Advancing the Potential of the Metastatistical Extreme Value Framework for Extreme Flood Estimation in German Catchments

Dissertation

vorgelegt von
Sumra Mushtaq

von der Naturwissenschaftlichen Fakultät III der Martin-Luther-Universität Halle-Wittenberg
zur Erlangung des Akademischen Grades

Doktorgrades der Naturwissenschaften
-Dr. rer. nat.-

Halle 2025

Promotionsausschuss:
Vorsitzender: Prof. Dr. Wolfgang Goßel
Gutachter: Prof. Dr. Ralf Merz
Gutachter: Prof. Dr. Kai Schröter
Tag der wissenschaftlichen Aussprache: 11. November 2025

Zusammenfassung

Hochwässer gehören zu den am weitesten verbreiteten und zerstörerischsten Naturkatastrophen, und es wird prognostiziert, dass ihr Risiko aufgrund sozioökonomischer und klimatischer Veränderungen zunehmen wird, was eine wachsende globale Bedrohung darstellt. Die Fähigkeit, Hoch- oder Extremhochwasser vorherzusehen, ist für eine wirksame Gefahrenabwehr unerlässlich. Ein entscheidender Aspekt zur Verbesserung der Hochwasserrisikoanalyse liegt in einem verbesserten wissenschaftlichen Verständnisses der Prozesse, die Überschwemmungen verursachen, insbesondere derjenigen, die mit extremen Ereignissen verbunden sind. Das Verhalten von Hochwasserhäufigkeitsverteilungen mit sogenannten „heavy tails“ – also endlastigen Verteilungen mit ausgeprägten Randbereichen – ist ein wichtiger Hinweis auf die Wahrscheinlichkeit extremer Überschwemmungen. Das unterstreicht, wie wichtig es ist, solche Muster genau zu erkennen und zuverlässig vorherzusagen.. Diese Dissertation zielt darauf ab, das Verständnis für stark heavy-tail Abfluss- und Hochwasserverteilungen zu verbessern und gleichzeitig neue Methoden für die Vorhersage und das Management extremer Hochwasserereignisse in der Praxis einzuführen.

Die Dissertation ist in drei Hauptabschnitte gegliedert, zusammen mit einer Einleitung und einem Schlusskapitel, das die wichtigsten Ergebnisse zusammenfasst und einen Ausblick auf zukünftige Forschungsrichtungen gibt. Im ersten Abschnitt wird untersucht, welche Rolle Verteilungen mit leichten und schweren Rändern bei der Schätzung von Hochwasserquantilen spielen. Hierzu wird ein nicht-parametrischer Ansatz entwickelt, der es ermöglicht, im Rahmen der metastatistischen Extremwerttheorie (MEV) zwischen diesen Verteilungstypen zu unterscheiden und mithilfe des Tail-Ratio-Index die jeweils am besten geeignete Verteilung für Abflussspitzen auszuwählen. Die vorgeschlagene Methode wird anhand täglicher Abflusszeitreihen von 182 Pegeln in Deutschland getestet. Für die Mehrheit der analysierten Einzugsgebiete konnte eine eindeutig geeignete Verteilung identifiziert werden, die zudem gegenüber Änderungen in den zugrunde liegenden Daten robust ist. Eine passende Verteilungsauswahl verbessert sowohl die Schätzung häufiger als auch seltener Hochwasserereignisse im Vergleich zu einem Ansatz, bei dem in allen Gebieten dieselbe Verteilungsfunktion verwendet wird. Schließlich zeigt die Studie, dass die entwickelte Methode im Vergleich zur Anwendung der verallgemeinerten Extremwertverteilung (GEV), die als Standard in der Hochwasserstatistik gilt, deutlich geringere Unsicherheiten bei der Schätzung hoher Hochwasserquantile aufweist. Die Ergebnisse dieser Studie tragen dazu bei, sowohl die Unterschätzung als auch die Überschätzung seltener Hochwasser zu minimieren, und verbessern damit unsere Fähigkeit, die von seltenen hydrologischen Extremereignissen ausgehenden Gefahren genau zu bewerten.

Im zweiten Teil dieser Dissertation werden Kenntnisse über Abflussbildungsprozesse bei Hochwasserereignissen mit der metastatistischen Extremwerttheorie (MEV) verknüpft. Diese Theorie nutzt Informationen aus gewöhnlichen Ereignissen, um Rückschlüsse auf das Ausmaß und die Häufigkeit extremer Hochwasser zu ziehen. Die Studie konzentriert sich auf Flusseinzugsgebiete, in denen ein Sprung in der Hochwasserwahrscheinlichkeitskurve zu beobachten ist – ein Verhalten, das mit herkömmlichen Methoden der Hochwasserstatistik nur schwer abgebildet werden kann. Die Ergebnisse zeigen, dass in diesen Einzugsgebieten Hochwasser aus unterschiedlichen Abflusserzeugungsprozessen resultieren, wobei mindestens einer dieser Prozesse eine endlastige Verteilung aufweist. Die vorgeschlagene Methode ist in der Lage, solche ungewöhnlichen Hochwasserwahrscheinlichkeitskurven realitätsnah zu reproduzieren. Diese Erkenntnisse liefern neue Einblicke in die Entstehung sogenannter Hochwasserscheiden und bieten eine Methodik zur verbesserten Schätzung von Hochwasserquantilen in besonders risikobehafteten Fällen.

Der dritte Abschnitt dieser Dissertation befasst sich mit den Herausforderungen bei der Identifizierung von nicht-stationärem Verhalten in der Hochwasserstatistik, insbesondere bei kurzen Datensätzen. Ansätze der metastatistischen Extremwerttheorie (SMEV), die auch gewöhnliche Hochwasserereignisse in die Hochwasserstatistik einbeziehen – anstatt sich ausschließlich auf Extremereignisse zu stützen – bieten potenziell Vorteile bei der Analyse nicht-stationären Verhaltens. Dennoch wurde dieser Aspekt bisher kaum untersucht. Diese Studie analysiert das Potenzial des nicht-stationären SMEV-Modells. Die vorgeschlagene Methode wird anhand täglicher Abflusszeitreihen von 19 Pegeln in Deutschland getestet, die alle ein nicht-stationäres Verhalten aufweisen. Als Referenz dienen asymptotische, GEV-basierte Modelle, die in der nicht-stationären Hochwasseranalyse weit verbreitet sind und auf der klassischen Extremwerttheorie (EVT) beruhen. Sowohl die SMEV- als auch die GEV-Modelle enthalten eine zeitliche Kovariate, sodass ihre Parameter im Verlauf der Zeit linear variieren können. Die Ergebnisse zeigen, dass GEV-basierte Modelle dazu neigen, Hochwasserquantile zu unterschätzen – insbesondere bei Wiederkehrperioden von über 10 Jahren – und eine höhere Schätzverzerrung aufweisen. Im Gegensatz dazu liefern SMEV-basierte Modelle für diese längeren Wiederkehrperioden genauere Quantilschätzungen und weisen geringere Unsicherheiten auf. Diese Ergebnisse zeigen, dass SMEV-basierte Modelle die nicht-stationäre Hochwasserhäufigkeitsanalyse verbessern können, was sie zu einem wertvollen Instrument für die Hochwasserrisikobewertung unter sich ändernden hydrologischen Bedingungen macht. Die Ergebnisse dieser Studie leisten einen Beitrag zu einer verbesserten und robusteren Schätzung extremer Hochwasser in Gebieten mit endlastigen Hochwasserverteilungen. Die vorgeschlagenen Methoden zeigen sich insbesondere als robust gegenüber begrenzten Datenreihen – einer häufigen Herausforderung in der Hochwasseranalyse – und erhöhen dadurch ihre Anwendbarkeit auch in Regionen mit unterschiedlichen hydroklimatischen Bedingungen. Diese Ergebnisse eröffnen zahlreiche Anwendungsmöglichkeiten in der Hochwasserrisikoanalyse und im Ingenieurwesen, darunter Richtlinien für die Ermittlung von Hochwasserhäufigkeiten, die Modellierung extremer Hochwasser, die multivariate Analyse von Hochwassergefahren, die Erforschung komplexer nichtstationärer Muster

und die Bewertung der Auswirkungen des Klimawandels. Darüber hinaus liefert diese Forschung wertvolle Erkenntnisse für ein effektiveres Hochwasserrisikomanagement in einem sich verändernden Klima.

Schlüsselwörter: Schätzung der Hochwasserhäufigkeit, metastatistische Extremwertverteilung, Schwanzeigenschaften, Hochwasserverteilung, Heavy-Tailed-Verteilung, Flusserzeugungsprozess, Extremwertstatistik,

Abstract

River floods are among the most widespread and destructive natural disasters, with their risk projected to increase due to socioeconomic and climate changes, posing a growing global threat. The ability to anticipate high or extreme floods is essential for effective hazard mitigation. A crucial aspect of improving flood risk analysis lies in deepening scientific understanding of the mechanisms driving floods, particularly those associated with extreme events. Heavy-tailed behavior in flood frequency distributions serves as a key indicator of extreme flood likelihood, emphasizing the need for accurate identification and prediction of such patterns. This dissertation aims to advance the understanding of heavy-tailed streamflow and flood distributions while introducing novel methods for predicting and managing extreme flood events in practical applications.

The dissertation is structured into three main sections, along with an introduction and a concluding chapter that summarizes key findings and provides an outlook on future research directions. The first section investigates the role of light and heavy-tailed streamflow distributions in high flood quantile estimation. A non-parametric approach is developed to select the most suitable distribution of ordinary peaks, distinguishing between light-tailed and heavy-tailed distributions within the Metastatistical Extreme Value (MEV) framework. The proposed method is tested using daily streamflow time series from 182 streamflow gauges in Germany. The approach accurately identifies the most suitable ordinary distribution in the majority of analyzed basins and is robust to changes of the considered dataset. The study reveals that the preliminary selection of the ordinary distribution, based on the tail ratio index, improves the estimation of both frequent and rare floods compared to MEV applied with a single, non-tailored distribution. Finally, a comparison between the developed methodology and the standard Generalized Extreme Value (GEV) distribution demonstrates its effectiveness in reducing estimation uncertainty for high flood quantiles. The findings of this study contribute to minimizing both underestimation and overestimation of rare floods, thereby enhancing our ability to accurately assess the hazards posed by infrequent hydrological events.

In the second section of this dissertation, knowledge of runoff-generation mechanisms across different events is combined with a statistical method that utilizes information from ordinary events to infer the magnitude and frequency of floods. First, river basins exhibiting a sharp increase in the magnitude of rare floods (i.e., a flood divide in their empirical flood magnitude-frequency curve), are identified within a large set of catchments. Then, a statistical approach that accounts for different runoff-generation processes is applied to predict the magnitude and frequency of extraordinarily high floods. The findings reveal that in catchments with a flood divide, ordinary peaks originate from

different runoff-generation processes, with at least one process exhibiting heavy-tailed behavior. By accounting for these distinct tail behaviors, the proposed method is able to reproduce flood-frequency curves in catchments exhibiting a flood divide. These results provide new insights into the origins of flood divides and establish a methodology to improve high flood quantile estimation in these high-risk cases.

The third section of this dissertation addresses the challenges of identifying non-stationary behavior in flood frequency analysis, particularly due to limited data records. The application of the non-asymptotic SMEV approach in non-stationary flood frequency analysis has been limited to date. This study explores the potential of the non-stationary SMEV model, which incorporates ordinary events rather than relying solely on extreme events, to improve flood quantile estimation in non-stationary contexts. The proposed method is tested using daily streamflow time series from 19 streamflow gauges in Germany, all of which exhibit non-stationary behavior. As a benchmark, asymptotic GEV-based models are employed, as they are widely used in non-stationary flood frequency analysis and rely on Extreme Value Theory (EVT). Both non-stationary SMEV and GEV models incorporate a temporal covariate (time), allowing their parameters to vary linearly over time. The performance assessment reveals that GEV-based models tend to underestimate flood quantiles and exhibit greater bias when estimating quantiles for return periods exceeding 10 years. In contrast, SMEV-based models demonstrate higher accuracy and lower uncertainty for these longer return periods. However, non-stationary GEV-based models show high variability in uncertainty, raising concerns about their reliability in predicting extreme floods. These results highlight that SMEV-based models can enhance non-stationary flood frequency analysis by improving accuracy and reducing uncertainty, making them a valuable tool for flood risk assessment in changing hydrological conditions.

The findings of this study significantly contribute to our understanding of the mechanisms driving heavy-tailed flood behavior while introducing a robust approach for predicting extreme floods. This method reduces sensitivity to limited data, a common challenge in flood analysis, and improves its applicability across regions with diverse hydroclimatic conditions. These outcomes open the door to numerous applications in disaster risk research and engineering, including flood frequency guidelines, modeling of extreme floods, multivariate flood hazard analysis, exploration of complex non-stationary patterns, and assessments of climate change-related impacts. Furthermore, this research provides valuable insights for environmental sustainability, supporting more effective flood risk management in a changing climate.

Keywords: Flood frequency estimation, Metastatistical extreme value distribution, Tail properties, Flood divide, Heavy-Tailed distribution, Flow Generation Process, Extreme value statistics

This PhD thesis is dedicated to my beloved late father, Muhammad Mushtaq, whose memory continues to inspire me every day, and to my dearest mother, whose unwavering love, support, and encouragement have been my greatest strength throughout this journey.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisors, Prof. Dr. Ralf Merz, Dr. Stefano Basso and Dr. Arianna Minussi, for their invaluable scientific guidance and unwavering support throughout my PhD journey in Germany. I would also like to extend heartfelt thanks to Francesco Marra for his significant contributions to my research and for his generous technical support.

My special thanks go to my mentor Dr. Stefano Basso, whose continuous encouragement and belief in my potential were pivotal in helping me navigate this challenging path. His mentorship, patience, and support made this journey possible, and I am truly grateful.

Special thanks to the incredible team in the Department of Catchment Hydrology. I am especially grateful to all my colleagues for their support, insightful discussions, and the many coffee breaks that provided both motivation and a sense of community during this journey. I would like to acknowledge the warm and collegial spirit shared with colleagues past and present—particularly Larisa, Felipe, Pietro, Domi, Hsing-Jui, Hong, Christina, Ying, Matteo, Safae, Deen, Daniela and Zhenyu .

I am deeply thankful to the Higher Education Commission of Pakistan (HEC) for initiating my PhD studies in Germany through the USTEP-HEC Scholarship. I also gratefully acknowledge the financial support from the Helmholtz Association, the Helmholtz Interdisciplinary Graduate School for Environmental Research (HIGRADE), and the German Research Foundation (DFG). Funding through the projects FOR 2416, *Space-Time Dynamics of Extreme Floods (SPATE)*, and *Propensity of Rivers to Extreme Floods: Climate-Landscape Controls and Early Detection (PREDICTED)* was instrumental in the successful completion of this dissertation, developed at the Department of Catchment Hydrology, Helmholtz Centre for Environmental Research (UFZ), Halle (Saale).

To my beloved families—both my physical family and my spiritual family within the household of God—I am deeply grateful for the support. My sincere thanks goes to my (late) Father and my mother who are always supportive of my life long struggle. Without their advices and encouragement, I would not have reached this milestone. I am equally thankful to my siblings and all my friends (especially Eshrat, Ayesha, and Masooma) for their emotional support and encouragement throughout this journey.

Finally, this PhD thesis is the result of countless sleepless nights and immense emotional and intellectual effort. I owe my deepest gratitude to my husband, Irfan Mahmood, for his steadfast support and encouragement. Your love, patience, and unwavering belief in me carried me through every challenge. I am truly the luckiest person to have you by my side.

Table of Contents

List of Figures

List of Tables

1

Introduction

Extreme events are a major natural hazard that poses serious threats to property damages and loss of human lives across the world (Liu et al., 2022; UNISDR, 2015). More than half a million deaths occurred due to extreme events and 2.8 billion people were affected globally between 1980 and 2009 (Doocy et al., 2013). Figure 1.1 compares the impact of natural disasters between different time periods 1980-1999 and 2000-2019, and shows that disasters have become more frequent and affecting more people while causing significantly higher economic losses (United Nations, 2020). The rising economic cost highlights the increasing financial vulnerability of societies to natural disasters.

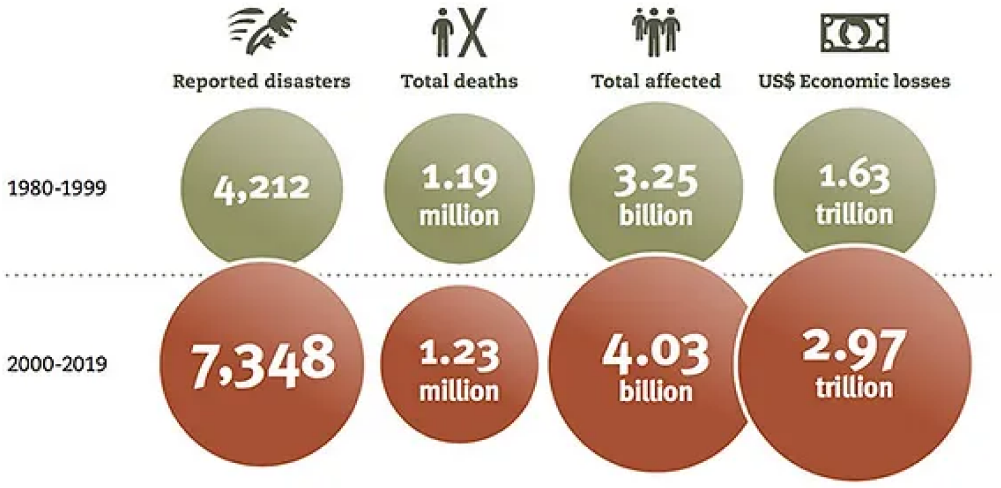


Figure 1.1: Disaster impacts in terms of deaths, total number of people affected, and economic losses (1980–1999) versus (2000–2019). (Source: United Nations, 2020).

Among various natural disasters, flood hazards have become increasingly severe worldwide, as the expansion of urban areas in flood-prone regions has increased since 1985, particularly after 2000 (Andreadis et al., 2022). Figures 1.2 and 1.3 illustrate the global distribution of disaster occurrences and the number of people affected by different event types from 2000 to 2019 (United Nations, 2020).

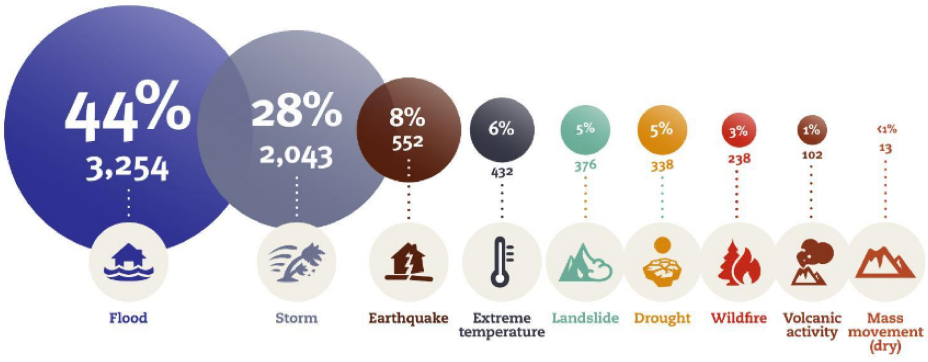


Figure 1.2: Global distribution of disaster occurrences by disaster type (2000–2019). (Source: United Nations, 2020).

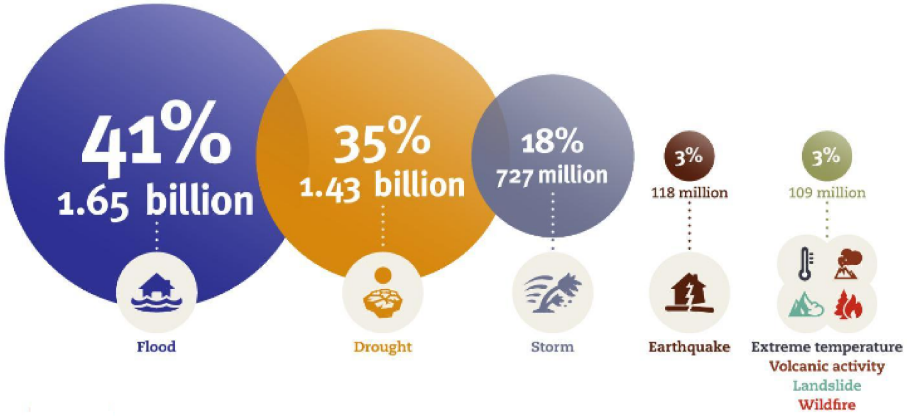


Figure 1.3: Percentage and number of people affected by different types of natural disasters globally (2000–2019). Floods affected the largest share of the population (41%), followed by droughts (35%) and storms (18%) (2000–2019). (Source: United Nations, 2020).

Flood risk needs to be quantified to ensure the design of resilient infrastructure, preventing failures that threaten human life and cause economic losses (Villarini et al., 2018). A reliable assessment of flood hazards improves preparedness, disaster response, and risk mitigation. However, extreme floods are difficult to quantify because they are rare events with limited historical data, making it challenging to establish a magnitude-frequency relationship using traditional flood estimation methods. The accuracy of extreme flood predictions is crucial for risk management, yet climate variability and human-induced changes further complicate estimation. This dissertation advances the understanding of extreme flood estimation by introducing robust methods that enhance prediction accuracy and improve flood risk assessment for practical applications in infrastructure planning, disaster preparedness, and climate adaptation.

1.0.1 Definition, significance, and current state

Flood frequency analysis relies on the key concept of the *return period*, which is defined as the average recurrence time of an event and quantitatively describes the frequency of occurrences exceeding a given magnitude. The return period (T) is a fundamental concept

in extreme event analysis and is widely applied in engineering regulations for mitigation and adaptation measures. Water infrastructure is designed to withstand events corresponding to specific return periods. For example, the design discharge for emergency spillways of concrete dams is typically based on $T = 1000$ years, while for earth-fill dams, it is around $T = 3000$ years. Similarly, levees are designed for $T \approx 200$ years, while bridge piers correspond to $T \approx 100$ years. However, these return periods far exceed the length of available observational records, which typically span only a few decades, with rare cases reaching 100 years or more. This discrepancy introduces substantial uncertainty in the estimation of extreme events. Since the design variable to be estimated often has a return period significantly longer than the available time series, extreme value estimation becomes an extrapolation exercise beyond the observed range. If one empirically assigns a return period higher or equal to the series length for the maximum recorded value, the result is an uncertain and unreliable extrapolation. As a result, statistical modeling becomes essential to bridge the gap between limited observations and the extreme events of interest. The key challenge lies in selecting a statistical model that can accurately capture the tail behavior and provide robust estimates of rare events.

Flood frequency analysis is typically performed using classical extreme value theory, which involves fitting a statistical distribution to either annual maxima (e.g., Cunnane, 1973; Villarini and Smith, 2010; Seckin et al., 2011) or a few values exceeding a threshold (i.e., the Peak Over Threshold (POT) approach (Davison & Smith, 1990). These methods allow for upper tail inference of a distribution without requiring complete knowledge of the underlying stochastic process. However, these approaches discard most of the information contained in the bulk of runoff events (Tarasova, Basso, Wendi, et al., 2020), often leading to estimations affected by large uncertainty (Miniussi et al., 2020). This limitation underscores the need for robust methods that effectively incorporate additional hydrological information to improve the accuracy of extreme flood quantile estimation.

1.1 Extreme Value Theory

1.1.1 A brief history of extreme value analysis

The statistical analysis of extreme values originated with the pioneering paper of Fisher and Tippett (1928), in which the extreme value theorem was first introduced, and the three fundamental limiting distributions were defined. Gnedenko (1943) later proved the theorem in its general form, and established the foundation for further developments in asymptotic theory. Subsequent contributions refined theoretical results, improving the understanding of convergence rates to the asymptotic distributions across a broad range of distribution functions. Notably, E. J. Gumbel (1958) applied extreme value methods to floods and drought modeling, and played a pivotal role in raising the awareness among hydrologists and engineering practitioners about heavy tailed distributions. A major advancement in extreme value analysis occurred with the development of the Peak Over Threshold (POT) method by Balkema and De Haan (1974) and Pickands (1975), which had a profound impact on hydrological applications.

In the last two decades EV analysis has seen a widening range of applications in diverse fields such as insurance, engineering design, finance, environmental and urban planning, where the traditional modelling based on light tails was proved to fail in several circumstances. The main question remains which of the three limiting distributions should be used to model a given data set, yet some authors proposed different approaches to tackle the problem. Recently Cook and Harris (2004) pointed out that relying on asymptotic models is unnecessary and often may lead to significant errors in high quantile estimation.

1.1.2 Classical Extreme Value Theory

Traditional Extreme Value Theory (EVT) defines extremes as block maxima—that is, the maximum value, x, observed among all "ordinary" values x_i $(i = 1, 2, 3, \ldots, n)$ that occurred within a specified time block, usually taken as 1 year. EVT assumes that the number of ordinary events in each block are independent and identically distributed (i.i.d.). Under this assumption, the cumulative distribution of block maxima is given by:

$H_n(x) = [F(x)]^n$. In this framework, the probability that the block maximum exceeds a specified value x in each year is: $p = 1 - H_n(x)$. This probability is used to define the return period T as: $T = \frac{1}{p} = \frac{1}{1-H_n(x)}$. The return period T of a given event magnitude x is a fundamental concept in extreme value analysis, representing the average time between two successive exceedances of x (please see Volpi et al., 2015, for a detailed discussion on different definitions of T).

According to the Three Types Theorem (Fisher and Tippett, 1928; Gnedenko, 1943; Gumbel, 2004), there are only three types of non-degenerate extreme value distributions: 1) Frechet distribution 2) Gumbel distribution and 3) Reverse-Weibull distribution. These three limiting distributions are unified within the Generalized Extreme Value (GEV) distribution (Coles, 2001), with their classification depending on the value of its shape parameter.

$$F(y;\boldsymbol{\theta}) = F(y;\xi,\alpha,\kappa) = \begin{cases} \exp\left(-\left[1+\kappa\frac{(y-\xi)}{\alpha}\right]^{\frac{1}{-\kappa}}\right), & \kappa \neq 0, \\ \exp\left(-\exp\left(-\frac{(y-\xi)}{\alpha}\right)\right), & \kappa = 0. \end{cases} \tag{1.1}$$

The parameter vector $\boldsymbol{\theta}$ comprises κ, α, and ξ, representing the shape, scale, and location parameters, respectively. The shape parameter κ determines the tail behavior of the distribution, classifying the Generalized Extreme Value (GEV) into three subfamilies: Extreme Value (EV) type I distribution when $\kappa = 0$; EV type II when $\kappa > 0$ (indicating heavy-tailed behavior); and EV type III when $\kappa < 0$ (indicating an upper-bounded distribution).

