Journal of Hydrology 636 (2024) 131289

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Incorporating historical flood events in type-based statistics

Svenja Fischer [a,b,*]

[a] Hydrology and Environmental Hydraulics Group, Wageningen University & Research, Droevendaalsesteeg 3, Wageningen, The Netherlands

[b] SPATE Research Unit, Department of Civil and Environmental Engineering, Ruhr-University Bochum, Universitätstraße 150, 44801 Bochum, Germany

ARTICLE INFO

Keywords:
Historical floods
Type-based flood statistics
MCMC
Partial probability weighted moments

ABSTRACT

Flood statistics form the basis of many hydrologic designs. However, short observation periods can lead to a high degree of uncertainty in the quantile estimates, especially for extreme floods. One way to expand the information used in the statistics and, in particular, to include other exceptional floods in the sample is to take historical floods into account. The use of these floods, which are usually reconstructed from chronicles or flood marks, has long been established for classic flood frequency analysis. However, the fact that historical floods can also have different origins and are therefore not statistically identically distributed has so far been ignored. To avoid this violation of the statistical assumptions, we show how historical floods, subdivided according to their genesis, can be taken into account in type-based flood statistics. For this purpose, the classical partial probability weighted moment (PPWM) method is extended for the type-based mixture model of partial duration series (TMPS). The result is compared to a Markov chain Monte Carlo method in order to compare the differences between the two methods. Type-based statistics can also be used to attribute the different influences of historical floods on the quantile estimate. An example shows that floods associated with long-lasting precipitation in particular are overestimated in short samples, while heavy rainfall floods are comparatively well represented. The results show that the consideration of historical events in type-based statistics not only allows a more balanced view of extreme floods, but also enables the influences of these to be attributed.

1. Introduction

Flood statistics are a key tool in hydrology. They usually form the basis for determining hydrological design variables, such as the design flood, that are used for appropriate dimensioning of a structure in order to fulfil a pre-defined purpose, e.g., flood protection. Yet, flood statistics are also often criticised for high uncertainty, especially when only short observation periods of the flood events in the systematic records are available (Arnaud et al., 2017; Fischer & Schumann, 2022). This sampling uncertainty can manifest in different ways. On the one hand, a short observation period that includes one or more extreme flood events might lead to an overestimation of the probability of such events, as their weight in the estimation is high. On the other hand, if the observation period is too short to include such extreme floods, design floods might be underestimated. To reduce this uncertainty related to the sample, information extension can be used, i.e., the incorporation of additional information in space and time. One example for such an extension is the consideration of historical floods or even paleo-flood data (Cohn and Stedinger, 1987; Gaume et al., 2010; Payrastre et al., 2011; Stedinger & Cohn, 1986; to name only a few). Such historical floods usually are extreme flood events that were recorded in chronicles or whose water levels were marked on historical buildings due to their severity, the caused damage or fatalities. Therefore, they can provide valuable information on what might be expected as extreme event in a catchment. A good example is the Ahrtal in Germany. This region was hit in July 2021 by a devastating flood event, causing at least 135 fatalities and billions of euros of damage (Lehmkuhl et al., 2022). This event was seen by many as a totally surprising event. However, historical flood events from the years 1804 and 1910 reveal that similar conditions in the past have led to similar flood peaks before (Kahle et al., 2022). Yet, these events were not part of the systematic records and hence the 2021 flood was not anticipated.

A challenge when aiming to incorporate historical flood events in flood statistics always lies in the information acquisition. Usually, extreme events are recorded in local chronicles or archives but often without precise information on the discharge at the flood peak, which is

* Corresponding author at: Hydrology and Environmental Hydraulics Group, Wageningen University & Research, Droevendaalesteeg 3, Wageningen, the Netherlands.

E-mail address: Svenja.fischer@wur.nl.

https://doi.org/10.1016/j.jhydrol.2024.131289
Received 27 November 2023; Received in revised form 12 April 2024; Accepted 17 April 2024
Available online 11 May 2024

required for flood statistics. Flood marks deliver water levels, though still the reconstruction of the discharge remains challenging, as the river profile and the hydraulic conditions might have changed over time. But even when the flood peak is available, the historical events remain a censored series since no information on ordinary flood events is available in the historical records. Therefore, statistical approaches have to be able to consider these censored series. There exist several approaches to do so, ranging from simple parametric or non-parametric re-sampling of the systematic records to “fill” the censored series, over Bayesian Markov Chain Monte Carlo (MCMC) approaches where prior information is included in the sampling (Reis and Stedinger, 2005) to an adapted parameter estimation for censored series. The latter originally is based on weighted moments, is rather easy to apply and was first proposed by the Bulletin 17B (USWRC, 1975). It was extended to a maximum likelihood estimation (Stedinger & Cohn, 1986), the expected moments algorithm (Cohn et al., 1997) and by Wang (1990) to the probability weighted moments. All these approaches have advantages and disadvantages in terms of efficiency, robustness or uncertainty. For a detailed review of approaches to include historical floods in flood statistics we refer to Kjeldsen et al. (2014) and the references therein. However, what is not considered so far when incorporating historical flood events in flood statistics is the flood genesis. Historical flood events might emerge from quite different atmospheric and catchment conditions, such as heavy rainfall, spatially extended and long-lasting rainfall, snowmelt or even ice jam. This information might be difficult to obtain, as chronicles usually only provide subjective descriptions of the weather and catchment conditions at that time. Yet, the joint consideration of historical flood events of different flood genesis leads to inhomogeneous samples and hence violates the statistical assumption of identically distributed data. In case of systematic records, Fischer et al. (2019) have proposed to use type-based flood statistics to overcome the issue of inhomogeneous samples. Type-based flood statistics distinguish between different flood types and consider each flood type separately with a unique distribution. The joint distribution of all flood events, the so-called type-based mixture model of partial duration series (TMPS), is then obtained by a mixture model. It was shown in several studies that this kind of information extension, the consideration of the deterministic causes of flood events, improves flood statistics in many ways (Fischer & Schumann, 2023; Yan et al., 2019). Similarly, Garavaglia et al. (2010) propose a mixture distribution to consider weather patterns in different samples for rainfall statistics, also with the aim to obtain more homogeneous samples.

Here, we will incorporate historical flood events in type-based flood statistics. For this purpose, the partial probability weighted moments are extended to the application to peak-over-threshold series and applied to different type-specific samples. This way, the change of the distribution of each flood type due to the consideration of historical events and their impact on the overall distribution function can be attributed. We make use of a hybrid causative-hydrograph-based classification of floods that allows us to classify even long systematic records without the need of extensive model results. Moreover, it can be easily applied also to historical floods, since only information on the rainfall events, its relative extension and intensity, is required. As a comparison, we will also apply a simple Bayesian MCMC approach to see how both approaches differ and to quantify the uncertainty of the estimation.

More precisely, we will address the following research questions:

- How can historical flood events be incorporated in type-based flood statistics?
- How does the consideration of historical flood events alter the type-specific distributions?
- What are the differences between a parameter-based approach and an MCMC approach?
- What is the uncertainty related to such a framework?

Besides the advantage of a larger data basis and hence a probably reduced uncertainty in the estimation of extreme floods, the incorporation of historic flood events in type-based statistics has further advantages. First, the awareness of extreme flood events increases, especially when the most critical flood-generating mechanisms are known. Second, certain flood types that might play no longer a significant role in the generation of flood events due to climate change, such as floods caused by ice jams in Central Europe, can be excluded a-priori from the estimation and hence do not impact flood frequency analysis. The approach has direct practical relevance, as the incorporation of historical floods in flood statistics is recommended in many flood guidelines as is the consideration of mixture distributions, e.g., for the US or for Germany.

2. Data

In this study we consider the Wechselburg catchment, located in Saxony in eastern Germany, with a catchment area of 2099 km^2. It is part of the Mulde river basin and records discharge of the Zwickauer Mulde river. The catchment has a mean elevation of 427 m a.s.l., while the gauge itself is located at 160 m a.s.l. (Fig. 1). The annual precipitation in the catchment is 861.5 mm with a mean annual discharge of 25.8 m^3/s. Within the catchment lies the city Chemnitz, such that 17.4 % of the land cover in the catchment are urban, while 46 % are agriculture and the rest is covered by forest. In the recent years, the Wechselburg catchment was affect by two extreme flood events. In August 2002, a flood event that exceeded all records to that date hit the catchment, leading to enormous damages. Only 11 years later, in June 2013, a flood event of similar peak size again hit the area (Merz et al., 2014). These extreme events make this catchment particularly interesting for the incorporation of historical flood events, as historical events might shed light on how rare extreme events like the ones in 2002 and 2013 in this area really are.

For this study, daily mean and monthly maximum discharges of the systematic records starting in November 1909 and ending in October 2013 were considered (Fig. 2). The daily discharges were used to identify and separate the flood events, while the monthly maximum peaks are used to capture the flood peak. Moreover, daily precipitation data from the E-OBS data set (Cornes et al., 2018) as well as from stations from the German Weather service (DWD) for the time period before 1920 have been used. The latter were interpolated to areal precipitation by simple Thiessen polygons (Okabe et al., 2000). Based on these precipitation data, together with daily temperature data, daily snowmelt was estimated by the degree-day approach with 100 m elevation steps (Rango & Martinec, 1995).

Additionally, information on historical flood events for the whole Mulde basin and in particular for the Wechselburg gauge were available. By intensive studying of chronicles from towns and churches in the region, Fickert (1934) collected rich information not only on the flood events itself but also on the driving mechanisms. He used this information to compare the historical events to the series of (at that time) extraordinary flood events in the years 1920 to 1930. Fickert reports that the five listed flood events (Table 1) are the largest in the period 1433 until the beginning of systematic records in 1910, and while smaller flood events occurred in this period, these were not at all comparable to these extraordinary five flood events. Due to Fickert’s work, the data basis for the Wechselburg is exceptional and provides an excellent basis for this study.

