A comprehensive uncertainty framework for historical flood frequency analysis: a 500-year-long case study

Mathieu Lucas, Michel Lang, Benjamin Renard, Jérôme Le Coz

Hydrol. Earth Syst. Sci., 28, 5031–5047, 2024
https://doi.org/10.5194/hess-28-5031-2024

Hydrology and Earth System Sciences
Open Access

A comprehensive uncertainty framework for historical flood frequency analysis: a 500-year-long case study

Mathieu Lucas[1], **Michel Lang**[1], **Benjamin Renard**[2], and **Jérôme Le Coz**[1]

[1]UR RiverLy, INRAE, Villeurbanne, France
[2]UR RECOVER, Aix Marseille Univ., INRAE, Aix-en-Provence, France

Correspondence: Michel Lang (michel.lang@inrae.fr)

Received: 20 February 2024 – Discussion started: 4 March 2024
Revised: 11 June 2024 – Accepted: 27 September 2024 – Published: 26 November 2024

Abstract. The value of historical data for flood frequency analysis has been acknowledged and studied for a long time. A specific statistical framework must be used to comply with the censored nature of historical data, for which only floods large enough to induce written records or to trigger flood marks are usually recorded. It is assumed that all floods which exceeded a given perception threshold were recorded as written testimonies or flood marks. Conversely, all years without a flood record in the historical period are assumed to have had a maximum discharge below the perception threshold. This paper proposes a binomial model that explicitly recognizes the uncertain nature of both the perception threshold and the starting date of the historical period. This model is applied to a case study for the Rhône River at Beaucaire, France, where a long (1816–2020) systematic series of annual maximum discharges is available along with a collection of 13 historical floods from documentary evidence over 3 centuries (1500–1815). Results indicate that the inclusion of historical floods reduces the uncertainty of 100- or 1000-year flood quantiles, even when only the number of perception threshold exceedances is known. However, ignoring the uncertainty around the perception threshold leads to a noticeable underestimation of flood quantile uncertainty. A qualitatively similar conclusion is found when ignoring the uncertainty around the historical period length. However, its impact on flood quantile uncertainty appears to be much smaller than that of the perception threshold.

1 Introduction

Flood frequency analysis (FFA) provides information on the magnitude and frequency of flood discharges. It is used to estimate the probability of flooding and manage the risk posed by floods to human health, the environment, the economy and cultural heritage (European Union, 2007). One of the main concerns in FFA is the difficulty of precisely estimating the parameters of the chosen distribution with discharge series of limited length (generally a few decades). This is particularly problematic when low-probability (e.g. an annual exceedance probability of 10^{-3} or 10^{-4}) design floods are required to ensure the safety of people and hydraulic structures (Apel et al., 2004; Kjeldsen et al., 2014). Fortunately, sampling uncertainty can be reduced by providing additional information beyond the flood sample obtained from discharge monitoring stations during a systematic period. Such information can be temporal (e.g. historical data on ancient floods), regional (e.g. discharge data from similar catchments), causal (e.g. rainfall data) (Merz and Blöschl, 2008) or a combination of these (Macdonald and Sangster, 2017). This paper focuses on the treatment of historical data in FFA.

Historical data can take a variety of forms: they may be issued from testimonials (Pichard, 1995; Kjeldsen et al., 2014); flood marks (Parkes and Demeritt, 2016; Piotte et al., 2016; Engeland et al., 2020; METS, 2023; Renard, 2023); or palaeoflood reconstructions derived from various proxies, such as sedimentary deposits or riparian tree rings (Stedinger and Cohn, 1986; Benito et al., 2004; Dezileau et al., 2014; St. George et al., 2020; Engeland et al., 2020). Using historical data in FFA has a long history and is now a well-

Published by Copernicus Publications on behalf of the European Geosciences Union.

established practice. Benson (1950) and Hirsch (1987) first focused on plotting position formulas for historical floods. Various types of historical data can be incorporated in FFA by selecting an adequate likelihood function (Stedinger and Cohn, 1986; Kuczera, 1999). Parameter estimation is often performed in a Bayesian way using Markov chain Monte Carlo (MCMC) algorithms (Reis and Stedinger, 2005). Most recent studies emphasize the need to take full account of uncertainties (Neppel et al., 2010; Kjeldsen et al., 2014; Parkes and Demeritt, 2016; Shang et al., 2021; Sharma et al., 2022). Although discharges from the systematic period are generally better known than those from the historical period, they are still affected by uncertainties. However, those uncertainties are often neglected when using historical data. Only a few works propose taking them into account: Reis and Stedinger (2005) or Parkes and Demeritt (2016) consider discharge uncertainty during the systematic period via the use of a fixed percentage error, while Neppel et al. (2010) use unknown multiplicative errors.

Historical flood data are not systematic: only floods large enough to induce written records or to trigger flood marks are recorded. Such censored data can be analysed statistically thanks to the perception threshold concept (Gerard and Karpuk, 1979; Stedinger and Cohn, 1986). The assumption is that all floods exceeding this perception threshold were recorded, thus ensuring the completeness of the historical flood record above the threshold. As a corollary, the annual maximum flood discharge can be assumed to be smaller than the perception threshold for all years in the historical period with no recorded flood. It is possible, although not mandatory, to reconstruct the discharge of historical floods above the perception threshold via the use of hydraulic models (Lang et al., 2004; Neppel et al., 2010; Machado et al., 2015). In cases where such reconstruction is too complex, the sole knowledge of the number of floods that exceeded the perception threshold during the historical period can be exploited by means of a binomial distribution, as described by Stedinger and Cohn (1986) or Payrastre et al. (2011). This method highlights two key quantities in historical FFA that constitute the main focus of this paper: the perception threshold and the length of the historical period.

The perception threshold is an empirical concept that only takes a physical meaning in specific cases. The ideal situation corresponds to the availability of a cross-section that has not changed over time, with overflows always occurring above the same discharge and systematically leaving a trace in written records or on infrastructure (flood marks) as a result of the damage caused. In such a situation, expressing the perception threshold as a precisely estimated discharge value is feasible. However, its estimation can be undermined by several factors, such as a complex river geometry or a temporally varying perception of flood damage among the populations living adjacent to the river. In spite of such uncertainties, the perception threshold is assumed to be perfectly known in the vast majority of studies, although the sensitivity of results to the perception threshold is often explored (Stedinger and Cohn, 1986; Viglione et al., 2013; Macdonald et al., 2014; Payrastre et al., 2011; Parkes and Demeritt, 2016).

In the same way, the length of the historical period is generally considered to be perfectly known in the historical FFA methods in the literature. In principle, the historical period (or surveying period) should start on the date when the source of the historical information came into existence. However, it is generally assumed to start with the first known flood and to finish with the starting date of the systematic period. Prosdocimi (2018) showed that this leads to a systematic underestimation of the length of the historical period and, thus, proposed an unbiased estimator of the starting date of the historical period. However, this unbiased estimator is still treated as a known value in the subsequent FFA procedure, even though it is affected by considerable uncertainty.

This paper presents an FFA probabilistic model that (1) uses the number of times a perception threshold is exceeded over a historical period and (2) accounts for the uncertainty in discharges during the systematic period. The key originality of this model is to recognize the imperfectly known nature of both the perception threshold and the length of the historical period by making them parameters of the probabilistic model. The aim is to correctly assess the uncertainties in flood quantiles, based on historical information.

This FFA model and several variants are applied to a case study based on the Rhône River at Beaucaire, France, offering a very long systematic record (1816–2020, 205 years) with carefully determined discharge uncertainties. An uncertainty propagation chain developed by Lucas et al. (2023) accounts for errors on stage and gauging measurements as well as rating curve estimation. In a first step, the 205-year systematic record is artificially subsampled in order to mimic a typical mixed dataset containing about 50 years of systematic data and about 150 years of censored historical data. This allows for the testing of the FFA models on a real-world dataset, with the full 205-year systematic dataset providing a precise baseline against which comparisons can be made. The added value of precisely knowing the discharge of historical floods vs. only knowing the number of perception threshold exceedances is also explored. In a second step, the same FFA models are then applied to the 1816–2020 systematic record and a collection of historical floods during the 1500–1815 period (Pichard and Roucaute, 2014). The impact of the various sources of uncertainty in the quantile estimates is discussed.

This remainder of the paper is organized as follows: methods for historical FFA are introduced in Sect. 2; the available data are presented in Sect. 3, and their stationarity is verified; the FFA models are applied and compared using the artificially subsampled record on the 1816–2020 period in Sect. 4 and the entire dataset on the 1500–2020 period in Sect. 5; and, finally, Sect. 6 discusses some key results of this work, while Sect. 7 summarizes the study's main conclusions.

2 Probabilistic models

We first present how censored historical floods can be included into a probabilistic model (Sect. 2.1); following this, we move towards more specific models accounting for uncertainties (Sect. 2.2).

2.1 Standard treatment of censored data

2.1.1 Likelihood function

We assume that the annual maximum (AMAX) discharge Q during systematic and historical periods is an independent and identically distributed (i.i.d.) random variable that follows a generalized extreme value (GEV) distribution, with the respective location, scale and shape parameters $\theta = (\mu, \sigma, \xi)$. When the shape parameter ξ is non-zero, the GEV cumulative distribution function (cdf) and the probability density function (pdf) are

$$F(q;\boldsymbol{\theta}) = \exp\left[-\left(1-\xi\frac{q-\mu}{\sigma}\right)^{1/\xi}\right]$$

and

$$f(q;\boldsymbol{\theta}) = \frac{\partial F(q;\boldsymbol{\theta})}{\partial q} = \frac{1}{\sigma}\left(1-\xi\frac{q-\mu}{\sigma}\right)^{1/\xi-1} F(q;\boldsymbol{\theta}). \quad (1)$$

Under this parameterization, a positive shape parameter ($\xi > 0$) corresponds to an upper-bounded distribution with quantiles lower than those of the corresponding Gumbel distribution. In the opposite case ($\xi < 0$), the distribution is heavy-tailed with above-Gumbel quantiles. The Gumbel case is obtained by continuity when ξ tends to zero.