Traditional Extreme Value Theory (EVT) offers a well-defined theoretical framework through the Generalized Extreme Value (GEV) distribution, but its strict assumptions, such as the justification for assuming a large number of events per year (ideally tending to infinity), limit its practical applicability. Which makes it impossible in practice to detect whether the actual distribution of extremes is close to an asymptotic Generalized Extreme Value form (see for example Cook and Harris (2004); Koutsoyiannis (2004a)). These constraints make it difficult to determine whether extreme value distributions exhibit heavy tails, leading to significant theoretical and practical challenges. Similarly, the Peak Over Threshold (POT) approach relies on strict asymptotic assumptions and requires

careful threshold selection, often making it ineffective in practice. Additionally, both EVT and POT do not make good use of the available information and consider only a small fraction of available observations (annual maxima or a few values over a high threshold) data, missing opportunities to optimize quantile estimation uncertainty. This limitation is particularly problematic in non-stationary processes, where understanding variations in extreme values over time is crucial. To address these issues, Marani and Ignaccolo (2015) introduced the Metastatistical Extreme Value Distribution (MEVD), which leverages all available data to construct the distribution of yearly maxima from the distribution of ordinary values. The MEVD has been primarily applied to rainfall (Zorzetto et al., 2016; Marra et al., 2018), where it has been shown to significantly reduce estimation uncertainty compared to traditional approaches. This improvement is particularly notable for return periods longer than the sample size used for distribution estimation.

To address these issues, Marani and Ignaccolo (2015) introduced the Metastatistical Extreme Value Distribution (MEVD), which leverages all available data to construct the distribution of yearly maxima from the distribution of ordinary values. The MEVD has been primarily applied to rainfall (Zorzetto et al., 2016; Marra et al., 2018; Zorzetto and Marani, 2019), where it has been shown to significantly reduce estimation uncertainty compared to traditional approaches.

1.1.3 The Metastatistical Extreme Value Distribution (MEVD)

Recently, developments in extreme value statistics and novel theoretical approaches to appraise extreme events have proposed estimating the maxima by leveraging the information content of the full distribution of ordinary events (Basso et al., 2016 ; M. Marani and Ignaccolo, 2015), thus making more effective use of the available data. In particular, the Metastatistical Extreme Value (MEV) distribution (M. Marani and Ignaccolo, 2015; E. 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 MEVD approach considers both the distribution describing the ordinary values, denoted by $F(x;\theta_j)$, and the number of occurrences in each year, n, as random variables. The MEVD assumes that ordinary events are independent and expresses the probability that a yearly maximum is smaller than or equal to a specified value x as follows (Miniussi et al., 2020):

$$\zeta(x) = \sum_{n=1}^{\infty} \int_{\Omega_{\vec{\theta}}} \left[F(x;\vec{\theta})\right]^n g(n,\vec{\theta})\, d\vec{\theta} \tag{1.2}$$

where $g(n,\boldsymbol{\theta})$ is the joint probability distribution of the random variables N and $\boldsymbol{\Theta}$ (discrete in n and continuous in $\boldsymbol{\theta}$) and $\Omega_{\boldsymbol{\Theta}}$ is the population of the parameter values. An approximation of the ensemble average can be obtained by computing the sample mean over the entire historical series (M), resulting in

$$\zeta_{\mathrm{MEVD}}(x) = \frac{1}{M} \sum_{j=1}^{M} |F(x,\theta_j)|^{n_j} \tag{1.3}$$

where $F(x;\theta_j)$ is the cumulative distribution of ordinary values. j denotes the year index in the MEVD framework, and n_j is the number of events in year j. M is the number of the years in historical period.

Theoretically, MEVD 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—such as robust parameter estimation and uncertainty reduction—of considering the whole sample for fitting the parameters of the ordinary distribution outweigh those obtained from the use of yearly varying parameters (Miniussi et al., 2020).

This dissertation aims to address the challenges related to understanding and quantifying flood tail behavior in flood frequency analysis, as well as to improve the modeling of extreme floods. The thesis is structured around three major topics: (i) reliable detection and prediction of light- and heavy-tailed flood distributions using the Modified Extreme Value (MEV) distribution, (ii) influence of heterogeneous flow generation processes on extreme flood estimations, and (iii) non-stationary flood frequency analysis using the MEV approach.

1.2 Challenges and Research Prospects

Drawing from an extensive literature review, the challenges and opportunities in understanding the flood tail behavior and predicting extreme floods are outlined into three key main areas:

1. **Reliable detection and prediction of light and heavy-tailed flood distributions**:

   A major challenge in flood frequency analysis lies in addressing the uncertainty associated with estimating heavy-tailed flood behavior, primarily due to the scarcity of accessible, long-term hydrological data. The importance of upper-tail behavior in describing extreme events has been highlighted by numerous studies (e.g., Smith et al., 2018; Wietzke et al., 2020). To investigate tail properties, previous studies have utilized various graphical methods, including the generalized Hill ratio plot, log-log plot (also known as the tail probability plot), and the mean excess function (El Adlouni et al., 2008). However, these graphical methods are inherently subjective, as they rely on visual interpretation and are not well-suited for analyzing large datasets. These limitations highlight the need for robust and objective methods (B. I. Cook et al., 2004). Recent efforts have been made to address this issue. For example, Nerantzaki and Papalexious (2019) developed an automated procedure for applying the mean excess function to investigate tail properties of precipitation on a global scale. Wietzke et al. (2020) conducted a detailed inter-comparison of various upper-tail indicators to identify the most appropriate method. Their study considered the shape parameter of the GEV distribution, the Gini Index (Eliazar and Sokolov, 2010), the obesity index (Cooke and Nieboer, 2011), and the upper-tail ratio (Smith et al., 2018). Their findings suggest that the GEV shape parameter is the most suitable indicator for distinguishing between different tail behaviors. However, this indicator has the drawback of being

parametric, meaning it requires a prior assumption of a distribution and fitting it to a sample of maxima, which introduces notable uncertainties (Basso et al., 2021). To address these uncertainties and improve the reliability of extreme flood estimations, developing robust, non-parametric approaches capable of objectively characterizing upper-tail behavior remains a significant challenge in hydrological research.

2. **Influence of heterogeneous flow generation processes on extreme flood estimations**:

   Understanding the impact of diverse flow generation processes on flood estimates remains a significant challenge in hydrology. The inherent complexity of these processes introduces substantial variability in the prediction of extreme floods (Merz and Blöschl 2003; Blöschl et., 2015), underscoring the need for innovative methodologies to enhance the accuracy of flood quantile estimations. This challenge is particularly pronounced when a small number of extreme events dominate observed flood records in terms of magnitude (Smith et al., 2018). These occurrences are evident in flood-frequency curves as a marked increase in the magnitude of rare events, a phenomenon recently termed "flood divides" (Basso et al., 2023). The emergence of flood divides has often been attributed to data scarcity, which hampers the effective characterization of the tail of flood distributions (e.g., Miniussi et al., 2023). However, previous research has also linked flood divides to nonlinear catchment responses, arising from unique physioclimatic characteristics of river basins (Rogger et al., 2012; Basso et al.,2023) and from the presence of multiple runoff-generation mechanisms within a catchment (Yu et al., 2022; Merz et al., 2022). Runoff events and subsequent floods may be triggered by various processes, including rainfall on dry or saturated soils, snowmelt, or a combination of these factors (Hirschboeck, 1987b; Villarini and Smith, 2010; Sikorska et al., 2015). Traditional flood magnitude-frequency analysis often proves inadequate for handling the heterogeneity of flood records, particularly when flood events differ significantly, leading to the manifestation of flood divides (Hirschboeck, 1987a). Mixed extreme value distributions have been proposed as a potential solution for such cases. However, these methods face challenges due to the limited availability of extreme flood observations, which introduces significant estimation uncertainty (Alila and Mtiraoui, 2002; Barth et al., 2019 ; Waylen and Woo, 1982). Therefore, there is a critical need to develop and implement innovative methodologies to improve the reliability and accuracy of flood quantile estimations.

3. **Challenges in reliable detection of non-stationarity in flood frequency analysis**:

   Traditional flood frequency analysis assumes a stationary hydrological process (Seckin et al., 2011; Villarini et al., 2009), failing to account for climate variability and human-induced changes that alter hydrological regimes. Stationary Flood Frequency Analysis (S-FFA) assumes that the statistical properties of extreme discharges (e.g., mean and variance) remain constant over time, relying on a time-invariant distribution (Haktanir et al., 2013; (Morrison and Smith, 2002). The widely used Generalized Extreme Value (GEV) distribution, based on Extreme Value Theory (EVT), assumes an infinite number

of events per block, making it unsuitable for limited hydrometeorological datasets and leading to significant uncertainty in flood quantile estimation. Traditional methods focus primarily on extreme events, neglecting valuable information from ordinary runoff events, which further amplifies predictive errors and estimation uncertainties (Miniussi et al., 2020; Vidrio-Sahagún and He, 2022). While non-stationary methods attempt to address hydrological variability, their reliance on asymptotic assumptions introduces additional challenges, particularly when long-term hydrological data are scarce, limiting their ability to capture temporal variations in flood distributions effectively. Non-asymptotic approaches, such as the Metastatistical Extreme Value (MEV) distribution, offer a promising alternative by utilizing data more efficiently and reducing uncertainty (Miniussi et al., 2020). However, their application within a non-stationary framework remains limited. These challenges highlight the need for innovative methods that enhance accuracy and practicality in non-stationary flood frequency analysis, ensuring reliable predictions for effective flood risk management. Developing an adaptable and robust analytical framework is essential for overcoming these limitations and improving flood hazard assessments in a changing climate.

1.3 Research objectives and dissertation structure

This research aims to address the challenges in extreme flood estimation, as outlined in Section 1.2. Specifically, Chapter 2 introduces a robust methodology for selecting the most suitable ordinary distribution within the Metastatistical Extreme Value (MEV) framework, ensuring improved reliability and accuracy in extreme value analysis. Chapter 3 presents an innovative approach for predicting extraordinarily high floods resulting from heterogeneous flow generation processes, addressing the complexities associated with diverse hydrological conditions. Chapter 4 explores non-stationary hydrological frequency analysis by implementing the Simplified Metastatistical Extreme Value (SMEV) distribution to effectively analyze and account for non-stationary hydrological conditions. Through these contributions, this research aims to advance both theoretical and practical understanding of extreme flood estimation under stationary and non-stationary conditions. The three chapters address different aspects of the topic by employing diverse methodologies and datasets, ensuring a comprehensive analysis of the specific issues posed in each chapter. Each chapter is preceded by a detailed literature review focused on its specific topic, discussing the state of the art and identifying research gaps. Based on this, each chapter formulates distinct research questions, which are explored in detail as follows:

Challenge 1: A method for selecting the optimal ordinary distribution within the Metastatistical Extreme Value Framework (Chapter 2) :

- How does the tail behavior of a distribution influence the reliability of extreme flood estimations?
- How can light-tailed and heavy-tailed distributions be distinguished within the Metastatistical Extreme Value framework?

- Can the development of a non-parametric index lead to more reliable estimation of flood tail behavior across diverse catchments?

Challenge 2: Modeling of extraordinarily high floods driven by heterogeneous flow generation processes (Chapter 3):

- How do heterogeneous flow generation processes contribute to the occurrence of extraordinarily high floods?
- How do multiple runoff-generation processes influence the tail behavior of flood distributions?
- How does integrating multiple runoff-generation processes within the mixed-SMEV framework enhance the accuracy of flood estimation in river basins with a flood divide?

Challenge 3: Non-Stationary flood frequency analysis by using the Simplified Metastatistical Extreme Value Framework (Chapter 4):

- How does non-stationarity in hydrological systems affect the accuracy of flood quantile estimation?
- How can we effectively evaluate prediction uncertainty across different return periods?
- How effective is the proposed framework in capturing temporal variations and improving flood risk predictions under changing climatic?

These refined challenges and research questions provide a clearer and more comprehensive direction for addressing critical issues in flood estimation studies. Individual results, discussions, and chapter conclusions corresponding to these focus areas are presented directly within each chapter. This dissertation concludes with a discussion on the potential applications and future research directions for predicting extreme floods within the Metastatistical Extreme Value (MEV) framework, aiming to enhance hazard prediction and prevention efforts.

2

A method for selecting the optimal ordinary distribution within the Metastatistical Extreme Value Framework

Floods are a major global hazard, causing extensive damage and significant loss of human life across the world (UNISDR, 2015). Between 1980 and 2009, floods claimed over half a million lives and affected approximately 2.8 billion people worldwide (Doocy et al., 2013). In Germany, severe flood events, such as those in June 2013 and July 2021, resulted in widespread devastation. The 2021 floods alone affected more than 40,000 people and caused 184 fatalities, primarily due to extreme precipitation (Fekete and Sandholz, 2021). The economic impacts of such events have been substantial, with damages ranging from several billion euros in 2002 (Thieken et al., 2005) to over 30 billion euros in 2021 (Deutsche Welle, 2021). Consequently, a reliable assessment of natural hazards underlying these disasters is critical for enhancing societal preparedness. In particular, accurate estimation of flood magnitude and frequency is essential for the design and management of engineering and hydraulic structures, which serve as an essential tools in reducing damage and protecting a wide range of social and economic activities (Eash, 1997, Blöschl et al., 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, Villarin 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 (Bobee et al., 1993, Morrison and Smith, 2002, Haktanir 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, Basso, and Merz, 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., Klemes, 1974, Iliopoulou and Koutsoyiannis, 2019, Lombardo et al., 2019). Among others, Hu et al. (2020) found that the estimation of 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 d (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 (e.g., M. 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 (M. Marani and Ignaccolo, 2015; E. 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 behaviour 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 (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 (B. I. Cook et al., 2004). Recent attempts towards this direction have been done. For example, Nerantzaki and Papalexious (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 Sokolov, 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, I 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. I thus intend to provide indications on how to suitably apply the MEV framework for the prediction of hydrologic extremes. The chapter is organized as follows: in Section 2.2.1, we describe the data and the methodology used in this study; Section 2.2.2 and 2.2.3 displays and discusses the results, which are finally summarized in Section 2.3.

2.1 Methods

2.1.1 Study area and data

I 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. I 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 2.1a). 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 2.1b). 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 beurton2009. 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 schadler2012.

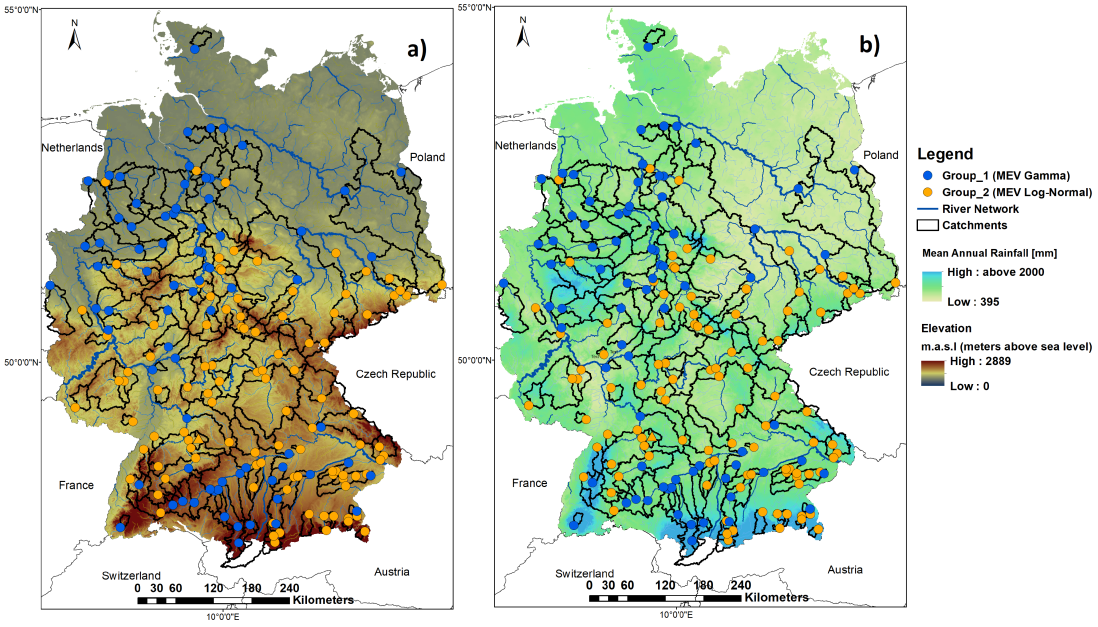


Figure 2.1: Locations within Germany of the 182 streamflow gauges analyzed in the present work (blue and orange dots). The two colors refer to the subdivision of catchments into two groups for which either Gamma or Log-Normal are the most suitable distributions of ordinary peaks, which is proposed in this study according to the methods explained in Section 2.1.3 and later discussed in Section 2.2 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.

2.1.2 The Metastatistical extreme value (MEV) distribution

2.1.2.1 Theoretical framework

In the present work I adopt a recently developed extreme value approach, the Metastatistical extreme value (MEV) distribution (M. Marani and Ignaccolo, 2015; E. Zorzetto et al., 2016; Marra, Nikolopoulos, 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; 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 E. Zorzetto et al. (2016).

$$\xi(x) = \frac{1}{M} \sum_{j=1}^{M} |F(x, \theta)|^{n_j} \tag{2.1}$$

where $F(x; \theta)$ is the ordinary distribution of peak flows, θ 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 θ) and 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 M. 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 miniussi2020.

In order to identify a set of independent ordinary events from daily streamflow records, which are typically highly correlated, I 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, M. Marani, and Villarini, 2020). At first, I identify one peak within each time block of length T = 5days + log(A), where A is the basin area in square miles, is the minimum lag time required by the independence criterion 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 6th to 9th 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 $F(x; \theta)$ in Equation (2.1) 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).

2.1.3 Ordinary distributions

Once the ordinary peaks are selected, we must choose a distribution to describe their statistical properties. miniussi2020 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, I 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; Hosking, 1990; Lancaster, 1966)

$$F(x;\alpha,\beta) = \frac{\beta^{-\alpha}\int_0^x t^{\alpha-1}e^{(-\frac{t}{\beta})}\,dt}{\Gamma(\alpha)} \tag{2.2}$$

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; Bobee et al., 1993)

$$F(x;\mu,\sigma) = \Phi\left(\frac{\log x - \mu}{\sigma}\right) \tag{2.3}$$

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

I estimate parameters of the Gamma and Log-Normal distributions by fitting them on the sample of ordinary peaks (selected as explained in Section 2.1.2) 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 hosking1997.

2.1.4 Tail ratio method and sensitivity analysis

I propose and apply in this study a methodology that enables the choice of the best ordinary distribution to be used in Equation (2.1), 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 behaviour 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 99^{th} and 95^{th} percentiles of ordinary peaks, similarly to the procedure suggested by **Nerantzaki2019** to evaluate the tail behavior of daily precipitation. In the present study, I tested different ratios (99^{th} to 90^{th} , as in **Nerantzaki2019** , 99^{th} to 95^{th} and intermediate ratios) with similar results. In the end, we found that in the case of flood peaks considering the ratio between the 99^{th} and 95^{th} percentile allows to describe best the upper tail of the distribution in our dataset. The method is inspired by the upper tail ratio of (J. A. 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, I 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) I 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) I calculate the tail ratio index of each series and seek a threshold to distinguish relatively lighter and heavier tailed distributions. I 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), I randomly select for 100 times 50% of the basins and calculate again the threshold value obtained from each resampling; (4) I 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.

2.2 Results and Discussion

I first apply the MEV approach by using either a Gamma or a Log-Normal distribution for each of the 182 German gauges. I then evaluate the capabilities of MEV-Gamma (Figure 2.2a) and MEV-Log-Normal (Figure 2.2b) to estimate observed maxima in the validation sample. I find that high quantiles are frequently underestimated by MEV-Gamma (see blue dots below the $45\circ$ line in Figure 2.2a). 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.

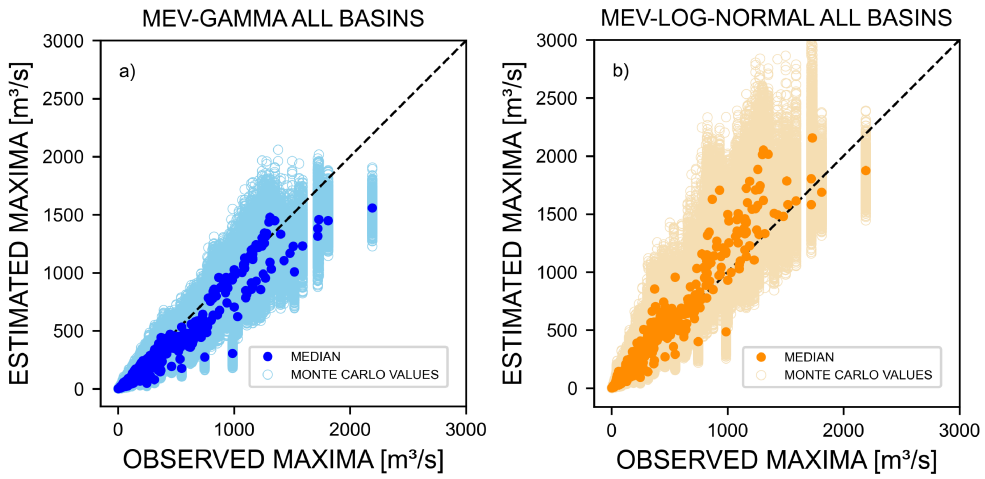


Figure 2.2: Quantile-quantile plots of streamflow maxima with return period longer than the length of the calibration sample (10 years) for MEV-Gamma (blue dots in panel a) and MEV-Log-Normal (orange dots in panel b). Observed annual maxima and return periods are empirically estimated by using the validation samples. Light and dark colors indicate respectively the 1000 resamples of the Monte-Carlo approach and their median values, for all basins.

These results are consistent with previous findings highlighting that flood records may exhibit both light and heavy-tail behaviors (Katz et al., 2002; Bernardara, Schertzer, Sauquet, et al., 2008). An example of this instance is provided in Figure 2.3. The blue lines in Figure 2.3a show that using a Gamma distribution to fit the empirical distribution of ordinary peaks (grey dots) would lead to an underestimation of the exceedance probability of the largest flows. This results into a significant underestimation of the largest annual maxima as shown in Figure 2.3c. 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 2.3d) because the underlying Log-Normal distribution is capable to correctly represent the empirical exceedance probability of the largest ordinary flows, as represented in Figure 2.3b. 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 (C. 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 ElAdlouni2008. To reach this goal, I calculate the tail ratio index introduced in Section 2.2.4 to evaluate the relative tail behavior of the distribution of ordinary peaks ex ante and independently from the MEV procedure. Figure 2.4 visualizes the results of the procedure detailed in Section 2.2.4 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.

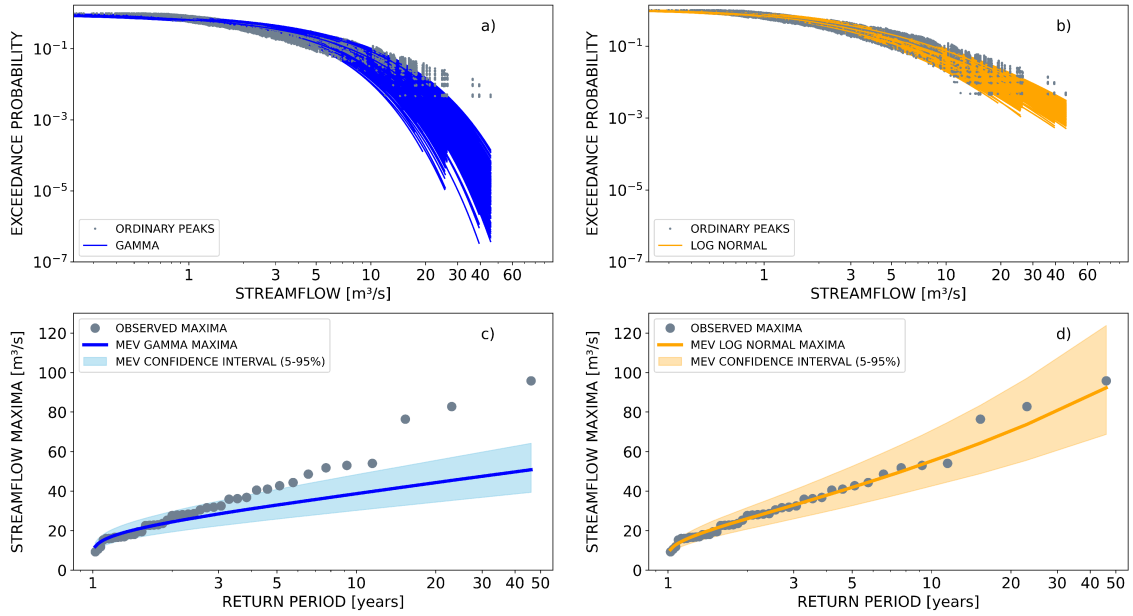


Figure 2.3: (a, b) Exceedance probability of ordinary peaks plotted as a function of streamflow values (in double logarithmic scale) for an exemplary case study indicated with a triangle in Figure 2.1 (the Neckar River at Rockenau, Area = 70 km^2). Gamma (panel a, blue solid lines) and Log-Normal (panel 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).

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 2.4). 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 2.4 refers to the number of catchments for which either distribution is the best, not to the whole dataset.

The crossing point of these two curves, identified with a black dashed line in Figure 2.4, 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 2.4) 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.

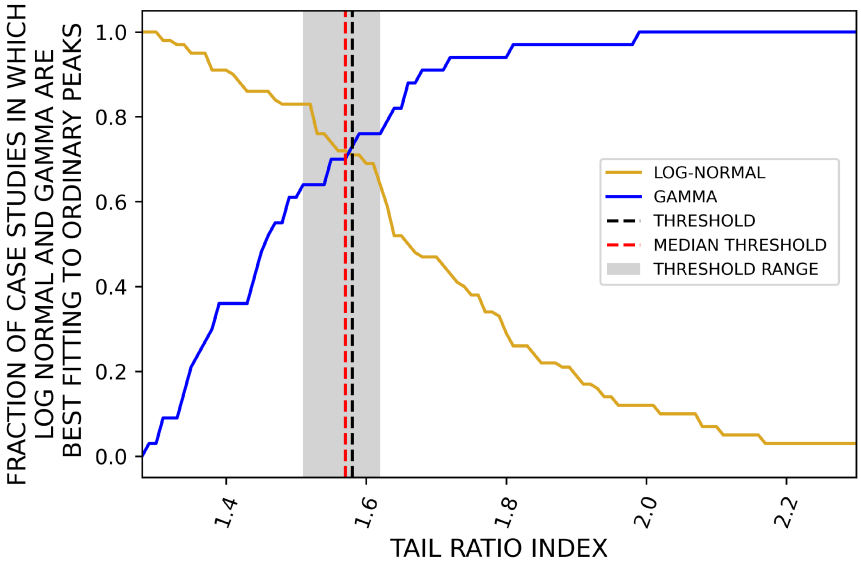


Figure 2.4: Identification of the threshold of the tail ratio index optimizing the number of catchments for which the most suitable distribution of ordinary peaks is selected ex ante based on the value of the tail ratio index itself. The most suitable distribution of ordinary peaks is identified on the 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.

