Marani, M., & Ignaccolo, M. (2015). A metastatistical approach to rainfall extremes. *Advances in Water Resources, 79*, 121-126. https://doi.org/10.1016/j.advwatres.2015.03.001
Marra, F., Nikolopoulos, E. I., Anagnostou, E. N., & Morin, E. (2018). Metastatistical Extreme Value analysis of hourly rainfall from short records: Estimation of high quantiles and impact of measurement errors. *Advances in Water Resources*, 117, 27–39. https://doi:10.1016/j.advwatres.2018.05.001
McAlister, D. (1879). The law of the geometric mean. *Proceedings of the Royal Society of London*, *29*, 367

–376. https://doi.org/10.1098/rspl.1879.0061
Merz, B., Elmer, F., Kunz, M., Mühr, B., Schröter, K., & Uhlemann-Elmer, S. (2014). The extreme flood in June 2013 in Germany. *La Houille Blanche*, *100*(1), 5–10. https://doi.org/10.1051/lhb/2014001
Miniussi, A., Marani, M., & Villarini, G. (2020). Metastatistical Extreme Value Distribution applied to floods across the continental United States. *Advances in Water Resources, 136*. https://doi.org/10.1016/j.advwatres.2019.103498
Miniussi, A., & Marani, M. (2020). Estimation of daily rainfall extremes through the metastatistical extreme value distribution: Uncertainty minimization and implications for trend detection. *Water Resources Research*, *56*(7). https://doi.org/10.1029%2F2019wr026535
Morrison, J. E., & Smith, J. A. (2002). Stochastic modeling of flood peaks using the generalized extreme value distribution. *Water Resources Research, 38*(12), 41-1-41-12. https://doi.org/10.1029/2001wr000502
Murphy, A. H., & Winkler, R. L. (1992). Diagnostic verification of probability forecasts. *International Journal of Forecasting*, *7*(4), 435–455. https://doi.org/10.1016/0169-2070(92)90028-8
Müller, M. F., & Thompson, S. E. (2016). Comparing statistical and process-based flow duration curve models in ungauged basins and changing rain regimes. *Hydrology and Earth System Sciences, 20*(2), 669-683. https://doi.org/10.5194/hess-20-669-2016
Nerantzaki, S. D., & Papalexiou, S. M. (2019). Tails of extremes: Advancing a graphical method and harnessing big data to assess precipitation extremes. *Advances in Water Resources, 134*. https://doi.org/10.1016/j.advwatres.2019.103448
Odry, J., & Arnaud, P. (2017). Comparison of Flood Frequency Analysis Methods for Ungauged Catchments in France. *Geosciences, 7*(3). https://doi.org/10.3390/geosciences7030088
Panorska, A.K., Gershunov, A., Kozubowski, T.J., 2007. From Diversity to Volatility: Probability of Daily Precipitation Extremes, in: Nonlinear Dynamics in Geosciences. Springer New York, New York, NY, pp. 465–484. https://doi.org/10.1007/978-0-387-34918-3_26
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 11*, 1455–1468. https://doi.org/10.5194/hess-11-1455-2007
Renard, B., Kochanek, K., Lang, M., Garavaglia, F., Paquet, E., Neppel, L., et al. (2013). Data-based comparison of frequency analysis methods: A general framework. *Water Resources Research*, *49*(2), 825-843. https://doi.org/10.1002/wrcr.20087
Rosbjerg, D., et al. (2013). Prediction of floods in ungauged basins, in Runoff Prediction in Ungauged Basins. A Synthesis Across Processes, Places and Scales, edited by G. Bloschl, et al., chapter. 9, pp. 189–226, Cambridge Univ. Press, Cambridge, U. K. https://doi.org/10.1017%2Fcbo9781139235761.001
Schädler, G., Berg, P., Düthmann, D., Feldmann, H., Ihringer, J., Kunstmann, H., et al. (2012). Flood Hazards in a changing climate: Project Report, pp 83. Center for Disaster Management and Risk Reduction Technology (CEDIM). http://www.cedim.de/download/Flood_Hazards_in_a_Changing_Climate.pdf
Seckin, N., Haktanir, T., & Yurtal, R. (2011). Flood frequency analysis of Turkey using L-moments Method. *Hydrological Processes*, *25*(22), 3499–3505. https://doi.org/10.1002/hyp.8077
Serinaldi, F., Lombardo, F., & Kilsby, C. G. (2020). All in order: Distribution of serially correlated order statistics with applications to hydrological extremes. *Advances in Water Resources*, *144*, 103686. https://doi.org/10.1016/j.advwatres.2020.103686
Smith, J. A., Cox, A. A., Baeck, M. L., Yang, L., & Bates, P. (2018). Strange Floods: The Upper Tail of Flood Peaks in the United States. *Water Resources Research, 54*(9), 6510-6542. https://doi.org/10.1029/2018wr022539
Stedinger, J. R. (1980). Fitting log normal distributions to hydrologic data. *Water Resources Research, 16*(3), 481-490. https://doi.org/10.1029/WR016i003p00481
Tarasova, L., Basso, S., Poncelet, C., & Merz, R. (2018). Exploring Controls on Rainfall-Runoff Events: 2. Regional Patterns and Spatial Controls of Event Characteristics in Germany. *Water Resources Research,*

54(10), 7688-7710. https://doi.org/10.1029/2018wr022588

Tarasova, L., Basso, S., Merz, R. (2020). Transformation of Generation Processes From Small Runoff Events to Large Floods. *Geophysical Research Letters* 47. https://doi.org/10.1029/2020gl090547

Thieken, A.H., Müller, M., Kreibich, H., Merz, B. (2005). Flood damage and influencing factors: New Insights from the August 2002 flood in Germany. *Water Resources Research 41*(12). https://doi.org/10.1029%2F2005wr004177

UNISDR (2015) Making development sustainable: the future of disaster risk management. Global Assessment report on disaster risk reduction. Centre for research on the epidemiology of Disasters United Nations Office for Disaster Risk Reduction (UNISDR) Geneva, Switzerland. https://doi.org/10.18356%2F919076d9-en

USWRC, 1976. Guidelines for Determining Flood Flow Frequency. United States Water Resources Council Bull.17, *Hydrology Committee*. https://doi.org/10.3133%2Ftm4b5

Villarini, G., & Smith, J. A. (2010). Flood peak distributions for the eastern United States. *Water Resources Research, 46*(6). https://doi.org/10.1029/2009wr008395

Vogel, R. M., Thomas, W. O., & McMahon, T. A. (1993). Flood-Flow Frequency Model Selection in Southwestern United States. *Journal of Water Resources Planning and Management*, *119* (3), 353–366. https://doi.org/10.1061/(asce)0733-9496(1993)119:3(353)

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

Wilson, P. S. & Toumi, R. (2005). A fundamental probability distribution for heavy rainfall. *Geophysical research letters*, *32* (14), s. n/a–n/a. https://doi:10.1029/2005gl022465

Zorzetto, E., Botter, G., & Marani, M. (2016). On the emergence of rainfall extremes from ordinary events. *Geophysical Research Letters, 43*(15), 8076-8082. https://doi.org/10.1002/2016gl069445