The sample of AMAX discharges during the systematic period covering j years is noted $\boldsymbol{q} = (q_t)_{t=1,j}$. For the time being, discharges are supposed to be perfectly known and not affected by any uncertainty. The historical sample is made of k events that exceeded the perception threshold S over a period of n years. Therefore, the perception threshold was not exceeded over the remaining $(n-k)$ years. The probability π of exceeding the threshold S can be written as follows:

$$\pi = (1-F(S;\boldsymbol{\theta})) = 1-\exp\left[-\left(1-\xi\frac{S-\mu}{\sigma}\right)^{1/\xi}\right]. \quad (2)$$

It is assumed that k, the number of exceedances of the perception threshold, follows a binomial distribution $\mathcal{B}(n,\pi)$. The likelihood function of a mixed sample of AMAX discharges $\boldsymbol{q}$ during the systematic period spanning j years and the number k of exceedances of the perception threshold S during the historical period spanning n years is as follows:

$$L(\boldsymbol{\theta};\boldsymbol{q},k) = \underbrace{\Pi_{t=1}^{j} f(q_t;\boldsymbol{\theta})}_{(a)} \underbrace{\left[\binom{n}{k}(F(S;\boldsymbol{\theta}))^{n-k}(1-F(S;\boldsymbol{\theta}))^{k}\right]}_{(b)}. \quad (3)$$

Here, term (a) in Eq. (3) represents the likelihood for systematic data and term (b) in Eq. (3) represents the likelihood for historical data. By applying Bayes' formula, the posterior distribution $p(\theta|\boldsymbol{q},k)$ of parameters θ given systematic and historical data is as follows:

$$p(\boldsymbol{\theta}|\boldsymbol{q},k) \propto L(\boldsymbol{\theta};\boldsymbol{q};k)\, p(\boldsymbol{\theta}). \quad (4)$$

The term $p(\theta)$ represents the prior distribution of the parameters and needs to be elicited before inference. The posterior distribution $p(\theta|\boldsymbol{q},k)$ is explored using an MCMC sampling algorithm (Renard et al., 2006), leading to a representation of sampling uncertainty by means of r parameter vectors $\Theta = (\theta_1, \ldots, \theta_r)$. The parameter vector $\hat{\theta}$ that maximizes the posterior distribution is called "maxpost". Thereafter, the prior distribution $p(\theta)$ of the GEV parameters will be as follows: a positive uniform distribution for μ and σ and a Gaussian distribution with a mean of zero and a standard deviation of 0.2 for ξ, as proposed by Martins and Stedinger (2000).

2.1.2 Starting date of the historical period

The starting date t^* of the historical period can be assessed by two methods. A first method proposed by Prosdocimi (2018) is based on the total number of exceedances of the perception threshold.

Let $\mathrm{NE} = \mathrm{NE_H}$ (historical period) $+ \mathrm{NE_C}$ (continuous period) denote the total number of exceedances of the perception threshold S recorded during NY years, with $\mathrm{NY} = \mathrm{NY_H}$ (historical period) $+ \mathrm{NY_C}$ (continuous period). Considering the date t_1 of the first known flood, which occurred $(\mathrm{NY}-1)$ years before the end of the systematic period, Prosdocimi (2018) proposed choosing the starting date as follows:

$$t^*_{(\mathrm{Prosdocimi})} = t_1 - (\mathrm{NY}-1)/\mathrm{NE}. \quad (5)$$

The idea behind this estimate is to start the historical period T_S years before the first known flood, where T_S is the return period of the perception threshold, estimated here as $(\mathrm{NY}-1)/\mathrm{NE}$.

The second method is based on the Poisson process paradox (Feller 1971). It takes advantage of the fact that the expected duration between the last T-year event and current time is equal to the expected duration between current time and the next T-year event. In some cases, the historical period (including flood and no-flood information) starts before the date t_1 of the first known flood (e.g. at the creation of the service in charge of surveying floods or at the date of bridge construction where historical data are available). We denote this date using t_start and consider the difference $(t_1 - t_\mathrm{start})$ between these two dates. Without any knowledge of the return period T_S of the threshold but using the difference $(t_1 - t_\mathrm{start})$, we obtain the following:

$$t^*_{(\mathrm{Poisson})} = t_\mathrm{start} - (t_1 - t_\mathrm{start}) = 2t_\mathrm{start} - t_1. \quad (6)$$

2.2 Models accounting for uncertainties

We first present a binomial model for the treatment of historical floods (a collection of historical floods larger than a

Table 1. Characteristics of the four binomial models.

Binomial model	Perception threshold S	Historical period length n
Model A	Fixed	Fixed
Model B	Uncertain	Fixed
Model C	Fixed	Uncertain
Model D	Uncertain	Uncertain

perception threshold S) and systematic floods (AMAX discharges from a gauging station). This first model includes a propagation procedure of hydrometric uncertainties from the systematic period (Model A). We then extend this model to take into account the perception threshold uncertainty (Model B), the historical period length uncertainty (Model C), or both (Model D) (Table 1). A fifth model (Model E) considers the case in which historical flood discharges are known within an interval.

2.2.1 Model A: binomial model for historical floods and propagation of systematic discharges uncertainties

In Eq. (3), the uncertainty of AMAX discharges for the systematic period is assumed to be negligible. As this uncertainty can reach 30 % at Beaucaire during the 19th century (see Sect. 3.1), it seems necessary to consider it. We use the propagation procedure accounting for both stage and rating curve uncertainties described by Lucas et al. (2023); this leads to $s = 500$ realizations of AMAX discharges: $\left(q_t^{(i)}\right)_{t=1,j;i=1,s}$. Each realization can be used to compute a posterior distribution with Eq. (4), and each posterior distribution can be explored with the MCMC sampler, with $r = 500$ MCMC simulations. This leads to a total of $r \times s$ parameter vectors $\left(\boldsymbol{\theta}_p^{(i)}\right)_{p=1,r;i=1,s}$, representing the combined effect of sampling uncertainty and hydrometric uncertainty from systematic data. The maxpost parameter vector is calculated using the maxpost sample of AMAX discharges. The model described above will be referred to as Model A. The propagation of hydrometric uncertainties from the systematic period described here will be carried out identically for all models defined in the following sections.

2.2.2 Model B: binomial model for historical floods, accounting for perception threshold uncertainty

A single perception threshold S for the entire sample is considered here. In order to take the imperfect knowledge of S into account, it is possible to consider it as an unknown parameter of the model and to represent this imperfect knowledge through a prior distribution. In the previous section, the perception threshold was already part of the model, but its value was assumed to be known, which is no longer the case here. Therefore, the left-hand side of Eq. (3) becomes $L(\boldsymbol{\theta}, S; \boldsymbol{q}, k)$ instead of $L(\boldsymbol{\theta}; \boldsymbol{q}, k)$, while the right-hand side remains unchanged. The posterior distribution of the parameters θ and S given the data is as follows:

$$p(\boldsymbol{\theta}, S|\boldsymbol{q}, k) \propto L(\boldsymbol{\theta}, S; \boldsymbol{q}, k)\, p(\boldsymbol{\theta}, S). \tag{7}$$

This posterior distribution takes the hydrometric uncertainty from the systematic period, the sampling uncertainty and the perception threshold's uncertainty into account. This model will be referred to as Model B in the following sections. Note that it is necessary to specify a prior distribution for the perception threshold S that reflects the knowledge on this parameter, which is highly case-specific and can range from very imprecise to nearly known.

2.2.3 Model C: binomial model for historical floods, accounting for the uncertainty in the historical period length

The uncertainty in the number n of years constituting the historical period can be treated in the same way as described in the previous section for the threshold S. Generally, the ending date of the historical period is perfectly known, as it also corresponds to the start of the systematic recordings. However, the starting date t^* of the historical sample, from which all floods above the perception threshold are supposed to be recorded, is generally poorly known. The number n of years constituting the historical period can hence be treated as an unknown parameter of the probabilistic model. The perception threshold S is assumed to be perfectly known in this case. Therefore, the left-hand side of Eq. (3) becomes $L(\boldsymbol{\theta}, n; \boldsymbol{q}, k)$. The posterior distribution of the parameters θ and n given the data is as follows:

$$p(\boldsymbol{\theta}, n|\boldsymbol{q}, k) \propto L(\boldsymbol{\theta}, n; \boldsymbol{q}, k)\, p(\boldsymbol{\theta}, n). \tag{8}$$

As previously stated, a prior distribution reflecting the partial knowledge of the length of the historical period has to be specified. The lack of knowledge on the length of the historical period is therefore taken into account in the model and has an impact on flood quantile uncertainty. This model will be referred to as Model C in the following sections.

2.2.4 Model D: binomial model for historical floods, accounting for both perception threshold and historical period length uncertainties

As the perception threshold S and the number n of years of the historical period are linked by definition (with a perception threshold being valid over a given duration), we finally consider a model that represents the lack of knowledge about both parameters. The left part of Eq. (3) becomes $L(\boldsymbol{\theta}, S; n; \boldsymbol{q}, k)$. The posterior distribution of the parameters θ, S and n given the data is as follows:

$$p(\boldsymbol{\theta}, S, n|\boldsymbol{q}, k) \propto L(\boldsymbol{\theta}, S, n; \boldsymbol{q}, k)\, p(\boldsymbol{\theta}, S, n). \tag{9}$$

This model, for which S and n are uncertain, will be called Model D in the following sections.

2.2.5 Model E: considering historical flood discharges within intervals

In some cases, the discharge of historical floods above the perception threshold can be reconstructed and taken into account in the probabilistic model (e.g. Stedinger and Cohn, 1986). As such reconstructions are typically obtained by means of hydraulic models affected by large uncertainties, it is also useful to consider that the reconstructed discharges are not perfectly known but lie within intervals. Several examples of such models exist in the literature (e.g. Payrastre et al., 2011, or Parkes and Demeritt, 2016). The corresponding likelihood can be written as follows:

$$L(\boldsymbol{\theta};\boldsymbol{q},\boldsymbol{y},k)=\prod\nolimits_{t=1}^{j}f(q_t;\boldsymbol{\theta})\prod\nolimits_{i=1}^{k}[F(y_i^{\text{sup}};\boldsymbol{\theta}) - F(y_i^{\text{inf}};\boldsymbol{\theta})](F(S;\boldsymbol{\theta}))^{n-k}, \tag{10}$$

where q_t corresponds to the j floods of the systematic period and y_i corresponds to the k floods of the historical period whose discharge values lie within the interval $\left[y_i^{\text{inf}}, y_i^{\text{sup}}\right]$. The posterior distribution of the model is as follows:

$$p(\boldsymbol{\theta}\,|\,\boldsymbol{q},\boldsymbol{y},k)\propto L(\boldsymbol{\theta};\boldsymbol{q},\boldsymbol{y},k)\,p(\boldsymbol{\theta}). \tag{11}$$

Here, the perception threshold and the length of the historical period are assumed to be perfectly known. This model will be referred to as Model E in the following sections. The quantiles can be compared with the results of the binomial models, for which only the number k of perception threshold exceedance S is known.

3 Case study: the Rhône River at Beaucaire

3.1 Discharge data over 5 centuries

We first consider the 205-year-long daily discharge series of the Rhône River at Beaucaire, France, from 1816 to 2020 (catchment area of 95 590 km^2). Daily stage measurements started in 1816. The gauging station was used until the construction of the Vallabrègues hydroelectric scheme in 1967, which led to the derivation of a part of the discharge. Consequently, the gauging station was relocated 2 km downstream, in order to measure the total discharge (i.e. downstream from the restitution of the diverted discharge). The Vallabrègues Dam has no impact on the discharge at the station, as it has a very limited storage capacity and is opened during floods to cancel the backwater effect that it creates for low flows. A set of 500 realizations of AMAX floods from 1816 to 2020 is available from Lucas et al. (2023), accounting for several sources of hydrometric uncertainty. The estimated 95 % discharge uncertainty varies from 30 % (19th century) to 5 % (1967–2020).