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 2.2.4. The thresholds obtained with this procedure span a narrow range between 1.51 and 1.62 (grey shaded area in Figure 2. 4), with a median value of 1.57 (red dashed line in Figure 2.4). 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, I hence split the whole set of case studies into two groups. Group 1 (76 basins; blue dots in Figure 2.1) includes catchments with a tail ratio index lower than 1.58, for which I apply the MEV-Gamma, whereas Group 2 (106 basins; yellow dots in Figure 2.1) 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 2.5a, 2.5b. Quantile-Quantile (QQ) plots in Figure 2.5 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 Figures 2.2a, 2.2b and 2.5a, 2.5b). These results agree to those of papalexiou2013, 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 2.2) 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.

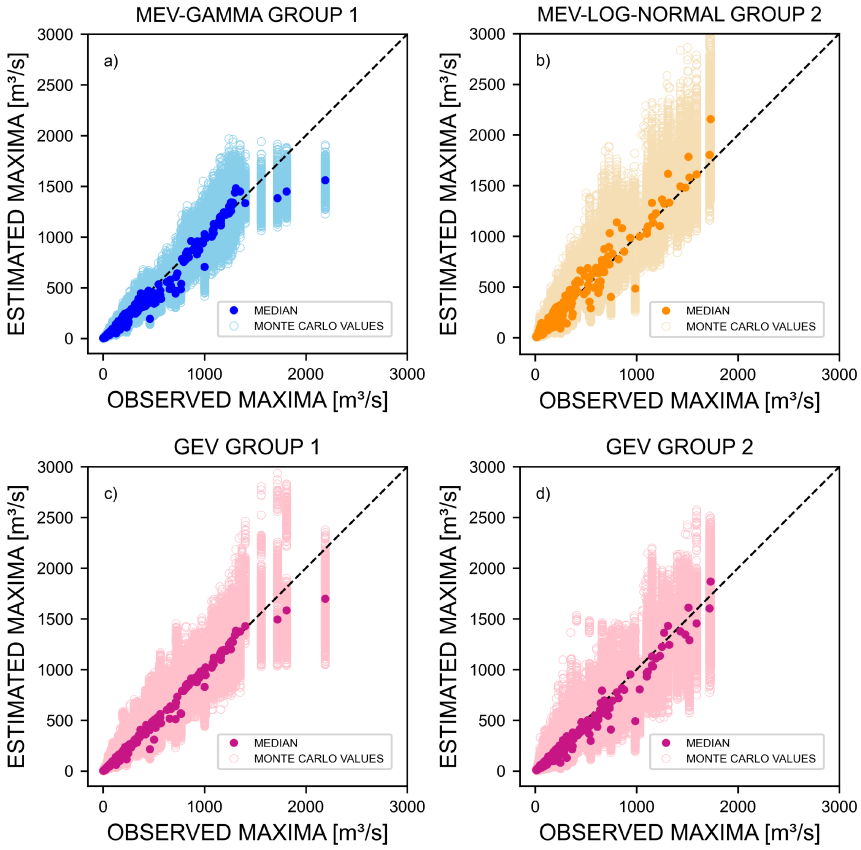


Figure 2.5: Quantile-quantile plots of flood magnitudes estimated for two subsets of river basins that have been separated ex ante based on the tail ratio index of the ordinary peaks. When the tail ratio index is lower than 1.58 we apply the MEV-Gamma (panel a, blue and light-blue dots), whereas when the tail ratio index is greater than this threshold we use the MEV-Log Normal (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.

I also compare these results with estimates obtained by means of the traditional GEV distribution (Figure 2.5c, 2.5d). 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 2.5a, 2.5b and 2.5c, 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 2.5a, 2.5c) and a remarkable decrease of the uncertainty of the estimate (light blue and pink dots in Figure 2.5a, 2.5c). These results are

supported by previous findings (S. Basso, Botter, 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.

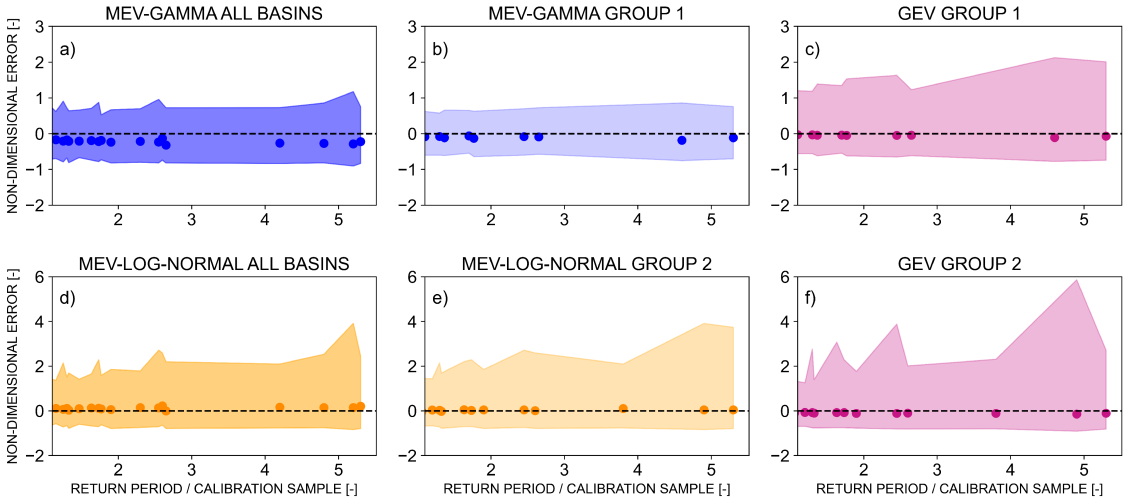


Figure 2.6: Non-dimensional error between observations and quantiles of the analyzed statistical distributions computed in the Monte Carlo simulation procedure (1000 realizations), plotted as a function of different ranges of the ratio between the return period and the calibration sample size (T/S). 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.

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 2.5a– 2.5d with Figure 2.6, which shows the non-dimensional error between observed and estimated quantiles for all random extractions in the Monte Carlo procedure and all the gauges zorzetto2016. 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.

MEV-Gamma (Figure 2.6a) and MEV-Log-Normal (Figure 2.6d) respectively under and overestimates when applied to the whole dataset. Figure 2.6a 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 2.6a 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 papalexiou2013. When comparing the relative error of MEV-Gamma (Figure 2.6b) and GEV (Figure 2.6c) for the basins in group 1 (or, likewise, the relative error of MEV-Log-Normal (Figure 2.6e) and GEV (Figure 2.6f) for the basins in group 2),

Table 2.1: 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 (All Basins)			MEV-Gamma (Group 1)		
T/S Range	Max$^-$	Max$^+$	Med	Max$^-$	Max$^+$	Med
$1 < \frac{T}{S} \leq 2$	-0.81	0.90	-0.20	-0.61	0.66	-0.08
$2 < \frac{T}{S} \leq 3$	-0.81	0.96	-0.23	-0.64	0.71	-0.10
$3 < \frac{T}{S} \leq 6$	-0.90	1.17	-0.28	-0.73	0.86	-0.12

	MEV-Log-Normal (All Basins)			MEV-Log-Normal (Group 1)		
T/S Range	Max$^-$	Max$^+$	Med	Max$^-$	Max$^+$	Med
$1 < \frac{T}{S} \leq 2$	-0.78	2.28	0.08	-0.78	2.28	0.02
$2 < \frac{T}{S} \leq 3$	-0.77	2.71	0.12	-0.77	2.71	0.06
$3 < \frac{T}{S} \leq 6$	-0.82	3.93	0.09	-0.82	3.91	0.15

	GEV distribution (Group 1)			GEV distribution (Group 2)		
T/S Range	Max$^-$	Max$^+$	Med	Max$^-$	Max$^+$	Med
$1 < \frac{T}{S} \leq 2$	-0.60	1.39	-0.04	-0.78	3.12	-0.07
$2 < \frac{T}{S} \leq 3$	-0.63	1.63	-0.05	-0.82	3.88	-0.09
$3 < \frac{T}{S} \leq 6$	-0.78	2.12	-0.08	-0.91	5.86	-0.12

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., odry2017, who stressed the importance of a reduced uncertainty when estimating high flow values. Notably, in group 1 the over estimation magnitude of MEV-Gamma is the half of GEV across all T/S (Figure 2.6b, 2.6c). 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.

I 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 2.6a, 2.6d) and of both MEV-Gamma (Figure 2.6b) and MEV-Log-Normal (Figure 2.6e) applied to their respective groups. In this way, I 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 2.6b are closer to the dashed line than in Figure 2.6a) and its uncertainty range (blue shaded area is narrower in Figure 2.6b than in Figure 2.6a), and to slightly reduce on average the overestimation issues of MEV-Log-Normal (Figure 2.6e).

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 2.1). Catchments for which a Gamma distribution (lighter tail) is the most suitable are primarily located in the Western part of Germany (Figure 2.1, blue dots), where winter floods triggered by precipitation on wet soils are dominant Tarasova, Basso, and Merz, 2020, with some catchments also gathered in the Alpine Fore lands. 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 2.1, 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, Basso, and Merz, 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.

2.3 Chapter conclusions

In this study, I adopted the Metastatistical extreme value (MEV) distribution for flood frequency analysis in a set of 182 catchments in Germany. My goal was to optimize this novel framework based on the statistical properties of the ordinary peaks. To this end, I 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. I 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. This 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. I 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, I reduce MEV-Gamma underestimation and Log-Normal overestimation issues of a median 57% and 58% respectively. Finally, I 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. The main findings of this study can be summarized as follows:

1. Tail Ratio Index for the Selection of Ordinary Distribution : The proposed tail ratio index offers an innovative approach to classify ordinary streamflow distributions into lighter or heavier tails (e.g., Gamma or Log-Normal). This method improves estimation accuracy by leveraging the information from ordinary events without relying on graphical evaluations.

2. Improved Flood Quantile Estimation within MEV Framework: The study optimizes the Metastatistical Extreme Value (MEV) distribution for flood frequency analysis by incorporating a non-parametric method to select the ordinary distribution of streamflow peaks, enhancing the estimation of high flow quantiles across 182 catchments in Germany.

3. Reduction of Estimation Errors: The tailored ordinary distribution selection within the MEV framework significantly reduces errors: it lowers underestimation issues in Gamma distributions by a median of 57% and overestimation in Log-Normal distributions by a median of 58%, compared to a blind application of MEV (i.e., without considering the tail behavior of the underlying ordinary distribution).

4. Robustness and Generalizability: The threshold used to discriminate between light and heavy tailed distributions is robust to random resampling and can reliably be applied to other regions, making the tail ratio index a versatile tool for flood frequency analysis.

5. Enhanced Accuracy over GEV Distribution: Benchmarking the MEV against the Generalized Extreme Value (GEV) distribution showed that tailoring the ordinary distribution in the MEV framework improves uncertainty reduction and flood estimation for long return periods in 135 out of 182 basins, demonstrating its utility for reliable assessments of extreme hydrological events.

2.4 Chapter statement

This chapter presents a formatted version of the original paper: Mushtaq, S., Miniussi, A., Merz, R., & Basso, S. (2022). *Reliable estimation of high floods: A method to select the most suitable ordinary distribution in the Metastatistical extreme value framework. Advances in Water Resources, 161* (104127). https://doi.org/10.1016/j.advwatres.2022.104127, with permission from the Authors and Elsevier Ltd. (© 2022 The Author(s). Published by Elsevier Ltd.)

Own contribution: The manuscript was primarily authored by Sumra Mushtaq, who was responsible for developing the methods and models, curating the data, conducting the analysis, and interpreting the results under the guidance and supervision of Stefano Basso, Ralf Merz, and Arianna Miniussi.

Acknowledgement:

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 advice. The first author is also thankful to the Higher Education Commission of Pakistan (HEC) and the German Academic Exchange Service (DAAD) for providing financial support to this study (DAAD-91716446).

Data Availability Statement

For providing the discharge data for Germany, we are grateful to the Bavarian State Office of Environment (LfU, https://www.gkd.bayern.de/de/fluesse/abfluss/tabellen). Readers can directly access the river discharge data used in this study by clicking on the provided link. We are also thankful to the Global Runoff Data Centre (GRDC) prepared by the Federal Institute for Hydrology (BfG, https://portal.grdc.bafg.de/applications/public.html?publicuser=PublicUserdataDownload/Home) for their valuable data resources. To access the data repository, please use the provided link and click on the “Download by Station” option, select Germany as the country of interest, and locate the table in the top-left corner to access discharge gauges for Germany.

3

Modeling of extraordinarily high floods driven by heterogeneous flow generation processes

3.1 Chapter Introduction

Reliable estimates of the magnitude and frequency of river floods are crucial for societal activities such as engineering design, urban planning and risk management (Barth et al., 2017; Smith et al., 2018). In particular, flood magnitude-frequency analyses establish a functional relation between the magnitude of floods and their exceedance probability, and represent a basic tool for quantifying flood hazard (e.g., Cunnane, 1973; Gotvald et al., 2012). However, river floods remain common perils that cause fatalities and damages worldwide (Bevere and Remondi, 2022; François et al., 2019).

Accurate estimation of flood magnitude and frequency is especially challenging when a few events are significantly larger in magnitude than the rest of the observed floods (Smith et al., 2018). These unexpectedly high floods can result in large socio-economic impacts (Davenport et al., 2021; Kreibich et al., 2022; Merz et al., 2021). This phenomenon, displayed in flood-frequency curves as a marked increase in the magnitude of rare events, was recently termed as flood divide (Basso et al., 2023) . This behavior is pervasive, although difficult to detect due to limited data availability. Past studies show that the number of flood divides increases with the length of the data series (e.g., Miniussi et al., 2023; Smith et al., 2018). The appearance of flood divides has been often related to data scarcity preventing a good characterization of the tail of flood distributions (e.g., Miniussi et al., 2023), an issue possibly exacerbated by non-stationarity of hydrologic processes caused by climate change (Yu et al., 2022). However, previous studies also linked them to non-linearities in the catchment response (Rogger et al., 2012) arising from distinct physioclimatic characteristics of river basins (Basso et al., 2023) and to the existence of different runoff-generation processes in a catchment (Merz et al., 2022). Runoff events and floods may indeed be triggered by different processes (e.g., rainfall on wet or dry soils, snowmelt, a combination of both; Hirschboeck, 1987b ; Villarini and Smith, 2010; Sikorska

et al., 2015) and thus be categorized into different types based on the inducing mechanisms (Stein et al., 2020; Tarasova, Basso, and Merz, 2020)

Traditional flood magnitude frequency analyses (e.g., Sivapalan and Samuel, 2009) are unsuitable to handle heterogeneous flood records (Hirschboeck, 1987b), especially when floods differ among each others to the point that flood divides are observed. Mixed extreme value distributions have been proposed for these cases, but the limited availability of large flood observations introduces considerable estimation uncertainty (Alila and Mtiraoui, 2002; Barth et al., 2019; Waylen and Woo, 1982). To tackle this issue, past studies (e.g., Fischer, 2018; Hirschboeck, 1987b) proposed a mixed distribution approach using peaks over a threshold (i.e., by using all observed peaks above a threshold).

These approaches improved hazard estimation compared to methods which assume homogeneity of the flood record. However, peaks over threshold may not be a representative sample of the processes that generate extraordinary events in catchments where the generation processes of the largest and most common floods are strikingly different (Tarasova, Basso, and Merz, 2020). In these cases, using ordinary peaks (i.e., all the independent streamflow peaks) may allow for sampling the full variety of generation processes while constraining prediction uncertainty. Recently, Miniussi et al. (2020) adopted such an approach to model floods occurring during different ENSO phases. However, studies that consider specific runoff-generation processes which produced ordinary peaks to estimate flood frequencies and magnitudes are lacking.

In this study, we aim to fill this gap by addressing the issue of reliably estimating flood frequencies and magnitudes in basins with a flood divide in the empirical flood magnitude-frequency curve. To do so, we relay on the analysis of ordinary peaks by explicitly considering the presence of multiple runoff-generation processes in a mixed non-asymptotic extreme value model. The advantage of this approach over traditional methods lies in the extension of information obtained by using a larger sample of ordinary peaks and accounting for the underlying physical processes leading to runoff generation in flood frequency analysis.

3.2 Method

3.2.1 Study Area and Data

This study uses daily streamflow records of 169 mesoscale river basins in Germany with drainage area between 30 and 23,000 km^2 (median: 581 km^2; Figure 3.5 in Supporting Information). The analyzed data series range from 1951 to 2013, encompassing 36–62 (median: 59) hydrological years per basin (November to October). From this data set, we identified river basins exhibiting a sharp increase in the magnitude of rare floods (i.e., a flood divide in their empirical flood magnitude-frequency curve, sensu Miniussi et al., 2023; Basso et al., 2023) by visually examining empirical flood-frequency curves. In order to facilitate their identification from a large data set, we calculated the L-skewness of the annual maxima sample (Hosking,1990), which is a general and universal index for tail heaviness and robust to the presence of outliers (Vogel and Fennessey, 1993), we noticed that, for Germany, all the cases with L-skewness exceeding 0.3 were also presenting this issue (Figure 3.4 in Supporting

Information). Notably, the threshold of 0.3 is selected for German basins and may vary for other regions. Visual inspection of the empirical flood-frequency curves remains the main tool to, confirm the presence of flood divides in a general case. In fact, flood frequency analyses are typically performed case by case by practitioners, who realize the presence of a flood divide in their data without relying on quantitative metrics, sometimes just because traditional approaches seem not to fit the data in any way. The selection yields 11 river basins with a clear flood divide in the empirical flood magnitude-frequency curve (Figure 3.5 in Supporting Information). All the following analyses are performed individually for each single basin.

3.2.2 Simplified Metastatistical Value Approach Applied to Heterogeneous Processes

We adopt the Simplified Metastatistical Extreme Value (SMEV) approach (M. Marani and Ignaccolo, 2015; Marra et al., 2019) to model the magnitude and frequency of extreme floods emerging from different runoff-generating processes (e.g., rain-on-wet or dry soils, snowmelt) in each single basin. The SMEV approach enables us to account for the presence of events triggered by multiple runoff-generation mechanisms (also termed event types in the following) through a parsimonious parametrization and to derive the compound extreme value distribution emerging from these types (Marra et al., 2019). Indicating with i = 1...S the event type in any of the basins, the mixed-SMEV cumulative distribution function ζ_{SMEV} can be written as

$$\zeta_{\mathrm{SMEV}}(x) = \prod_{i=1}^{S} [F_i(x;\theta_i)]^{n_i} \tag{3.1}$$

where θ_i are the parameters of the cumulative distribution F_i of ordinary events of the ith type and n_i is the average yearly number of ordinary events of type i. Notice that only two event types are considered in each basin.

To evaluate the model performance, I compared estimates obtained through the mixed-SMEV approach against those of a single- SMEV and Generalized Extreme Value (GEV) distributions, a widely used approach for flood magnitude-frequency analyses (Katz et al., 2002; Petrow et al., 2007). To this end, I estimate the GEV distribution parameters from the sample of annual maxima of each basin using the L-moments method (Hosking,1990). For the single SMEV approach, I used a log-normal distribution to describe all ordinary events in each basin.

The choice of a log-normal distribution is supported by previous study on the same data set of river basins in Germany (Mushtaq et al., 2022). I performed resampling with replacement (i.e., bootstrap) across years (1,000 realizations of all available years; Overeem et al., 2008) to assess the estimation uncertainty of all the methods. We finally quantified the model accuracy by means of non-dimensional error, computed as $e = \frac{x_{\mathrm{est}} - x_{\mathrm{obs}}}{x_{\mathrm{obs}}}$ between estimated (x_{est}) and observed maxima (x_{obs}) (E. Zorzetto et al., 2016). Notice that such an error metric favors the GEV distribution, which is explicitly parameterized to match the observed maxima, as opposed to mixed-SMEV and single-SMEV approaches, whose parameter estimation is performed on a larger sample, of which just a subset belongs to the annual maxima.

3.3 Process-Based Classification of Ordinary Events

In this study, I employed the method of Lang et al. (1999) to identify the independent ordinary peaks from the stream flow record required to apply the SMEV approach, as previously done by Miniussi et al. (2020) and Mushtaq et al. (2022). The number of independent ordinary peaks obtained for each catchment ranges from 370 to 933 (median value: 796).

The ordinary peaks were further classified into process-based types by using the classification of Tarasova, Basso, Wendi, et al. (2020), which labels streamflow events corresponding to the identified ordinary peaks according to their runoff-generation processes, which are assessed based on the nature of the inducing events (i.e., rainfall vs. snowmelt) and the catchment wetness states (i.e., wet or dry). This process-based approach uses observed daily precipitation (Rauthe et al., 2013) as well as daily snow water equivalent and soil moisture simulated by the mesoscale Hydrological Model (Kumar et al., 2013; Samaniego et al., 2010). Tarasova, Basso, Wendi, et al. (2020) employed dimensionless indicators to differentiate between inducing events and catchment wetness states and observed the small uncertainties associated with model structure and parametric uncertainty (Figure 4 in Tarasova, Basso, Wendi, et al. (2020)). However, it is worth noting that employing different indicators or classification frameworks could lead to substantial differences in event classification (Tarasova, et al., 2019), consequently affecting the accuracy of flood estimates.

In order to efficiently incorporate distinct event types in flood magnitude-frequency analyses and to ensure sufficient sample sizes for each of them, we aggregate the event types by Tarasova, Basso, Wendi, et al. (2020) into two major groups (S = 2): processes related to dry antecedent conditions (rain-on-dry events—Type-1) and wet antecedent conditions (rain-on-wet and snowmelt events—Type-2). The probability distributions of the magnitude of these event types are significantly different ($p < 0.05$), as evaluated through a pairwise two-sided Kolmogorov-Smirnov test (Massey Jr, 1951).

3.4 Selection of Ordinary Distributions

For each event type, we choose a suitable ordinary distribution. The presence of a clear flood divide in the empirical flood magnitude-frequency distribution suggests that one of the event types is characterized by a heavy-tailed distribution(Merz et al., 2022). We used Weibull plotting positions (Weibull, 1939) to derive the empirical cumulative distributions of ordinary events of each type and distinguish between those either exhibiting or not a heavy-tailed behavior, here intended as a power-law tail (i.e., a linear behavior in double-logarithmic coordinates; Newman, 2005).

This process can in principle be automated, for example, by using tests like the one applied by Marra et al. (2023). However, since the number of basins considered in this study is limited, it is preferred here to proceed with the accuracy of human supervision. Visual inspection of the empirical cumulative distribution functions of ordinary events of the two types showed that, for each of the 11 study catchments, either Type-1 or Type-2 events clearly exhibit power-law behavior (see respectively Figures 3.2a and 3.2e), as manifested by the linear form of the distribution in a log-log probability plot.

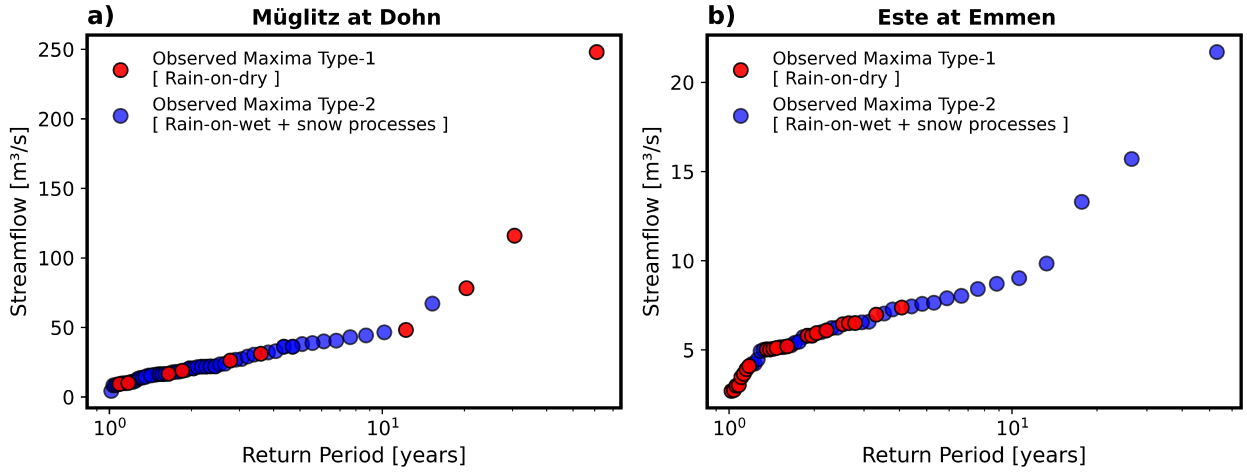


Figure 3.1: Empirical flood magnitude-frequency curves for two exemplary case studies exhibiting a flood divide: (a) the Müglitz River at Dohn (Gauge-ID: 550940, area = 196 km^2) and (b) the Este River at Emmen (Gauge-ID: 6338260, area = 171 km^2). Floods are classified into two major event types: rain-on-dry (red dots— Type-1) and combination of rain-on-wet and snow processes (blue dots—Type-2).

If the empirical distribution of ordinary events of the examined type has a heavy tail, we use a power-law distribution (Malamud and Turcotte, 2003, 2006). The cumulative distribution function of the power-law is

$$P(x) = \left(\frac{x}{x_{\min}}\right)^{-\alpha} \tag{3.2}$$

where x is the analyzed variable, $x_{\min}$ is a left-censoring threshold (i.e., the threshold above which the power-law behavior is manifest), and α is the scaling parameter. We estimated the parameters α and $x_{\min}$ through the method of Clauset et al. (2009). If the empirical cumulative distribution function of the ordinary events of the examined type is not heavy-tailed (i.e., it does not show a linear behavior in double-logarithmic coordinates), we model it using a two parameter log-normal distribution (Bobee et al., 1993). The cumulative distribution function of the The log-normal distribution is expressed as

$$F(x;\mu,\sigma) = \Phi\left(\frac{\log(x)-\mu}{\sigma}\right) \tag{3.3}$$

where σ and μ are respectively its shape and scale parameters, and Φ is the standard normal distribution function. where σ and μ are respectively its shape and scale parameters and σ is the standard normal distribution function. I fit the log-normal distribution using the method of L-moments (Hosking,1990) by left-censoring the lower portion of the ordinary events for the mixed SMEV (i.e., considering Type-1 and Type-2) and single-SMEV (i.e., by considering all ordinary events without types). The left-censoring method enables us to characterize the tail of the distribution with few parameters, by ignoring the magnitudes of the censored part while retaining their probability (Marra et al., 2019). For both the power-law and log-normal distributions, I selected the left-censoring threshold by minimizing the root mean square error between predicted and observed magnitudes of ordinary events in the upper twentieth percentile (Ritter and Munoz-Carpena, 2013).