3. Methodology

Here we present the methodology applied in this manuscript to derive type-based flood statistics under the consideration of historical events. The applied flood event separation and classification as well as the statistical model for type-based statistics are described in greater detail also in Fischer and Schumann (2023).

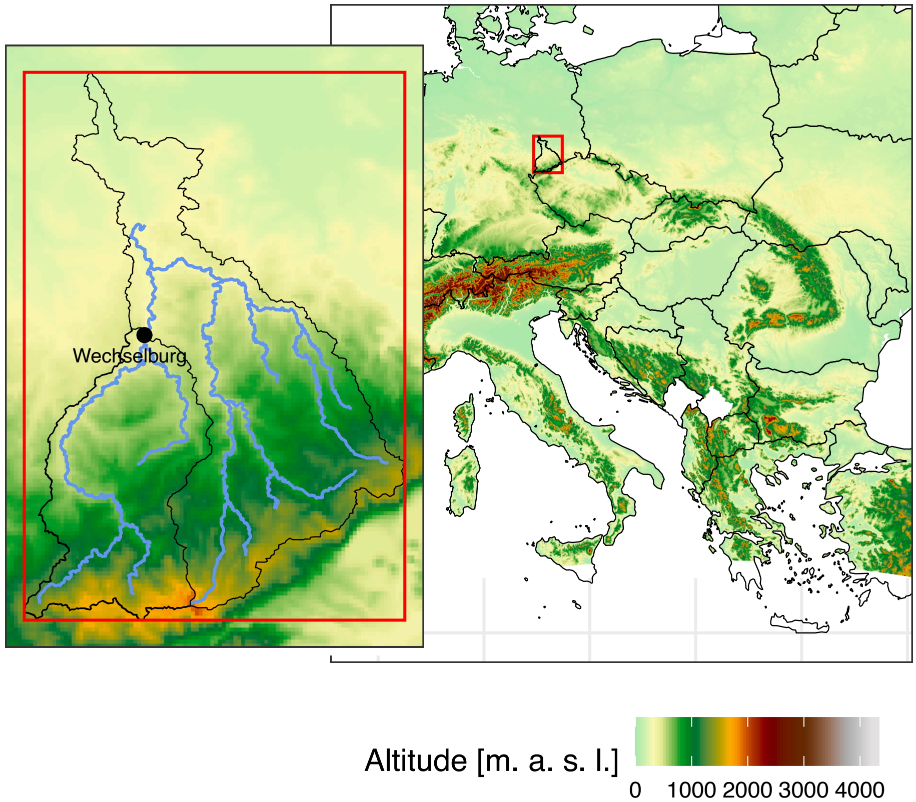

Fig. 1. Location of the catchment in the Mulde river basin (left) and in Europe (right).

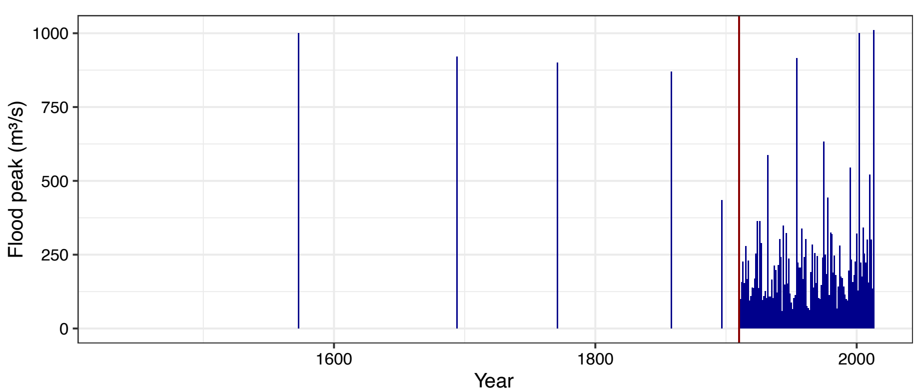

Fig. 2. Annual maximum series for the years 1910 (red vertical line) until 2013 (systematic records) and the historical flood events from the year 1433 on. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Historical flood events in the period 1433 until 1910.

Year	Date	Flood peak (m^3/s)	Rank
1573	1573–08-14	1000	Second largest (together with 2002)
1694	1694–06-20	920	Fourth largest
1771	1771–06-30	900	Sixth largest (after 2013, 2002, 1573,1694 and 1954)
1858	1958–08-01	870	Seventh largest
1897	1897–07-31	435	12th largest

3.1. Flood event classification

In a first step, the systematic records have been scanned for flood events. The identification and separation of flood events from daily discharge series was done by the automatic statistics-based flood event separation by Fischer et al. (2021). This approach uses a moving three-day-window variance-based threshold to identify flood events. The beginning of the corresponding flood event is then defined by the day where the discharge rises for the first time. The end of the event is defined as the point in time when the baseflow, i.e., the discharge at the beginning of the flood event, is reached again after the flood peak, while also small increases in the baseflow are allowed. The baseflow is then separated by a straight-line method, such that flood volume and flood duration can be obtained. The flood peak was obtained from the instantaneous flood records, the monthly maximum discharges, to omit an underestimation of the flood peaks in flood statistics due to a smoothening by daily time steps (Bartens & Haberlandt, 2023).

In a second step, the flood events have been classified according to their genesis. Since we aim to use as long records as possible, a classification based on only few input data available for a long period in time was applied. We chose the hybrid hydrograph-based-causative classification proposed by Fischer et al. (2019), which makes use of daily rainfall and snowmelt data of the catchment only. First, the flood events were classified into rainfall-driven and snow-induced floods. This was done by comparing the proportion of snowmelt on the total amount of flood-generating water to a pre-defined threshold. For Germany, we chose this threshold as 20 %, following recommendations given by Kampf and Lefsky (2016) among others. Second, the flood events were further distinguished. The rainfall-driven flood events were classified into three flood types according to their peak-volume relationship, the so-called flood timescale. More precisely, the events were sorted into three groups such that the linear regression between flood peak and volume is optimal (according to the coefficient of determination). The snowmelt-induced flood events were further distinguished into rain-on-snow floods and snowmelt-driven floods by application of a kMeans clustering to the rainfall and snowmelt sums of the considered events. Therefore, five flood types were considered for this study:

- R1 flood events, which have a small flood timescale and hence are characterised by steep hydrographs of short duration, usually associated with heavy rainfall,
- R2 flood events with a medium flood timescale and therefore a balanced relation between flood peak and volume, usually associated with long-duration rainfall,
- R3 flood events with large flood timescales, characterised by hydrographs of very long duration with often several peaks, usually associated with sequences of rainfall over a long period,
- S1 flood events, which are characterised as rain-on-snow floods, where rainfall falls down on a snow cover,
- S2 flood events, which are characterised as snowmelt-induced floods, i.e., with a high amount of snowmelt and only a negligible amount of rainfall contributing to the generation of the flood event.

These flood types have been utilised before for different studies in Europe and were shown to represent the flood-generating processes well (Brunner & Fischer, 2022; Fischer & Schumann, 2022; Fischer et al., 2019). Yet, for the approach proposed here, basically any flood classification could be applied, as long as it is available for long records and applicable to historical events.

3.2. Type-based flood statistics

For type-based flood statistics, we applied the type-based mixture model of partial duration series (TMPS) developed by Fischer (2018) and Fischer et al. (2019). This model considers the flood types in different POT-samples and combines these in a multiplicative mixture model. Assume that we have M flood types (here: $M = 5$). Each flood type $j = 1, \ldots, M$ is represented by a sample $X_{j;1}, \ldots, X_{j;n_j}$ of flood events. First, each flood type is modelled separately by a POT-approach with threshold u_j. The exceedances $X_{j;i} > u_j$ are assumed to follow a Generalized Pareto distribution G_j with parameters $\theta_j = \left(\kappa_j, \beta_j\right)$, where κ_j is the shape parameter and β_j is the scale parameter, a typical extreme value model for exceedances (Balkeema-de Haan-Theorem, Balkeema & de Haan, 1974). The threshold u_j was chosen as three times the long-term mean discharge weighted by the frequency of the respective flood type in each month, an empirical hydrology-based threshold that makes the obtained distribution comparable to the one of the annual maxima as already discussed in Fischer (2018). Theoretically, other thresholds could be taken into account, though it has to be noticed that the choice of the threshold may critically impact the results. To obtain the distribution of the whole sample of flood type j, besides the distribution of the exceedances also the probability of non-exceedance of the threshold must be known. Here, we estimated this probability by fitting a Generalized Extreme Value Distribution, H_j, to all flood events of type j. H_j has shape parameter ξ_j, scale parameter σ_j and location parameter μ_j. Therefore, the distribution function of flood events of type j is given by

$$F_j(x) = H_j(u_j) + (1 - H_j(u_j))G_j(x)$$

$$= \exp\left(-\left(1+\xi_j\frac{u_j-\mu_j}{\sigma_j}\right)^{-\frac{1}{\xi_j}}\right) + \left(1-\left(1+\kappa_j\left(\frac{x-u_j}{\beta_j}\right)\right)^{\frac{-1}{\kappa_j}}\right)\left(1 - \exp\left(-\left(1+\xi_j\frac{u_j-\mu_j}{\sigma_j}\right)^{-\frac{1}{\xi_j}}\right)\right).$$

To consider the annual distribution of flood type j, i.e., the one required to estimate annual return periods, one can simply combine the distribution G_j with a discrete distribution modelling the probability that the number of events per year of flood type j, λ_j, is equal to r:

$$\widetilde{F}_j(x) = \sum_{r=0}^{\infty} \mathbb{P}_j(\lambda_j = r)G_j(x)^r$$

We chose a Poisson distribution for this purpose.