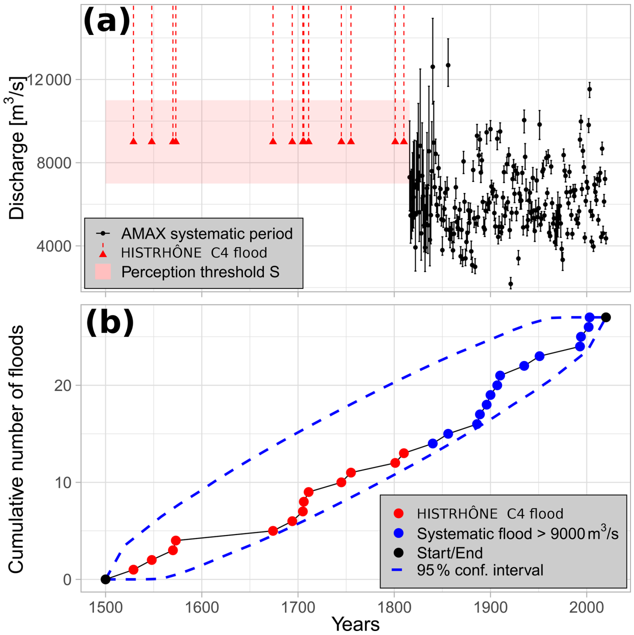

Figure 1. (a) The Rhône River at Beaucaire, AMAX flood discharges with their 95 % uncertainty intervals (1816–2020, systematic period; Lucas et al., 2023) and C4-class floods from 1500 to 1815 (HISTRHÔNE database). **(b)** The cumulative number of C4-class floods and peak-over-threshold floods (systematic period) with the 95 % Poisson process confidence interval.

Secondly, a collection of historical flood testimonies from 1500 to 1815 is available from the HISTRHÔNE database (Pichard and Roucaute, 2014). We focus on the 13 extreme floods (in 1529, 1548, 1570, 1573, 1674, 1694, 1705, 1706, 1711, 1745, 1755, 1801 and 1810), referenced as C4-class events: "extreme flood and inundation". This ensemble is considered to be a comprehensive survey of the most damaging floods of the historical period. The perception threshold S is about 9000 m^3 s^{-1} according to Pichard et al. (2017). Figure 1a shows the available flood discharge sample with the corresponding uncertainties. Note that we used a very uncertain prior for the perception threshold in order to more clearly highlight its impact.

3.2 Stationarity tests

As the probabilistic models described in Sect. 2 assume that AMAX values are independent and identically distributed (i.i.d.), statistical tests should be applied to check the stationarity of both systematic and historical periods.

3.2.1 Systematic data

The Pettitt step-change test (Pettitt, 1979) and Mann–Kendall trend test (Mann, 1945; Kendall, 1948) were applied to the maxpost series of AMAX discharges during the 1816–2010 period. The respective p values of 0.15 and 0.4 indicate no significant change. The segmentation procedure proposed by Darienzo et al. (2021) was also applied, as it allows one to account for the uncertainty around AMAX discharges. This procedure indicates that the optimum number of segments is equal to 1, confirming the absence of significant change.

3.2.2 Historical data

The historical data used here (13 extreme C4-class floods) can be interpreted as peak-over-threshold (POT) values, as they correspond to all floods that exceeded the perception threshold S. Moreover, there is no year with more than one C4-class flood during the 1500–1815 period. AMAX values from the continuous period larger than $9000\,\mathrm{m^3\,s^{-1}}$ can also be viewed as POT values, as the 14 largest values from the daily record (in 1840, 1856, 1886, 1889, 1896, 1900, 1907, 1910, 1935, 1951, 1993, 1994, 2002 and 2003) are from different years. Assuming that the number of occurrences of POT discharges follows a Poisson process, it is possible to compute a confidence interval for the cumulative number N_t of POT values during a period 0–t (Lang et al., 1999) and to verify that the experimental curve is inside the limits of the interval. The Poisson test is applied to the whole period 1500–2020, using POT values from both historical and systematic periods (1500–1815 and 1816–2020). Figure 1b shows that the experimental curve is within the 95 % confidence interval. Hence, the whole sample can be considered to be stationary.

4 Flood frequency analysis on the 1816–2020 period

4.1 Subsampled datasets

In this section, the GEV distribution fitted with AMAX discharges from the 1816–2020 period (including the propagation of hydrometric uncertainty described in Sect. 2.2, Model A) is used as a reference and is referred to as "AMAX long". It can be compared with GEV distributions estimated with subsampled data, where only part of the available information is used:

– The first subsample contains a short sample of AMAX discharges during the 1970–2020 period, referred to as "AMAX short". It corresponds to the typical length of hydrometric series in France, about 50 years of record, leading to a large extrapolation of the estimated distribution towards large quantiles (100-year or 1000-year return period).

– The second subsample is a mixed sample, with AMAX discharges for the 1970–2020 period and a collection of historical values for the 1816–1969 period, referred to as "Mixed A"–"Mixed E", according to the model used. A perception threshold $S = 9000\,\mathrm{m^3\,s^{-1}}$ leads to a collection of 10 historical floods. When using Model B (or Model D), we consider a vague normal prior distribution on the perception threshold: $N\,(9000; 2000)$, with 2000 being the standard deviation. When using Model C (or Model D), we consider a uniform prior distribution on the starting date of the historical period: $U\,[1316; 1816]$, corresponding to a large uncertainty (500 years).

Results with systematic data only (AMAX long and AMAX short) or with a mixed sample (AMAX short + 10 historical floods) are presented for the estimation of Q_{100} and Q_{1000} floods (Fig. 2b), parameters (ξ, S and t^*) (Table 2) and GEV distributions (Fig. 2a). The plotting position formula proposed by Hirsch (1987) for a mixed sample with AMAX values from a continuous period and historical discharges larger than a perception threshold is applied. The Appendix provides a procedure for a case in which historical flood discharges are unknown, using only exceedances of the threshold.

4.2 Value of adding historical information from the 1816–1969 period

Unsurprisingly, when the length of the systematic record (~ 50 years) is too short compared with the target return period (~ 100 or 1000 years), the results are highly uncertain (AMAX short in Fig. 2). A binomial model exploiting historical flood events notably reduces uncertainty when the perception threshold S is known (Mixed A and Mixed C in Fig. 2b), although not achieving the precision obtained with 205 years of systematic records (AMAX long in Fig. 2b). Accounting for the uncertainty in the threshold S (Mixed B and Mixed D in Fig. 2b) increases the uncertainty in flood quantiles and nearly annihilates the interest of historical flood occurrences.

The main part of the uncertainty comes from the estimation of the shape parameter ξ, which governs the behaviour of the tail of the distribution. Note that all of the estimates are close to zero and slightly positive (Table 2), corresponding to an upper-bounded distribution. As might be expected, the estimate of parameter ξ is much more precise with a long series (AMAX long) of 2 centuries than with a short series (AMAX short) of 5 decades. The use of historical data through a binomial model is not very efficient with respect to reducing uncertainty in the shape parameter ξ (Table 2). Overall, the maxpost estimate of the Q_{100} and Q_{1000} quantiles are very close for all models (Fig. 2b). In the next sections, the impact of accounting for the uncertainties in the perception threshold S and the starting date t^* of the historical period is assessed in more detail.

4.3 Impact of considering that the perception threshold is uncertain

The use of Model B reflects a lack of knowledge on the perception threshold, which becomes a parameter of the model. Figure 2b shows that the quantile uncertainty estimated with

Table 2. Maxpost estimation ± the posterior standard deviation (expressed as a percentage) for Q_{100} and Q_{1000} floods and the ξ, S and t^* parameters (1816–2020 period).

Data set		AMAX values		AMAX short + historical data (1816-1969)				
		AMAX long (1816-2020)	AMAX short (1970-2020)	Mixed A	Mixed B	Mixed C	Mixed D	Mixed E
Quantiles (m³/s)	Q_{100}	11451 ± 6%	11076 ± 23%	11132 ± 11%	11302 ± 21%	11517 ± 7%	11147 ±18%	11286 ± 8%
	Q_{1000}	13919 ± 10%	13154 ± 50%	13367 ± 23%	13622 ± 43%	14069 ± 15%	13262 ± 36%	13827 ± 16%
Parameters	ξ	0.058 ± 76%	0.077 ± 132%	0.062 ± 142%	0.058 ± 176%	0.041 ± 202%	0.074 ± 130%	0.035 ± 191%
	S (m³/s)	/	/	/	9163 ± 8%	/	9332 ± 9%	/
	t^*	/	/	/	/	1833 ± 4%	1785 ± 6%	/

Model B is much greater than with Model A and that it is close to the one obtained with systematic data only (AMAX short). Therefore, poor knowledge of the perception threshold has major consequences for the quantile estimates, as it greatly reduces the value of using historical occurrences. The true value of the perception threshold is $S = 9000\,\mathrm{m^3\,s^{-1}}$. The prior and posterior distributions of the threshold S are shown in Fig. 3a. It can be seen that the posterior estimate for Model B ($9163\,\mathrm{m^3\,s^{-1}}$) is close to the true value ($9000\,\mathrm{m^3\,s^{-1}}$) and that the model has effectively improved the knowledge of the threshold compared with the prior distribution $N(9000; 2000)$. The posterior uncertainty in the shape parameter ξ for Model B is greater than that of Model A and, thus, becomes almost identical to that of AMAX short (Table 2). In real-world case studies, specifying a more precise prior should limit this impact and should, hence, be considered to be a priority objective for historical FFA.

4.4 Impact of considering that the historical period length is uncertain

Model C is used to represent the lack of knowledge on the length of the historical period. In Fig. 2b, the maxpost quantile estimates for Model C have slightly higher values than the estimates for Model A. This may be due to the underestimation of the length of the historical period, as can be seen in Fig. 3b. The maxpost date is 1833, whereas the series actually begins in 1816. This 17-year underestimation can be explained by a greater frequency of floods above the threshold S during the systematic period (4 floods during 50 years, i.e one exceedance every 12.5 years) than during the historical period (10 floods during 153 years, i.e one exceedance every 15 years). This imbalance is probably due to sampling variability, as no break or trend was detected by the stationarity tests in Sect. 3.2. The posterior distribution of the starting date t^* for Model C (Fig. 3b) is much more precise than the prior distribution and is strongly asymmetric. The uncertainty around the quantiles estimated by Model C is very similar to that estimated by Model A (Fig. 2b), as is the distribution of the shape parameter (Table 2). Overall, these results indicate that a poor knowledge of the length of the historical period has less impact on the precision of quantile estimates than poor knowledge of the perception threshold.