3.5 Results and Discussion

In this study, we investigate river basins that experienced extraordinarily high floods with much larger magni tudes than the bulk of recorded annual maxima. To better illustrate the properties of these rare floods we focus on two exemplary case studies, the Müglitz and Este River basins (Figure 3.1), which exhibit marked flood divides in their empirical flood magnitude-frequency curves. In fact, the largest annual maxima (roughly those with a return period exceeding 10 years) grow to considerably larger magnitudes than the smaller ones (e.g., the largest observed annual maxima are 3–8 times and 2–4 times larger than their mean values for the Müglitz and Este River basins, respectively).

In Figure 3.1, annual maxima are color-coded based on the type of processes that generated the events. We observe that floods with runoff-generation processes, different from those mostly observed for common floods (i.e., those on the left hand sides of panels a, b) may strongly affect the upper tail behavior, as also highlighted by Tarasova, Basso, and Merz (2020). For instance, extraordinarily high floods in the Müglitz River (Figure 3.1a) mainly belong to the rain-on-dry type (4 out of 6 maxima with a return period greater than 10 years are characterized as Type-1), while this type is hardly present in lower flood peaks.

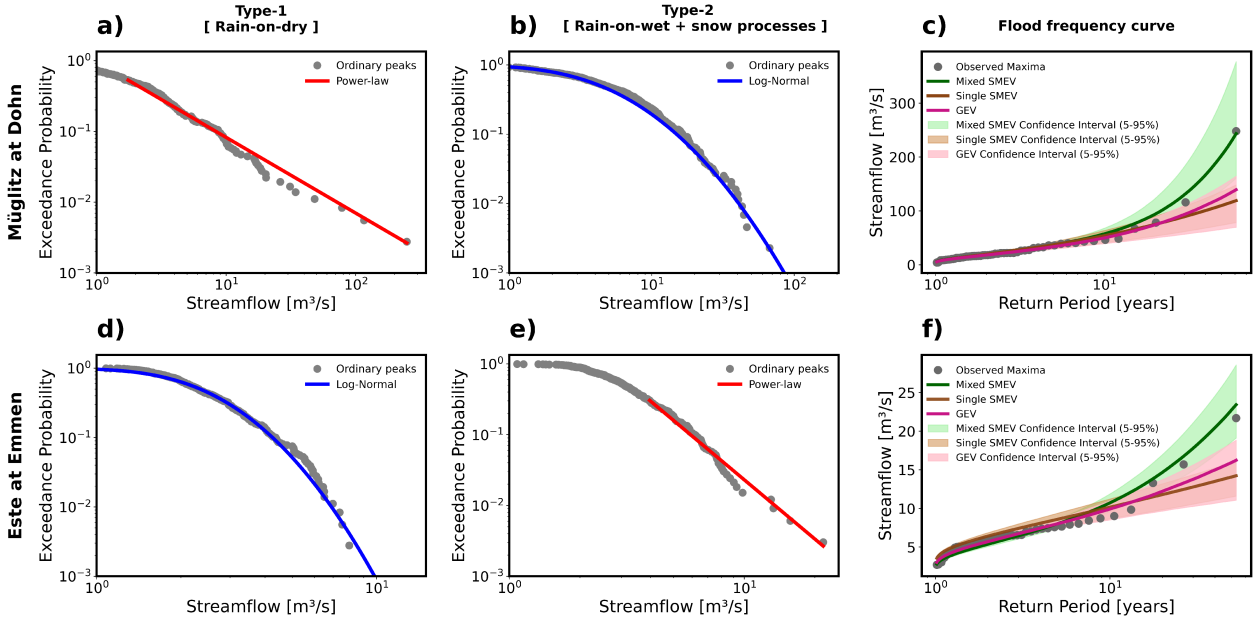


Figure 3.2: Exceedance cumulative distributions of ordinary peaks (a, b, d, e) and flood magnitude-frequency curves resulting from a mixed-SMEV, single-SMEV and a standard Generalized Extreme Value (GEV) approach (c, f) for the Müglitz River at Dohn (a–c) and the Este River at Emmen (d–f). Panels (a, d) show the ordinary distributions for Type-1 events, and panels (b, e) show the ordinary distributions for Type-2 events. Blue (log-normal) and red (power-law) lines in panels (a, b, d, e) display the probability distributions describing ordinary events of Type-1 and Type-2. Green, orange and pink curves in panels c and f show the median values for the corresponding quantiles of 1,000 resample values with replacement for mixed-SMEV, single-SMEV and GEV estimates, respectively. Green, orange and pink shaded areas indicate the related confidence intervals (5th–95th percentiles).

Conversely, extraordinarily high floods in the Este basin (Figure 3.1b) are caused by rain-on-wet conditions and snowmelt processes (5 maxima with a return period greater than 10 years are characterized as Type-2). As described in Section 2.4, we examine the empirical cumulative distribution functions of the different types of ordinary events (gray dots in Figures 3.2a, 3.2b, 3.2d, and 3.2e). We notice that in all cases in which a flood divide is present one event type displays heavy-tailed behavior (Figures 3.2a and 3.2e). Hence, we fit the empirical ordinary distribution of these events with a power-law distribution (see red lines in Figures 3.2a and 3.2e), and use a log-normal distribution for the other type (blue lines in Figures 3.2b and 3.2d). While previous studies identified both heavy-tailed and light-tailed behaviors in flood records (Bernardara et al., 2008; Mushtaq et al., 2022), here we move a step further and show that different tail behaviors may be associated with distinct runoff-generation processes (Yu et al., 2022).

These results demonstrate that relying solely on a pre-defined statistical distribution and disregarding the heterogeneity of the sample arising from different underlying physical processes may lead to an erroneous estimation of extreme floods. The proposed approach could thus improve the estimation of upper tail quantiles in basins where a flood divide is observed in the flood magnitude-frequency curves. These results confirm that heavy tails (i.e., flood divides) can originate from a mixture of flood-generating processes (Merz et al., 2022). It is worth noting that rain-on-dry events dominate the upper tail in the Müglitz River basin (Figure 3.2a), while rain-on-wet and snowmelt events dominate the upper tail in the Este River basin (Figure 3.2e).

Past studies indicate that various runoff-generation processes might be associated with heavier tails of the distribution of floods in different regions (Tarasova et al., 2023), as we observe in our set of case studies. The flood divide arises from the fact that the mixture contains a heavy-tailed process, whose upper tail tends to dominate the mixture distribution at large quantiles (Figures 3.2c and 3.2f). These results are coherent with previous studies, which show that the tail of a mixture distribution is influenced by the component with the most pronounced tail (e.g., Cavanaugh et al., 2015). Results for all other case studies are reported in Figures 3.6–3.14 in Supporting Information. These results are broadly consistent with what was found for the exemplary case studies discussed in Figures 3.1 and 3.2, despite a few cases showing a less clear distinction among event types. As expected, this leads to comparable performances of the three methods in these catchments.

Figure 3.3 summarizes the findings obtained for all 11 cases which exhibit flood divides in the set of 169 German river basins. Observed versus estimated annual maxima resulting from the proposed mixed-SMEV, the single SMEV and the GEV approaches are evaluated in a bootstrap fashion (see Methods) and respectively displayed in panels a, b and c. This overall comparison confirms the results discussed above for the two exemplary case studies and highlights the capability of the mixed-SMEV approach to provide reliable estimates of floods for a wide range of quantiles in all river basins with a flood divide (Figure 3.3a). The comparison with the single-SMEV (Figure 3.3b) and the GEV (Figure 3.3c) distribution reveals that mixed-SMEV estimates (Figure 3.3a) are characterized by a considerably smaller bias, especially for the upper tail quantiles. To summarize the performance of the mixed-SMEV approach to estimate flood magnitude - frequency in catchments with a flood divide, I computed the non-dimensional errors (see Section 2.2) between observed annual maxima and estimates of the corresponding empirical quantiles (Figures 3.3d and 3.3e).

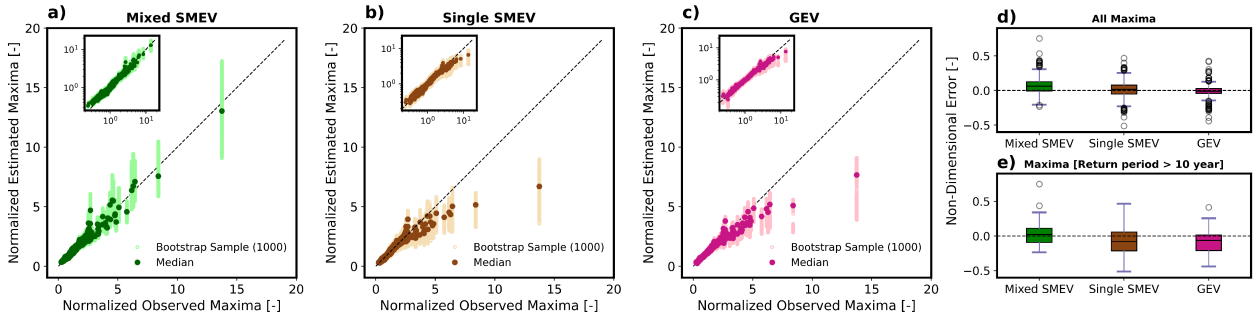


Figure 3.3: Estimated versus observed normalized (i.e., divided by their median value) annual maxima for (a) mixed-SMEV (green dots) (b) single-SMEV (orange dots) and (c) Generalized Extreme Value (pink dots) for 11 river basins in the data set which exhibit flood divides. Light and dark colors indicate results for 1,000 realizations of the bootstrap and their median values, respectively. Insets of panels (a–c) show the same results plotted on a double-logarithmic scale. Panels d and e show non-dimensional error between observed and estimated maxima of the analyzed statistical distributions computed for the median of 1,000 bootstrap values (i.e., dark green, orange and pink dots in a–c respectively) with the same return period: (d) all quantiles; (e) the quantiles corresponding to return periods >10 years.

Figure 3.3d displays a boxplot of non-dimensional errors for all observed maxima in the 11 case studies. Here, the mixed-SMEV approach tends to overestimate the bulk of floods, mainly due to small floods with a return period less than 10 years, whereas single-SMEV and GEV distributions show the same degrees of over and underestimation. However, when we focus our analysis on quantiles with return period greater than 10 years (Figure 3.3e), which are the most relevant for flood hazard assessment, the non-dimensional errors provided by the mixed-SMEV approach are lower compared to single-SMEV and GEV distributions.

Separately fitting different probability distributions to samples of ordinary peaks with distinct tail behaviors increases the physical basis of flood magnitude-frequency analyses. In fact, the proposed method allows for accounting for different statistical properties of events triggered by various runoff-generation processes. It is, however, worth noting that separating the processes yields significant improvements only if event types are characterized by markedly different tail behaviors (e.g., Marra et al., 2019)). If the different processes have similar tail behaviors, the mixed distribution will not be distinguishable from the one obtained by using all the data together, since uncertainty will be predominant due to the need to estimate a larger number of parameters on a smaller amount of data. The practice of classifying event types evolved in recent years (Tarasova, Basso, Wendi, et al., 2020; Turkington et al., 2016; Vormoor et al., 2016) and is deemed to grow (Merz et al., 2022; Tarasova et al., 2019), increasing the availability of information required to apply the proposed approach.

A substantial difference among runoff-generation processes was suggested as a possible cause of flood divides by past studies. Rogger et al. (2012, 2013) showed, by means of extensive field surveys and modelling analyses, that flood divides may emerge from strong non-linearities in runoff-generation processes resulting from progressive saturation of the catchment (i.e., a shift from dry to wet conditions). Basso et al. (2016) provided a mechanistic explanation of this phenomenon and the resulting appearance of flood divides in flood magnitude-frequency curves by linking it to the catchment water balance. Catchments in wetter climates experience sustained water supply which determines unvaried runoff-generation processes. Conversely, river basins in drier areas, where longer lag times between rainfall events allow for the catchment to dry, undergo transient conditions

leading to varied runoff-generation processes (Basso et al., 2023). Our results confirm that explicitly accounting for the existence of different runoff-generation processes enables us to adequately model the tail of flood distributions in catchments exhibiting flood divides.

The above discussion hints at the interplay between various hydro-meteorological drivers and the hydroclimatic settings leading to the occurrence of extraordinarily high floods in a region. In fact, river basins exhibiting flood divides (i.e., the exemplary Müglitz River and other eight case studies in our data set) are mostly located in the Central-Alpine region of Europe (Blöschl et al., 2017). In this region, the distributions of precipitation volumes for rain-on-wet and snowmelt floods have slightly heavier tails than for rain-on-dry events, leading to the possible occurrence of high floods of Type-2 (as exemplified by one river basin in our data set, see Figure 3.7 in Supporting Information). However, the distribution of precipitation intensity of rain-on-dry events in this area is remarkably heavier than for the other event types (Tarasova et al., 2023), mostly due to the occurrence of Vb-cyclones (Hofstätter et al., 2016). This feature likely underlies the occurrence of heavy-tailed distributions of rain-on-dry ordinary events and extraordinarily high floods of Type-1, which are mostly responsible (8 case studies) for the appearance of flood divides in flood-magnitude frequency curves. In contrast, the Este River and one additional catchment among the 11 showing a flood divide are located in the Atlantic region, where rain-on-wet flood events exhibit heavier tails than rain-on-dry events (Tarasova et al., 2023).

Variable distributions of key factors contributing to runoff generation, such as rainfall intensity, volume, soil moisture, snow accumulation and release are mirrored by differences in the tail behavior of ordinary distributions across regions, which can be leveraged by means of the proposed approach to improve the estimation of the hazard posed by extraordinarily high floods. However, climate change induces shifts in the mixture of flood generation processes, such as a decrease in snowmelt events and an increase in intense rainfall events (Hall et al., 2014; Huo et al., 2022). These changes inevitably influence the flood probabilities. Our method explicitly incorporates the presence of diverse flow-generating processes, and can therefore be used to predict changes in flood frequency based on the projected changes in the frequency of the runoff-generation processes.

3.6 Chapter conclusions

We provide a framework to derive accurate estimates of flood magnitude and frequency for basins where extraordinarily high floods occur. The approach leverages knowledge of the heterogeneity of runoff-generation processes by means of a process-informed mixed non-asymptotic statistical method, the SMEV framework. We employ the mixed-SMEV to estimate flood magnitude and frequency for 11 river basins in Germany featuring a flood divide. In these cases, at least one runoff-generation process is characterized by a heavy-tailed empirical distribution. Thanks to the explicit consideration of various runoff-generation processes, the proposed approach enables us to accurately predict the magnitude of rare floods, outweighing the performance of single-SMEV and GEV distributions.

The approach relies on a classification of event types and is only worth using when runoff events generated by different hydrologic processes are characterized by distributions with distinct tail behaviors. These requirements may constrain the use of the approach in practice. Nonetheless, classifications of event types are becoming more common, and the method provides a process-based solution to estimate large flood quantiles in contexts for which current methods fail. As climate change may alter the frequency of different event types, our approach also offers a way to account for climate change impacts on flood hazards in a physically sound manner. The key conclusions of the study are:

1. Applicability to Heavy-Tailed Runoff Processes: The mixed-SMEV framework is particularly well-suited for basins where at least one runoff-generation process follows a heavy-tailed empirical distribution, making it effective in predicting the magnitude of extreme floods in these scenarios.

2. Enhanced Flood Estimation with SMEV Framework: Nonetheless, the ability of accounting for mixtures of distributions in the flood-peak SMEV formulation has significant practical potential: several flood drivers can be identied (e.g., snowmelt, rain-on-snow, atmospheric rivers) whose role can be studied using the approach developed. The mixed-SMEV framework significantly improves the estimation of flood magnitudes and frequencies for river basins where extraordinary floods occur, particularly by accounting for the heterogeneity in runoff-generation processes. This process-informed statistical approach outperforms traditional methods such as GEV distribution in predicting rare flood events.

3. Utility of Process-Based Classification: The effectiveness of the approach is strengthened by its ability to classify runoff events into distinct types based on their generation processes, particularly when these processes exhibit different tail behaviors in their statistical distributions. As process-based classification becomes increasingly adopted in hydrological studies, it greatly enhances the practical applicability and robustness of the framework.

4. Adaptability to Climate Change: The SMEV framework offers a physically grounded approach for integrating the potential impacts of climate change into flood hazard assessments. By explicitly accounting for changes in the frequency and characteristics of different runoff-generation processes, it allows for a more realistic representation of evolving flood risks. This adaptability makes the SMEV framework a robust and forward-looking solution for improving the reliability of flood predictions under changing climatic.

3.7 Supporting Information

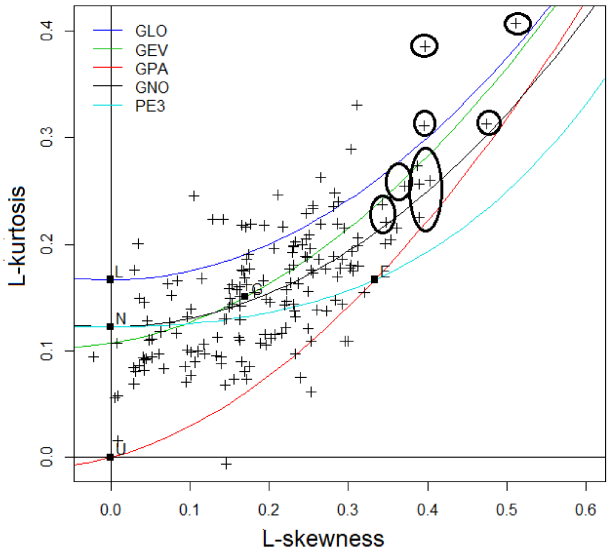


Figure 3.4: L-moments diagram for 169 catchments (represented by + sign). Black circles represent the 11 catchments with L-skewness value greater than 0.3, which are used in the analysis.

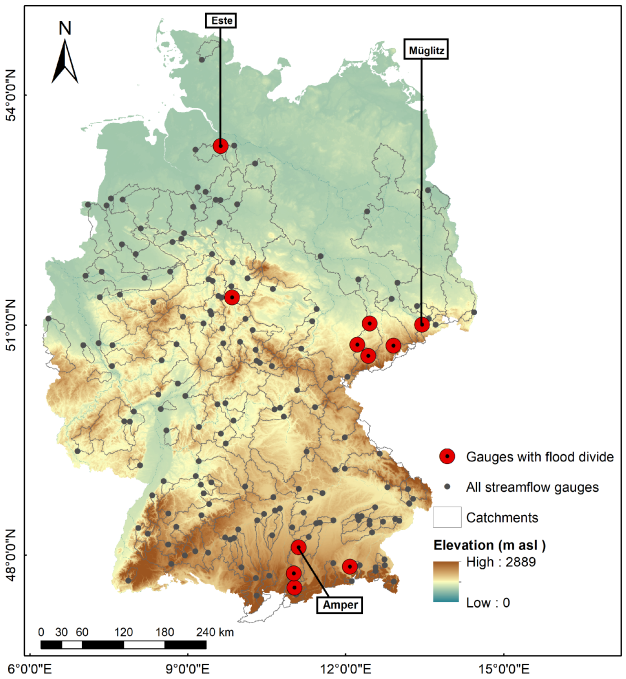


Figure 3.5: Map of the 169 streamflow gauges in Germany (grey dots) initially screened to select the 11 basins exhibiting a flood divide (red filled circles) and used as study cases for the methodology application. The minimum, median, and maximum lengths of the daily streamflow records for 169 basins are 36, 59, and 62 years and 49, 60, 62 for the 11 catchments with flood divides respectively, between 1951 and 2013. The map background indicates the country elevation (m a.s.l.) and catchments are represented with gray outlines.

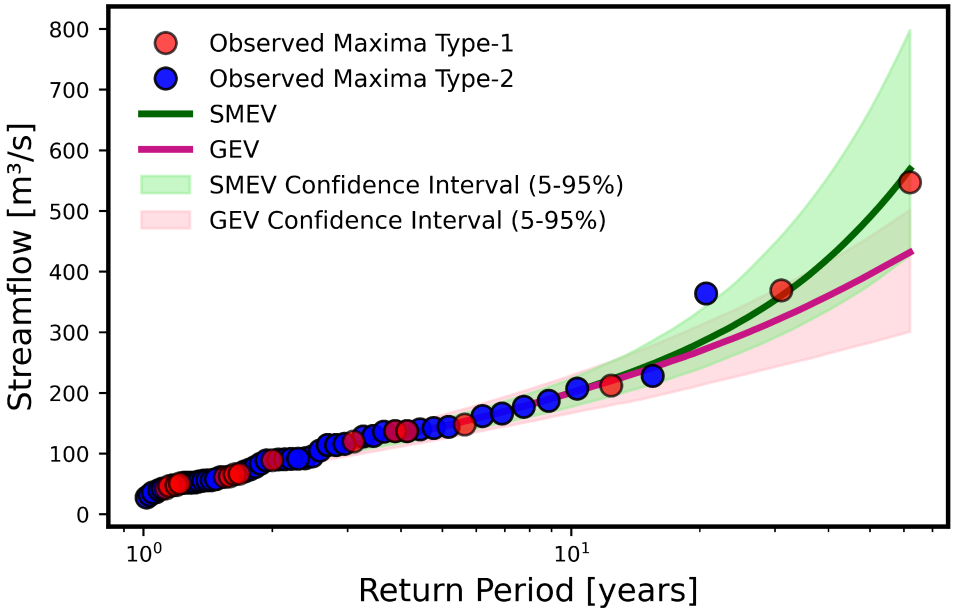


Figure 3.6: Flood magnitude-frequency curves resulting from a mixed SMEV (green curve) and a standard GEV (pink curve) distribution for the Zwickauer Mulde at Zwickau-Poelbitz (Gauge ID: 562070, area = 1034 km^2). Floods are generated from different runoff generation processes and classified into two major event types: rain-on-dry events (red dots – Type 1) and combination of rain-on-wet and snow processes (blue dots – Type 2). Green and pink shaded areas indicate the related confidence intervals (5th–95th percentiles) for SMEV and GEV distributions, respectively.

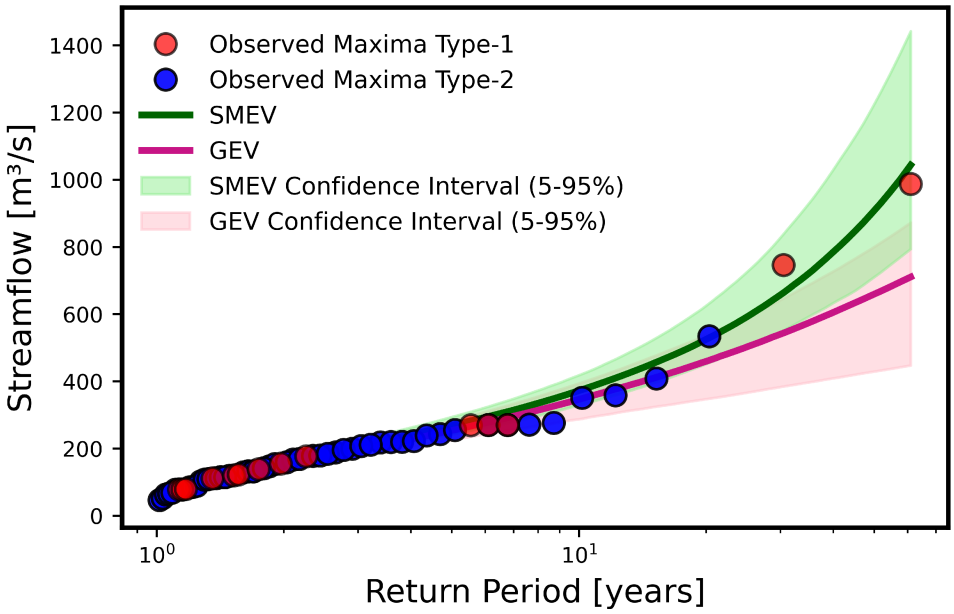


Figure 3.7: Flood magnitude-frequency curves resulting from a mixed SMEV (green curve) and a standard GEV (pink curve) distribution for the Zwickauer Mulde at Wechselburg (Gauge ID: 562115, area = 2107 km^2). Floods are generated from different runoff generation processes and classified into two major event types: rain-on-dry events (red dots – Type 1) and combination of rain-on-wet and snow processes (blue dots – Type 2). Green and pink shaded areas indicate the related confidence intervals (5th-95th percentiles) for SMEV and GEV distributions, respectively.

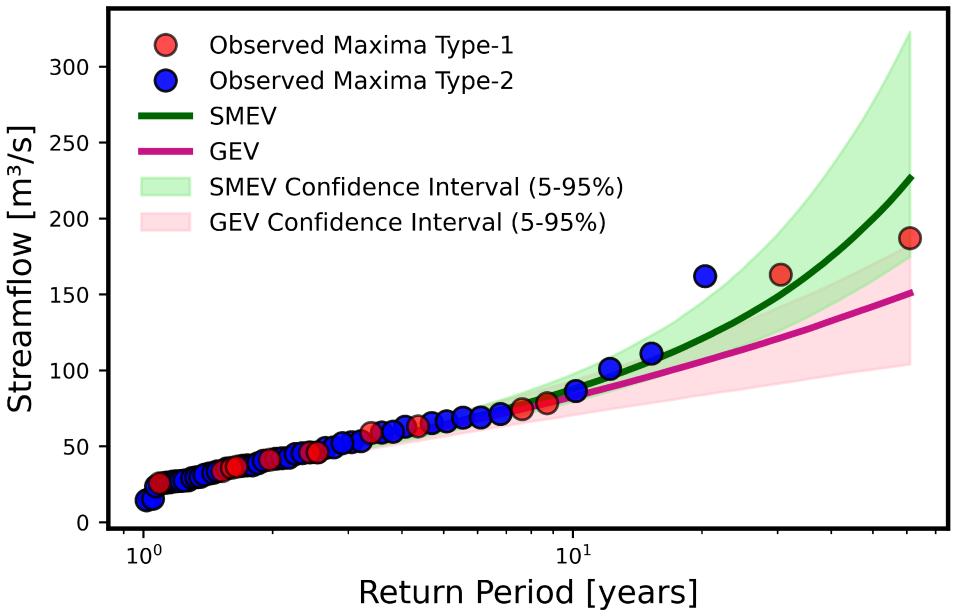


Figure 3.8: Flood magnitude-frequency curves resulting from a mixed SMEV (green curve) and a standard GEV (pink curve) distribution for the Schwarzwasser at Aue (Gauge ID: 563790, area = 362 km^2). Floods are generated from different runoff generation processes and classified into two major event types: rain-on-dry events (red dots – Type 1) and combination of rain-on-wet and snow processes (blue dots – Type 2). Green and pink shaded areas indicate the related confidence intervals (5th-95th percentiles) for SMEV and GEV distributions, respectively.