Finally, to consider all flood types jointly, a mixture model is applied to receive the TMPS model with distribution

$$F(x) = \prod_{j=1}^{M} F_j(x)$$

3.3. Partial probability weighted moments (PPWM)

As stated before, there exist various possibilities on how to include historical flood events in flood statistics. One that includes these events directly in the parameter estimation and is at the same time robust and efficient was proposed by Wang (1990). The idea is to split the sample of all flood events into those above a given threshold and those below this threshold. The threshold is assumed to be the flood peak from which on it is highly probably that a flood event was recorded in some chronicles or similar, therefore becoming a historical flood event. Those times that do not appear in the chronicles are assumed to only have flood events below the threshold. The sample of events above the threshold therefore consists of all historical flood events as well as those events of the systematic records that are greater than the threshold.

Statistically speaking, one has a sample of n flood events $X_1, \ldots, X_n$ in the systematic records. Furthermore, one has knowledge on flood events over a historical period of m years, though only for those events greater than a threshold u_H the flood peak is known. Let's assume that this applies to l event in the historical series denoted with X_h, while k events in the systematic record exceed the threshold, too. Wang (1990) then defines a censored sample $X_{(n-k+1)}, \ldots, X_{(n)}, X^h_{(m-l+1)}, \ldots, X^h_{(m)}$ with the corresponding order statistic $X'_{(n+m-k-l+2)} \leqslant \ldots \leqslant X'_{(n+m)}$. Additionally, we have a second sample $X_{(1)}, \ldots, X_{(n-k)}$ with values from the systematic records below the threshold. Based on these two samples, now parameter estimators can be obtained. Wang (1990) proposes the probability weighted moments (PWM, Greenwood et al., 1979). These are the basis for L-moment estimators, which are widely used in flood statistics due to their robustness and efficiency even for small samples (Hosking, 1990). Probability weighted moments of order i for a random variable X with distribution $F(x) = \mathbb{P}(X \leqslant x)$ are defined as

$$\beta_i = \int_0^1 x(F)F^i dF.$$

Unbiased estimators for probability weighted moments of order i,b_i, of a sample $X_1,\ldots,X_n$ are given by

$$b_i = \frac{1}{n}\sum_{j=1}^{n}\frac{(j-1)(j-2)\cdot\ldots\cdot(j-i)}{(n-1)(n-2)\cdot\ldots\cdot(n-i)}X_{(i)}$$

Probability weighted moments β_i including historical events can be obtained from the two censored samples by the so-called partial probability weighted moments (PPWM)

$$\beta_i = \int_0^1 x(F)F^i dF = \int_0^{F(u_H)} x(F)F^i dF + \int_{F(u_H)}^1 x(F)F^i dF = \beta_i^{'} + \beta_i^{''}$$

Unbiased estimators based on the censored sample below the threshold are given by

$b_i^{'} = \frac{1}{n}\sum_{j=1}^{n}\frac{(j-1)(j-2)\cdot\ldots\cdot(j-i)}{(n-1)(n-2)\cdot\ldots\cdot(n-i)}X_{(j)}^{*}$ with

$$X_{(i)}^{*} = \begin{cases} X_{(i)}, \text{if } X_{(i)} \leqslant u_H \\ 0, \text{else} \end{cases}$$

and based on the censored sample above the threshold by

$$b_i^{''} = \frac{1}{n+m}\sum_{j=n+m-k-l+2}^{n+m}\frac{(j-1)(j-2)\cdot\ldots\cdot(j-i)}{(n+m-1)(n+m-2)\cdot\ldots\cdot(n+m-i)}X_{(j)}^{'}$$

Wang (1990) showed that the use of $b_i^{'}$ and $b_i^{''}$ leads to unbiased estimates of $\beta_i^{'}$ and $\beta_i^{''}$.

In the context of type-based flood statistics, things become a bit more complicated. In this case, we want to apply the PPWM approach to the parameter estimation for each type-specific distribution $F_j, j = 1,\ldots,M$, therefore for a POT-sample for G_j and the whole uncensored sample for H_j. Please note that the equations above apply to typical flood samples, mainly annual series. For the type-specific samples, however, we deal with POT samples, which are censored already. Moreover, the number of events in the historical period is unknown, as it is no longer equal to the number of years, since in POT series more than one event per year can occur. Therefore, we assume here a constant number of flood events of a given flood type per year. This number can thus be estimated from the systematic records and be transferred to the historical period. Let us assume that for flood type j we have an average number of λ_j events per year in the systematic records. Then, for the historical period of m years we would have a number of $m{\cdot}\lambda_j$ events, where l_j of the events exceed the type-specific threshold $u_{H;j}$. In the systematic sample of flood type j, $X_{j;1}$, $\ldots,X_{j;n_j}$, k_j events exceed the threshold $u_{H;j}$.

The PPWMs for a censored series like the POT series considered for the distribution function G_j were provided by Wang (1990). While the sample above the historical threshold $u_{H;j}$ remains the same as before, the sample below this threshold is now censored from above by $u_{H;j}$ and from below by the type-specific threshold of the POT sample, u_j. Therefore, the PPWMs of order i for each flood type j are given by

$$\beta_{i;j} = \int_{F_j(u_j)}^{F_j(u_{H;j})} x(F_j)F_j^i dF_j + \int_{F_j(u_{H;j})}^1 x(F_j)F^i dF = \beta_i^{'''} + \beta_i^{''}$$

The corresponding unbiased estimators then would be

$b_{i;j}^{'''} = \frac{1}{n_j}\sum_{t=1}^{n_j}\frac{(t-1)(t-2)\cdot\ldots\cdot(t-i)}{(n_j-1)(n_j-2)\cdot\ldots\cdot(n_j-i)}X_{j;(t)}^{*}$ with

$$X_{(k)}^{*} = \begin{cases} X_{j;(k)}, \text{if } u_j < X_{j;(k)} \leqslant u_{H;j} \\ 0, \text{else} \end{cases}$$

and

$$b_{i;j}^{''} = \frac{1}{n_j+m{\cdot}\lambda_j-k_j-l_j}\sum_{t=n_j+m\cdot\lambda_j-k_j-l_j+2}^{n_j+m\cdot\lambda_j}\frac{(t-1)(t-2)\cdot\ldots\cdot(t-i)}{(n_j+m{\cdot}\lambda_j-1)(n_j+m{\cdot}\lambda_j-2)\cdot\ldots\cdot(n_j+m{\cdot}\lambda_j-i)}X_{j;(t)}^{'}$$

With this extended version of the PPWM, the parameters of the distribution G_j can be estimated. We use here the PPWMs to obtain L-Moments in the traditional way.

The parameter estimation for the distribution function H_j is analogously to the case of annual maximum series described before though with an adjusted number of historical events equal to $m{\cdot}\lambda_j$.

For visual comparison in the distribution plots, plotting positions of the annual maximum series were estimated. When considering systematic data only, i.e. n annual maximum floods X_i, the traditional Weibull plotting positions

$$pp(i) = \frac{rank(X_i)}{n+1}$$

were considered, where $pp(i)$ is the plotting position of the event i.

In the case of h historical data Y_i in a period of m years and n systematic annual maximum data X_i, whereof in total $n^{'} = h+l$ exceed the threshold of historical data u_H, the plotting positions $pp^h(i)$ described in Hirsch (1987) were considered:

$$pp^h(i) = \begin{cases} \dfrac{rank(X_i;X,Y)}{n+1}, \text{if } X_i < u_H \\[2ex] \dfrac{n^{'}}{m+n} + \dfrac{rank(X_i;X,Y)-n^{'}}{n+h-n^{'}+1}\cdot\dfrac{n+m-n^{'}}{n+m}, \text{if } X_i \geqslant u_H. \end{cases}$$

where $rank(X_i;X,Y)$ refers to the rank of the observation X_i conditional on the systematic and historical records.

When considering POT series, as it is done for the type-specific distributions, the plotting positions for flood type j are changed according to the relationship $T = 1-\frac{1}{\lambda_j(1-p)}$.

3.4. Bayesian Markov Chain Monte Carlo (MCMC)

As a second approach to consider historical flood events in the TMPS model, we apply the Bayesian Markov Chain Monte Carlo (MCMC) approach. This approach has been used before in the context of historical floods for annual maximum series and regionalisation approaches (Reis & Stedinger, 2005; Payrastre et al., 2011; Gaume et al., 2010; Gaál et al., 2010) or for rainfall (Isikwue et al., 2015). The basic idea is to include prior information in the estimation. In traditional Bayesian estimation approaches, it is crucial to estimate the normalisation constant, which is difficult in many cases, especially when having complex models. MCMC, instead, samples values of the parameters from the posterior distribution without computing the normalisation constant. Here, we apply the Metropolis-Hasting algorithm, which generates data with a distribution converging to the posterior distribution (Reis and Stedinger, 2005). Roughly speaking, this approach is similar to sampling techniques, yet not with independent draws. Here, we even assume no a-priori information, avoiding assumptions on a suitable distribution, a suitable variance and an acceptance rate (see MacPherson et al., 2023). The MCMC approach has the advantage that also confidence intervals of the flood quantiles can be obtained easily.