4.5 Impact of considering that both the perception threshold and the historical period length are uncertain

Model D assumes that both S and n are uncertain in the probabilistic model. The maxpost quantiles estimated in Fig. 2b are close to the reference values. In contrast, the width of the credibility interval is large and lies between that of Model B and Model C. Although the estimate is more accurate than with a short series (AMAX short), it remains very imprecise for the 1000-year flood. Figure 3 helps to understand the origin of this large uncertainty. The posterior distribution of the perception threshold S, although with a maxpost value ($9332\,\mathrm{m^3\,s^{-1}}$) close to the true value ($9000\,\mathrm{m^3\,s^{-1}}$), is very imprecise with a large standard deviation ($883\,\mathrm{m^3\,s^{-1}}$). The perception threshold S appears to be slightly less precisely estimated than with Model B (Table 2), with a posterior standard deviation of 9 % and 8 %, respectively. The starting date of the historical period is even more difficult to estimate, particularly in comparison with the estimate from Model C. It can be seen that the posterior distribution t^* of Model D is very similar to the prior uniform distribution (Fig. 3b), although it is slightly asymmetrical and shows a maximum not far from the true value (the year 1816). However, the flood discharge quantiles are less uncertain for Model D than for Model B. The precise reasons for this are unclear at this stage but might be due to some correlations between parameters. In particular, the Pearson correlation coefficient ρ is equal to

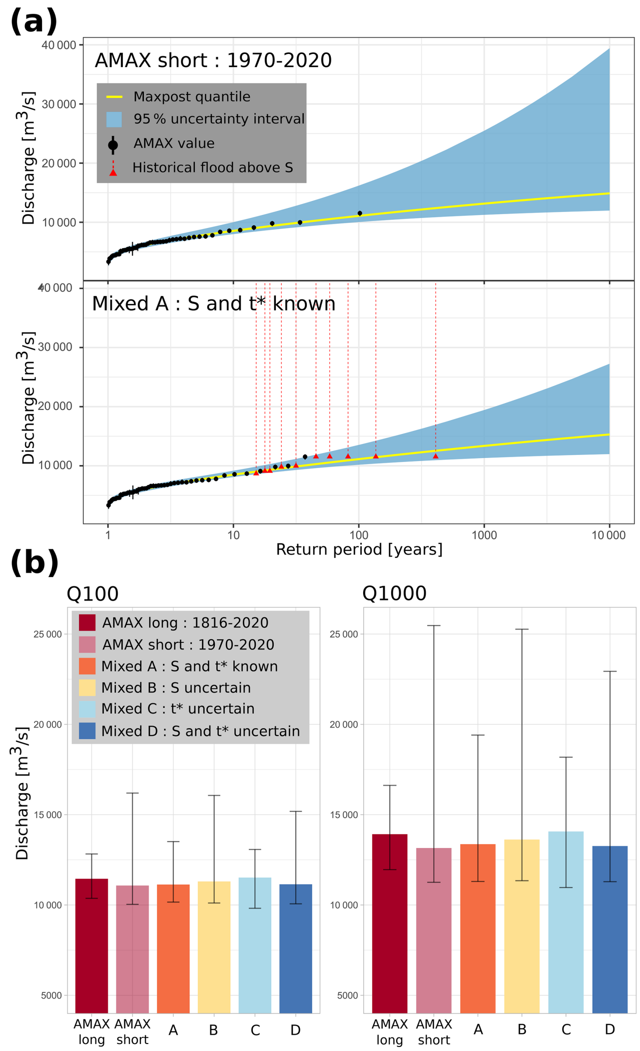

Figure 2. **(a)** GEV quantiles with their 95 % credibility intervals, displaying an example of two different models and datasets: the GEV model on AMAX values (AMAX short 1970–2020) and the binomial Model A on a mixed sample (1816–2020). **(b)** The Q_{100} and Q_{1000} floods with their 95 % credibility intervals displayed as error bars. AMAX long refers to the sample for the 1816–2020 period; AMAX short refers to the sample for the 1970–2020 period; and Mixed A, B, C and D refer to a mixed sample ("historical" floods in the 1816–1969 period and AMAX 1970–2020) for various statistical models.

0.44 and 0.42 between the length of the historical period n and the perception threshold S and between the perception threshold S and the shape parameter ξ, respectively.

4.6 Value of estimating the peak discharge of the historical flood

Binomial models A, B, C and D only use information on the number of times k that a perception threshold S is exceeded over a period of n years. The discharge of historical floods that have exceeded the threshold is therefore ignored. Model E allows peak discharge estimates (with uncertainty) to be taken into account. The results are shown in Fig. 4. There is a reduction in uncertainty of around 25 % for Q_{1000} with Model E compared with binomial Model A (posterior standard deviations of 2255 and 3019 $m^3\,s^{-1}$, respectively). However, the uncertainty in Model E remains around 65 % greater than that of the GEV 1816–2020 model for Q_{1000}. Although it is not a necessary condition for using historical data, knowledge of the discharge of historical floods does reduce the uncertainty around extreme quantiles. However, these results are only valid for the perception threshold S used here, which has a return period of about 15 years (with 14 exceedances during 205 years). Stedinger and Cohn (1986) and Payrastre et al. (2011) showed that the difference in uncertainty between the results of these two types of models tends to decrease as the return period of the perception threshold increases towards 50 years or so, until it becomes negligible above this magnitude. This encourages the use of the number of exceedances of a perception threshold when it is not possible to obtain better information on historical floods.

5 Flood frequency analysis on the 1500–2020 period

In the previous section, we used a synthetic case study from a 205-year systematic record (1816–2020), which gives a baseline to compare the performance of five proposed models (A, B, C, D and E) with known parameters (S and n). The systematic record has been artificially subsampled into a mixed dataset that contains 51 years of systematic data (1970–2020) and 154 years of censored historical data larger than a known perception threshold (1816–1969). In this section, binomial models (A, B, C and D) are applied to a 500-year-long case study, using the 205-year systematic record (1816–2020) and a collection of historical floods from HISTRHÔNE database (1500–1815). This time, S and n are not perfectly known.

5.1 Prior on the perception threshold S and the starting t^* of the historical period

Binomial models A, B, C and D are now applied to a mixed sample over the period from 1500 to 2020, with AMAX values for the systematic period from 1816 to 2020 and occurrences of flood above the perception threshold for the historical 1500–1815 period. The perception threshold and the starting date of the historical period are not known precisely, and a first analysis is carried out with vague priors, with $S \sim N\,(9000; 2000)$ and $t^* \sim U\,[1129; 1529]$. By definition, the historical period begins, at the latest, on the date of the first known historical flood in 1529. The lower limit of the uniform distribution is arbitrarily set 400 years before the date of the first historical flood in order to represent the lack of knowledge on t^*.

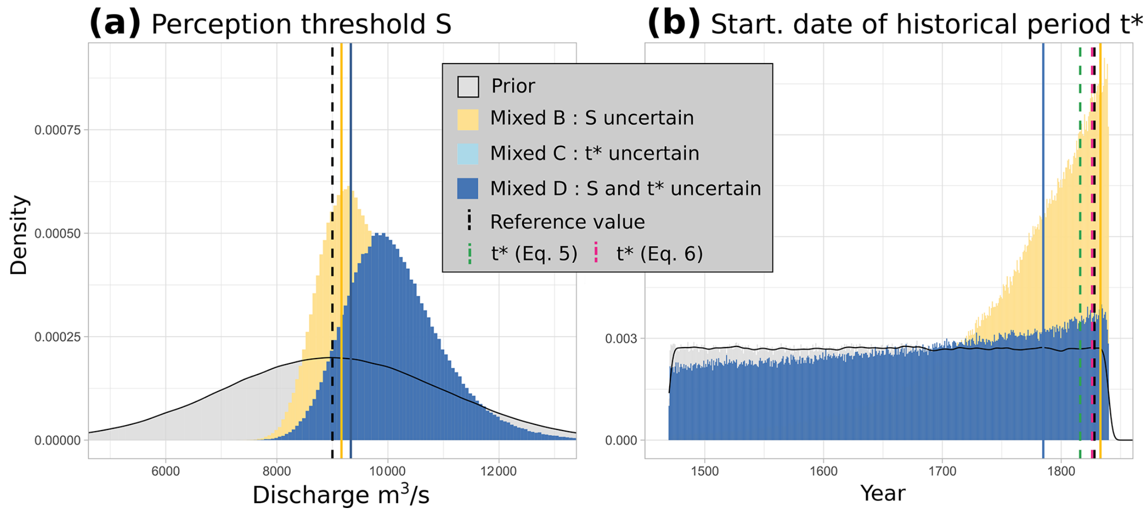

Figure 3. Prior and posterior distributions of **(a)** the perception threshold S and **(b)** the starting date t^* of the historical period (1816–2020 period). The solid vertical lines represent the maxpost estimate of the parameter for each of the models, whereas the dashed black lines represent the reference values ($S = 9000\,\mathrm{m^3\,s^{-1}}$ and $t^* = 1816$). The dashed vertical green and pink lines in panel **(b)** represent the estimates of t^* from Eqs. (5) and (6), respectively.

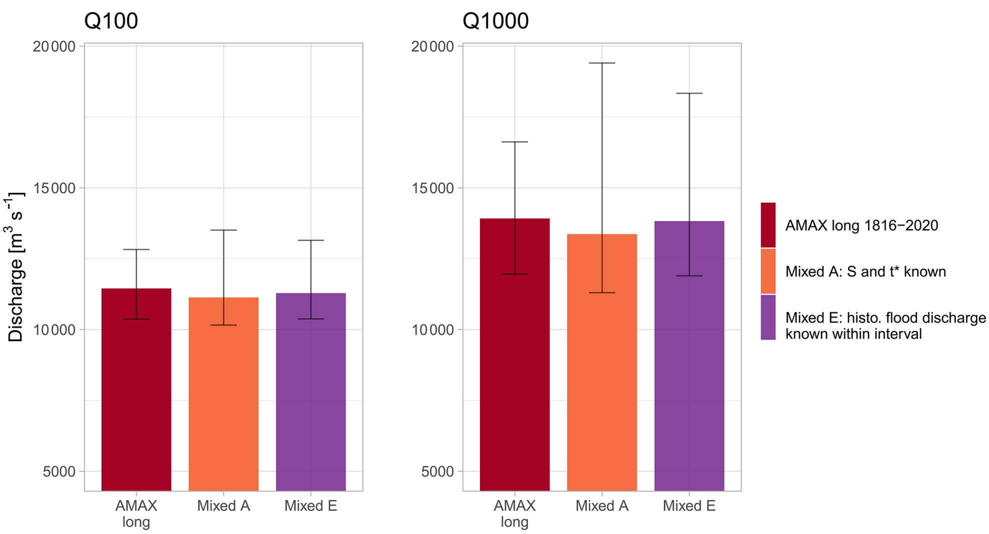

Figure 4. The Q_{100} and Q_{1000} floods with their 95 % credibility intervals displayed as error bars. AMAX long refers to an annual maximum sample for the 1816–2020 period; Mixed refers to a mixed sample ("historical" floods in the 1816–1969 period and AMAX 1970–2020). Model A uses only the number of times that the perception threshold has been exceeded, whereas Model E considers the peak discharge (and its uncertainty) of each historical flood that exceeded threshold S. Perception threshold S and the start date of historical period t^* are considered perfectly known (models A and E).