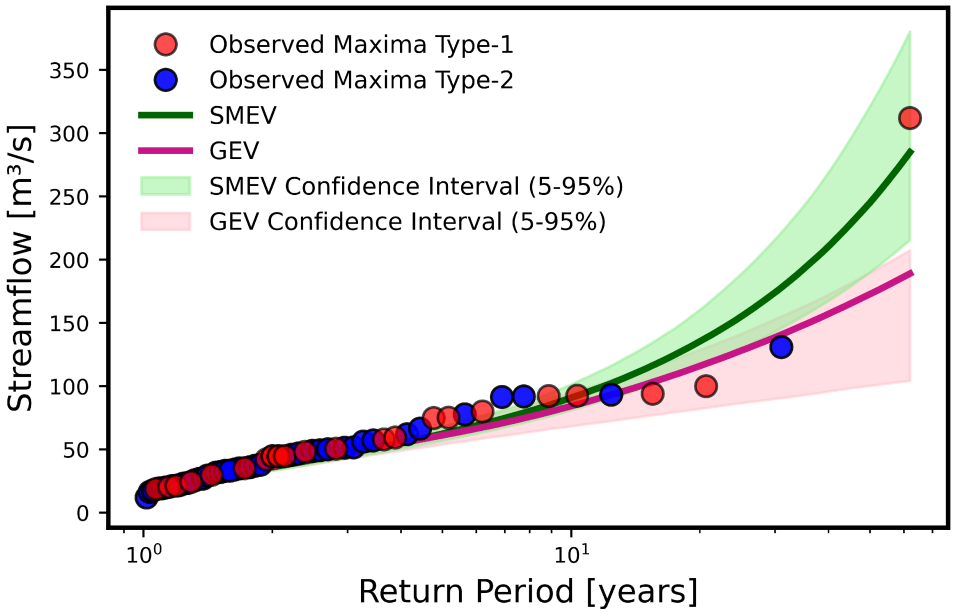


Figure 3.9: Flood magnitude-frequency curves resulting from a mixed SMEV (green curve) and a standard GEV (pink curve) distribution for the Flöha at Pockau (Gauge ID: 568140, area = 385 km^2). Floods are generated from different runoff generation processes and classified into two major event types: rain-on-dry events (red dots – Type 1) and combination of rain-on-wet and snow processes (blue dots – Type 2). Green and pink shaded areas indicate the related confidence intervals (5th-95th percentiles) for SMEV and GEV distributions, respectively.

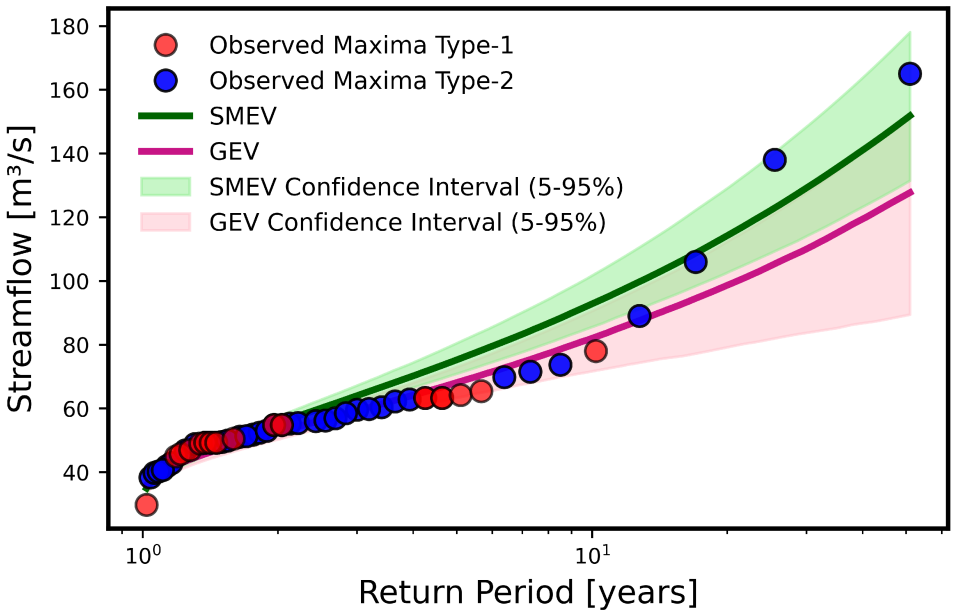


Figure 3.10: Flood magnitude-frequency curves resulting from a mixed SMEV (green curve) and a standard GEV (pink curve) distribution for the Amper at Grafrath (Gauge ID: 16603000, area = 1192 km^2). Floods are generated from different runoff generation processes and classified into two major event types: rain-on-dry events (red dots – Type 1) and combination of rain-on-wet and snow processes (blue dots – Type 2). Green and pink shaded areas indicate the related confidence intervals (5th-95th percentiles) for SMEV and GEV distributions, respectively.

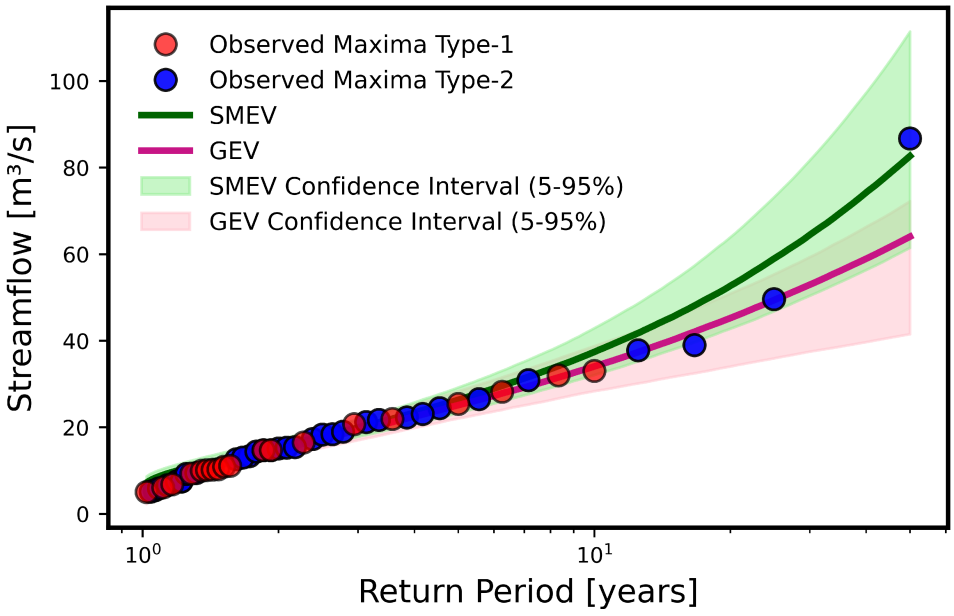


Figure 3.11: Flood magnitude-frequency curves resulting from a mixed SMEV (green curve) and a standard GEV (pink curve) distribution for the Leine at Reckershausen (Gauge ID: 6337541, area = 317 km^2). Floods are generated from different runoff generation processes and classified into two major event types: rain-on-dry events (red dots – Type 1) and combination of rain-on-wet and snow processes (blue dots – Type 2). Green and pink shaded areas indicate the related confidence intervals (5th-95th percentiles) for SMEV and GEV distributions, respectively.

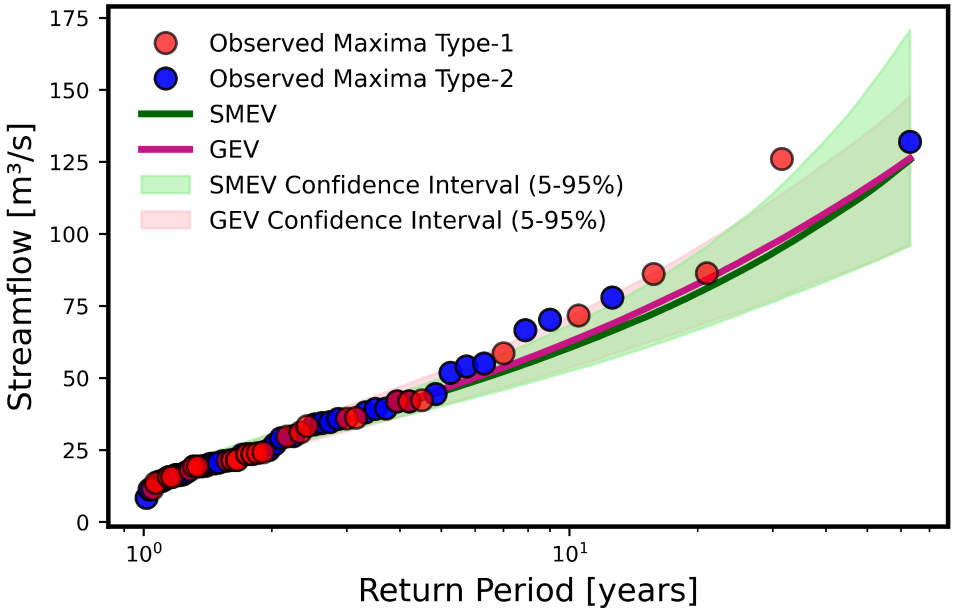


Figure 3.12: Flood magnitude-frequency curves resulting from a mixed SMEV (green curve) and a standard GEV (pink curve) distribution for the Ammer at Oberammergau (Gauge ID: 16610709, area = 114 km^2). Floods are generated from different runoff generation processes and classified into two major event types: rain-on-dry events (red dots – Type 1) and combination of rain-on-wet and snow processes (blue dots – Type 2). Green and pink shaded areas indicate the related confidence intervals (5th-95th percentiles) for SMEV and GEV distributions, respectively

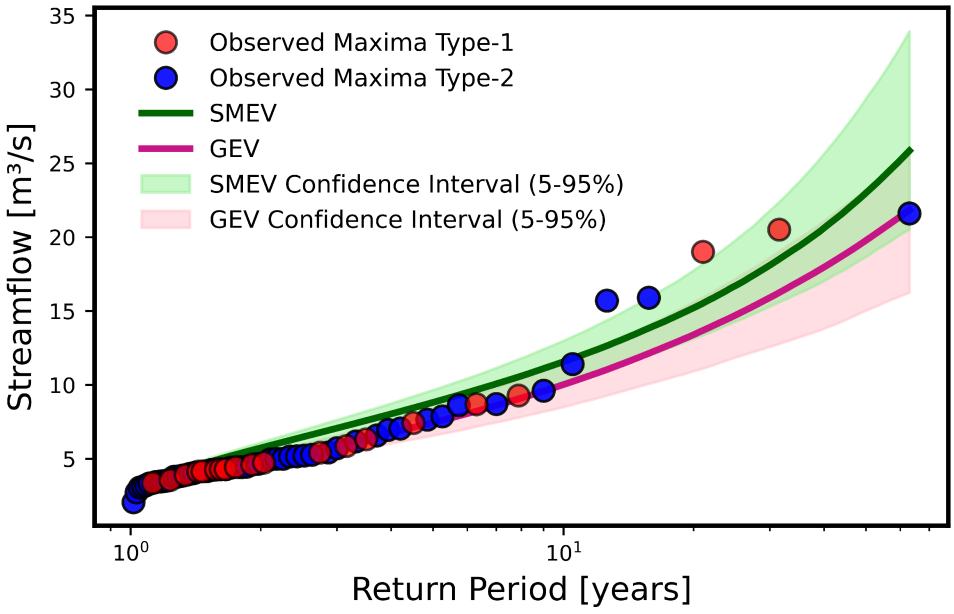


Figure 3.13: Flood magnitude-frequency curves resulting from a mixed SMEV (green curve) and a standard GEV (pink curve) distribution for the Sims at Stephanskirchen (Gauge ID: 18199008, area = 182 km^2). Floods are generated from different runoff generation processes and classified into two major event types: rain-on-dry events (red dots – Type 1) and combination of rain-on-wet and snow processes (blue dots – Type 2). Green and pink shaded areas indicate the related confidence intervals (5th-95th percentiles) for SMEV and GEV distributions, respectively.

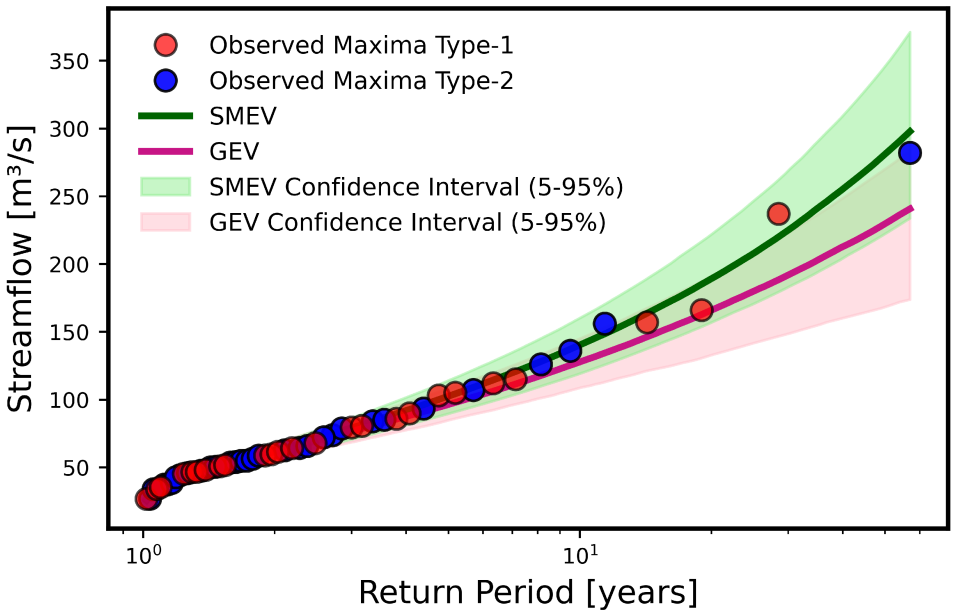


Figure 3.14: Flood magnitude-frequency curves resulting from a mixed SMEV (green curve) and a standard GEV (pink curve) distribution for the Ammer at Peißenberg (Gauge ID: 16612001, area = 291 km^2). Floods are generated from different runoff generation processes and classified into two major event types: rain-on-dry events (red dots – Type 1) and combination of rain-on-wet and snow processes (blue dots – Type 2). Green and pink shaded areas indicate the related confidence intervals (5th-95th percentiles) for SMEV and GEV distributions, respectively.

Chapter statement

This chapter presents a formatted version of the original paper: Mushtaq, S., Miniussi, A., Merz, R., Tarasova, L., Marra, F., & Basso, S. (2023). *Prediction of extraordinarily high floods emerging from heterogeneous flow generation processes.* Geophysical Research Letters, 50, e2023GL105429. https://doi.org/10.1029/2023GL105429 , with permission from the Authors and AGU. (© 2023 The Author(s). Published by AGU.)

Own contribution: The manuscript was primarily authored by Sumra Mushtaq, who was responsible for developing the methods and models, curating the data, conducting the analysis, and interpreting the results under the supervision of Stefano Basso, Ralf Merz, and Arianna Miniussi and with the guidance from Larisa Tarasova and Francesco Marra.

Acknowledgement:

This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project number 421396820 “Propensity of rivers to extreme floods: climate-land scape 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 also well acknowledged. 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). Francesco Marra was partially funded by the CARIPARO Foundation through the Excellence Grant 2021 to the “Resilience” project. Open Access funding enabled and organized by Projekt DEAL

Data Availability Statement

For providing the discharge data for Germany, we are grateful to the Bavarian State Office of Environment (LfU, https://www.gkd.bayern.de/de/fluesse/abfluss/tabellen). Readers can directly access the river discharge data used in this study by clicking on the provided link. We are also thankful to the Global Runoff Data Centre (GRDC) prepared by the Federal Institute for Hydrology (BfG, https://portal.grdc.bafg.de/applications/public.html?publicuser=PublicUserdataDownload/Home) for their valuable data resources. To access the data repository, please use the provided link and click on the “Download by Station” option, select Germany as the country of interest, and locate the table in the top-left corner to access discharge gauges for Germany.

4

Non-Stationary flood frequency analysis by using the Simplified Metastatistical Extreme Value Framework

4.1 Chapter Introduction

Traditional flood frequency analysis estimates the flood quantiles corresponding to specific return periods, under the core assumption that the underlying hydrological process is stationary (Seckin et al., 2011; Villarini, Serinaldi, et al., 2009). This approach is often referred to as the stationary flood frequency analysis (S-FFA), assuming that no temporal changes occur in the statistics of extreme discharges (e.g., mean or variance) and thus employs a time-invariant distribution (Haktanir et al., 2013; Morrison and Smith, 2002). However, there is broad consensus that the magnitude and frequency of extremes such as floods are changing dynamically due to the changes in climate, and intensive human activities (AghaKouchak et al., 2020, Blum et al., 2020; Hirabayashi et al., 2013; Vogel et al., 2011; Yin et al., 2018). Because of these significant shifts in the hydrological regime, such an assumption of stationarity can be violated in flood frequency analysis (FFA), making it necessary to adopt non-stationary approaches, for accurately modeling the flood risks (Luke et al., 2017; Vogel et al., 2011). Nonstationarities in hydroclimatic extremes may be expressed through a significant shift in the mean, variance, or shape of a given time series (Slater et al., 2021). It has been acknowledged that disregarding non-stationarities in frequency analysis can result in significant underestimation or overestimation of high flood quantiles (Huang et al., 2021; Vu and Mishra, 2019; O'Brien and Burn, 2014). Therefore, improving the accuracy of flood frequency estimation is essential to ensure the safety of hydraulic engineering structures and for flood risk management (Merz et al., 2008; Viglione et al., 2013).

Recent studies have utilized non-stationary methods to estimate the changes in the properties of the extremes, such as magnitude, frequency, duration, or timing of flooding (Archfield et al., 2016; Eastoe, 2019; Faulkner et al., 2023; Hecht and Vogel, 2020; Salas and Obeysekera, 2019; Slater and Villarini, 2016). The advantage of these non-stationary methods is their ability to model the changes

in the distribution of extreme events as a function of explanatory variables such as time, climate variability indices or land cover (e.g., Prosdocimi et al., 2015; Silva et al., 2015; Villarini et al., 2009). Non-Stationary flood frequency analysis (NS-FFA) is usually performed by fitting a distribution on the annual maxima (AM) or peak-over-threshold (POT) streamflow observations, which constitutes the basis of traditional extreme value analysis (e.g., Cunnane, 1973; Seckin et al., 2011; Villarini and Smith, 2010).

The Generalized Extreme Value distribution (GEV) has been widely used to estimate the changes in the magnitude of extremes, such as the annual or seasonal series of maximum streamflow (D. Faulkner et al., 2019; Katz et al., 2002; Prosdocimi et al., 2015). However, the GEV distribution relies on the asymptotic assumption that the number of events per block tends to infinity (Gnedenko, 1943; E. J. Gumbel, 1958). This assumption may not be suitable for the hydrometeorological variables such as rainfall and streamflow, as the number of events per block is generally far from infinity. Hydrometeorological records are often limited in length, resulting in short annual maxima or extremes series. The use of these short datasets in non-stationary flood frequency analyses (NS-FFA) based on asymptotic distributions would introduce non-negligible uncertainty to the analysis. Furthermore, these methods focus on a small portion of measurements corresponding to extreme events, disregarding the rest of the available information contained in the bulk of runoff events (Tarasova, Basso, and Merz, 2020), thus typically providing estimations affected by large uncertainty (Miniussi, M. Marani, and Villarini, 2020; Vidrio-Sahagún and He, 2022). To facilitate the broader adoption of NS-FFA in real-world applications, it is essential to reduce this uncertainty, which currently limits its practicality. As a result, the validity of the asymptotic assumptions and the sample size limitations, as highlighted by several studies (e.g., Iliopoulou and Koutsoyiannis, 2019; Lombardo et al., 2019), poses significant challenges to producing reliable estimations in traditional non-stationary frequency analysis.

Non-asymptotic distributions have gained more attention in stationary flood frequency analyses (S-FFA) recently (De Michele, 2019; Lombardo et al., 2019; M. Marani and Ignaccolo, 2015; Marra et al., 2019). Unlike the distributions based on the Extreme Value Theory (EVT), non-asymptotic distributions such as MetaStatistical (MEV) and Simplified MetaStatistical Extreme Value (SMEV) distributions offer significant advantageous in exploiting more available information, as they use the ordinary events (i.e., all the independent peaks) rather than focusing on the extremes (i.e., annual maxima). Furthermore, these methods relax the requirement of asymptotic convergence.The non-asymptotic distributions have shown superiority in using the available data more efficiently, resulting in lower predictive errors and reduced estimation uncertainty in S-HFA, particularly for return periods longer than the sample size (Marra et al., 2019; (Miniussi et al., 2020); Mushtaq et al., 2022; E. Zorzetto et al., 2016). However, the application of non-asymptotic distributions within a non-stationary context remains limited in the current literature.

In this study, I introduce the non-stationary SMEV approach as an alternative method for estimating flood quantiles corresponding to specific return periods in non-stationary context. The rationale behind SMEV is to exploit all the information provided by observational data (M. Marani and Ignaccolo, 2015; Mushtaq et al., 2022; E. Zorzetto et al., 2016), with the objective of better estimating the return period in a wider range of values, not only at the largest extremes that are the focus of extreme value theory

Therefore, this study aims to explore the potential advantages of using the non-asymptotic approach (i.e., the SMEV) within a non-stationary framework, focusing on aspects of accuracy and uncertainty. We implement the non-Stationary flood frequency analysis (NS-FFA) by coupling the probability distribution of floods within a non-stationary structure that governs their temporal evolution. The performance of the non-asymptotic distribution is assessed against its asymptotic counterpart as benchmarks. We also benchmark SMEV estimations with the traditional GEV approach, which constitutes the suggested distribution for flood frequency analysis in Germany (Petrow et al., 2007). The SMEV and GEV distributions were applied under various non-stationary scenarios within the non-asymptotic and asymptotic modeling frameworks, respectively.

The study is structured as follows: Sections 4.2 - 4.6 introduces the study area, the data used, and the methodologies for both non-asymptotic and asymptotic models. Section 4.7 presents the results and discussion, while Section 4.8 concludes with a summary of the findings.

4.2 Materials and methods

4.2.1 Study area and data

The initial step involved analyzing the 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 4.1a).

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 4.1b). 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).

4.2.2 Identification of trends in Annual Maxima series (AMs) and ordinary events

In the first step of our analysis, I examined the trends in the annual maxima of streamflow data across 169 catchments in Germany. I applied the non-parametric Mann-Kendall test at the 5% significance level to the annual maxima series (AMs) (Kendall, 1975; Mann, 1945) and evaluated the presence of temporal patterns in the streamflow data (Slater et al., 2021). The feasibility of the stationary assumption is assessed considering the presence of statistically significant non-stationary

signatures, such as deterministic temporal trends in the mean or standard deviation of the AMs. By analysing the annual maxima series, we identified 40 river basins with significant trends in their mean or standard deviation in annual maxima. In the second step, we applied the likelihood ratio (LR) test to these 40 river basins that exhibited significant trends to determine whether a one-parameter or two-parameter non-stationary model was required. The LR test is a widely used statistical method for assessing whether the inclusion of additional parameters significantly improves a model's fit by comparing the likelihood of two maximum likelihood estimates (Zakaria et al., 2021). Specifically, it was used to evaluate the statistical significance of the time dependence of the model parameters and to determine whether a stationary or non-stationary model should be applied. Deciding whether to use stationary or non-stationary methods to model extreme events is a major challenge for practitioners, especially with the arising uncertainty due to climate change (e.g., (Faulkner et al., 2019, 2023).

I performed the LR test for both the GEV and SMEV across all 40 catchments under different scenarios (as detailed Section 4.2.2), using the p-values at a 5% significance ($p < 0.05$) to determine whether to apply a one- or two-parameter non-stationary model. This second-step identification method provides a more robust approach for detecting trends in time series, as it incorporates all available ordinary peaks in addition to annual maxima.

In contrast, analyzing trends in the annual maxima series relies on fewer data points, making it more difficult to detect such signals. Luke et al. (2017) highlighted that applying Mann-Kendall trend tests to peak discharges alone is insufficient to justify the selection of non-stationary model for prediction. By using the two-step trend identification method, our analysis reveals that 19 out of the 40 catchments had statistically significant results, leading to the selection of either one-parameter or two-parameter non-stationary non-asymptotic models (Vidrio-Sahagún, Ruschkowski, et al., 2024), as detailed in Sections 4.2.2 and 4.3. Figure 4.1 displays the locations of 19 streamflow gauges exhibiting non-stationary behavior.

4.3 Non Asymptotic Models

We adopt the Simplified Metastatistical Extreme value (SMEV) approach to model the magnitude of extreme floods in a non-stationarity context (Vidrio-Sahagún and He, 2022). In non-stationarity approach,the SMEV is allowed to vary over time by using the covariate-dependent distribution parameters, where the parameters of the distribution are expressed as a function of covariate (e.g., time), This approach is sometimes referred to as a time-varying moments model (Zhang et al., 2018). In the SMEV approach, the interannual variability of F(x; θ) and n are neglected and the temporal dependence of the distribution parameters can be deterministically described (Marra et al., 2019); (Mushtaq., 2023). The cumulative distribution function for the SMEV distribution is define as (Marra et al., 2019).

$$\zeta_{\mathrm{SMEV}}(x,t) = \prod_{i=1}^{S} [F_i(x;\theta_{i,t})]^{n_i} \tag{4.1}$$

where S is the number of different types of x, which is set to 1 in this study; $\theta_{i,t} = (\lambda_{i,t}, \beta_{i,t})$, is the time-dependent θ of the ith x type; and n_i is the time-invariant average n of the ith type. $F_i(x;\theta_{i,t})$ is the cumulative distribution function for the ordinary distribution.

The non-asymptotic SMEV adopted a parsimonious linear nonstationary structures, in which the scale parameter (λ_t), shape parameter (β_t) and both the scale and shape parameters (λ_t, β_t) are expressed as a linear functions of the covariate time t. In this study, three different scenarios of non-asymptotic models are implemented, which are selected based on the likelihood ratio test : i) a non-stationary SMEV, with the shape parameter kept constant and the scale parameter that changes linearly in time; ii) a non-stationary SMEV model with the scale parameter kept constant and the shape parameter that changes linearly in time; iii) a fully non-stationary SMEV model in which both the parameters are allowed to change linearly in time.

These non-stationary models are denoted as $\text{SMEV}_{[1,0]}$ $\text{SMEV}_{[0,1]}$ and $\text{SMEV}_{[1,1]}$ respectively, and are given by:

$$\text{SMEV}_{[1,0]} \sim \text{SMEV}(\lambda_t, \beta, \tilde{n}); \quad \begin{cases} \lambda_t = \lambda_0 + \lambda_1 t \\ \beta = \text{constant} \\ \tilde{n} = \text{constant} \end{cases} \tag{4.2}$$

$$\text{SMEV}_{[0,1]} \sim \text{SMEV}(\lambda, \beta_t, \tilde{n}); \quad \begin{cases} \beta_t = \beta_0 + \beta_1 t, \\ \lambda = \text{constant}, \\ \tilde{n} = \text{constant}. \end{cases} \tag{4.3}$$

$$\text{SMEV}_{[1,1]} \sim \text{SMEV}(\lambda_t, \beta_t, \tilde{n}); \quad \begin{cases} \lambda_t = \lambda_0 + \lambda_1 t, \\ \beta_t = \beta_0 + \beta_1 t, \\ \tilde{n} = \text{constant}. \end{cases} \tag{4.4}$$

Where, λ and β represent the scale and shape parameters of the selected ordinary distributions (e.g., Gamma or Log-normal distributions, as detailed in Section 4.5). Under nonstationarity conditions, λ and β were modeled as linear functions of time. Specifically, λ_0 and λ_1 and β_0 and β_1 are the regression coefficients of λ and β, respectively.