To make Bayesian MCMC applicable to the TMPS model, we make a few assumptions. First, we assume that the probability of exceedance of the threshold u_j is constant over time. Therefore, it does not have to be estimated via Bayesian MCMC, keeping the algorithm simple. Second, if we do not observe historical flood events of a given flood type, we assume that all historical floods lie below the historical threshold $u_{H;j}$. This is reasonable, as large historical floods would have been recorded.

The Likelihood function of the TMPS model for the systematic series $X_1, \ldots, X_n$ with parameter set θ then is obtained as follows

$$L(X|\theta) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} (F(x_i))'$$

$$= \prod_{i=1}^{n} \prod_{j=1}^{5} \left((p_{u;j}) + (1 - p_{u;j}) G_j(x_{i;j})' \right)$$

$$= \prod_{i=1}^{n} \sum_{j=1}^{5} (1 - p_{u;j}) g_j(x) \prod_{\{k \in 1, \ldots, 5\}/j} F_k(x_{i;j})$$

where $p_{u;j} = H_j(u_j)$ is assumed to be known.

When considering systematical and historical records $Y_1, \ldots, Y_h$ jointly, the Likelihood function is given by the product of both Likelihoods (MacPherson et al., 2023):

$$L(X, Y|\theta) = L(X|\theta) \cdot L(Y|\theta)$$

where $L(Y|\theta) = \prod_{j=1}^{h} f(y_j) \cdot P_j(h|\theta)$ and $P_j(h|\theta)$ is assumed to be known.

The likelihood function for the AMS series are derived analogously, though solely based on the GEV distribution, i.e. $f(x)$ would be replaced by $h(x)$ in this case.

4. Results and discussion

4.1. Classification of systematic and historical flood events

After application of the flood event separation and flood classification to the systematic records of Wechselburg gauge, 248 flood events have been identified for the years 1910 until 2013, i.e., an average of 2.38 flood events per year. Thereof, 74 events have been classified as type R1 (on average 0.712 events per year), 41 as type R2 (on average 0.394 events per year), 74 as type R3, 23 as type S1 and 36 as type S2. The majority of events therefore are rainfall driven. To be able to understand the meteorological drivers of these flood events and to later on classify the historical events according to these drivers, the characteristics of the flood-related rainfall were compared (Fig. 3). Besides rainfall sum, rainfall duration and runoff coefficient of each flood events, also the relative intensity of the flood-related rainfall was considered, which is defined as the proportion of the maximum 1-day rainfall sum on the total sum of rainfall. For flood types S1 and S2, it has to be considered for this comparison that the contributing snowmelt is not considered. The results prove significant differences between the flood-related rainfall events of different flood types. R1 flood events are associated with rainfall events of short duration of only one or two days (probably shorter, but that would be hidden by the daily resolution of the input data). While the overall amount of rainfall is not large, the relative intensity is: On average, 60 % of the total precipitation falls within one day. This leads to surface runoff, since the soil often cannot infiltrate such high intensities in such short time. R1 flood events therefore can be clearly associated with heavy rainfall. The rainfall sums as well as the duration of R2 floods are larger than for R1 floods, while the rain intensities are more uniform and smaller. With an average rainfall duration of three days and larger runoff coefficients, a larger amount of rainfall is transferred to subsurface runoff. Obviously, long-duration synoptic rainfall events can be associated with this flood type. Flood type R3 is associated with the longest rainfall events (on average 10 days) with largest rain sums. The intensity is low and uniformly distributed over the whole rainfall event, leading to large runoff coefficients. The rainfall associated with flood type S1, the rain-on-snow events, is comparable to that of the R1 floods, though with lesser intensity. This is reasonable, as the energy transferred by the rainfall into the snow cover that leads to a rapid snow melt has to be large. Finally, the rainfall associated with the snowmelt-driven floods (S2) can be seen as negligible. The sum is comparably small (and might even contain snow) and distributed over a long period of time. Here, clearly snow melt from the snow cover is the main driver.

The annual distribution of the flood events is given in Fig. 4. There, a circular statistic was applied (Burn, 1997), where the coordinates in the unit circle were multiplied with the empirical non-exceedance

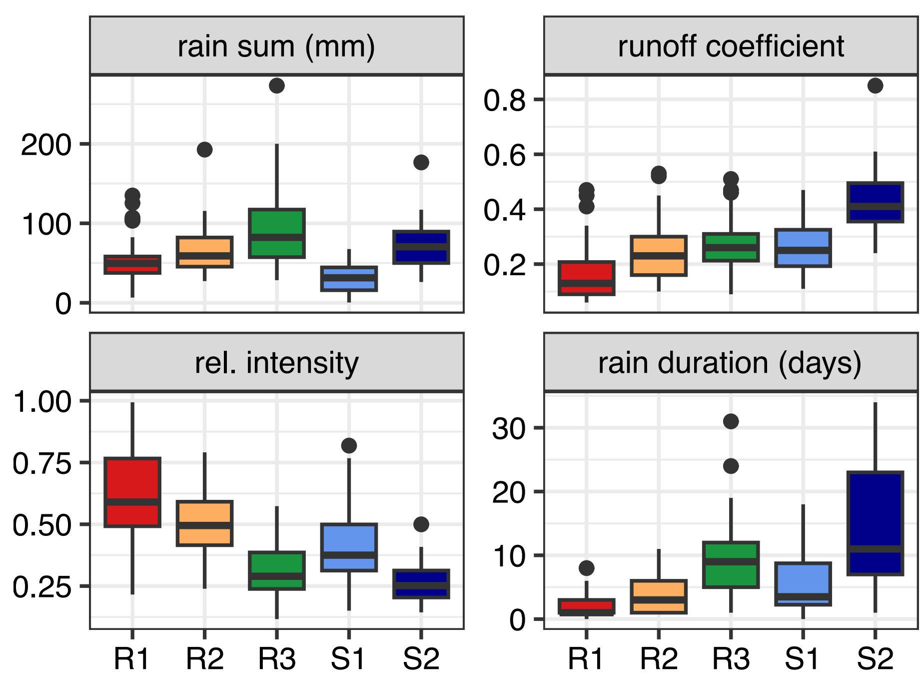

Fig. 3. Characteristics of flood-related rainfall events differentiated by the flood type.

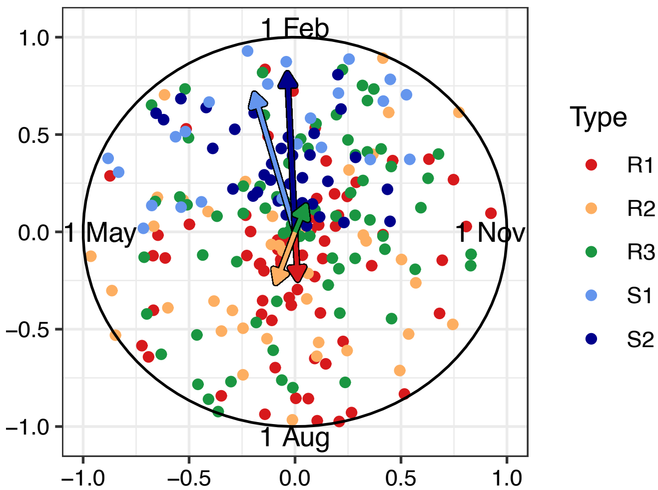

Fig. 4. Circular statistics for the identified flood events at Wechselburg gauge. The distance of each dot to the unit circle shows the peak magnitude of the event, i.e. the closer to the unit circle the larger the flood peak. Arrows indicate the mean day of occurrence of a flood event of the respective flood type. The length of the arrow indicates the probability of an event of this type to occur at that day.

probability. The flood events with largest flood peaks are thus displayed closest to unit circle, while flood events with small flood peaks are displayed close to the centre. The largest events (close to the unit circle) clearly occur in summer and early winter. In August, most of the flood events are caused by heavy rainfall (type R1), while in July sequences of rainfall and therefore probably a wet soil led to the most extreme floods (type R3). In spring and early winter, instead, long-duration rainfall led to the largest flood events (type R2). Snow-induced flood event only play a minor role for the generation of extreme floods, and if they do, then only in combination with heavy rainfall (type S1). The mean day of occurrence of flood types S1 and S2 is clearly in January, while heavy-rainfall floods (R1) and long-duration-rainfall floods (R2) tend to occur in August respectively July, though less uniformly.

To include the historical events in Table 1 in the parameter estimation, these have to be classified first. From the previous analyses it is obvious that the historical events must be rainfall-driven, since all flood events occur in summer. However, from the flood peak alone it is not clear into which of the three rainfall-driven flood types these events should be classified. If the flood volume would be known, a classification would be easily possible with the given classification as simply the peak-volume-relationship would be compared to those of the systematic records. However, flood volume is rarely available for historical floods or afflicted with high uncertainty due to changing hydraulic conditions in flood plains. However, information on the corresponding rainfall might be available, though not as specific rainfall amounts but as descriptions of the rainfall events, i.e., their duration, intensity and extension. For this purpose, information from chronicles is often valuable. For the Wechselburg gauge, Fickert (1934) has summarised such rainfall information that is now used here for the classification. For the 1573 flood event, it is known that the flood was caused by “catastrophic downpour” which concentrated on August 12 on small cells of only few square kilometres in the catchments in the Mulde river basin. Therefore, rainfall of a duration of less than two days led to an extreme flood. Given this information and having in mind Fig. 2, we classified the 1573 as heavy-rainfall-driven flood and therefore R1. For the 1694 flood, it is known that it was caused by short, intensive rainfall that only occurred for one day. It was therefore classified as R1 event, too. The 1771 flood was quite different. The whole June of the year 1771 was characterised by several long-duration rainfall events, leading to several smaller floods in early and mid-June. Finally, on June 30th and Juli 1st, a rainfall event with extraordinary large rainfall sums hit large parts of several sub-basins in the Mulde river basin. The wet soil led to a fast reaction of the catchment and the catastrophic flood event. The temporal and spatial extension of the flood-generating rainfall event leads to the conclusion that it must have been an R2-flood event. The flood event 1858 followed a period of drought in July 1858. On July 28th, rainfall started that lasted for several days with constant intensity. Finally, on August 1st, the intensity intensified for large areas of the Mulde river basin, causing the flood event. Therefore, again it can be concluded that the 1858 was an R2 flood event. The best documented historical flood event in the Mulde river basin is the one occurring in 1897, since there already rainfall records (though in small scale) were available. These records prove that rainfall over four days with only rare extreme intensities were the trigger for the 1897 flood event. Therefore, this flood event was classified as R2 flood, too.