A second analysis will refine the results of Model D, with more accurate prior estimates of S and t^* used for the historical 1500–1815 period, based on information of the systematic 1816–2020 period. The application of Model D with these more informative priors will be referred to as Model D*. Figure 5 cross-references C4-class (extreme) floods occurring between 1816 and 2000 according to the HISTRHÔNE database (Pichard et al., 2017) and the estimated AMAX discharge values on the same period (Lucas et al., 2023). A total of 5 amongst 14 C4-class floods are below the threshold $S = 9000\,\mathrm{m^3\,s^{-1}}$. Even accounting for discharge uncertainty, three C4-class floods are still fully below the threshold S. As the flood ranking of the HISTRHÔNE database is based on physical impacts (e.g. morphologic river changes and city submersions), it is not possible to have a direct match between C4-class floods and an exact discharge threshold. We refine the prior distribution $N(9000; 500)$, with a standard deviation of $500\,\mathrm{m^3\,s^{-1}}$ (in-

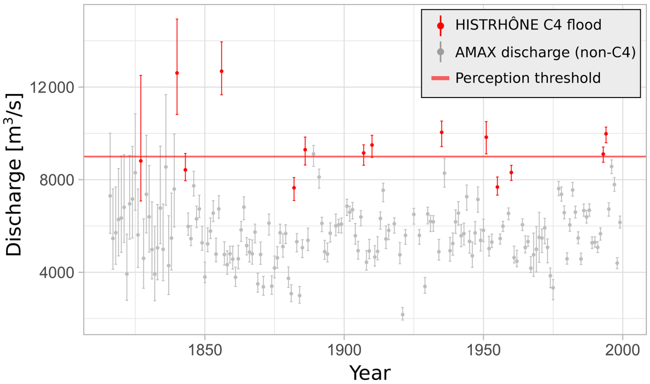

Figure 5. AMAX flood discharges (1816–2000) from Lucas et al. (2023) (in grey) cross-referenced with C4-class floods from the HISTRHÔNE database (in red). The horizontal line corresponds to the estimated perception threshold $S = 9000\,\mathrm{m^3\,s^{-1}}$.

stead of $2000\,\mathrm{m^3\,s^{-1}}$ with Model D). No C4-class flood is fully below the 95 % prior 8000–10 000 interval.

Considering the 13 C4 floods in the HISTRHÔNE database (1500–1815) and the 14 floods higher than a threshold $S = 9000\,\mathrm{m^3\,s^{-1}}$ during the 1816–2020 period, we have two possible estimates of the starting date t^* of the historical period:

- $t^*_{\text{(Prosdocimi)}} = 1511$ (from Eq. 5), with knowledge of the date of the first known flood ($t_1 = 1529$), the total number of threshold exceedances (NE = 13 (C4 floods) + 14 (AMAX > S) = 27) and the total number of years (NY = 2020–1529 = 491 years);
- $t^*_{\text{(Poisson)}} = 1471$ (from Eq. 6), with knowledge of the starting date of the surveying period ($t_{\text{start}} = 1500$).

We refine the prior distribution of t^* as $U\,[1471; 1529]$, with a width of 58 years (instead of 400 years with Model D).

5.2 Results with a vague prior on the perception threshold and the historical period length

Results with systematic data only (AMAX) on the 1816–2020 period or with a mixed sample (AMAX + 13 historical floods) on the 1500–2020 period are presented for the estimation of Q_{100} and Q_{1000} floods (Fig. 6) and parameters (ξ, S and t^*) (Table 3).

The results with a mixed sample for the 1500–2020 period show that the uncertainty in Q_{100} and Q_{1000} floods (Fig. 6a) is lower than with AMAX values for the 1816–2020 period for models assuming a known perception threshold (models A and C). For these two models (Model A and Model C), the maxpost quantiles are also slightly lower (by around 5 %) than with AMAX values for the 1816–2020 period (Fig. 6a). As for the subsamples in Sect. 4, this suggests that poor knowledge of the perception threshold (models B and D) is more detrimental to the precision of estimated quantiles than poor knowledge of the historical period length (models C and D). In particular, these differences can be explained by looking at the posterior distributions of the parameters S and t^* (Fig. 6b).

The posterior standard deviations for the perception threshold (models B and D) are relatively small (around $500\,\mathrm{m^3\,s^{-1}}$ for both models) and the distributions mostly lie above the prior value of $9000\,\mathrm{m^3\,s^{-1}}$ (maxpost values around $9600\,\mathrm{m^3\,s^{-1}}$; Table 3). The starting date t^* of the historical period is more precisely estimated with Model C than with Model D, whose posterior distribution is very close to the prior distribution. For both models, the maxpost estimates of t^* are almost 30 years higher than the assumed value of 1500. In particular, the posterior distribution for Model C shows a maximum for the year 1529, which corresponds to the date of the first flood in the sample.

This trend towards a higher threshold and a shorter historical period could be a symptom of the non-exhaustiveness of the extreme floods (C4 category) in the HISTRHÔNE database, despite the fact that the stationarity hypothesis of the Poisson test over the 1500–2020 period was not rejected (Fig. 1b). Once again, we can compare the rate of occurrence of floods above the threshold $S = 9000\,\mathrm{m^3\,s^{-1}}$ for each of the two samples. For the historical sample, 13 floods were observed over 316 years, i.e. one exceedance every 24 years. For the systematic sample, there were 14 floods over a period of 205 years, i.e. one exceedance every 15 years. This larger frequency of S exceedances of the systematic period, whether due to sampling variability, climatic variability or the non-exhaustiveness of the historical data, leads to the estimation of a higher perception threshold and/or a shorter historical period length.

5.3 Refining prior distributions of the perception threshold and the historical period length

The previous analysis is refined using narrower prior distributions of the perception threshold S and the starting date t^* of the historical period. A comparison of the binomial models D and D* and the AMAX GEV 1816–2020 model is presented in Fig. 7a. It can be seen that the uncertainty in the quantiles is about 15 % lower than the reference for Q_{100} and Q_{1000}. Maxpost estimates are also reduced by approximately 3 % for both return periods. Therefore, the use of historical floods appears relevant to reduce the uncertainty in the quantiles, even in the case where S and n are uncertain. It can also be noted that the elicitation of more informative priors (see Falconer et al., 2022, for a methodological review) reduced the standard deviation of the posterior distribution for Q_{1000} by about 25 % (comparison of Model D with vague priors for S and t^* and Model D* with refined priors).

The posterior distributions of S and t^* are shown in Fig. 7b. Once again, the posterior distribution of the perception threshold is shifted towards values higher than the

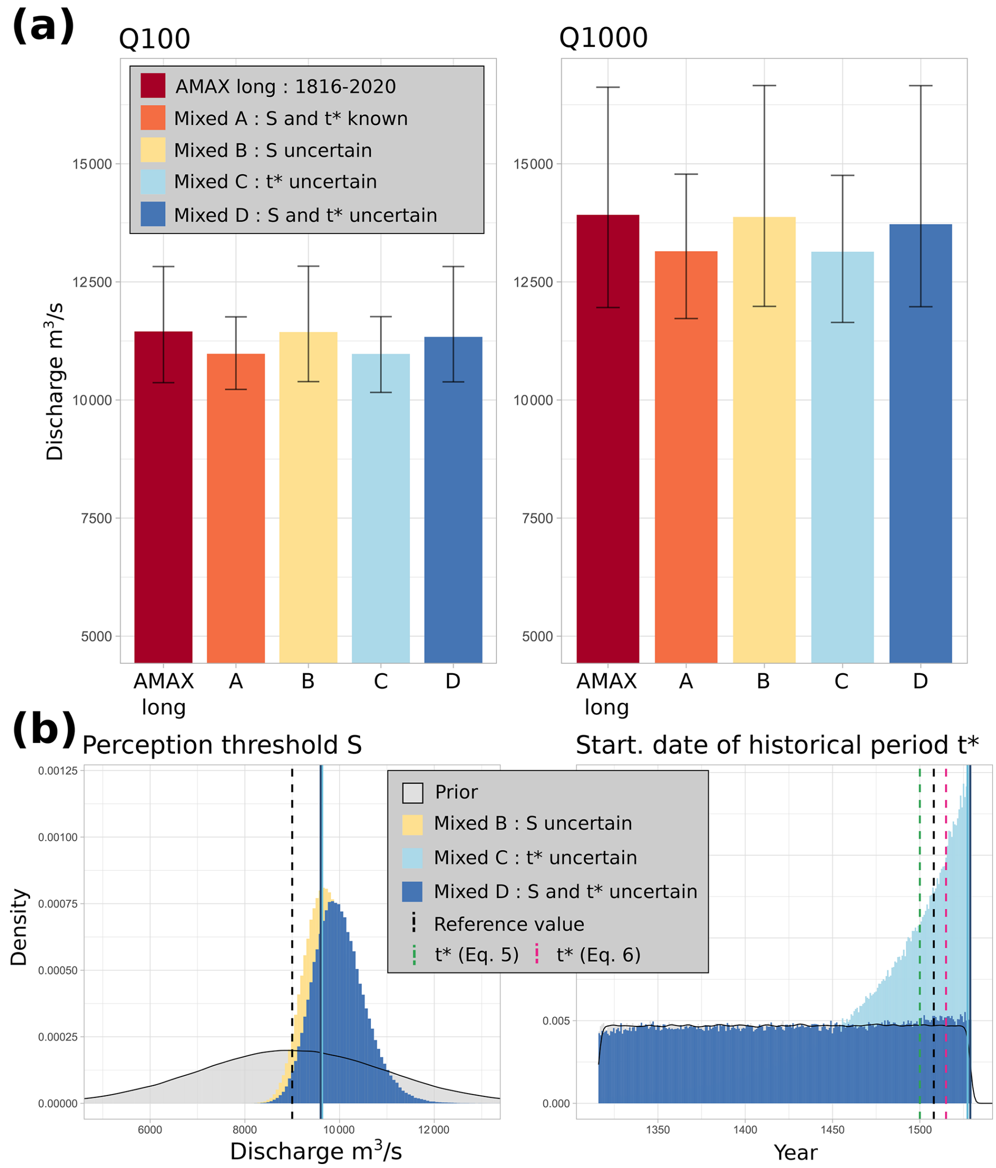

Figure 6. (a) The Q_{100} and Q_{1000} floods with their 95 % credibility intervals displayed as error bars. AMAX long refers to the annual maximum sample for the 1816–2020 period; Mixed A, B, C and D refer to a mixed sample ("historical" floods in the 1500–1815 period and AMAX for 1816–2020) for various statistical models. **(b)** Posterior distribution of (left) the perception threshold S and (right) the starting date t^* of the historical period (1500–2020 period). The solid vertical lines represent the parameter maxpost estimates for each model, whereas the dashed black lines represent the reference values ($S = 9000\,\mathrm{m^3\,s^{-1}}$ and $t^* = 1500$). The dashed vertical green and pink lines (right) represent the estimates of t^* by Eqs. (5) and (6), respectively.

assumed value of $9000\,\mathrm{m^3\,s^{-1}}$, with a maxpost threshold at $9386\,\mathrm{m^3\,s^{-1}}$. The posterior distribution of t^* is again very close to the prior distribution, with a slightly higher density for the years close to the date of the first flood. Here, the maxpost estimate of t^* is 1526, i.e. a historical period length that is 26 years shorter than expected. Therefore, doubt remains as to the completeness of the historical sample or the inter-sample stationarity as described in the previous section.