4.4 Asymptotic Models

The cumulative GEV function of the random variable, y, which corresponds to the annual maxima of every year is defined as (e.g., Coles, 2001; Hosking, 1990).

$$F(y;\boldsymbol{\theta}) = F(y;\xi,\alpha,\kappa) = \begin{cases} \exp\left(-\left[1+\kappa\frac{(y-\xi)}{\alpha}\right]^{\frac{1}{-\kappa}}\right), & \kappa \neq 0, \\ \exp\left(-\exp\left(-\frac{(y-\xi)}{\alpha}\right)\right), & \kappa = 0. \end{cases} \tag{4.5}$$

Where the parameter vector $\boldsymbol{\theta}$ comprises κ, α, and ξ, representing the shape, scale, and location parameters, respectively. The shape parameter κ determines the tail behavior of the distribution, classifying the GEV into three subfamilies:

- Extreme Value (EV) type I distribution when $\kappa = 0$,
- Extreme Value (EV) type II, when $\kappa > 0$ (indicating heavy-tailed behavior), and
- Extreme Value (EV) type III, when $\kappa < 0$ (indicating an upper-bounded distribution).

In this study, two non-stationary models based on the GEV distribution were considered, each incorporating different non-stationary structures. In the first model, only the location parameter ξ_t varies linearly with the covariate t, while in the second model, both the location and scale parameters (ξ_t, α_t) change as a linear function of the covariate t.

The shape parameter κ was assumed constant, considering its difficult estimation due to constraints in the data length and, consequently, its unrealistic time-variant characterization (Coles, 2001; Salas and Obeysekera, 2019). These nonstationary models are denoted as $\text{GEV}_{[1,0,0]}$ and $\text{GEV}_{[1,1,0]}$, respectively, and are given by:

$$\text{GEV}_{[1,0,0]} \sim \text{GEV}(\xi_t, \alpha, \kappa); \quad \begin{cases} \xi_t = \xi_0 + \xi_1 t, \\ \alpha = \text{constant}, \\ \kappa = \text{constant}. \end{cases} \tag{4.6}$$

$$\text{GEV}_{[1,1,0]} \sim \text{GEV}(\xi_t, \alpha_t, \kappa); \quad \begin{cases} \xi_t = \xi_0 + \xi_1 t, \\ \alpha_t = \alpha_0 + \alpha_1 t, \\ \kappa = \text{constant}. \end{cases} \tag{4.7}$$

Where the parameters ξ, α, and κ, namely the location, scale, and shape parameters of the GEV distribution, respectively. The variations in ξ_t and α_t allow these models to capture potential time-dependent trends in the data, such as shifts in the central tendency or spread of extreme values.

4.5 Selection of ordinary distributions and Parameters Estimation

We applied the method proposed by Lang et al. (1999) to identify the independent ordinary peaks from the streamflow records, a prerequisite for implementing the Simplified Metastatistical Extreme Value (SMEV) approach, as similarly performed by Miniussi et al. (2020) and Mushtaq et al. (2022). This method involves to identify one peak within each time block of length T = 5days + log(A), where A is the basin area in square miles, is the minimum lag time required by the independence criterion and the value of T is rounded off to the nearest integer number. The number of independent ordinary peaks identified for each catchment ranged from 370 to 933, with a median value of 796. Once the ordinary peaks are selected, we must choose a distribution to describe their statistical properties. For the selected 19 catchments, we applied the best fitted ordinary distribution to be either the Gamma or Log-Normal distribution, following the methodology implemented by (Mushtaq et al., 2022). The parameters of stationary and non-stationary asymptotic and non-asymptotic models were estimated by using Maximum Likelihood (Hosking, 1990) by fitting them to the Annual Maxima series (AMs) and to the sample of ordinary peaks, respectively. Moreover, the practical applications of the non-asymptotic models in the stationary frequency analysis have shown that the ordinary events of lower magnitude negatively impact the evaluation of the distribution's right tail by diverging the estimates (Marra et al., 2019); Miniussi and Marra, 2021). This issue has been solved by left-censoring, by ignoring the magnitudes of the small ordinary events while retaining their existence in the distribution (Marra et al., 2019, Marra et al., 2020).

4.6 Model performance assessment

In order to evaluate the estimation accuracy of both stationary and non-stationary asymptotic and non-asymptotic models, we employ the non-dimensional error metric (E. Zorzetto et al., 2016), which quantifies the performance of different models. We calculate the non-dimensional error: $\varepsilon = \frac{x_{\text{est}} - x_{\text{obs}}}{x_{\text{obs}}}$ between the estimated and observed maxima, where x_{est} and x_{obs} represent the estimated and observed maxima, respectively. It is important to note that we quantify errors with respect to the empirical values, as the performance of an extreme value approach cannot be evaluated when quantiles with return periods higher than the empirical ones are taken into account. We adopt a parametric bootstrap method to calculate the confidence level in this framework, and the uncertainty bounds are calculated by using the 100 bootstrap samples.The parametric bootstrap is considered superior to the non-parametric bootstrap, as it typically provides empirical coverage frequencies that more closely align the pre-selected nominal confidence level and and is less likely to underestimate uncertainty (Kyselý, 2008; Panagoulia et al., 2013). The uncertainty is evaluated by using the average bandwidth (AW), where a higher AW value indicates greater uncertainty in the model. The uncertainty performance of the non-asymptotic distributions is assessed against their asymptotic counterparts as benchmarks in the assessment.

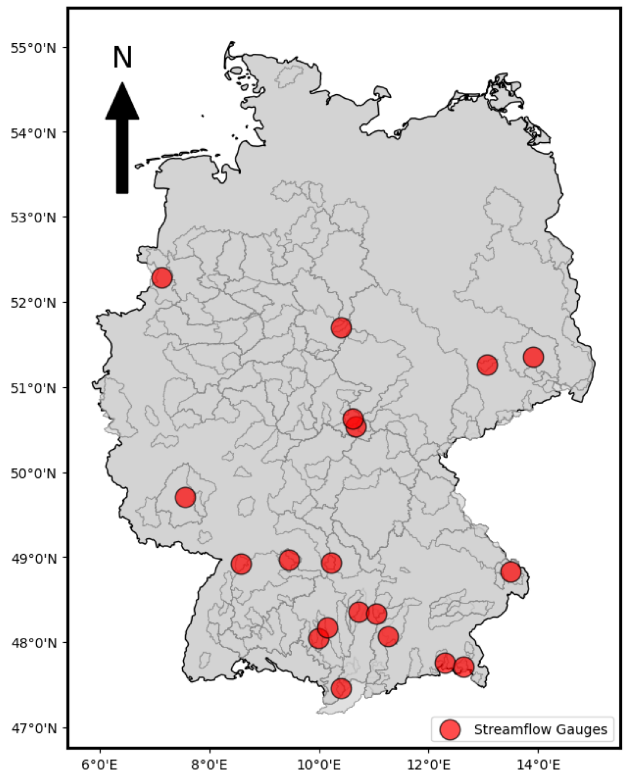
$$\text{Average Bandwidth (AW)} = \frac{1}{n} \sum \left(m_r^U - m_r^L \right) \tag{4.8}$$

Where m_r^U and m_r^L are the upper and lower uncertainty bounds associated with the return period. A model with a lower value of Average Bandwidth is less uncertain.

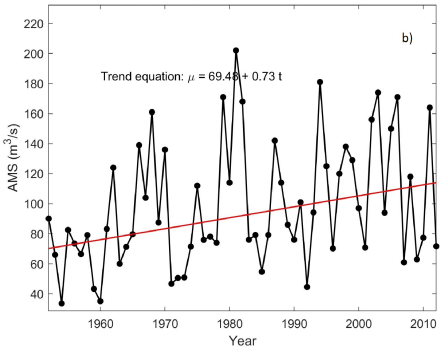
4.7 Results and Discussions

In this study, we examine the river basins that exhibit non-stationarity in their mean and variance as detailed in section 4.2.2. We implemented the non-stationary SMEV and GEV distributions to model the flood frequency curves, incorporating the linear dependence on time as a covariate. To better illustrate the properties of these floods, I selected 19 case studies exhibiting the non-stationary behavior (displayed in Figure 4.1). As shown in Figure 4.1, the two representative gauging stations display significant temporal trends in the mean of their AMSs. These long-term trends indicate the presence of non-stationarity in extreme events. Specifically, the trends suggest that streamflow maxima are increasing in some cases (Figure 4.1b) and decreasing in others (Figure 4.1c) over time. This behavior implies potential changes in hydrological conditions or other influencing factors within the system, highlighting the importance of accounting for non-stationarity in flood frequency analysis.

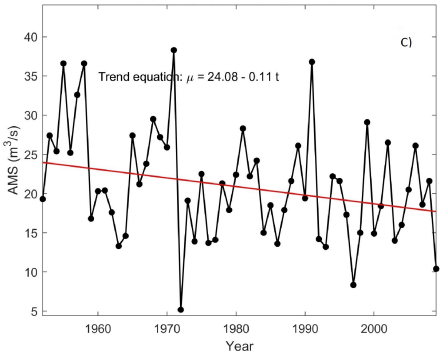
The flood frequency curves for both non-stationary and stationary SMEV and GEV models, for a representative case study, located at Murr River at Murr (Gauge-ID: 6335675, area = 504 km^2), are presented in Figure 4.2. Red dots represent the empirical frequencies of the sample of annual maxima. The flood frequency curves for the non-stationary SMEV and GEV are calculated by integrating the values across the entire available time period (represented by solid lines in Figures 4.2a and 4.2c) and uncertainty bounds are calculated by using the 100 bootstrap samples (shaded areas in Figures 4.2a, 4.2b, 4.2c and 4.2d).



(a) Study Area Map



(b) Increasing Trend in AMS



(c) Decreasing Trend in AMS

Figure 4.1: (a) Map of 19 streamflow gauges in Germany (red dots) selected from an initial screening of 169 gauges. These catchments exhibited non-stationarity in the mean and standard deviation and were used as case studies for the methodology. The daily streamflow records for these 19 catchments, spanning the period from 1951 to 2013, have a minimum, median, and maximum length of 50, 60, and 62 years, respectively. Catchment boundaries are shown with gray outlines. Figure 4.1b and Figure 4.1c show the trend estimator for the two representative basins. (b) shows the trend in annual maximum streamflow (AMS) with an increasing trend equation, and (c) shows a basin with a decreasing trend in AMS over the study period from 1951 to 2013. Trend lines and corresponding equations highlight the direction and magnitude of changes in the streamflow over time.

These analysis shows that accounting for temporal changes in the distribution properties allows the non-stationary SMEV to provide improved estimates of upper tail quantiles (Figure 4.2a; values greater than 10 years return period) compared to stationary SMEV (Figure 4.2b), highlighting the potential utility of incorporating ordinary events (Vidrio-Sahagún and He, 2022; Vidrio-Sahagún, Ruschkowski, et al., 2024). In contrast, non-stationary GEV tends to overestimate the largest floods (results for the NS-GEV and S-GEV are shown for comparison in Figures 2c and 2d). The shaded areas in Figures 4.2a-4.2d show the 95% confidence intervals calculated via bootstrap for stationary and non-stationary SMEV and GEV distributions. The NS-SMEV exhibits slightly greater uncertainty compared to the S-SMEV. This is primarily because the NS-SMEV involves estimating a larger number of parameters than the S-SMEV, which inherently leads to slightly higher uncertainty.

To compare the performance of the non-stationary SMEV and GEV distributions across all the return periods, we present the Quantile-Quantile (Q-Q) plots of the normalized observed versus estimated maxima for the 19 river basins, exhibiting the non-stationarity (Figure 4.4). Figure 4.3a shows the results for the NS-SMEV (blue dots), while Figure 4.3b depicts NS-GEV (pink dots). The

dashed line represents the 1:1 line, between observed and estimated values. Both the non-stationary SMEV and GEV demonstrate similar performance across all return periods.

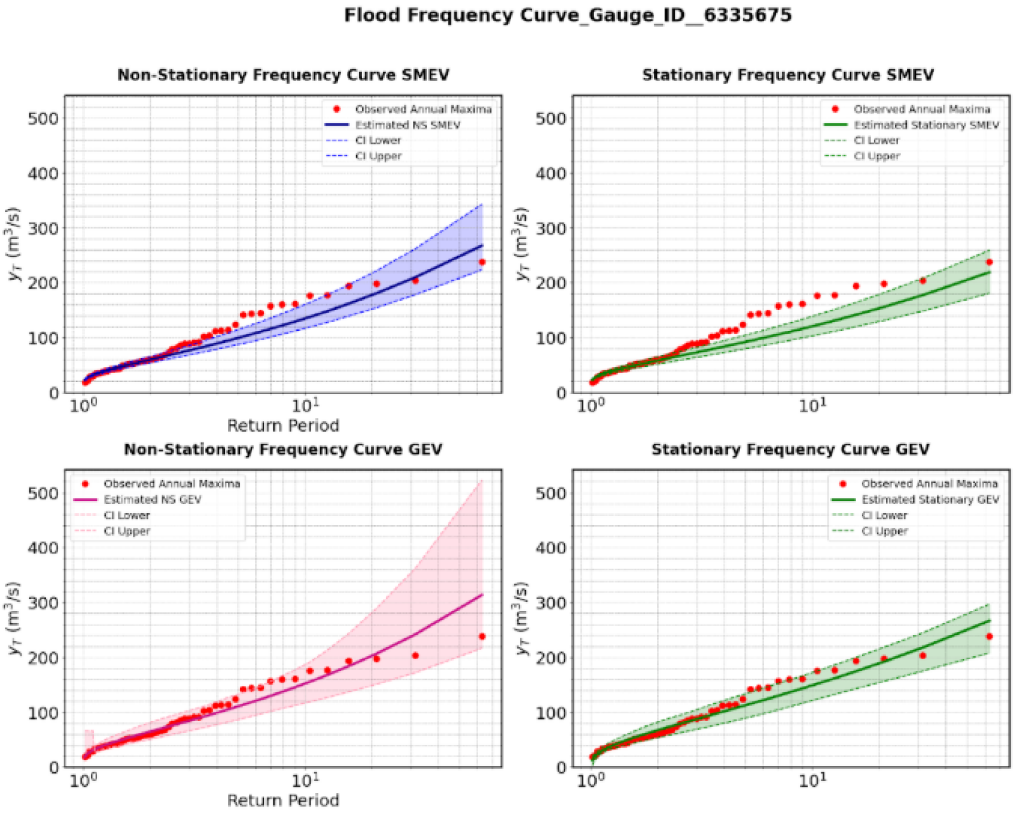


Figure 4.2: Comparison of Non-Stationary and Stationary Frequency Curves plots for SMEV (a, b) and GEV Models (c,d). The first two rows display frequency curves for SMEV (top row, Gauge-ID: 6335675) and GEV (bottom row, Gauge-ID: 6335675) models. The left column shows the non-stationary models (blue for SMEV and pink for GEV) with observed annual maxima and confidence intervals (CI), while the right column represents the stationary models (green) for both SMEV and GEV with corresponding CI bounds.

However, the GEV distribution shows greater consistency across the return periods, while the NS-SMEV shows slightly higher variability and a tendency to overestimate maxima, particularly for small to moderate values, as observed in Figure 4.3a. Figure 4.3c presents boxplots of the non-dimensional error across different models for all the return periods, illustrating the spread and median values of the error. The dashed line at zero represents the target error value. The models compared include NS-SMEV (Non-Stationary SMEV), S-SMEV (Stationary SMEV), NS-GEV (Non-Stationary GEV), and S-GEV (Stationary GEV). The S-SMEV provides results comparable to NS-SMEV, but with slightly more variability, indicating that both models may be less reliable when considering all return periods. In contrast, the S-GEV and NS-GEV generally perform equivalently, as their interquartile ranges appear similar, although NS-GEV exhibits slightly higher variability across all the return periods (Figure 4.3c). However, when we focus our analysis on quantiles with return period greater than 10 years (Figure 4.4), which are the most relevant for flood hazard assessment, NS-GEV appears to be biased, as it tends to underestimate the maxima. Figure 4.4 shows the Quantile-Quantile (Q-Q) plots for all 19 river basins, focusing on high return periods (T $>$ 10 years).

For NS-SMEV, most of the points cluster relatively close to the 1:1 line, indicating a reasonable agreement between observed and estimated values, though some scatter is still present (Figure 4.4a).

In contrast, NS-GEV exhibits more pronounced scatter, especially at higher maxima (Figure 4.4b), indicating greater deviation for extreme values.Figures 4.4c and 4.4d present boxplots of the relative error for the four different distributions across the 19 cases, focusing on return periods of T > 10 years and T > 20 years, respectively.

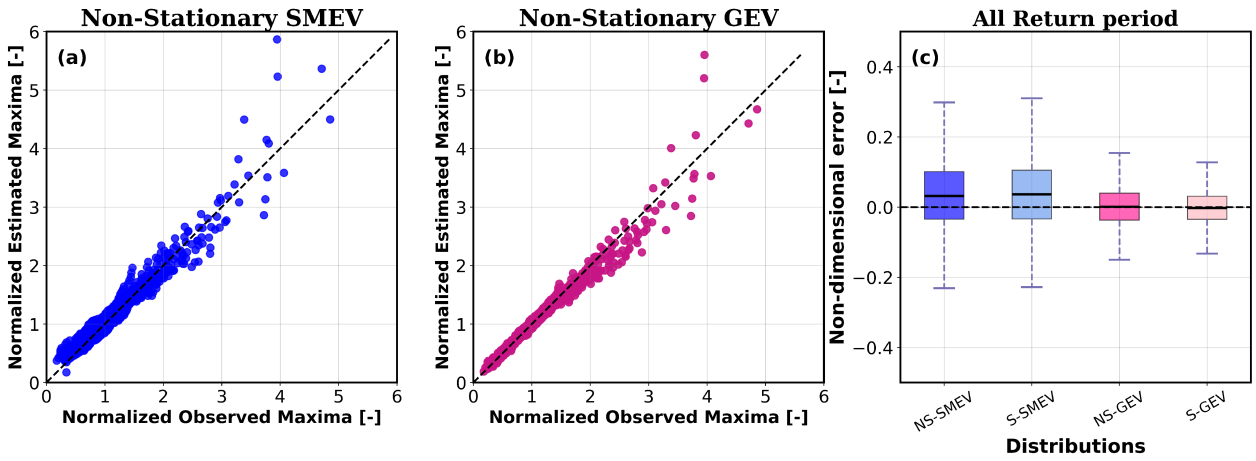


Figure 4.3: (Left) Comparison of normalized observed versus estimated maxima for (a) Non-Stationary SMEV and (b) Non-Stationary GEV, with the 1:1 line (black dotted line). Panel (c) presents box plots of the non-dimensional error across different models (NS-SMEV, S-SMEV, NS-GEV, and S-GEV) for all return periods for all 19 river basins, highlighting the performance differences between stationary and non-stationary models.

Figures 4.4c and 4.4d present boxplots of the relative error for the four different distributions across the 19 cases, focusing on return periods of T > 10 years and T > 20 years, respectively. The results show that NS-SMEV improves quantile estimation and is, on average, less biased compared to the other models. For example, for the return period T >10 years, NS-SMEV demonstrates a smaller average bias compared to GEV-based models (Figure 4.4c), while S-SMEV provides similar results but with slightly higher variability. Although GEV-based distributions exhibit lower variability overall, they tend to underestimate maxima, as evident in the Q-Q plot (Figure 4.4b). This underestimation becomes more pronounced for large return periods, likely due to the limitations of small sample sizes. For even larger return periods (T > 20 years), the advantage of NS-SMEV becomes more apparent. Notably, SMEV-based models generally outperform GEV-based models in terms of bias , particularly for large return periods. These results suggest that while GEV tends to underestimate maxima at large return periods, NS-SMEV provides reliable estimates, particularly for these extreme events.These findings are particularly significant, as both underestimation and overestimation present critical challenges when estimating high return levels of hydrological variables. Underestimation poses a serious concern due to the potential for underestimating risks (Papalexiou et al., 2013), while overestimation can lead to substantial economic costs associated with implementing design values which are larger than needed (Cho et al., 2004).

To evaluate uncertainty, we calculated the average bandwidth (as described in Section 4.6) for each basin and presented a combined boxplot for all 19 cases (Figure 4.5). Among the models, NS-GEV exhibited the greatest variability, with a wide range of bandwidth values. In contrast, NS-SMEV demonstrated the most consistent performance, with the lowest variability and narrower bandwidth. Overall, SMEV-based models showed smaller bandwidths and less variability across cases compared to GEV-based modelsNotably, NS-GEV displayed higher bandwidth values on average,

indicating a larger spread in uncertainty. This increasing variability was particularly pronounced for higher return periods (e.g., T > 10 years, T > 20 years) as presented in Figure 4.5a, and 4.5b respectively. However, NS-SMEV exhibited comparable uncertainty to S-SMEV and showed clear improvements over both NS-GEV and S-GEV.

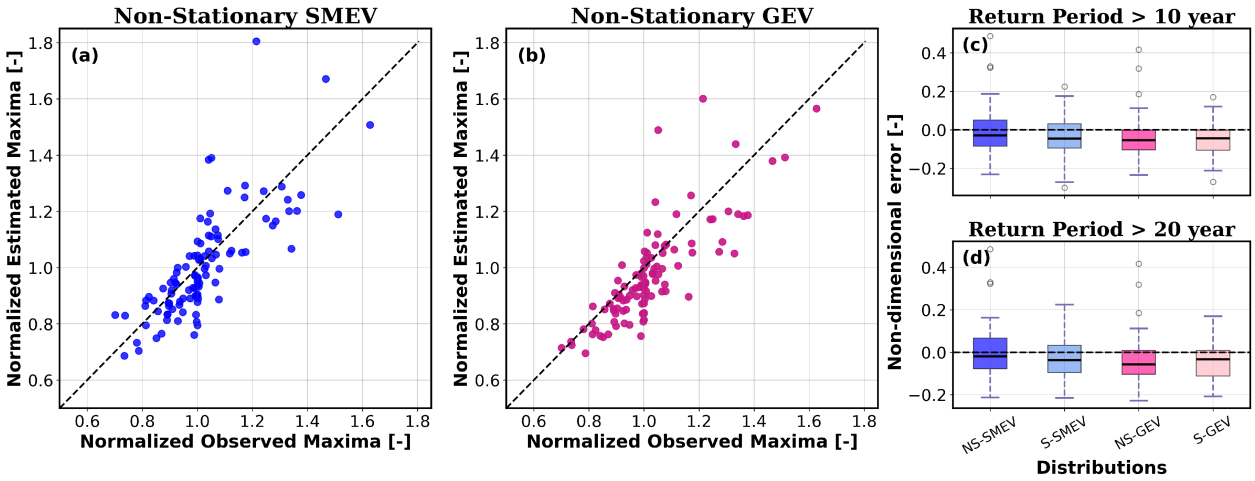


Figure 4.4: Comparison of normalized observed versus estimated maxima for return periods greater than 10 years: (a) Non-Stationary SMEV and (b) Non-Stationary GEV models, with the 1:1 line indicating perfect agreement. Panel (c) presents box plots of the non-dimensional error for the same return periods across different models for all 19 river basins togeher (NS-SMEV, S-SMEV, NS-GEV, and S-GEV), illustrating the performance differences between stationary and non-stationary models for long return periods.

The lower uncertainty of the Stationary SMEV and GEV-based models compared to their counterparts, is consistent with the notion that adding more parameters to the NS-FFA models would increase the prediction uncertainty (Vidrio-Sahagún and He, 2022). In terms of overall uncertainty, the SMEV models consistently produced lower average bandwidths (AWs) than the GEV-based models and demonstrated greater robustness, as evidenced by their narrower ranges (Figure 4.5). These findings highlight the advantages of SMEV in achieving reduced uncertainty and improved reliability.

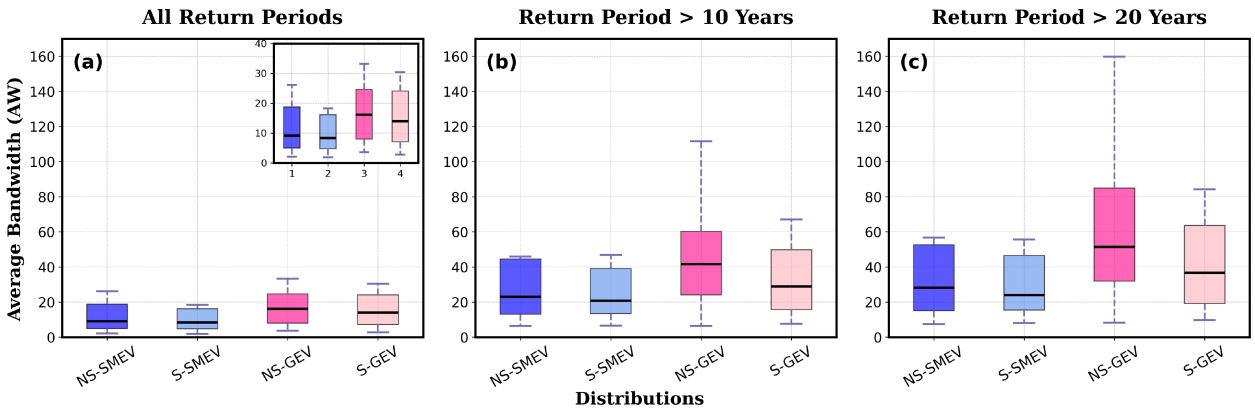


Figure 4.5: Boxplots comparing the average bandwidth (AW) for four different types of distributions: NS-SMEV, S-SMEV, NS-GEV, and S-GEV. under varying return periods. (a) shows the average bandwidth for all return periods, with a zoomed inset highlighting lower bandwidth values. (b) shows the average bandwidth for the return periods greater than 10 years, and (c) represents return periods over 20 years. The median AW is indicated by the bold horizontal lines within each box, while the interquartile range (IQR) and whiskers represent the spread and variability of the data. .

The assessment from the uncertainty point of view showed prominent superiority of the SMEV over GEV models. In particular, the superiority of the SMEV-based models compared to the GEV-based models becomes more prominent as the return period T increases (Figure 4.5b and 4.5c), which is consistent with previous findings for the S-FFA (e.g., Schädler et al., 2012). The reduced uncertainty of the SMEV-based models would be primarily ascribed to the increased sample size due to the use of ordinary events for estimating the distribution parameters (Vidrio-Sahagún and He, 2022). The disadvantage of the GEV model from the uncertainty perspective further argues that the implementation of this model might not be promising for the NS-HFA. Thus, the non-asymptotic approach could potentially advance the NS-FFA in improving their overall performance by harnessing the information of the ordinary events.