4.2. *Incorporation of historical flood events in type-based flood statistics*

With the previous flood classification, the application of the PPWMs to the type-specific flood samples (historical and systematic) was possible. In the following, we will evaluate how the distributions including historical events changed compared to when only the systematic records are used for flood statistics. A special focus is laid on the shape parameter since this parameter is mainly responsible for the shape of the right tail and is known to be most impacted by the extreme floods and also by uncertainty in estimation.

First, the input parameters of the PPWM approach must be defined. The historical records start in the year 1433, while the systematic records cover the period 1910 to 2013. The historical period therefore covers $m = 478$ years. The beginning of the historical period was set to 1433, as since then it can be assumed that all historical flood events have been noted (Fickert, 1934). Next, the threshold for the historical events had to be defined. Here, we chose the threshold type-specifically, i.e., the threshold may vary with the flood type. The reason behind this assumption is that spatially limited events, as caused, e.g., by heavy rainfall, might only appear in chronicles if they were extreme. Only in this case they might have affected enough people or caused enough damage to be reported. Spatially extensive events, like R2 or R3 floods which are caused by large-scale rainfall events, instead, automatically

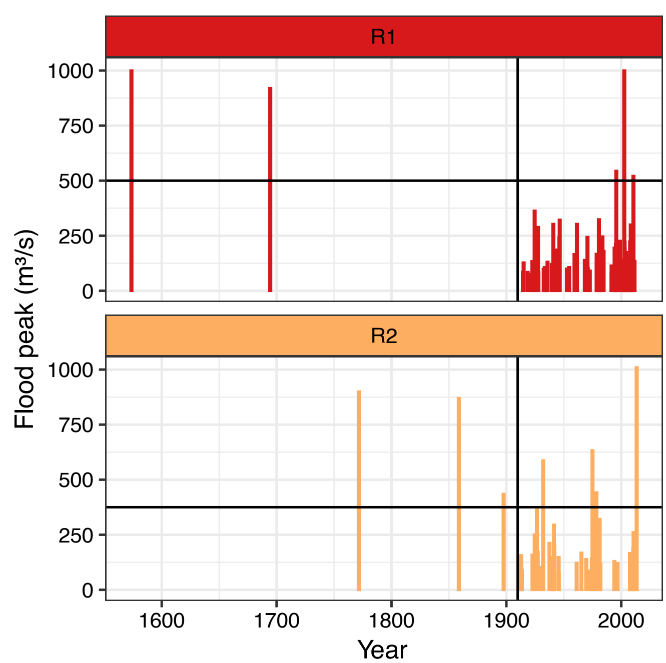

Fig. 5. Flood events of the systematic and the historical period, differentiated by the flood type. The black vertical line denotes the beginning of the systematic records, the black horizontal lines are the type-specific thresholds $u_{H;j}$.

affect more people and might thus be recorded even if the flood peak itself was not that extreme. We chose $u_{H;1} = 500$ m^3/s and $u_{H;2} = 375$ m^3/s such that historical events as well as only the most extreme events of the systematic records exceed the thresholds (Fig. 5). The choice of these thresholds is a subjective one and might affect the goodness of fit critically. For example, Wang (1990) uses x_{100} as the historical threshold such that on average five flood events would lie above the threshold when having historical and systematic records of 500 years. In our example, we have 481 years of historical and systematic records and with the chosen thresholds 5 (R1) respectively 7 (R2) flood events exceed the thresholds. Therefore, the chosen thresholds are comparable to the choice of Wang (1990). However, other thresholds are possible, and the choice might depend on the catchment under consideration. One could check the assumption made for the threshold by application of a Poisson test (Lang et al., 1999). To allow to investigate the impact of the choice of the type-specific threshold, we have repeated the estimation with one unique threshold $u_H = u_{H;1} = u_{H;2} = 500$ (Figure A.1 in the appendix). The results did not differ much.

As a comparison, the PPWM approach was applied to the annual maximum series (AMS), too. Here, we used the GEV distribution and a historical threshold of 800 m^3/s, leading to five flood events in the historical and systematic records to lie above the threshold. This approach is analogous to the one proposed by Wang (1990).

When comparing the distribution functions based on the systematic records only (Fig. 6a) and of systematic and historical records (Fig. 6b), differences become visible. First of all, the resulting joint TMPS model results in significantly smaller flood quantiles for return period from 100 years on. Additionally, the right tail seems to be less heavy, leading to a less exponential shape in the plot. At the same time, the consideration of historical events leads to slightly larger flood quantiles for return periods in the range of 5 to 50 years. Comparing the results to those obtained when using the annual maximums series (AMS) and a GEV distribution, it appears that both models change quite similarly. When comparing the 100-year floods, one can see that both approaches, AMS and TMPS, deliver quite similar flood quantiles when considering the systematic data only. The 100-year flood in this case is in a range of 914 m^3/s (AMS) to 975 m^3/s (TMPS). When considering the historical and the systematic records jointly, the 100-year flood decreases to 764 m^3/s (AMS) respectively 899 m^3/s (TMPS). Thus, the flood quantiles of the AMS approach decrease similar to those of the TMPS model.

This implies that the PPWMs have a similar effect on the TMPS model and on the classical AMS approach. However, for the TMPS model it is now possible to attribute the changes by having a look at the type-specific distribution. Since there have been no historical flood events of flood types R3, S1 and S2, the distribution functions of these flood types do not change. For the two flood types with historical events, R1 and R2, however, significant changes become visible. For the heavy-rainfall floods (R1), the flood quantiles become larger for most return periods. This is probably caused by the two extreme historical floods that have been added to the record and that are among the largest floods of all time in this catchment. Yet, the right tail seems to be less heavy. For the R2 floods, a clear decrease of the tail heaviness of the right tail can be observed. Instead of an exponential increase in the plot, a converging behaviour occurs for the distribution function when adding the historical floods to the records. It can therefore be concluded that the general decrease of R2 flood quantiles leads to a decrease of flood quantiles in the TMPS model. Yet, the right tail of the TMPS model is still dominated by the R1 flood events, leading to a slightly heavy tail. The plotting positions for the AMS confirm the benefit of incorporating historical flood events. The largest floods are better captured by the distribution function. However, plotting positions depend much on the observation period. Therefore, their explanatory power is limited, which has to be considered here. Plotting positions for the type-specific distributions can be found in Figure A.2 in the appendix.

The observations regarding the TMPS model are confirmed when having a look at the distribution parameters (Table 2). Indeed, for both flood types, R1 and R2, the shape parameter κ_j of the GPD distribution, which serves as an indicator for skewness and tail heaviness, has decreased for the PPWM approach. For heavy-rainfall floods (R1), it decreased from 0.201 to 0.051 and therefore only a slight heavy tail remains. For the R2 floods, the heavy tail ($\kappa_2 = 0.189$) changed to a light tail ($\kappa_2 = -0.428$). At the same time, the variability indicated by the scale parameter β_j increased significantly. The GEV distribution included in the type-specific distribution is less affected. Since the GEV is only used to estimate the probability of exceedance of the threshold u_j, it has limited effect on the overall distribution and is only given here for sake of completion.

What is striking for the PPWM approach is that the variability increases significantly for the GPD distribution. This is because in the PPWM approach, for the historical period all events below the threshold

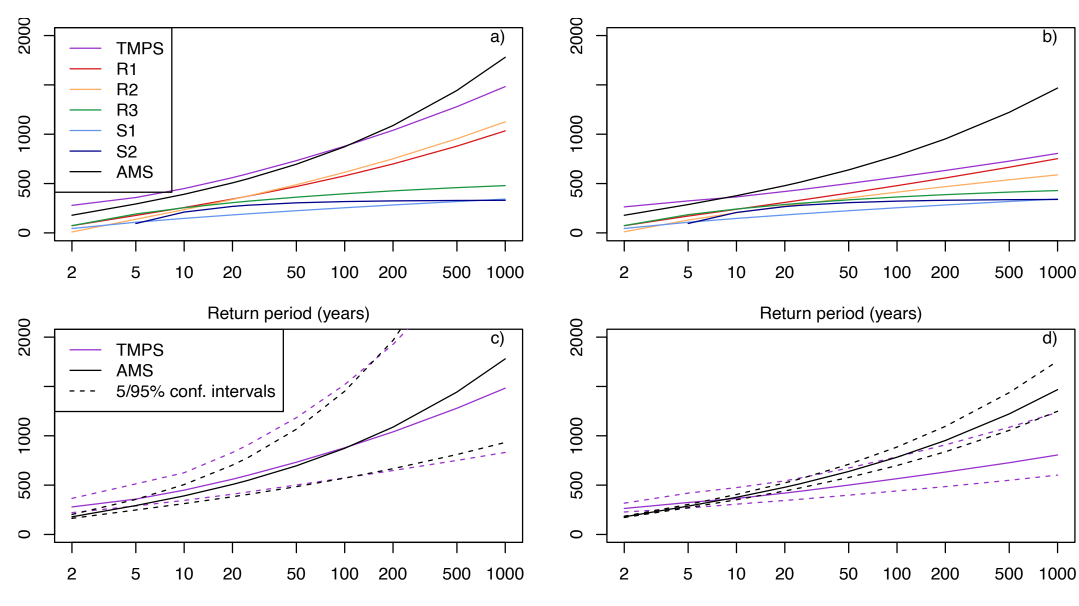

Fig. 6. Distribution functions for given return periods based on a) the systematic records only and b) the systematic and historical records derived with PPWM. Black dots indicate the plotting positions.