6 Discussion

6.1 Main findings for the 1816–2020 period

By using the probabilistic models described in Sect. 2 on an artificially degraded sample whose characteristics are well known, it is possible to assess the impact of limited knowledge of the perception threshold S and the length n of the historical period on the estimation of extreme quantiles. The results show that poor knowledge of the perception thresh-

Table 3. Maxpost estimation ± the posterior standard deviation (expressed as a percentage) for Q_{100} and Q_{1000} floods and the ξ, S and t^* parameters (1816–2020 and 1500–2020 periods).

Data set		*AMAX 1816-2020*	*AMAX + historical data (1500-1815)*				
			Mixed A	*Mixed B*	*Mixed C*	*Mixed D*	*Mixed D**
Quantiles (m³/s)	Q_{100}	11451 ± 6%	10977 ± 4%	11438 ± 6%	10975 ± 4%	11336 ± 7%	11118 ± 5%
	Q_{1000}	13919 ± 10%	13149 ± 6%	13875 ± 10%	13139 ± 6%	13721 ± 11%	13421 ± 8%
Parameters	ξ	0.058 ± 76%	0.073 ± 52%	0.060 ± 73%	0.074 ± 51%	0.061 ± 72%	0.063 ± 63%
	S (m³/s)	/	/	9628 ± 5%	/	9613 ± 6%	9386 ± 4%
	t^*	/	/	/	1527 ± 3%	1529 ± 4%	1526 ± 1%

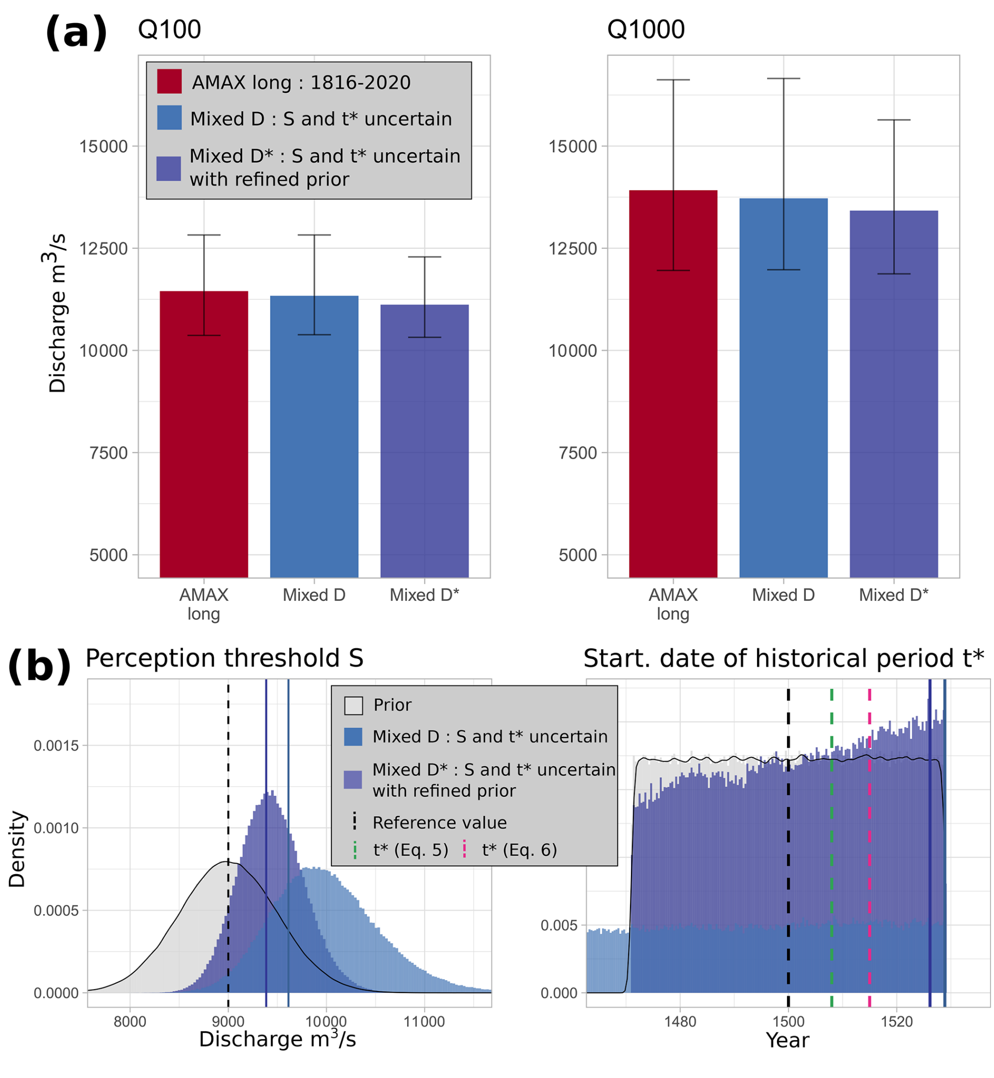

Figure 7. (a) The Q_{100} and Q_{1000} floods with their 95 % credibility intervals displayed as error bars. AMAX long refers to an annual maximum sample for the 1816–2020 period; Mixed D* refers to a mixed sample ("historical" floods in the 1500–1815 period and AMAX for 1816–2020), with refined priors for S and t^*. **(b)** Posterior distributions of (left) the perception threshold S and (right) the starting date t^* of the historical period for the two mixed models, Mixed D and Mixed D*. The solid vertical lines represent the maxpost estimate of the parameter for each of the models, whereas the dashed black lines represent the reference values ($S = 9000\,\mathrm{m^3\,s^{-1}}$ and $t^* = 1500$).

old has a greater impact than poor knowledge of the historical period length. Even if the maxpost estimates of the perception threshold for models B and D are close to the true value ($9000\,\mathrm{m^3\,s^{-1}}$), the uncertainty resulting from determining the threshold has a strong impact on quantile uncertainty. Furthermore, the estimation of the historical period length in the case of Model C is also quite imprecise, but this has little impact on the uncertainty in the results when compared with those of Model A. The comparison of Model A, for which only the number of exceedances of the perception threshold is known, with Model E, for which the discharge of historical floods is known within an uncertainty interval, demonstrated the value of reconstructing the discharge of each historical flood.

Finally, the results for the 1816–2020 period suggest that the quantile uncertainty may be underestimated when the perception threshold and the historical period length are unduly considered to be perfectly known. The models proposed in this paper allow us to account for imperfect knowledge when estimating extreme quantiles.

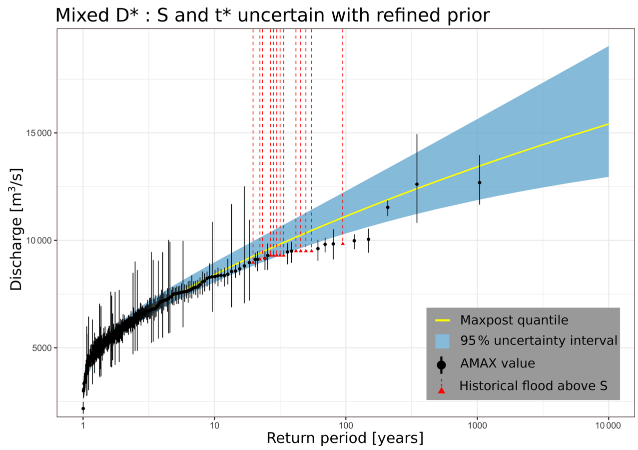

Figure 8. Flood distribution and the 95 % confidence interval of Model D* (mixed sample: systematic period 1816–2020 + 13 historical exceedances on 1500–1815, refined prior on S and t^*). The experimental distribution is shown in black (AMAX values) or red (exceedances of the perception threshold).

6.2 Main findings for the 1500–2020 period

Application of the binomial Model D to a mixed sample with the discharge estimate of AMAX values for the systematic 1816–2020 period and a collection of 13 historical floods from 1500 to 1815 allows the uncertainty around the perception threshold S and the historical period length to be considered. Priors of Model D were refined, in order to have a more realistic assessment of threshold S and the starting date t^* of the historical period. This refined model, called Model D*, gives the following results:

- Despite the fact that the available AMAX flood series for the 1816–2020 period is really long (205 years), it is possible to reduce the uncertainty in the flood quantiles (Fig. 7a) by adding information on 13 exceedances of a threshold $S = 9000\,\mathrm{m^3\,s^{-1}}$ during 3 prior centuries (period 1500–1815) and prior knowledge about S and t^*.

- The refinement of the prior distributions for the threshold S and the starting date t^* with Model D* gives a more precise assessment of flood quantiles than with Model D (Fig. 7a). The posterior standard deviation (expressed as a percentage) of the Q_{1000} quantile decreases from 11 % to 8 % (Table 3). In both cases, considering the perception threshold S to be uncertain has much more of an impact on the uncertainty in the results than considering a lack of knowledge about the length of the historical period.

- The combination of an increased perception threshold ($S_{\mathrm{maxpost}} = 9386\,\mathrm{m^3\,s^{-1}}$ vs. $S_{\mathrm{prior}} = 9000\,\mathrm{m^3\,s^{-1}}$) and a reduced span of the historical period ($t^*_{\mathrm{maxpost}} = 1526$ vs. $t^*_{\mathrm{prior}} = 1500$) may be a symptom of the non-exhaustiveness of floods in the historical samples in the HISTRHÔNE database, even though no non-stationarity of the frequency of floods was detected (Fig. 1b). As the historical flood inventory is based on physical impacts, it may be sensitive to some changes in land settlement and flood protection.