4.7.1 Flood magnitude changes

The magnitude of flood flow estimates for different return periods for the non-stationary SMEV shows both increased and decreased in the different regions of Germany (Figure 4.6).

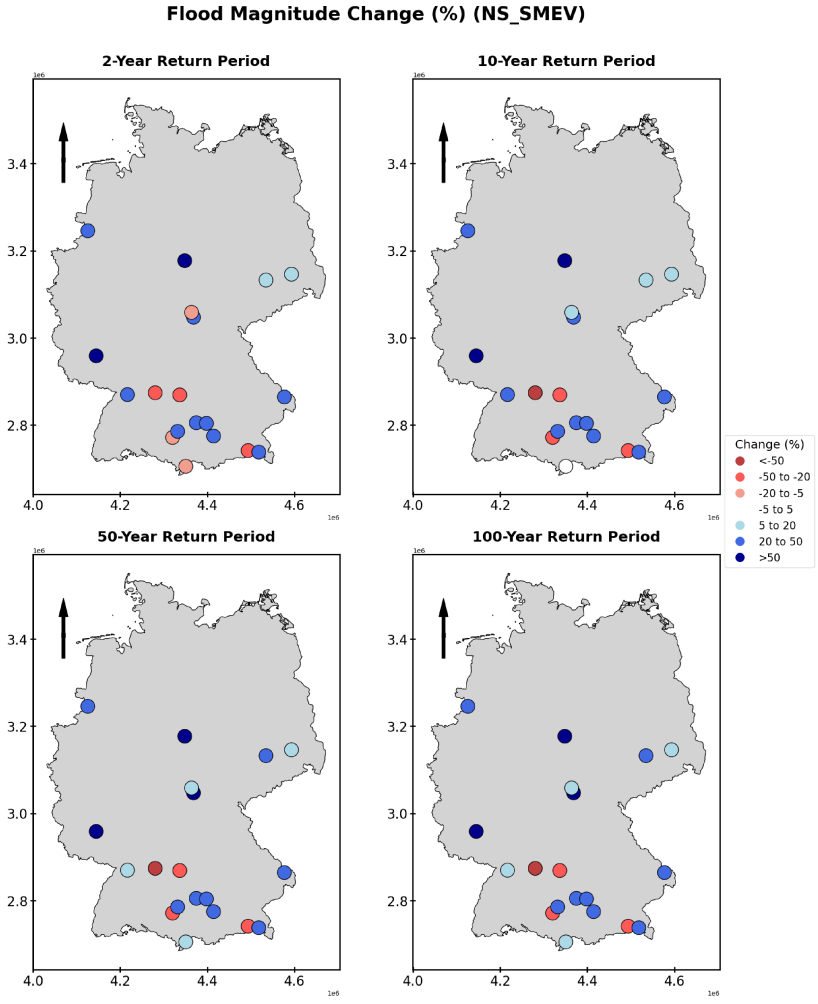


Figure 4.6: Flood magnitude changes (%) for 2-year, 10-year. 50-year and a 100-year return period across Germany, by using the NS-SMEV model. The color of each circle represents the percentage change in flood magnitude, with blue color indicating increases and red indicating decreases. Darker shades correspond to more extreme changes, with dark blue representing increases greater than 50% and dark red representing decreases greater than 50%.

The map illustrates the percentage change in flood magnitude for different return periods (2-year, 10-year, 50-year, and 100-year) across various locations in Germany, Some locations exhibit increases in flood magnitudes (represented by blue circles), while others show decreases (represented by red circles). The southern and southwestern regions of Germany exhibit greater variability in flood magnitude changes, with both increases and decreases observed. In contrast, the eastern regions, with fewer data points, show smaller and less variable changes. The non-stationary SMEV model highlights regional differences in flood risks, indicating potential increases in southern areas while other regions may experience reductions in flood magnitudes.

The spatial distribution of changes varies slightly by return period. For shorter return periods (2- and 10-year), most locations show modest changes, with a mix of reductions (red) and increases (blue). For longer return periods (50- and 100-year), there is a slight increase in flood magnitudes, particularly in the northern and eastern parts of Germany, while decreases are concentrated in the southern regions.

The direction and magnitude of change at each site inherently depend on factors such as the choice of beginning and end dates (Harrigan et al., 2017), dominant flood-generating mechanisms, and human influences. Changes in extreme floods are likely driven by a combination of climate variations, and anthropogenic activity (Blum et al., 2020; Hirabayashi et al., 2013; Prosdocimi et al., 2015; Yin et al., 2018). Future studies should explore the underlying causes of this non-stationarity, though such an analysis is beyond the scope of this work.

4.8 Chapter conclusions

In this study, we explored the potential application of the non-stationary SMEV, which incorporates ordinary events rather than relying on extreme events. The use of the non-asymptotic SMEV approach in the context of Non-Stationary Flood Frequency Analysis (NS-FFA) has been limited to date. Therefore, in this study, we employ the non-stationary SMEV across 19 river basins in Germany, exhibiting the non-stationary behavior. As a benchmark, we used the asymptotic GEV-based models, which are widely used in NS-FFA and rely on Extreme Value Theory (EVT). Both the non-stationary SMEV and GEV models incorporated a temporal covariate (i.e., time). We fitted Gamma and Log-normal distributions for ordinary events, allowing the scale and shape parameters to vary linearly over time. The performance assessment revealed that the GEV model tended to underestimate flood quantiles and exhibited greater bias when estimating quantiles for return periods exceeding 10 years. In contrast, SMEV-based models demonstrated better accuracy and lower uncertainty for higher return periods. However, NS-GEV exhibited high varibility in terms of uncertainty, raising concerns about its reliability for predicting extreme floods. Our analysis also showed that flood magnitude changes vary across Germany, with some areas experiencing reductions while others see significant increases. Southern Germany displayed the most notable variability, with both increases and decreases in flood magnitudes. These results show that the SMEV-based models could advance the NS-FFA by improving the accuracy and reducing the uncertainty in a non-stationary context.

Future research on the practical implementation of SMEV-based models should explore more complex non-stationary patterns, such as non-linear relationships and the integration of additional physical covariates. Efforts should also aim to refine the attribution of non-stationarity and expand the

application of these models to a broader range of hydrometeorological variables. Such advancements would further enhance the use of SMEV-based models in Non-Stationary Flood Frequency Analysis. The key findings of the study are as follows:

1. Superiority of Non-Stationary SMEV Models: The non-stationary SMEV framework outperforms traditional GEV-based models in Non-Stationary Flood Frequency Analysis (NS-FFA). GEV-based models tend to underestimate flood quantiles and exhibit higher variability and bias, particularly for return periods exceeding 10 years, raising concerns about their reliability in predicting extreme floods.

2. Spatial Variability in Flood Magnitudes: Flood magnitude trends show considerable spatial variability across Germany, with southern regions experiencing the most pronounced changes, including both increases and decreases. This highlights the need for spatially tailored flood management and adaptation strategies.

3. Reduced Uncertainty with SMEV-Based Models: SMEV-based models significantly reduce uncertainty compared to traditional GEV-based models, particularly for longer return periods ($T > 20$ years). This reduction in uncertainty is primarily attributed to the larger effective sample size achieved by incorporating ordinary events into the estimation of distribution parameters.

4. Potential for Broader SMEV Applications: The incorporation of ordinary events and temporal covariates into the SMEV framework demonstrates significant potential for advancing NS-FFA, offering a more flexible approach to flood frequency analysis under changing conditions.

Data Availability Statement

For providing the discharge data for Germany, we are grateful to the Bavarian State Office of Environment (LfU, https://www.gkd.bayern.de/de/fluesse/abfluss/tabellen). Readers can directly access the river discharge data used in this study by clicking on the provided link. We are also thankful to the Global Runoff Data Centre (GRDC) prepared by the Federal Institute for Hydrology (BfG, https://portal.grdc.bafg.de/applications/public.html?publicuser=PublicUserdataDownload/Home) for their valuable data resources. To access the data repository, please use the provided link and click on the "Download by Station" option, select Germany as the country of interest, and locate the table in the top-left corner to access discharge gauges for Germany.

5

Conclusions and Future Avenues

This chapter provides an overview of the key findings obtained from the previous chapters. It highlights the progress made in understanding flood tail behavior and accurately modeling flood quantiles within the Metastatistical Extreme Value (MEV) framework, showcasing the contributions of this study. Additionally, practical applications, and directions for further research stemming from this work are outlined.

5.1 Conclusions

This dissertation focuses on the extension and optimization of the Metastatistical Extreme Value Distribution (MEVD) within both stationary and non-stationary contexts. A non-parametric approach was developed to identify the optimal ordinary distribution of streamflow peaks across a large set of catchments in Germany, enabling accurate estimation of high-flow quantiles in catchments with diverse climatic and physiographic conditions (Chapters 2 and 4). Additionally, this study integrates multiple runoff-generation processes within the mixed-SMEV framework to enhance flood estimation accuracy in river basins exhibiting a flood divide (Chapter 3). The next section briefly summarises the main results of the dissertation:

1. The study enhances flood quantile estimation by introducing a non-parametric method for selecting the ordinary distribution of streamflow peaks. This method enables binary classification of ordinary distribution into lighter-tailed (Gamma) or heavier-tailed (Log-Normal) distributions, improving the accuracy of flood frequency analysis across 182 catchments in Germany (Section 2.14). It significantly reduces estimation errors—lowering underestimation in Gamma distributions by 57% and overestimation in Log-Normal distributions by 58% compared to a blind MEV application (Section 2.2). Additionally, by leveraging ordinary streamflow data and eliminating the need for graphical evaluations, the method allows for efficient large-scale analysis and leading to more reliable extreme flood quantile estimates.

2. By incorporating information from ordinary streamflow peaks, the MEV approach—when benchmarked against traditional methods significantly reduces uncertainty in flood estimation for long return periods (i.e., T > 10 years) across the majority of river basins, specifically in 135 out of 182 catchments (Section 2.2). The findings of this study contribute to minimizing both underestimation and overestimation of rare floods, thereby enhancing the reliability of hazard assessments for infrequent hydrological events.

3. Furthermore, the findings indicate that catchments best represented by a Gamma distribution (lighter tail) are mainly located in western Germany, where winter floods triggered by precipitation on wet soils are dominant, with some catchments also clustered in the Alpine Forelands. In contrast, catchments where a Log-Normal distribution (which has a relatively heavier tail) is the best choice are located in the eastern and southern areas of the regions which are influenced by Vb-cyclones that produce rare but intense rainfall on dry soils, often resulting in severe floods (Section 2.1, Figure 2.1).

4. The findings of this study reveal that river basins experiencing exceptionally high floods (i.e., a flood divide) exhibit heavy-tailed behavior for one type of flood event and light-tailed behavior for the other, suggesting that differing tail behaviors are associated with distinct runoff-generation processes (Section 3.4, Figure 3.1).

5. The mixed-SMEV framework provides a physically grounded method to incorporate the potential impacts of climate change on flood hazards by explicitly considering the changes in the nature of runoff-generation processes, offering a robust solution for addressing future flood risks (Section 3.4). The mixed-SMEV outperforms in river basins where at least one runoff-generation process follows a heavy-tailed distribution, enhancing extreme flood prediction. It significantly improves flood magnitude and frequency estimation in basins prone to extraordinary floods (Figure 3.2). This process-informed approach outperforms single-SMEV and GEV distributions in predicting rare flood events (Section 3.5).

6. The non-stationary SMEV approach enhances flood quantile estimation compared to traditional non-stationary methods. While GEV-based non-stationary models tend to underestimate flood quantiles and exhibit higher bias for return periods exceeding 10 years (Figure 4.4). Non-stationary SMEV improves accuracy with reduced uncertainty. The high variability observed in NS-GEV models raises concerns about their reliability in predicting extreme floods, underscoring the potential of SMEV for more robust flood risk assessment under non-stationary conditions (Section 4.7).

The findings contribute to a better understanding of heavy-tailed flood behavior and introduce a robust approach for assessing flood tail heaviness and for the prediction of extreme floods . This method helps mitigate sensitivity to limited data, a common challenge in such analyses, and enhances applicability across different regions. These outcomes lay the foundation for future applications in flood hazard research and engineering, as outlined in the following section.

5.2 Future avenues

1. **Application in flood frequency guideline**: A key application of this research lies in the selection of appropriate candidate probability distributions for flood frequency analysis—an essential step for accurately estimating return periods and enhancing flood risk management (Nerantzaki and Papalexiou, 2022). Misinterpreting the tail behavior of the underlying distribution can result in significant underestimation or overestimation of extreme events (Huang et al., 2021; O'Brien and Burn, 2014; Papalexiou et al., 2013; Vu and Mishra, 2019). To mitigate such risks, current guidelines recommend pre-estimating tail behavior and carefully selecting suitable probability distributions (DWA, 2012). This study contributes to that objective by introducing a method for identifying and potentially predicting heavy-tailed flood behavior.

2. **Application in Multivariate Flood Hazard**: The increasing concern over the impacts of environmental changes on extreme floods has led to a growing global focus on multivariate flood risk assessment (Wang et al., 2022). Unlike traditional flood models that analyze single flood drivers, multivariate flood hazards arise from the interaction of multiple contributing factors, such as compound events that occur simultaneously or sequentially, leading to more severe flooding than individual events alone. Future research could extend the mixed-MEV approach to address multivariate flood hazards, thereby improving the comprehensiveness of flood risk assessments. For example, storm surges driven by strong winds and low atmospheric pressure may coincide with riverine flooding from heavy rainfall, producing widespread inundation in both coastal and inland regions. Incorporating multiple hydrological drivers—such as river discharge and storm surge—into the mixed-SMEV framework would allow for better modeling of the complex dependencies between flood-generating processes. This advancement could lead to more accurate and robust flood predictions, supporting improved risk mitigation strategies and more informed climate adaptation planning.

3. **Exploration of complex non-stationary patterns**:

   Future research should focus on refining MEV-based models by exploring more complex non-stationary patterns, enhancing the attribution of non-stationarity, and extending their application to a broader range of hydrometeorological variables. For example, flood distribution parameters could be modeled as functions of temporally varying hydrometeorological drivers, thereby increasing the models' relevance and impact in flood hazard assessments. The integration of ordinary events and temporal covariates in SMEV models demonstrates their potential for advancing non-stationary flood frequency analysis. Expanding the model to include a broader range of physical covariates, such as climate indices, land use changes, or precipitation patterns, would provide a deeper understanding of the process driving non-stationarity. This expansion could improve the attribution of changes in flood behavior and enhance the model's robustness across diverse environmental conditions.

4. **Extension of developed framework to diverse geographies regions**

   Future research could expand the application of MEV-based models beyond Germany to assess their generalizability across diverse climatic and physiographic conditions. Extending the proposed method to various climates would help evaluate its robustness and adaptability. Additionally, global-scale validation could provide valuable insights into its effectiveness for large-scale flood hazard assessments under both stationary and non-stationary conditions.

5. **Assessing the climate change projections**

   The intensification of extreme flood events due to climate change and human intervention highlights the need for an accurate assessment of evolving flood likelihood (Tabari, 2020). In a changing climate, flood risks are expected to vary across different magnitudes, presenting ongoing challenges for flood control and disaster mitigation. These assessments are critical for informing flood risk management and decision-making. However, traditional methods rely on long-term data records, which are often unavailable or insufficient for climate change studies. Consequently, their ability to capture future climate-driven changes is constrained by the short length of historical time series. The MEV approach addresses this limitation by reducing sensitivity to data scarcity and providing a more robust analytical framework. Future research should integrate climate models and projections to better understand how shifting hydrological patterns affect flood quantile estimation, thereby improving the predictive capacity of flood risk assessments in a changing climate

These research avenues would further refine and expand the utility of the proposed MEV framework, enhancing its accuracy, adaptability, and predictive power. By integrating diverse climatic conditions, advanced hydrological models, and automated classification methods, the approach could become a more robust tool for flood frequency analysis. Additionally, incorporating non-stationary flood behavior would improve its long-term reliability, ensuring its continued relevance in addressing emerging challenges in flood hazard estimation and management.

References

AghaKouchak, A. et al. (2020). "Climate Extremes and Compound Hazards in a Warming World". In: *Annual Review of Earth and Planetary Sciences* 48.1, pp. 519–548. DOI: 10.1146/annurev-earth-071719-055228.

Alila, Y. and A. Mtiraoui (2002). "Implications of heterogeneous flood-frequency distributions on traditional stream-discharge prediction techniques". In: *Hydrological Processes* 16.5, pp. 1065–1084. DOI: 10.1002/hyp.346.

Andreadis, K. M. et al. (2022). "Urbanizing the floodplain: Global changes of imperviousness in flood-prone areas". In: *Environmental Research Letters* 17.10, p. 104024. DOI: 10.1088/1748-9326/ac9197. URL: https://doi.org/10.1088/1748-9326/ac9197.

Archfield, S. A. et al. (2016). "Fragmented patterns of flood change across the United States". In: *Geophysical Research Letters* 43.19. DOI: 10.1002/2016gl070590.

Balkema, August A and Laurens De Haan (1974). "Residual life time at great age". In: *The Annals of probability* 2.5, pp. 792–804.

Barth, N. A., G. Villarini, and K. White (2019). "Accounting for mixed populations in flood frequency analysis: Bulletin 17C perspective". In: *Journal of Hydrologic Engineering* 24.3, p. 04019002. DOI: 10.1061/(ASCE)HE.1943-5584.0001762.

Barth, Nancy A et al. (2017). "Mixed populations and annual flood frequency estimates in the western United States: The role of atmospheric rivers". In: *Water Resources Research* 53.1, pp. 257–269. DOI: doi.org/10.1002/2016WR019064.

Basso, S, M Schirmer, and Gianluca Botter (2016). "A physically based analytical model of flood frequency curves". In: *Geophysical Research Letters* 43.17, pp. 9070–9076.

Basso, S., G. Botter, et al. (2021). "PHEV! the physically-based extreme value distribution of river flows". In: *Environmental Research Letters*. DOI: 10.1088/1748-9326/ac3d59. URL: https://doi.org/10.1088/1748-9326/ac3d59.

Basso, S., R. Merz, et al. (2023). "Extreme flooding controlled by stream network organization and flow regime". In: *Nature Geoscience* 16.4, pp. 339–343. DOI: 10.1038/s41561-023-01155-w.

Bernardara, P., D. Schertzer, S. Eric, et al. (2008). "The flood probability distribution tail: How heavy is it?" In: *Stochastic Environmental Research and Risk Assessment* 22.1, pp. 107–122. DOI: 10.1007/s00477-006-0101-2.

Bernardara, P., D. Schertzer, E. Sauquet, et al. (2008). "The flood probability distribution tail: how heavy is it?" In: *Stochastic Environmental Research and Risk Assessment* 22.1, pp. 107–122. DOI: 10.1007/s00477-006-0101-2. URL: https://doi.org/10.1007/s00477-006-0101-2.

Beurton, S. and A.H. Thieken (2009). "Seasonality of floods in German". In: *Hydrological Sciences Journal* 54.1, pp. 62–76. DOI: 10.1623/hysj.54.1.62. URL: https://doi.org/10.1623/hysj.54.1.62.

Bevere, L. and F. Remondi (2022). *Natural catastrophes in 2021: The floodgates are open*. Swiss Re Institute. URL: https://www.swissre.com/institute/research/sigma-research/sigma-2022-01.html.

Blöschl, G. et al. (2017). "Changing climate shifts timing of European floods". In: *Science* 357.6351, pp. 588–590. DOI: 10.1126/science.aan2506.

Blum, A. G. et al. (2020). "Causal effect of impervious cover on annual flood magnitude for the United States". In: *Geophysical Research Letters* 47.5. DOI: 10.1029/2019gl086480.

Bobee, B et al. (1993). "Towards a systematic approach to comparing distributions used in flood frequency analysis". In: *Journal of Hydrology* 142.1-4, pp. 121–136.

Cavanaugh, N. R. et al. (2015). "The probability distribution of intense daily precipitation". In: *Geophysical Research Letters* 42.5, pp. 1560–1567. DOI: 10.1002/2015GL063238.

Chen, Y. et al. (2004). "Study on L-moment estimations for log normal distribution with historical flood data". In: *GIS and RS in Hydrology, Water Resources and Environment*. Vol. 1. Wallingford, U.K.: IAHS Press, pp. 107–113.

Cho, H.-K., K. P. Bowman, and G. R. North (2004). "A comparison of gamma and lognormal distributions for characterizing satellite rain rates from the tropical rainfall measuring mission". In: *Journal of Applied Meteorology* 43.11, pp. 1586–1597. DOI: 10.1175/jam2165.1.

Clauset, A., C. R. Shalizi, and M. E. J. Newman (2009). "Power-law distributions in empirical data". In: *SIAM Review* 51.4, pp. 661–703. DOI: 10.48550/arXiv.0706.1062.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer London. DOI: 10.1007/978-1-4471-3675-0. URL: http://dx.doi.org/10.1007/978-1-4471-3675-0.

Cook, B. I. et al. (2004). "Statistical simulation of the influence of the NAO on European winter surface temperatures: Applications to phenological modeling". In: *Journal of Geophysical Research: Atmospheres* 109.D16. DOI: 10.1029/2003jd004305. URL: https://doi.org/10.1029/2003jd004305.

Cook, Nicholas J and R Ian Harris (2004). "Exact and general FT1 penultimate distributions of extreme wind speeds drawn from tail-equivalent Weibull parents". In: *Structural Safety* 26.4, pp. 391–420.

Cooke, R. M. and D. Nieboer (2011). "Heavy-Tailed distributions: Data, diagnostics, and new developments". In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.1811043. URL: https://doi.org/10.2139/ssrn.1811043.

Cunnane, C (1973). "A particular comparison of annual maxima and partial duration series methods of flood frequency prediction". In: *Journal of hydrology* 18.3-4, pp. 257–271.

— (1985). "Factors affecting choice of distribution for flood series". In: *Hydrological Sciences Journal* 30.1, pp. 25–36. DOI: 10.1080/02626668509490969. URL: https://doi.org/10.1080/02626668509490969.

Davenport, F.V., M. Burke, and N.S. Diffenbaugh (2021). "Contribution of historical precipitation change to US flood damages". In: *Proceedings of the National Academy of Sciences of the United States of America* 118.4, e2017524118. DOI: 10.1073/pnas.2017524118.

Davis, P.J. (1959). "Leonhard Euler's integral: a historical profile of the gamma function: in memoriam: Milton Abramowitz". In: *American Mathematical Monthly* 66, p. 849. DOI: 10.2307/2309786. URL: https://doi.org/10.2307/2309786.

De Michele, C. (2019). "Advances in deriving the exact distribution of maximum annual daily precipitation". In: *Water* 11.11, p. 2322. DOI: 10.3390/w11112322.

Deutsche Welle (2021). *German floods: Up to 30 billion needed for recovery fund*. Accessed: 2024-06-27. URL: https://www.dw.com/en/german-floods-up-to-30-billion-needed-for-recovery-fund/a-58807147.

Doocy, Shannon et al. (2013). "The human impact of floods: a historical review of events 1980-2009 and systematic literature review". In: *PLoS currents* 5.

DWA (2012). *Merkblatt DWA-M 552 - Ermittlung von Hochwasserwahrscheinlichkeiten*. Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall e. V. (DWA). Hennef.

Eash, David A (1997). *Effects of the 1993 Flood on the Determination of Flood Magnitude and Frequency in Iowa*. Vol. 1120. US Department of the Interior, US Geological Survey.

Eastoe, E. F. (2019). "Nonstationarity in peaks-over-threshold river flows: A regional random effects model". In: *Environmetrics* 30.5. DOI: 10.1002/env.2560.

El Adlouni, S., B. Bobée, and T. B. M. J. Ouarda (2008). "On the tails of extreme event distributions in hydrology". In: *Journal of Hydrology* 355.1–4, pp. 16–33. DOI: 10.1016/j.jhydrol.2008.02.011. URL: https://doi.org/10.1016/j.jhydrol.2008.02.011.

Eliazar, I. I. and I. M. Sokolov (2010). "Gini characterization of extreme-value statistics". In: *Physica A: Statistical Mechanics and Its Applications* 389.21, pp. 4462–4472. DOI: 10.1016/j.physa.2010.07.005. URL: https://doi.org/10.1016/j.physa.2010.07.005.

Faulkner, D. et al. (2019). "Can we still predict the future from the past? Implementing non-stationary flood frequency analysis in the UK". In: *Journal of Flood Risk Management* 13.1. DOI: 10.1111/jfr3.12582.

Faulkner, D. S. et al. (2023). "Modelling non-stationary flood frequency in England and Wales using physical covariates". In: *Hydrology and Earth System Sciences*. DOI: 10.5194/hess-2022-401.

Fekete, Alexander and Simone Sandholz (2021). "Here comes the flood, but not failure? Lessons to learn after the heavy rain and pluvial floods in Germany 2021". In: *Water* 13.21, p. 3016.

Fischer, S. (2018). "A seasonal mixed-POT model to estimate high flood quantiles from different event types and seasons". In: *Journal of Applied Statistics* 45.15, pp. 2831–2847. DOI: 10.1080/02664763.2018.1441385.

Fisher, Ronald Aylmer and Leonard Henry Caleb Tippett (1928). "Limiting forms of the frequency distribution of the largest or smallest member of a sample". In: *Mathematical proceedings of the Cambridge philosophical society*. Vol. 24. 2. Cambridge University Press, pp. 180–190.

François, B. et al. (2019). "Design considerations for riverine floods in a changing climate – a review". In: *Journal of Hydrology* 574, pp. 557–573. DOI: 10.1016/j.jhydrol.2019.04.068.

Gnedenko, B. (1943). "Sur La distribution limite du Terme maximum D'Une Serie 788 aleatoire". In: *Annals of Mathematics* 44.3, pp. 423–453. DOI: 10.2307/1968974. URL: https://doi.org/10.2307/1968974.

Gotvald, A.J. et al. (2012). *Methods for determining magnitude and frequency of floods in California, based on data through water year 2006*. Tech. rep. 2012-5113. U.S. Geological Survey. URL: http://pubs.usgs.gov/sir/2012/5113/.

Greenwood, J.A. et al. (1979). "Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form". In: *Water Resources Research* 15, pp. 1049–1054. DOI: 10.1029/WR015i005p01049. URL: https://doi.org/10.1029/WR015i005p01049.

Gumbel, E. J. (1958). *Statistics of Extremes*. Columbia University Press. DOI: 10.7312/gumb92958.

— (2004). *Statistics of extremes Dover Publications*.

Haddad, K. and A. Rahman (2011). "Selection of the best fit flood frequency distribution and parameter estimation procedure: a case study for Tasmania in Australia". In: *Stochastic Environmental Research and Risk Assessment* 25.3, pp. 415–428. DOI: 10.1007/s00477-010-0412-1. URL: https://doi.org/10.1007/s00477-010-0412-1.