Table 2
Distribution parameters of the AMS and TMPS model for the PWM approach for systematic records only (Syst.) and the PPWM approach for historical and systematic records (Hist.).

Model	Generalized Pareto Distribution						Generalized Extreme Value Distribution					
	κ_j		β_j		u_j		ξ_j		σ_j		μ_j	
	Syst.	Hist.	Syst.	Hist.	Syst.	Hist.	Syst.	Hist.	Syst.	Hist.	Syst.	Hist.
AMS	–	–	–	–	–	–	0.304	0.212	76.5	80.3	147.6	138.7
R1	0.201	0.051	80.89	165.5	73.6	73.6	0.434	0.254	45.3	51.9	113.6	104.4
R2	0.189	−0.428	111.74	291.2	77.7	77.7	0.477	0.279	54.8	66.7	128.5	101.1
R3	−0.196	–	111.1	–	81.4	–	0.273	–	45.2	–	123.6	–
S1	−0.045	–	53.3	–	90.9	–	0.213	–	29.4	–	110.7	–
S2	−0.796	–	193.6	–	98.7	–	−0.015	–	56.8	–	174.5	–

are assumed to be equal to the mean. Together with few extreme events this increases the variability rapidly, especially when considering a high threshold. Still, the resulting distributions seem to be reasonable as they represent the empirical frequency of the extreme events well. For example, a flood event with flood peak 1000 m^3/s has a return period of about 200 years. Therefore, the three most extreme floods with about that flood peak magnitude would have a theoretical probability close the empirical probability when considering the systematic and historic records, as can be seen also by the plotting positions.

4.3. Comparison to MCMC approach

When comparing the results of the MCMC approach to those of the PPWM approach, a difference of the resulting distribution functions can be seen (Fig. 7a and b). This applies mainly to the flood types R1 and R2, as can be expected since only there additional information in terms of historical events is used. For these two flood types, only the shape parameter of the GPD changes significantly (Table 3). For both flood types it reduces to values close to zero and hence to a light-tailed distribution, which is visible in Fig. 5a. Such behaviour is then inherited by the TMPS model which, compared to the results of the systematic data only (Fig. 6a), also has a light tail and much smaller flood quantiles for return periods from 10 years on. Even for a return period of 1000 years, the corresponding flood quantile of the TMPS model lies at 804 m^3/s and thus below the five largest flood events in the records. Compared to the results of the PPWM approach, the smaller flood quantiles up to 5 years are similar but for larger return periods the flood quantiles derived with the MCMC approach are much smaller. This is mainly due to the smaller scale parameters derived by the MCMC approach. It appears that the MCMC approach assumes a smaller variability of the flood events. This is especially crucial for the heavy-rainfall floods of type R1. Here, a higher variability is able to better capture the extreme events that occurred in the past. The MCMC approach, however, results in a distribution that would assign these events with return periods much higher than 1000 years which is questionable as such events have appeared a couple of times in the historical and systematic records. When comparing the 100-year flood events, similar differences become evident. While for the systematic period, both approaches, AMS and TMPS, result in almost the same flood quantile (877 m^3/s) with confidence bands between 577 and 1468 m^3/s (AMS) respectively 594 and 1532 m^3/s (TMPS), the difference increases when considering the historical events. The 100-year flood of the TMPS model decreases to 564 m^3/s (434–756 m^3/s), while the 100-year flood of the AMS decreases to 783 m^3/s (701–905 m^3/s).

Table 3
Distribution parameters of the AMS and TMPS model for the MCMC approach for systematic records only (Syst.) and for historical and systematic records (Hist.).

Model	Generalized Pareto Distribution					
	κ_j		β_j		u_j	
	Syst.	Hist.	Syst.	Hist.	Syst.	Hist.
R1	0.152	0.072	82.4	81.7	73.6	73.6
R2	0.125	−0.068	120.2	104.0	77.7	77.7
R3	−0.218	−0.262	114.1	111.0	81.4	81.4
S1	−0.073	−0.060	54.8	53.9	90.9	90.9
S2	−0.810	−0.811	197.4	196.6	98.7	98.7
	Generalized Extreme Value Distribution					
	ξ_j		σ_j		μ_j	
AMS	0.304	0.212	76.5	80.3	147.6	138.7

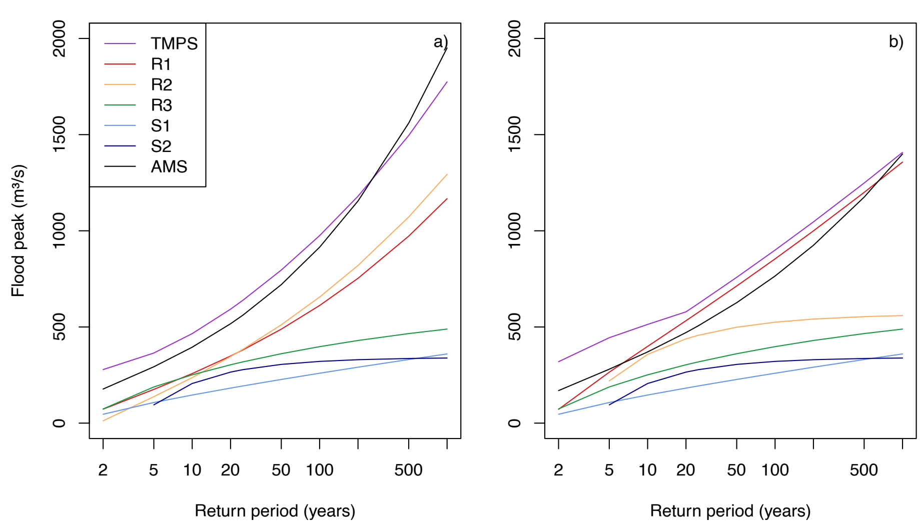

Fig. 7. Distribution functions for given return periods based on a) the systematic and b) systematic and historical records derived with MCMC; c) and d): Corresponding confidence bands for TMPS and AMS models.

Comparing the results of the PPWM and the MCMC approach for the AMS series demonstrates that both approaches have almost no difference. Here, the variability is reduced similarly. From these results it can be concluded that the MCMC approach seems to underestimate the variability of the historical period. The obtained flood quantiles for the TMPS model seem to be too small compared to the observed extreme floods. Especially heavy-rainfall floods seem to be underestimated by this approach, while the PPWM approach better captures this flood type. For the AMS approach, no difference between the two approaches becomes apparent. This is due to a different weighting in the MCMC approach. Such a behaviour should be noticed when considering flood types in flood statistics in the context of historical floods.

The MCMC approach also allows to investigate the uncertainty related to the estimation of the models. In theory, this is also possible for the PPWM method, e.g. by nonparametric bootstrap. However, in such a case the resampling with replacement for historical floods would not be meaningful, as the variability is limited. Moreover, when considering type-specific distributions, it has to be taken into account that neither the frequencies nor the event magnitudes of the flood types should be altered by the approach. Thus, a resampling would only be possible within the type-specific samples. For these reasons, such an approach was not considered in Section 4.2. The MCMC approach does not have these issues.

It can be seen in Fig. 7c and d that the uncertainty related to the TMPS is slightly higher than for the AMS approach. This mainly due to the higher number of parameters used in this model. Though the higher number of flood events considered by the POT approach balances this a bit, the uncertainty remains high especially for large return periods. This was already reported by Fischer and Schumann (2023). In the upper range of the confidence bands of the TMPS model, flood quantiles of up to 1000 m^3/s for a return period of 1000 years could be obtained. However, the 95 % confidence interval of the TMPS model does not even reach the 5 % interval of the AMS approach, demonstrating again that the TMPS model is significantly smaller than the AMS model when using the MCMC approach for large return periods. Yet, it can also be seen that the incorporation of historical events significantly reduces the uncertainty related to the estimation of both approaches.

4.4. Limitations of the approach

The previous investigations have demonstrated that both approaches, PPWM and MCMC can be applied to the TMPS model. However, the PPWM approach resulted in quantile estimates that seem to better represent the observed return periods of the most extreme floods. Still, there are limitations of this approach that have to be discussed.

One crucial choice that impacts the model results is the threshold. As always for POT approaches, the choice of the threshold can increase or decrease the flood quantile estimates. Here, we made two assumptions on the threshold. First the type-specific POT-threshold in the TMPS model was chosen as three times the weighted long-term mean discharge, where the weights were chosen according to the frequency of the respective flood type in each month. This choice was made based on the empirical, hydrological consideration that a flood event should be several times higher than the mean discharge (Dyck et al., 1980). It was discussed in comparison to other, statistical approaches before by Fischer (2018). Depending on the catchment and the climatic conditions, other choices might be appropriate, too. For central Europe, this threshold has been proven to deliver reliable flood estimates. The second choice of a threshold was made for the historical threshold, i.e., the flood peak from which on it is assumed that historical flood events have been recorded. Here, we used a visual inspection of the type-specific samples and chose the threshold such that it clearly separated extreme flood events from ordinary flood events. As a result, in nearly 600 years of records, five to seven flood events exceeded this threshold, a number quite close to what was assumed by Wang (1990) as a reasonable threshold. Wang (1990) used the empirical 100-year flood and argued that a return period of 100 years indicates an extreme level. Again, other choices might be reasonable here and should be investigated further in future studies. The assumption of type-specific historic thresholds did not influence the results much but might be hydrologically more reasonable than one fixed threshold for all flood types, as the severity and extension of a flood event and thus the probability to be reported on varies much with the flood type. Spatially extended flood events like R2 or R3 floods affect more people and hence might be recorded more frequently, even though their magnitude might be smaller than that of spatially limited heavy-rainfall floods. The assumption of these thresholds applied to both approaches, PPWM and MCM, and hence affects both models alike.