- The flood distribution and 95 % credibility interval of Model D* are represented in Fig. 8. AMAX values are reported with their uncertainty (from 5 % to 30 %), and historical floods are reported as exceedances. Information on floods during 3 prior centuries (1500–1815) reduces the level of extrapolation towards extreme floods (flood of record has a plotting position around the 1000-year return period in Fig. 8, rather than a 400-year return period, as shown in Fig. 2a, for the 1816–2020 period).

7 Conclusion

This paper proposes binomial models for the inclusion of historical data into FFA; these models explicitly recognize the uncertain nature of both the perception threshold and the starting date of the historical period.

The models are first tested with a 205-year-long series of AMAX values for the outlet of the Rhône River at Beaucaire, France. The time series has been artificially subsampled in order to mimic a historical context, considering AMAX values for a 50-year period (1970–2020) and a collection of 10 "historical" floods during the 1816–1969 period. The estimated quantiles were compared with estimates from a GEV model with AMAX values for the entire period (1816–2020). Considering that the perception threshold is perfectly known when this is not the case can lead to a significant underestimation of the uncertainty in flood quantiles. This also holds for the length of the historical period, although to a much lesser extent. In the case of this subsample, the use of historical data makes it possible to reduce the uncertainty in the quantiles compared with the sole use of the short systematic sample (1970–2020), considering uncertainties in the threshold S and the starting date t^* of the historical period. The binomial model estimate with known S and t^* (Model A) was then compared to an estimate for which historical flood discharges are known within an interval (Model E). In Beaucaire, the use of the historical flood discharges turned out to be slightly more informative than the use of the sole number of exceedances of the perception threshold.

The paper also presents the results of the binomial model with a mixed sample of 205 AMAX values (1816–2020 period) and 13 occurrences of historical floods (1500–1815 period). The addition of historical information for 3 centuries reduces the uncertainty in the Q_{100} and Q_{1000} flood quantiles (about 15 %), despite only the number of exceedances being known. However, some doubts remain about the completeness of the historical sample, as the posterior estimation of S and t^* are larger than the prior.

The stationarity hypothesis may be challenged by climatic variability at Beaucaire, as trends in flood magnitudes have been identified in several regions of Europe (Hall et al., 2014; Blöschl et al., 2020) and France (Giuntoli et al., 2019). To date, there are no rules in France for accounting for the impact of climate change on flood risk estimates. However, it is still possible to integrate temporal changes in climate processes or watershed characteristics into the probabilistic model itself, as increasingly described in the literature (see Salas et al., 2018, for an overview). It is also important to note that, although outside of the FFA scope, such long series remain interesting for the study of the long-term variability in floods over several centuries and of their value for risk awareness and memory.

Appendix A: Plotting position for unknown historical floods

The exceedance probability of the ith value $q(i)$ of a sample $(q(1) \geq \ldots \geq q(N))$, sorted by decreasing value, is as follows:

$$p'_i = \mathrm{Prob}\left[Q > q(i)\right] = \frac{i-a}{N+1-2a}, \tag{A1}$$

using, for example, $a = 0.44$, the optimum value for a Gumbel distribution (Cunnane, 1978).

Hirsch (1987) proposed splitting a mixed sample, formed by $\left(q_1, \ldots, q_{\mathrm{NY_C}}\right)$ AMAX values during $\mathrm{NY_C}$ years (continuous period) and $\mathrm{NE_H}$ historical values larger than a perception threshold S during $\mathrm{NY_H}$ years (historical period), into two subsamples:

– NE exceedances of the threshold S for the whole period (divided into $\mathrm{NE_H}$ and $\mathrm{NE_C}$ exceedances for the historical and continuous periods, respectively), during NY years, with $\mathrm{NY} = \mathrm{NY_H} + \mathrm{NY_C}$ years,

$$\mathrm{Prob}\left[Q > q(i)\right] = \left(\frac{\mathrm{NE}}{\mathrm{NY}}\right) p'_i = \left(\frac{\mathrm{NE}}{\mathrm{NY}}\right) \frac{i-0.44}{\mathrm{NE}+0.12} \quad i = 1, \mathrm{NE}; \tag{A2}$$

– floods lower than S for the continuous period,

$$\mathrm{Prob}\left[Q > q(i)\right] = \left(\frac{\mathrm{NE}}{\mathrm{NY}}\right) + \left(1-\frac{\mathrm{NE}}{\mathrm{NY}}\right) \frac{(i-\mathrm{NE})-0.44}{\mathrm{NY_C}-\mathrm{NE_C}+0.12} \quad i = \mathrm{NE}+1, \mathrm{NY_C}+\mathrm{NE_H}. \tag{A3}$$

In the current case study, as the discharge of historical floods is not known (only threshold exceedance), it is not possible to rank all values of the mixed sample. A way to circumvent this problem is to randomly rank the historical unknown floods amongst the $\mathrm{NE_C}$ floods larger than S during the continuous period. This is done using the following steps:

– *Step 1.* Randomly sample (without replacement) the rank of the $\mathrm{NE_C}$ floods of the continuous period within the whole period using sample ($x = 1$: NE, *size* $= \mathrm{NE_C}$, *replace* = FALSE) (R code).

– *Step 2.* As we know the values of the $\mathrm{NE_C}$ floods larger than S during the continuous period, apply the ranks just sampled to them (i.e. the smallest sample rank is assigned to the largest flood, etc.).

– *Step 3.* Assign the remaining ranks to the $\mathrm{NE_H}$ floods larger than S during the historical period.

As we assigned ranks to all exceedances of the threshold S for the whole period, we are able to compute their plotting position with Eq. (A2). Let $q_1 \geq \ldots \geq q_{\mathrm{NE_C}} \geq S$ denote the known discharges of the continuous period larger than the threshold S, with their corresponding ranks $r_1 < \ldots < r_{NE_C}$. We now assign an interval to the unknown historical discharges (see Fig. 8):

– If $r_1 > 1$, we have $(r_1 - 1)$ historical flood discharges larger than q_1. They will be plotted with vertical dashed lines larger than q_1.

– If $(r_{i+1} - r_i) > 1$, we have $(r_{i+1} - r_i)$ historical flood discharges within the interval $[q_{i+1}, q_i]$. They will be plotted with vertical dashed lines larger than $1/2\,(q_i + q_{i+1})$.

– If $r_{\mathrm{NE_C}} < \mathrm{NE}$, we have (NE - $r_{\mathrm{NE_C}}$) historical flood discharges within the interval $[S, q_{\mathrm{NE_C}}]$. They will be plotted with vertical dashed lines larger than $1/2\,(S + q_{\mathrm{NE_C}})$.

This order is random, but it makes it possible to draw the empirical distribution of floods and to compare it with the estimated GEV distributions.

Code and data availability. The continuous time series (1816–2016), the historical data (1500–1815) and the codes for flood frequency analysis are available at https://doi.org/10.5281/zenodo.14138600 (Lucas, 2024).

Author contributions. ML: data curation, analysis, original draft (in French) and figures. ML: review and editing, supervision, original draft (in English), and project administration. BR: conceptualization, review and editing, and supervision. JLC: review and editing and supervision.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. This study was conducted within the Rhône Sediment Observatory (OSR), a multi-partner research programme funded through Plan Rhône by the European Regional Development Fund (ERDF), Agence de l'eau RMC (France), CNR (France), EDF (France), and three regional councils (Auvergne-Rhône-Alpes, PACA and Occitanie; France). Data and expert knowledge on the Rhône River at Beaucaire were provided by Gilles Pierrefeu from CNR, the Rhône Sediment Observatory, Pascal Billy and Helene Decourcelle from the DREAL Auvergne-Rhône-Alpes (Ministry of Ecology), and the HISTRHÔNE database from the CEREGE (Georges Pichard). Finally, we thank Neil Macdonald and Helen Hooker for their constructive comments that helped us improve the paper.

Financial support. This research has been supported by the Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (co-funding of Mathieu Lucas's PhD), the Compagnie Nationale du Rhône (co-funding of Mathieu Lucas's PhD) and the Université de Lyon (grant no. ANR-17-EURE-0018).

Review statement. This paper was edited by Frederiek Sperna Weiland and reviewed by Neil Macdonald and Helen Hooker.

References

Apel, H., Thieken, A. H., Merz, B., and Blöschl, G.: Flood risk assessment and associated uncertainty, Nat. Hazards Earth Syst. Sci., 4, 295–308, https://doi.org/10.5194/nhess-4-295-2004, 2004.

Benito, G., Lang, M., Barriendos, M., Llasat, M. C., Francés, F., Ouarda, T., Thorndycraft, V., Enzel, Y., Bardossy, A., Cœur, D., and Bobée, B.: Use of Systematic, Palaeoflood and Historical Data for the Improvement of Flood Risk Estimation. Review of Scientific Methods, Nat. Hazards, 31, 623–643, https://doi.org/10.1023/B:NHAZ.0000024895.48463.eb, 2004.

Benson, M. A.: Use of historical data in flood-frequency analysis, Eos T. Am. Geophys. Un., 31, 3, 419–424, https://doi.org/10.1029/TR031i003p00419, 1950.

Blöschl, G., A. Kiss, A. Viglione, M. Barriendos, O. Böhm, R. Brázdil, D. Coeur, G. Demarée, M. C. Llasat, N. Macdonald, D. Retsö, L. Roald, P. Schmocker-Fackel, I. Amorim, M. Bělínová, G. Benito, C. Bertolin, D. Camuffo, D. Cornel, R. Doktor, L. Elleder, S. Enzi, J. C. Garcia, R. Glaser, J. Hall, K. Haslinger, M. Hofstätter, J. Komma, D. Limanówka, D. Lun, A. Panin, J. Parajka, H. Petrić, F. S. Rodrigo, C. Rohr, J. Schönbein, L. Schulte, L. P. Silva, W. H. J. Toonen, P. Valent, J. Waser and Wetter, O.: Current European flood-rich period exceptional compared with past 500 years, Nature, 583, 560–566, https://doi.org/10.1038/s41586-020-2478-3, 2020.

Cunnane, C.: Unbiased plotting position – a review, J. Hydrol., 37, 205–222, https://doi.org/10.1016/0022-1694(78)90017-3, 1978.

Darienzo, M., Renard, B., Le Coz, J., and Lang, M.: Detection of Stage-Discharge Rating Shifts Using Gaugings: A Recursive Segmentation Procedure Accounting for Observational and Model Uncertainties, Water Resour. Res., 57, e2020WR028607, https://doi.org/10.1029/2020WR028607, 2021.

Dezileau, B., Terrier, L., Berger, J., Blanchemanche, P., Latapie, A., Freydier, R., Bremond, L., Paquier, A., Lang, M., and Delgado, J.: A multidating approach applied to historical slackwater flood deposits of the Gardon River, SE France, Geomorphology, 214, 56–68, https://doi.org/10.1016/j.geomorph.2014.03.017, 2014.