Haktanir, T., M. Cobaner, and B. Gorkemli (2013). "Assessment of right-tail prediction ability of some distributions by Monte Carlo analyses". In: *Journal of Hydrologic Engineering* 18.5, pp. 499–517. DOI: 10.1061/(ASCE)HE.1943-5584.0000620.

Hall, J. et al. (2014). "Understanding flood regime changes in Europe: A state-of-the-art assessment". In: *Hydrology and Earth System Sciences* 18.7, pp. 2735–2772. DOI: 10.5194/hess-18-2735-2014.

Harrigan, S. et al. (2017). "Designation and trend analysis of the updated UK Benchmark Network of river flow stations: The UKBN2 dataset". In: *Hydrology Research* 49.2, pp. 552–567. DOI: 10.2166/nh.2017.058.

Hashino, T., A.A. Bradley, and S.S. Schwartz (2006). "Evaluation of bias-correction methods for ensemble streamflow volume forecasts". In: *Hydrology and Earth System Sciences Discussions* 3, pp. 561–606. DOI: 10.5194/hessd-3-561-2006. URL: https://doi.org/10.5194/hessd-3-561-2006.

Hirabayashi, Y. et al. (2013). "Global flood risk under climate change". In: *Nature Climate Change* 3.9, pp. 816–821. DOI: 10.1038/nclimate1911.

Hirschboeck, K.K. (1987a). "Catastrophic flooding and atmospheric circulation anomalies". In: *Catastrophic Flooding*. Ed. by L. Mayer and D. Nash. Allen Unwin, pp. 23–56. DOI: 10.4324/9781003020325-2.

— (1987b). "Hydroclimatically-defined mixed distributions in partial duration flood series". In: *Hydrologic Frequency Modeling: Proceedings of the International Symposium on Flood Frequency and Risk Analyses*. Ed. by V.P. Singh. Baton Rouge, Louisiana State University: Springer Netherlands, pp. 199–212. DOI: 10.1007/978-94-009-3953-0_13.

Hofstätter, M. et al. (2016). "A new classification scheme of European cyclone tracks with relevance to precipitation". In: *Water Resources Research* 52.9, pp. 7086–7104. DOI: 10.1002/2016WR019146. URL: https://doi.org/10.1002/2016WR019146.

Hosking, J.R.M. (1990). "L-moments: analysis and estimation of distributions using linear combinations of order statistics". In: *Journal of the Royal Statistical Society Series B* 52.1, pp. 105–124. DOI: 10.1111/j.2517-6161.1990.tb01775.x. URL: https://doi.org/10.1111/j.2517-6161.1990.tb01775.x.

Hu, L. et al. (2020). "Sensitivity of flood frequency analysis to data record, statistical model, and parameter estimation methods: an evaluation over the contiguous United States". In: *Journal of Flood Risk Management* 1, p. 13. DOI: 10.1111/jfr3.12580. URL: https://doi.org/10.1111/jfr3.12580.

Huang, H., H. Cui, and Q. Ge (2021). "Will a nonstationary change in extreme precipitation affect dam security in China?" In: *Journal of Hydrology* 603, p. 126859. DOI: 10.1016/j.jhydrol.2021.126859. URL: https://doi.org/10.1016/j.jhydrol.2021.126859.

Huo, R. et al. (2022). "Changing flood dynamics in Norway since the last millennium and to the end of the 21st century". In: *Journal of Hydrology* 613, p. 128331. DOI: 10.1016/j.jhydrol.2022.128331.

Iliopoulou, T. and D. Koutsoyiannis (2019). "Revealing hidden persistence in maximum rainfall records". In: *Hydrological Sciences Journal* 64.14, pp. 1673–1689. DOI: 10.1080/02626667.2019.1657578. URL: https://doi.org/10.1080/02626667.2019.1657578.

Katz, R.W., M.B. Parlange, and P. Naveau (2002). "Statistics of extremes in hydrology". In: *Advances in Water Resources* 25.8, pp. 1287–1304. DOI: 10.1016/S0309-1708(02)00056-8. URL: https://doi.org/10.1016/S0309-1708(02)00056-8.

Kendall, M. G. (1975). *Rank Correlation Methods*. London, UK: Charles Griffin, Oxford University Press.

Klemes, V. (1974). "Some problems in pure and applied stochastic hydrology". In: *Proceedings of the Symposium on Statistical Hydrology*. Vol. 1275. miscellaneous publication No., pp. 2–15.

Kobierska, F., K. Engeland, and T. Thorarinsdottir (2018). "Evaluation of design flood estimates – a case study for Norway". In: *Hydrology Research* 49.2, pp. 450–465. DOI: 10.2166/nh.2017.068. URL: https://doi.org/10.2166/nh.2017.068.

Koutsoyiannis, D. (2020). *Stochastics of Hydroclimatic Extremes—A Cool Look At Risk*. Tech. rep. Athens, Greece: National Technical University of Athens.

KOUTSOYIANNIS, DEMETRIS (n.d.). "rainfall: I. Theoretical investigation". In: ().

Kreibich, H. et al. (2022). "The challenge of unprecedented floods and droughts in risk management". In: *Nature* 608.7921, pp. 80–86. DOI: 10.1038/s41586-022-04917-5.

Kumar, R., L. Samaniego, and S. Attinger (2013). "Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations". In: *Water Resources Research* 49.1, pp. 360–379. DOI: 10.1029/2012WR012195.

Kyselý, J. (2008). "A cautionary note on the use of nonparametric bootstrap for estimating uncertainties in extreme-value models". In: *Journal of Applied Meteorology and Climatology* 47.12, pp. 3236–3251. DOI: 10.1175/2008JAMC1763.1.

Lancaster, H.O. (1966). "Forerunners of the Pearson chi squared". In: *Australian Journal of Statistics* 8, pp. 117–126. DOI: 10.1111/j.1467-842x.1966.tb00262.x. URL: https://doi.org/10.1111/j.1467-842x.1966.tb00262.x.

Lang, M., T.B.M.J. Ouarda, and B. Bobée (1999). "Towards operational guidelines for over-threshold modeling". In: *Journal of Hydrology* 225.3-4, pp. 103–111. DOI: 10.1016/S0022-1694(99)00167-5. URL: https://doi.org/10.1016/S0022-1694(99)00167-5.

Liu, J. et al. (2022). "Global changes in floods and their drivers". In: *Journal of Hydrology* 614, p. 128553. DOI: 10.1016/j.jhydrol.2022.128553. URL: https://doi.org/10.1016/j.jhydrol.2022.128553.

Lombardo, F. et al. (2019). "On the exact distribution of correlated extremes in hydrology". In: *Water Resources Research* 55.12, pp. 10405–10423. DOI: 10.1029/2019WR025547. URL: https://doi.org/10.1029/2019WR025547.

Luke, A. et al. (2017). "Predicting nonstationary flood frequencies: Evidence supports an updated stationarity thesis in the United States". In: *Water Resources Research* 53.7, pp. 5469–5494. DOI: 10.1002/2016WR019676.

Macdonald, N. et al. (2006). "Historical and pooled flood frequency analysis for the River Tay at Perth, Scotland". In: *Area* 38.1, pp. 34–46. DOI: 10.1111/j.1475-4762.2006.00673.x. URL: https://doi.org/10.1111/j.1475-4762.2006.00673.x.

Malamud, B.D. and D.L. Turcotte (2003). "Shelf record of climatic changes in flood magnitude and frequency, north-coastal California: Comment". In: *Geology* 31.3, p. 288. DOI: 10.1130/0091-7613(2003)031<0288:SROCCI>2.0.CO;2.

— (2006). "The applicability of power-law frequency statistics to floods". In: *Journal of Hydrology* 322.1–4, pp. 168–180. DOI: 10.1016/j.jhydrol.2005.02.032.

Mann, H. B. (1945). "Nonparametric tests against trend". In: *Econometrica: Journal of the Econometric Society*, pp. 245–259.

Marani, M. and M. Ignaccolo (2015). "A Metastatistical approach to rainfall extremes". In: *Advances in Water Resources* 79, pp. 121–126. DOI: 10.1016/j.advwatres.2015.03.001. URL: https://doi.org/10.1016/j.advwatres.2015.03.001.

Marra, F., W. Amponsah, and S.M. Papalexiou (2023). "Non-asymptotic Weibull tails explain the statistics of extreme daily precipitation". In: *Advances in Water Resources* 173, p. 104388. DOI: 10.1016/j.advwatres.2023.104388.

Marra, F., E.I. Nikolopoulos, et al. (2018). "Metastatistical extreme value analysis of hourly rainfall from short records: estimation of high quantiles and impact of measurement errors". In: *Advances in Water Resources* 117, pp. 27–39. DOI: 10.1016/j.advwatres.2018.05.001. URL: https://doi.org/10.1016/j.advwatres.2018.05.001.

Marra, F., D. Zoccatelli, et al. (2019). "A simplified MEV formulation to model extremes emerging from multiple nonstationary underlying processes". In: *Advances in Water Resources* 127, pp. 280–290. DOI: 10.1016/j.advwatres.2019.04.002.

Massey, F.J. (1951). "The Kolmogorov-Smirnov test for goodness of fit". In: *Journal of the American Statistical Association* 46.253, pp. 68–78. DOI: 10.1080/01621459.1951.10500769.

McAlister, D. (1879). "The law of the geometric mean". In: *Proceedings of the Royal Society of London* 29, pp. 367–376. DOI: 10.1098/rspl.1879.0061. URL: https://doi.org/10.1098/rspl.1879.0061.

Merz, B., S. Basso, et al. (2022). "Understanding heavy tails of flood peak distributions". In: *Water Resources Research* 58.6, e2021WR030506. DOI: 10.1029/2021WR030506.

Merz, B., G. Blöschl, et al. (2021). "Causes, impacts and patterns of disastrous river floods". In: *Nature Reviews Earth and Environment* 2.9, pp. 592–609. DOI: 10.1038/s43017-021-00195-3. URL: https://doi.org/10.1038/s43017-021-00195-3.

Merz, R. and G. Blöschl (2008). "Flood frequency hydrology: 2. Combining data evidence". In: *Water Resources Research* 44.8. DOI: 10.1029/2007WR006745.

Miniussi, A. and M. Marani (2020). "Estimation of daily rainfall extremes through the Metastatistical extreme value distribution: uncertainty minimization and implications for trend detection". In: *Water Resources Research* 56.7. DOI: 10.1029/2019wr026535. URL: https://doi.org/10.1029/2019wr026535.

Miniussi, A., M. Marani, and G. Villarini (2020). "Metastatistical extreme value distribution applied to floods across the continental United States". In: *Advances in Water Resources* 136. DOI: 10.1016/j.advwatres.2019.103498. URL: https://doi.org/10.1016/j.advwatres.2019.103498.

Miniussi, A. and F. Marra (2021). "Estimation of extreme daily precipitation return levels at-site and in ungauged locations using the simplified MEV approach". In: *Journal of Hydrology* 603, p. 126946. DOI: 10.1016/j.jhydrol.2021.126946.

Miniussi, A., R. Merz, et al. (2023). "Identifying discontinuities of flood frequency curves". In: *Journal of Hydrology* 617, p. 128989. DOI: 10.1016/j.jhydrol.2022.128989.

Morrison, J.E. and J.A. Smith (2002). "Stochastic modeling of flood peaks using the generalized extreme value distribution". In: *Water Resources Research* 38.12, pp. 41-1–41-12. DOI: 10.1029/2001wr000502. URL: https://doi.org/10.1029/2001wr000502.

Murphy, A.H. and R.L. Winkler (1992). "Diagnostic verification of probability forecasts". In: *International Journal of Forecasting* 7.4, pp. 435–455. DOI: 10.1016/0169-2070(92)90028-8. URL: https://doi.org/10.1016/0169-2070(92)90028-8.

Mushtaq, S. et al. (2022). "Reliable estimation of high floods: A method to select the most suitable ordinary distribution in the metastatistical extreme value framework". In: *Advances in Water Resources* 161, p. 104127. DOI: 10.1016/j.advwatres.2022.104127.

Newman, M.E. (2005). "Power laws, Pareto distributions and Zipf's law". In: *Contemporary Physics* 46.5, pp. 323–351. DOI: 10.1080/00107510500052444.

O'Brien, N. L. and D. H. Burn (2014). "A nonstationary index-flood technique for estimating extreme quantiles for annual maximum streamflow". In: *Journal of Hydrology* 519, pp. 2040–2048. DOI: 10.1016/j.jhydrol.2014.09.041.

Overeem, A., A. Buishand, and I. Holleman (2008). "Rainfall depth-duration-frequency curves and their uncertainties". In: *Journal of Hydrology* 348.1–2, pp. 124–134. DOI: 10.1016/j.jhydrol.2007.09.044.

Panagoulia, D., P. Economou, and C. Caroni (2013). "Stationary and nonstationary generalized extreme value modelling of extreme precipitation over a mountainous area under climate change". In: *Environmetrics* 25.1, pp. 29–43. DOI: 10.1002/env.2252.

Papalexiou, S.M., D. Koutsoyiannis, and C. Makropoulos (2013). "How extreme is extreme? An assessment of daily rainfall distribution tails". In: *Hydrology and Earth System Sciences* 17.2, pp. 851–862. DOI: 10.5194/hess-17-851-2013. URL: https://doi.org/10.5194/hess-17-851-2013.

Petrow, T. et al. (2007). "Aspects of seasonality and flood generating circulation patterns in a mountainous catchment in south-eastern Germany". In: *Hydrology and Earth System Sciences* 11.4, pp. 1455–1468. DOI: 10.5194/hess-11-1455-2007.

Pickands III, James (1975). "Statistical inference using extreme order statistics". In: *the Annals of Statistics*, pp. 119–131.

Prosdocimi, I., T. R. Kjeldsen, and J. D. Miller (2015). "Detection and attribution of urbanization effect on flood extremes using nonstationary flood-frequency models". In: *Water Resources Research* 51.6, pp. 4244–4262. DOI: 10.1002/2015WR017065.

Rauthe, M. et al. (2013). "A Central European precipitation climatology – Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS)". In: *Meteorologische Zeitschrift* 22.3, pp. 235–256. DOI: 10.1127/0941-2948/2013/0436.

Renard, B. et al. (2013). "Data-based comparison of frequency analysis methods: a general framework". In: *Water Resources Research* 49.2, pp. 825–843. DOI: 10.1002/wrcr.20087. URL: https://doi.org/10.1002/wrcr.20087.

Ritter, A. and R. Munoz-Carpena (2013). "Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments". In: *Journal of Hydrology* 480, pp. 33–45. DOI: 10.1016/j.jhydrol.2012.12.004.

Rogger, M., H. Pirkl, et al. (2012). "Step changes in the flood frequency curve: Process controls". In: *Water Resources Research* 48.5. DOI: 10.1029/2011wr011187. URL: https://doi.org/10.1029/2011wr011187.

Rogger, M., A. Viglione, et al. (2013). "Quantifying effects of catchments storage thresholds on step changes in the flood frequency curve". In: *Water Resources Research* 49.10, pp. 6946–6958. DOI: 10.1002/wrcr.20553.

Salas, J. D. and J. Obeysekera (2019). "Probability Distribution and Risk of the First Occurrence of k Extreme Hydrologic Events". In: *Journal of Hydrologic Engineering* 24.10. DOI: 10.1061/(ASCE)HE.1943-5584.0001809.

Samaniego, L., R. Kumar, and S. Attinger (2010). "Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale". In: *Water Resources Research* 46.5, W05523. DOI: 10.1029/2008WR007327.

Schädler, G. et al. (2012). *Flood Hazards in a Changing Climate: Project Report*. Tech. rep. Center for Disaster Management and Risk Reduction Technology (CEDIM), p. 83. URL: http://www.cedim.de/download/Flood_Hazards_in_a_Changing_Climate.pdf.

Seckin, N., T. Haktanir, and R. Yurtal (2011). "Flood frequency analysis of Turkey using L-moments method". In: *Hydrological Processes* 25.22, pp. 3499–3505. DOI: 10.1002/hyp.8077. URL: https://doi.org/10.1002/hyp.8077.

Serinaldi, F., F. Lombardo, and C.G. Kilsby (2020). "All in order: distribution of serially correlated order statistics with applications to hydrological extremes". In: *Advances in Water Resources* 144, p. 103686. DOI: 10.1016/j.advwatres.2020.103686. URL: https://doi.org/10.1016/j.advwatres.2020.103686.

Sikorska, A. E., D. Viviroli, and J. Seibert (2015). "Flood-type classification in mountainous catchments using crisp and fuzzy decision trees". In: *Water Resources Research* 51.10, pp. 7959–7976. DOI: 10.1002/2015wr017326. URL: https://doi.org/10.1002/2015wr017326.

Silva, A. T., M. Naghettini, and M. M. Portela (2015). "On some aspects of peaks-over-threshold modeling of floods under nonstationarity using climate covariates". In: *Stochastic Environmental Research and Risk Assessment* 30.1, pp. 207–224. DOI: 10.1007/s00477-015-1072-y.

Sivapalan, M. and J. M. Samuel (2009). "Transcending limitations of stationarity and the return period: Process-based approach to flood estimation and risk assessment". In: *Hydrological Processes* 23.11, pp. 1671–1675. DOI: 10.1002/hyp.7292. URL: https://doi.org/10.1002/hyp.7292.

Slater, L. J. and G. Villarini (2016). "Recent trends in U.S. flood risk". In: *Geophysical Research Letters* 43.24. DOI: 10.1002/2016GL071199.

Smith, J. A. et al. (2018). "Strange floods: The upper tail of flood peaks in the United States". In: *Water Resources Research* 54.9, pp. 6510–6542. DOI: 10.1029/2018wr022539. URL: https://doi.org/10.1029/2018wr022539.

Stedinger, J.R. (1980). "Fitting log normal distributions to hydrologic data". In: *Water Resources Research* 16.3, pp. 481–490. DOI: 10.1029/WR016i003p00481. URL: https://doi.org/10.1029/WR016i003p00481.

Stein, L., F. Pianosi, and R. Woods (2020). "Event-based classification for global study of river flood generating processes". In: *Hydrological Processes* 34.7, pp. 1514–1529. DOI: 10.1002/hyp.13678.

Tabari, Hossein (2020). "Climate change impact on flood and extreme precipitation increases with water availability". In: *Scientific reports* 10.1, p. 13768.

Tarasova, L, S Basso, and R Merz (2020). "Transformation of generation processes from small runoff events to large floods". In: *Geophysical Research Letters* 47.22, e2020GL090547.

Tarasova, L, S Basso, Dadiyorto Wendi, et al. (2020). "A process-based framework to characterize and classify runoff events: The event typology of Germany". In: *Water Resources Research* 56.5, e2019WR026951.

Tarasova, L., D. Lun, et al. (2023). "Shifts in flood generation processes exacerbate regional flood anomalies in Europe". In: *Communications Earth & Environment* 4.1, p. 49. DOI: 10.1038/s43247-023-00714-8.

Tarasova, L., R. Merz, et al. (2019). "Causative classification of river flood events". In: *Wiley Interdisciplinary Reviews: Water* 6, e1353. DOI: 10.1002/wat2.1353.

Thieken, A.H. et al. (2005). "Flood damage and influencing factors: New Insights from the August 2002 flood in Germany". In: *Water Resources Research* 41.12. DOI: 10.1029/2005wr004177. URL: https://doi.org/10.1029/2005wr004177.

Turkington, T. et al. (2016). "A new flood type classification method for use in climate change impact studies". In: *Weather and Climate Extremes* 14, pp. 1–16. DOI: 10.1016/j.wace.2016.10.001.

UNISDR (2015). *Making Development Sustainable: The Future of Disaster Risk Management*. Tech. rep. Global assessment report on disaster risk reduction. Geneva, Switzerland: Centre for Research on the Epidemiology of Disasters, United Nations Office for Disaster Risk Reduction (UNISDR). DOI: 10.18356/919076d9-en. URL: https://doi.org/10.18356/919076d9-en.

United Nations (2020). *Human cost of disasters*. DOI: 10.18356/79b92774-en. URL: https://doi.org/10.18356/79b92774-en.

USWRC (1976). *Guidelines for determining flood flow frequency*. Tech. rep. United States Water Resources Council Bulletin 17, hydrology committee. DOI: 10.3133/tm4B5. URL: https://doi.org/10.3133/tm4B5.

Vidrio-Sahagún, C. T. and J. He (2022). "Hydrological frequency analysis under nonstationarity using the Metastatistical approach and its simplified version". In: *Advances in Water Resources* 166, p. 104244. DOI: 10.1016/j.advwatres.2022.104244.

Vidrio-Sahagún, C. T., J. Ruschkowski, et al. (2024). "A practice-oriented framework for stationary and nonstationary flood frequency analysis". In: *Environmental Modelling & Software* 173, p. 105940. DOI: 10.1016/j.envsoft.2024.105940.

Viglione, A. et al. (2013). "Flood frequency hydrology: 3. A Bayesian analysis". In: *Water Resources Research* 49.2, pp. 675–692. DOI: 10.1029/2011WR010782.

Villarini, G., F. Serinaldi, et al. (2009). "On the stationarity of annual flood peaks in the continental United States during the 20th century". In: *Water Resources Research* 45.8. DOI: 10.1029/2008WR007645.

Villarini, G. and J. A. Smith (2010). "Flood peak distributions for the eastern United States". In: *Water Resources Research* 46.6. DOI: 10.1029/2009wr008395. URL: https://doi.org/10.1029/2009wr008395.

Vogel, R.M. and N.M. Fennessey (1993). "L moment diagrams should replace product moment diagrams". In: *Water Resources Research* 29.6, pp. 1745–1752. DOI: 10.1029/93WR00341.

Volpi, E et al. (2015). "One hundred years of return period: Strengths and limitations". In: *Water Resources Research* 51.10, pp. 8570–8585.

Vormoor, K. et al. (2016). "Evidence for changes in the magnitude and frequency of observed rainfall vs. snowmelt driven floods in Norway". In: *Journal of Hydrology* 538, pp. 33–48. DOI: 10.1016/j.jhydrol.2016.03.066.

Vu, T. M. and A. K. Mishra (2019). "Nonstationary frequency analysis of the recent extreme precipitation events in the United States". In: *Journal of Hydrology* 575, pp. 999–1010. DOI: 10.1016/j.jhydrol.2019.05.090. URL: https://doi.org/10.1016/j.jhydrol.2019.05.090.

Wang, Hsing-Jui et al. (2023). "Inferring heavy tails of flood distributions from common discharge dynamics". In: *EGUsphere* 2023, pp. 1–24.

Waylen, P. and M. Woo (1982). "Prediction of annual floods generated by mixed processes". In: *Water Resources Research* 18.4, pp. 1283–1286. DOI: 10.1029/wr018i004p01283. URL: https://doi.org/10.1029/wr018i004p01283.

Weibull, W. (1939). *A statistical theory of strength of materials*. IVB-Handl. URL: https://cir.nii.ac.jp/crid/1573668925268278656.

Wietzke, L. M. et al. (2020). "Comparative analysis of scalar upper tail indicators". In: *Hydrological Sciences Journal* 65.10, pp. 1625–1639. DOI: 10.1080/02626667.2020.1769104. URL: https://doi.org/10.1080/02626667.2020.1769104.

Yin, J. et al. (2018). "Large increase in global storm runoff extremes driven by climate and anthropogenic changes". In: *Nature Communications* 9.1. DOI: 10.1038/s41467-018-06765-2.

Yu, G., D. B. Wright, and F. V. Davenport (2022). "Diverse physical processes drive upper-tail flood quantiles in the US Mountain West". In: *Geophysical Research Letters* 49.10. DOI: 10.1029/2022gl098855. URL: https://doi.org/10.1029/2022gl098855.

Zakaria, S. A. et al. (2021). "Stationary and non-stationary models of extreme ground-level ozone in peninsular Malaysia". In: *Mathematics and Statistics* 9.3, pp. 357–370. DOI: 10.13189/ms.2021.090318.

Zhang, T. et al. (2018). "Nonstationary flood frequency analysis using univariate and bivariate time-varying models based on GAMLSS". In: *Water* 10.7, p. 819. DOI: 10.3390/w10070819.

Zorzetto, E., G. Botter, and M. Marani (2016). "On the emergence of rainfall extremes from ordinary events". In: *Geophysical Research Letters* 43.15, pp. 8076–8082. DOI: 10.1002/2016gl069445. URL: https://doi.org/10.1002/2016gl069445.

Zorzetto, Enrico and Marco Marani (2019). "Downscaling of rainfall extremes from satellite observations". In: *Water Resources Research* 55.1, pp. 156–174.

6

List of Publications and Author Contributions

This section outlines the chapters in this dissertation, which are based on published papers and have been modified. The author contributions, aligned with the Contributor Roles Taxonomy suggested by Allen et al. (2019)[1], are detailed below.

Chapter 2

In a modified version of the published paper in AWR: Mushtaq, S., Miniussi, A., Merz, R., & Basso, S. (2022). Reliable estimation of high floods: A method to select the most suitable ordinary distribution in the Metastatistical extreme value framework.

Advances in Water Resources, 161 (104127). https://doi.org/10.1016/j.advwatres.2022.104127, with permission from the Authors and Elsevier Ltd. (© 2022 The Author(s). Published by Elsevier Ltd.) with copyright permission.

Author contributions

- **SM**: Conceptualization, Methodology, Validation, Software, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review and editing, Visualization, Funding acquisition
- **AM**: Conceptualization, Writing - review and editing, Supervision, Project administration
- **RM**: Conceptualization, Writing - review and editing, Supervision, Project administration, Funding acquisition
- **SB**: Conceptualization, Methodology, Writing - review and editing, Supervision, Project administration, Funding acquisition

Chapter 3

This chapter presents a formatted version of the original paper: Mushtaq, S., Miniussi, A., Merz, R., Tarasova, L., Marra, F., & Basso, S. (2023). *Prediction of extraordinarily high floods emerging from heterogeneous flow generation processes.* Geophysical Research Letters, 50, e2023GL105429. , with permission from the Authors and AGU. (© 2023 The Author(s). Published by AGU.)

Author contributions

- **SM**: Conceptualization, Methodology, Validation, Software, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review and editing, Visualization, Funding acquisition
- **AM**: Conceptualization,Methodology, Writing - review and editing, Supervision, Project administration
- **RM**: Conceptualization, Writing - review and editing, Supervision, Project administration, Funding acquisition
- **LR**: Data curation,Conceptualization, Writing - review and editing,
- **FM**: Conceptualization,Methodology, Writing - review and editing
- **SB**: Conceptualization, Methodology, Writing - review and editing, Supervision, Project administration, Funding acquisition

Allen[1], L., O'Connell, A., and Kiermer, V. (2019). How can we ensure visibility and diversity in research contributions? How the Contributor Role Taxonomy (CRediT) is helping the shift from authorship to contributorship. *Learned Publishing, 32*(1), 71–74. https://doi.org/10.1002/leap.1210