Another assumption made for the models was that of stationarity. It was assumed that the frequency of flood types did not change over the last 600 years and that therefore it can be transferred to the historical period. Moreover, it was assumed that the probability distributions were stationary. Since in the last centuries some flood types such as heavy-rainfall floods have become more frequent (Fischer & Schumann, 2023 & 2024) and other flood types might have vanished (such as ice floods in Germany), such an assumption is at least questionable. Yet, it is difficult to obtain information on these past changes solely from the systematic records. Moreover, the impact of the frequency of events in the historical period on the resulting flood is small, as it is considered only indirectly in the ranks used for the PPWM. For the MCMC approach, the impact is assumed to be greater as here the whole period of historical floods is simulated under this assumption. Alternatively, one could think about having time-varying weights in the mixture model. Such an approach would take into account that the impact of certain flood types changes over time. Besides a varying frequency, it is also argued sometimes that the magnitude of flood events has changed over time. In such a case, non-stationarity of the data would be considered by time-varying distribution parameters (Yan et al., 2017). Of course, such assumptions add uncertainty to the approach and should be only made based on solid studies. Future studies should therefore focus on analysing the changes that affected flood (types) in the past years and the future and consider time-varying parameters and frequencies that could be supported by climate scenarios.

One limitation that only applies to the PPWM approach is that flood types R3, S1 and S2 are not considered in the historical period, as no historical floods of these types have been observed. However, the absence of historical floods also contains information, namely that no extreme floods of these types have occurred in the past. Such information is not used in the PPWM approach. However, in the systematic records no extreme floods of these event types have been observed, too, as indicated by the light-tailed distributions. Therefore, there would not have changed much in the distributions if the information on no extreme events in the past would have been used. This has also been demonstrated by the MCMC approach, where indeed the distributions of R3, S1 and S2 floods did not change much when the historical period is considered.

In this study, five flood types have been considered. This decision is based on previous studies for Europe (Brunner & Fischer, 2022; Fischer & Schuman, 2023), where this number of flood types proved to result in clearly distinguishable flood types, each with different characteristics of rainfall intensity, rainfall duration, snowmelt as well as flood volume and duration. Therefore, this number of flood types was adapted here, too. However, as shown in Fig. 7, a larger number of flood types increases the number of parameters of the TMPS model and thus the related uncertainty. One could therefore think of reducing the number of flood types and apply an objective criterion such as the BIC (Spiegelhalter et al., 2002) to determine the optimal number of flood types. However, this might limit the deterministic interpretability of the flood types, which was one major focus point in this study. Therefore, it was decided to keep the five flood types here. Additionally, one could also think about a different flood classification, e.g. a hydroclimatic one or a more process-based classification (Tarasova et al., 2019).

Finally, one of the major limitations of both approaches, PPWM and MCMC, in the context of type-based flood statistics lies in the required information for the flood classification. Compared to traditional approaches based on annual maximum series, type-based statistics require knowledge on the genesis of each flood to allow for a flood classification. Such information might be hard to obtain, especially when chronicles or similar sources are not available. Here, we demonstrated that sufficient information can be obtained when the season of occurrence of the event is known (which is usually the case as at least the month of occurrence is recorded) and the extension and relative magnitude of the corresponding rainfall event is described. However, in other parts of the world such information might be harder to obtain, limiting the application of the proposed approach. As for all historical flood records, there is always uncertainty related to the flood peak. Historical records are mostly based on water stages and the discharge is thus obtained by stage-discharge curves. These curves are usually not derived for such conditions and extreme flood events as considered in the historical period. Thus, these curves are extrapolated and their validity for extreme events remains unclear.

5. Conclusions

In this work we have demonstrated how historical flood events can be incorporated into type-based flood statistics. For this purpose, the partial probability weighted moments were applied, and the historical flood events were classified according to their flood genesis for a catchment in Germany. The results were compared to an MCMC approach. Both approaches resulted in smaller flood quantiles and less heavy-tailed up to light-tailed distributions when the historical flood events were considered. The flood quantiles of the TMPS model decreased similarly to those of the annual maximum series. With the type-based flood statistics, it was possible to attribute this decrease. It was shown that the distribution of heavy-rainfall floods changes only slightly, as extreme floods similar to those in the last century have occurred for 600 years. This is in particular interesting for the 2002 extreme flood in this catchment, since a similar flood event occurred almost 600 years before. The 2002 flood event is therefore neither unexpected nor exceptional in its genesis. Moreover, the variability of this flood type is constantly high over the whole period, as small flood events and large flood events occur alike. In contrast, the distribution of the long-duration-rainfall floods decreased significantly and especially the right tail changed from a heavy to a light tail. From the historical events it can be seen that indeed only in the last century the most extreme events of this type occurred with the most exceptional event occurring in 2013. Previous flood events have been smaller and extreme floods of this type only occurred since the late 18th century. The incorporation of flood genesis therefore did not only help to improve the estimation of flood quantiles for large return periods by extending the available information, it also allows to evaluate the impact of certain rainfall patterns over time. In this study, the PPWM approach proved to better capture the variability of the flood types compared to the MCMC approach. The MCMC tends to result in small flood quantiles for large return periods in case of type-based statistics which should be kept in mind for future studies. We have also discussed the limitations of both approaches, which mainly lie in the assumptions of thresholds and stationarity. Future studies should address the assumption of stationarity, which might limit the applicability of these approaches for long historic periods.

CRediT authorship contribution statement

Svenja Fischer: Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The link to access the data is given in the acknowledgement section

Acknowledgements

The financial support of the Deutsche Forschungsgemeinschaft (DFG) in terms of the research unit SPATE („Space-Time Dynamics of Extreme Floods", FOR 2416) is gratefully acknowledged. The author would like to thank Uwe Büttner from the Saxony State Office for Environment, Agriculture and Geology for the support on historical floods. The considered discharge data of the Wechselburg gauge are available at https://www.umwelt.sachsen.de/umwelt/infosysteme/ida/index.xhtml or in the related report at https://publikationen.sachsen.de/bdb/artikel/13700/documents/41602 (pages 104-105)

The author would like to thank Lei Yan and one anonymous referee for their insightful and valuable comments that helped to improve the manuscript.

Appendix

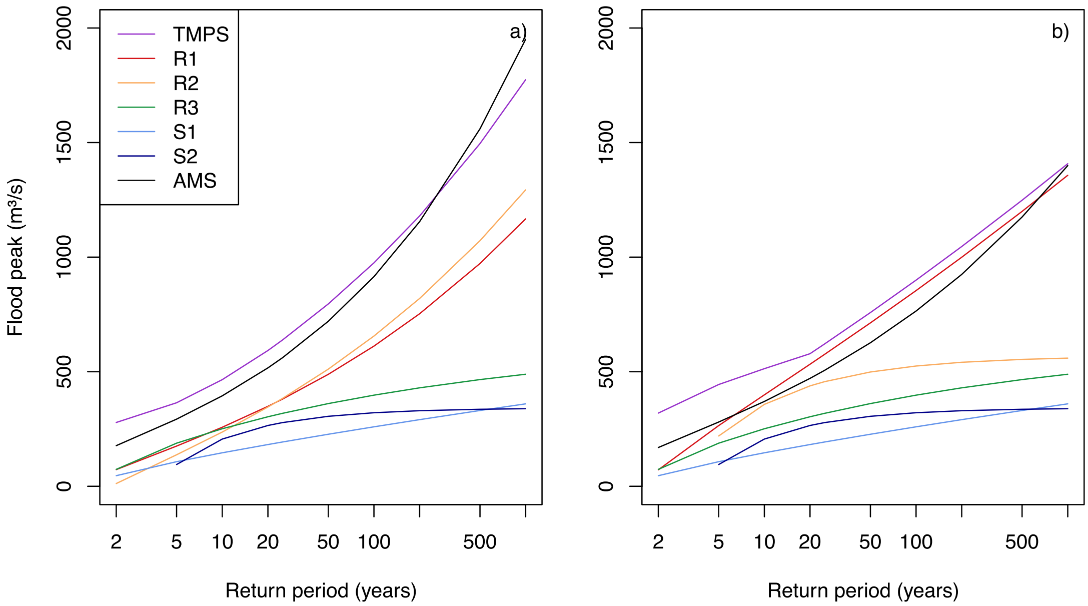

Fig. A1. Distribution functions for given return periods based on a) the systematic records only and b) the systematic and historical records derived with PPWM with unique threshold for all flood types.