Engeland, K., Aano, A., Steffensen, I., Støren, E., and Paasche, Ø.: New flood frequency estimates for the largest river in Norway based on the combination of short and long time series, Hydrol. Earth Syst. Sci., 24, 5595–5619, https://doi.org/10.5194/hess-24-5595-2020, 2020.

European Union: Directive 2007/60/EC of the European Parliament and of the council of 23 October 2007 on the assessment and management of flood risks, Official Journal of the European Union, 12 pp., https://eur-lex.europa.eu/eli/dir/2007/60/oj (last access: 20 November 2024), 2007.

Falconer, J. R., Frank, E., Polaschek, D. L. L., and Joshi, C.: Methods for Eliciting Informative Prior Distributions: A Critical Review, Decis. Anal., 2022, 189–204, https://doi.org/10.1287/deca.2022.0451, 2022.

Feller, W.: An Introduction to Probability Theory and Its Applications, Vol. II, 2nd ed., New York, Wiley, MR0270403, ISBN-10 9780471257097, ISBN 13 978-0471257097, 1971.

Gerard, R. and Karpuk, E. W.: Probability Analysis of Historical Flood Data, J. Hydraul. Eng.-ASCE, 105, 1153–1165, https://doi.org/10.1061/JYCEAJ.0005273, 1979.

Giuntoli, I., Renard, B., and Lang, M.: Floods in France, Changes in Flood Risk in Europe, 1st ed., CRC Press, 199–211, ISBN 9780203098097, 2019.

Hall, J., Arheimer, B., Borga, M., Brázdil, R., Claps, P., Kiss, A., Kjeldsen, T. R., Kriaučiūnienė, J., Kundzewicz, Z. W., Lang, M., Llasat, M. C., Macdonald, N., McIntyre, N., Mediero, L., Merz, B., Merz, R., Molnar, P., Montanari, A., Neuhold, C., Parajka, J., Perdigão, R. A. P., Plavcová, L., Rogger, M., Salinas, J. L., Sauquet, E., Schär, C., Szolgay, J., Viglione, A., and Blöschl, G.: Understanding flood regime changes in Europe: a state-of-the-art assessment, Hydrol. Earth Syst. Sci., 18, 2735–2772, https://doi.org/10.5194/hess-18-2735-2014, 2014.

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

Kendall, M.: Rank Correlation Methods, Charles Griffin Book Series, London, Oxford University Press, ISBN 10 0195208374, ISBN 13 978-0195208375, 1948.

Kjeldsen, T., Macdonald, N., Lang, M., Mediero, L., Albuquerque, T., Bogdanowicz, E., Brázdil, R., Castellarin, A., David, V., Fleig, A., Gül, G., Kriauciuniene, J., Kohnová, S., Merz, B., Nicholson, O., Roald, L., Salinas, J., Sarauskiene, D., Šraj, M., Strupczewski, W., Szolgay, J., Toumazis, A., Vanneuville, W., Veijalainen N., and Wilson D.: Documentary evidence of past floods in Europe and their utility in flood frequency estimation, J. Hydrol., 517, 963–973, https://doi.org/10.1016/j.jhydrol.2014.06.038, 2014.

Kuczera, G.: Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference, Water Resour. Res., 35, 1551–1557, https://doi.org/10.1029/1999WR900012, 1999.

Lang, M., Ouarda, T., and Bobée T.: Towards operational guidelines for over-threshold modelling, J. Hydrol., 225, 103–117, https://doi.org/10.1016/S0022-1694(99)00167-5, 1999.

Lang, M., Fernandez, Bono, J. F., Recking, A., Naulet, R., and Grau Gimeno, P.: Methodological guide for paleoflood and historical peak discharge estimation, Systematic, Palaeoflood and Historical Data for the Improvement of Flood Risk Estimation: Methodological Guidelines, edited by: Benito, G. and Thorndycraft, V., CSIC Madrid, Spain, 43–53, ISN 84-921958-3-5, 2004.

Lucas, M.: MatLcs/HistoFloods: Codes and data related to the article: Lucas et al. A comprehensive uncertainty framework for historical flood frequency analysis: a 500-year-long case study. Hydrology and earth sciences sciences (HESS), Zenodo [data set] and [code], https://doi.org/10.5281/zenodo.14138601, 2024.

Lucas, M., Renard, B., Le Coz, J., Lang, M., Bard, A., and Pierrefeu, G.: Are historical stage records useful to decrease the uncertainty of flood frequency analysis? A 200-year long case study, J. Hydrol., 624, 129840, https://doi.org/10.1016/j.jhydrol.2023.129840, 2023.

Macdonald, N. and Sangster, H.: High-magnitude flooding across Britain since AD 1750, Hydrol. Earth Syst. Sci., 21, 1631–1650, https://doi.org/10.5194/hess-21-1631-2017, 2017.

Macdonald, N., Kjeldsen, T. R., Prosdocimi, I., and Sangster, H.: Reassessing flood frequency for the Sussex Ouse, Lewes: the inclusion of historical flood information since AD 1650, Nat. Hazards Earth Syst. Sci., 14, 2817–2828, https://doi.org/10.5194/nhess-14-2817-2014, 2014.

Machado, M. J., Botero, B. A., López, J., Francés, F., Díez-Herrero, A., and Benito, G.: Flood frequency analysis of historical flood data under stationary and non-stationary modelling, Hydrol. Earth Syst. Sci., 19, 2561–2576, https://doi.org/10.5194/hess-19-2561-2015, 2015.

Mann, H. B.: Nonparametric Tests Against Trend, Econometrica, 13, 245–259, https://doi.org/10.2307/1907187, 1945.

Martins, E. S. and Stedinger, J. R.: Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data, Water Resour. Res., 36, 737–744, https://doi.org/10.1029/1999WR900330, 2000.

Merz, R. and Blöschl, G.: Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information, Water Resour. Res., 44, W08432, https://doi.org/10.1029/2007WR006744, 2008.

METS: Repères de crues, plateforme collaborative de référence pour le recensement des repères de crues en France, https://www.reperesdecrues.developpement-durable.gouv.fr (last access: 20 November 2024), 2023.

Neppel, L., Renard, B., Lang, M., Ayral, P. A., Coeur, D., Gaume, E., Jacob, N., Payrastre, O., Pobanz, K., and Vinet, F.: Flood frequency analysis using historical data: accounting for random and systematic errors, Hydrolog. Sci. J., 55, 192–208, https://doi.org/10.1080/02626660903546092, 2010.

Parkes, B. and Demeritt, D.: Defining the hundred year flood: A Bayesian approach for using historic data to reduce uncertainty in flood frequency estimates, J. Hydrol., 540, 1189–1208, https://doi.org/10.1016/j.jhydrol.2016.07.025, 2016.

Payrastre, O., Gaume, E., and Andrieu, H.: Usefulness of historical information for flood frequency analyses: Developments based on a case study, Water Resour. Res., 47, W08511, https://doi.org/10.1029/2010WR009812, 2011.

Pettitt, A. N.: A non-parametric approach to the change-point problem, Appl. Stat.-J. Roy. St. C, 28, 126–135, https://doi.org/10.2307/2346729, 1979.

Pichard, G.: Les crues sur le bas Rhône de 1500 à nos jours. Pour une histoire hydro-climatique, Méditerranée, 82, 105–116, https://doi.org/10.3406/medit.1995.2908, 1995.

Pichard, G. and Roucaute, E.: Sept siècles d'histoire hydroclimatique du Rhône d'Orange à la mer (1300–2000). Climat, crues, inondations, Presses Universitaires de Provence (Hors-série de la revue Méditerranée), 194 pp., https://journals.openedition.org/geocarrefour/9491 (last access: 20 November 2024), 2014.

Pichard, G., Arnaud-Fassetta, G., Moron, V., and Roucaute E.: Hydroclimatology of the Lower Rhône Valley: historical flood reconstruction (AD 1300–2000) based on documentary and instrumental sources, Hydrolog. Sci. J., 62, 1772–1795, https://doi.org/10.1080/02626667.2017.1349314, 2017.

Piotte, O., Boura, C., Cazaubon, A., Chaléon, C., Chambon, D., Guillevic, G., Pasquet, F., Perherin, C., and Raimbault, E.: Collection, storage and management of high-water marks data: praxis and recommendations, E3S Web Conf., https://doi.org/10.1051/e3sconf/20160716003, 2016.

Prosdocimi, I.: German tanks and historical records: the estimation of the time coverage of ungauged extreme events, Stoch. Env. Res. Risk A., 32, 607–622, https://doi.org/10.1007/s00477-017-1418-8, 2018.

Reis, D. S. and Stedinger, J. R.: Bayesian MCMC flood frequency analysis with historical information, J. Hydrol., 313, 97–116, https://doi.org/10.1016/j.jhydrol.2005.02.028, 2005.

Renard, B.: Use of a national flood mark database to estimate flood hazard in the distant past, Hydrolog. Sci. J., 68, 1078–1094, https://doi.org/10.1080/02626667.2023.2212165, 2023.

Renard, B., Garreta, V., and Lang, M.: An application of Bayesian analysis and Markov chain Monte Carlo methods to the estimation of a regional trend in annual maxima, Water Resour. Res., 42, W12422, https://doi.org/10.1029/2005WR004591, 2006.

Salas, J. D., Obeysekera, J., and Vogel R. M.: Techniques for assessing water infrastructure for nonstationary extreme events: a review, Hydrolog. Sci. J., 63, 325–352, https://doi.org/10.1080/02626667.2018.1426858, 2018.

Shang, X., Wang, D., Singh, V. P., Wang, Y., Wu, J., Liu, J., Zou, Y., and He, R.: Effect of Uncertainty in Historical Data on Flood Frequency Analysis Using Bayesian Method, J. Hydrol. Eng., 26, 04021011, https://doi.org/10.1061/(ASCE)HE.1943-5584.0002075, 2021.

Sharma, S., Ghimire, G. R., Talchabhadel, R., Panthi, J., Lee, B. S., Sun, F., Baniya, R., and Adhikari, T. R.: Bayesian characterization of uncertainties surrounding fluvial flood hazard estimates, Hydrolog. Sci. J., 67, 277–286, https://doi.org/10.1080/02626667.2021.1999959, 2022.

St. George, S., Hefner, A. M., and Avila, J.: Paleofloods stage a comeback, Nat. Geosci., 13, 766–768, https://doi.org/10.1038/s41561-020-00664-2, 2020.

Stedinger, J. R. and Cohn, T. A.: Flood Frequency Analysis With Historical and Paleoflood Information, Water Resour. Res., 22, 785–793, https://doi.org/10.1029/WR022i005p00785, 1986.

Viglione, A., Merz, R., Salinas, J. L., and Blöschl, G.: Flood frequency hydrology: 3. A Bayesian analysis, Water Resour. Res., 49, 675–692, https://doi.org/10.1029/2011WR010782, 2013.