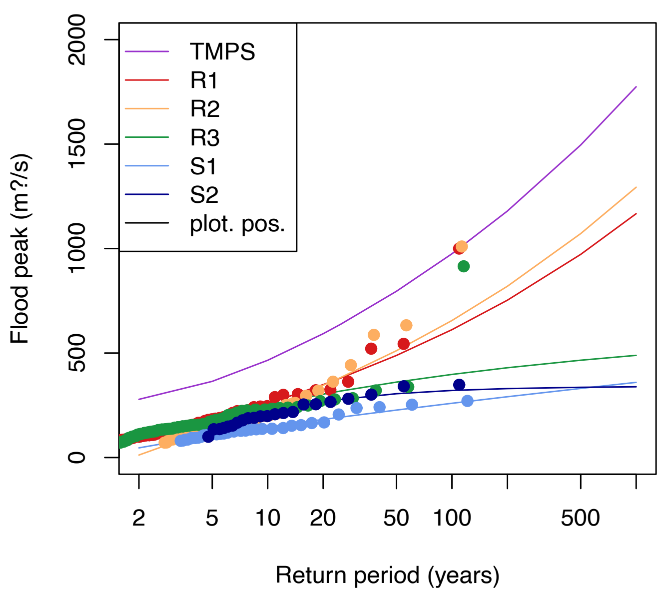

Fig. A2. Distribution functions for given return periods based on the systematic records for the different flood types and the TMPS model. Dots indicate the plotting positions.

References

Arnaud, P., Cantet, P., Odry, J., 2017. Uncertainties of flood frequency estimation approaches based on continuous simulation using data resampling. J. Hydrol. 554, 360–369.

Balkeema, A.A., de Haan, L., 1974. Residual life time at great age. Ann Prob 2, 792–804.

Bartens, A., Haberlandt, U., 2023. Flood frequency analysis using mean daily flows vs. instantaneous peak flows. Hydrol. Earth Syst. Sci. Discuss. 2023, 1–28.

Brunner, M.I., Fischer, S., 2022. Snow-influenced floods are more strongly connected in space than purely rainfall-driven floods. Environ. Res. Lett. 17 (10), 104038.

Burn, D.H., 1997. Catchment similarity for regional flood frequency analysis using seasonality measures. J. Hydrol. 202 (1–4), 212–230.

Cohn, T.A., Lane, W.L., Baier, W.G., 1997. An algorithm for computing moments-based flood quantile estimates when historical flood information is available. Water Resour. Res. 33 (9), 2089–2096.

Cohn, T.A., Stedinger, J.R., 1987. Use of historical information in a maximum-likelihood framework. J. Hydrol. 96 (1–4), 215–223.

Cornes, R., van der Schrier, G., van den Besselaar, E.J.M., Jones, P.D., 2018. An Ensemble Version of the E-OBS Temperature and Precipitation Datasets. J. Geophys. Res. Atmos. 123 https://doi.org/10.1029/2017JD028200.

Dyck, S., Grünewald, U., Kluge, C., Lauterbach, D., Schramm, M. (1980). Angewandte Hydrologie: Teil 1: Berechnung und Regelung des Durchflusses der Flüsse (Applied Hydrology . Wilhelm Ernst & Sohn. ISBN 10: 3433008825.

Fickert, R., 1934. Die größten Sommerhochwasser des sächsischen Muldegebiets in den letzten Jahrzehnten (The largest summer floods of the Saxon Mulde basin in the last decades). Beilage Zum Jahrbuch Des Sächsischen Landesamtes Für Gewässerkunde. Available Online.

Fischer, S., 2018. A seasonal mixed-POT model to estimate high flood quantiles from different event types and seasons. J. Appl. Stat. https://doi.org/10.1080/02664763.2018.1441385.

Fischer, S., Schumann, A., Bühler, P., 2019. Timescale-based flood typing to estimate temporal changes in flood frequencies. Hydrol. Sci. J. https://doi.org/10.1080/02626667.2019.1679376.

Fischer, S., Schumann, A., Bühler, P., 2021. A statistics-based automated flood event separation. Journal of Hydrology X 10, 100070.

Fischer, S., Schumann, A.H., 2022. Handling the stochastic uncertainty of flood statistics in regionalization approaches. Hydrol. Sci. J. 67 (10), 1449–1465.

Fischer, S., Schumann, A.H., 2023. Type-Based Flood Statistics – An Interlink Between Stochastic and Deterministic Flood Hydrology. Springer. https://doi.org/10.1007/978-3-031-32711-7.

Gaál, L., Szolgay, J., Kohnová, S., Hlavcová, K., Viglione, A., 2010. Inclusion of historical information in flood frequency analysis using a Bayesian MCMC technique: a case study for the power dam Orlík, Czech Republic. Contrib. Geophys. Geod. 40, 121–147.

Garavaglia, F., Gailhard, J., Paquet, E., Lang, M., Garcon, R., Bernardara, P., 2010. Introducing a rainfall compound distribution model based on weather patterns sub-sampling. Hydrol. Earth Syst. Sci. 14 (951–964), 2010.

Gaume, E., Gaál, L., Viglione, A., Szolgay, J., Kohnová, S., Blöschl, G., 2010. Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites. J. Hydrol. 394 (1–2), 101–117.

Greenwood, J.A., Landwehr, J.M., Matalas, N.C., Wallis, J.R., 1979. Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form. Water Resour. Res. 15 (5), 1049–1054.

Hirsch, R.M., 1987. Probability plotting position formulas for flood records with historical information. J. Hydrol. 96 (1), 185–199. https://doi.org/10.1016/0022-1694(87)90152-1.

Hosking, J.R., 1990. L-moments: analysis and estimation of distributions using linear combinations of order statistics. J. R. Stat. Soc. Ser. B Stat Methodol. 52 (1), 105–124.

Isikwue, M.O., Onoja, S.B., Naakaa, D.S., 2015. Classical and Bayesian Markov Chain Monte Carlo (MCMC) Modeling of Extreme Rainfall (1979–2014) in Makurdi. Nigeria, International Journal of Water Resources and Environmental Engineering 7, 123–131. https://doi.org/10.5897/IJWREE2015.0588.

Kahle, M., Kempf, M., Martin, B., Glaser, R., 2022. Classifying the 2021 'Ahrtal' flood event using hermeneutic interpretation, natural language processing, and instrumental data analyses. Environmental Research Communications 4 (5), 051002.

Kampf, S.K., Lefsky, M.A., 2016. Transition of dominant peak flow source from snowmelt to rainfall along the Colorado Front Range: Historical patterns, trends, and lessons from the 2013 Colorado Front Range floods. Water Resour. Res. 52 (1), 407–422.

Kjeldsen, T.R., Macdonald, N., Lang, M., Mediero, L., Albuquerque, T., Bogdanowicz, E., Wilson, D., 2014. Documentary evidence of past floods in Europe and their utility in flood frequency estimation. J. Hydrol. 517, 963–973.

Lang, M., Ouarda, T.B., Bobée, B., 1999. Towards operational guidelines for over-threshold modeling. J. Hydrol. 225 (3–4), 103–117.

Lehmkuhl, F., Schüttrumpf, H., Schwarzbauer, J., Brüll, C., Dietze, M., Letmathe, P., Hollert, H., 2022. Assessment of the 2021 summer flood in Central Europe. Environ. Sci. Eur. 34 (1), 107.

MacPherson, L.R., Arns, A., Fischer, S., Méndez, F.J., Jensen, J., 2023. Bayesian extreme value analysis of extreme sea levels along the German Baltic coast using historical information. Nat. Hazards Earth Syst. Sci. 23, 3685–3701. https://doi.org/10.5194/nhess-23-3685-2023.

Merz, B., Elmer, F., Kunz, M., Mühr, B., Schröter, K., Uhlemann-Elmer, S., 2014. The extreme flood in June 2013 in Germany. La Houille Blanche 100 (1), 5–10. https://doi.org/10.1051/lhb/2014001.

Okabe, A., Boots, B., Sugihara, K., Chiu, S.N., 2000. Concepts and applications of Voronoi diagrams. John Wiley, Chichester.

Payrastre, O., Gaume, E., Andrieu, H., 2011. Usefulness of historical information for flood frequency analyses: Developments based on a case study. Water Resour. Res. 47 (8).

Rango, A., Martinec, J., 1995. Revisiting the degree-day method for snowmelt computations 1. JAWRA Journal of the American Water Resources Association 31 (4), 657–669.

Reis Jr, D.S., Stedinger, J.R., 2005. Bayesian MCMC flood frequency analysis with historical information. J. Hydrol. 313 (1–2), 97–116.

Spiegelhalter, D.J., Best, N., Carlin, B.P., Linde, A., 2002. Bayesian measures of model complexity and fit (with discussion). J. R. Stat. Soc. Ser. B 64 (1), 34. https://doi.org/10.1111/1467-9868.00353.

Stedinger, J.R., Cohn, T.A., 1986. Flood frequency analysis with historical and paleoflood information. Water Resour. Res. 22 (5), 785–793.

Tarasova, L., Merz, R., Kiss, A., et al., 2019. Causative classification of river flood events. WIREs Water 6, e1353.

USWRC - Water Resources Council (US). Hydrology Committee. (1975). Guidelines for determining flood flow frequency (No. 17). US Water Resources Council, Hydrology Committee.

Wang, Q.J., 1990. Unbiased estimation of probability weighted moments and partial probability weighted moments from systematic and historical flood information and their application to estimating the GEV distribution. J. Hydrol. 120 (1–4), 115–124.

Yan, L., Xiong, L., Liu, D., Hu, T., Xu, C.-Y., 2017. Frequency analysis of nonstationary annual maximum flood series using the time-varying two-component mixture distributions. Hydrol. Process. 31, 69–89. https://doi.org/10.1002/hyp.10965.

Yan, L., Xiong, L., Ruan, G., Xu, C.Y., Yan, P., Liu, P., 2019. Reducing uncertainty of design floods of two-component mixture distributions by utilizing flood timescale to classify flood types in seasonally snow covered region. J. Hydrol. 574, 588–608.