A COMPREHENSIVE UNCERTAINTY FRAMEWORK FOR HISTORICAL FLOOD FREQUENCY ANALYSIS: A 500-YEAR LONG CASE STUDY.

Mathieu Lucas [a], Michel Lang [a], Benjamin Renard [b], Jérôme Le Coz [a]
[a] INRAE, UR Riverly, Villeurbanne, France
[b] INRAE, Aix Marseille Univ., UR RECOVER, Aix-En-Provence, France

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 having 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 a maximum discharge below the perception threshold. This paper proposes a Binomial model which 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 evidences over three 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 quantiles uncertainty. A qualitatively similar conclusion is found when ignoring the uncertainty around the historical period length. However, its impact on flood quantiles uncertainty appears to be much smaller than that of the perception threshold.

Keywords

Flood frequency analysis
Historical data
Perception threshold
Historical period length
Uncertainty
Rhône River

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 to precisely estimate the parameters of the chosen distribution with discharge series of limited length (a few decades, generally). This is particularly problematic when low-probability (e.g. annual exceedance probability 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 Bloschl, 2008), or a combination of these (Macdonald and Sangster, 2017). This paper focuses on the first option and on the treatment of historical data in FFA.

Historical data can take a variety of forms. Historical data 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 paleoflood 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-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 much 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 to take 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 can be assumed to be smaller than the perception threshold for all years in the historical period with no recorded flood. It is possible, albeit 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 having 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 description 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 infrastructures (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, this ideal situation rarely holds, and the estimation of the perception threshold can be undermined by many factors including the difficulty to precisely estimate discharge, the temporally-varying perception of flood damages by populations living adjacent to the river, etc. 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).

The length of the historical period is generally considered to be perfectly known in the historical FFA methods of the literature. In principle, the historical period (or surveying period) should start at the date when the source of the historical information started to exist. 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 proposes 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, whereas it is affected by considerable uncertainty.

This paper presents an FFA probabilistic model that uses the number of times a perception threshold is exceeded over an historical period, and takes into account the uncertainty of 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 of 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 discharge uncertainties carefully determined. An uncertainty propagation chain developed by Lucas *et al.* (2023) accounts for errors on stage and gauging measurements, and 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 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 on quantile estimates is discussed.

This paper is organized as follows. Methods for historical FFA are introduced in section 2. Available data are presented in section 3 and their stationarity is verified. The FFA models are then applied and compared using the artificially subsampled record on the 1816-2020 period (section 4), and then to the entire dataset on the 1500-2020 period (section 5). Section 6 discusses some key results of this work and section 7 summarizes its main conclusions.

2. Probabilistic models

We first present how censored historical floods can be included into a probabilistic model (section 2.1) and then move towards more specific models accounting for uncertainties (section 2.2).

2.1 Standard treatment of censored data

Likelihood function

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

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

Under this parametrization, 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 having 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:

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

It is assumed that k, the number of exceedances of the perception threshold, follows a Binomial distribution $\mathscr{B}(n, \pi)$. 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:

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

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 $p(\boldsymbol{\theta}|\boldsymbol{q}, k)$ of parameters $\boldsymbol{\theta}$ given systematic and historical data is:

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

The term $p(\boldsymbol{\theta})$ represents the prior distribution of the parameters and needs to be elicited before inference. The posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{q}, k)$ is explored via a MCMC method (Renard *et al.*, 2006), leading to a representation of sampling uncertainty by means of r parameter vectors $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_r)$. The parameter vector $\widehat{\boldsymbol{\theta}}$ that maximizes the posterior distribution is called *maxpost*. Thereafter, the prior distribution $p(\boldsymbol{\theta})$ of the GEV parameters will be as follows: a positive Uniform distribution for μ and σ, and a Gaussian distribution with mean zero and standard deviation 0.2 for ξ, as proposed by Martins and Stedinger (2000).

Starting date of the historical period

The starting date t^* of the historical period can be assessed by two methods.

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

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

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 $(NY - 1) / NE$.

In some cases, the historical period (including flood and no-flood information) starts before the date t_1 of the first known flood (for instance, at the creation of the service in charge of surveying floods, or at the date of bridge construction where historical data is available). Let denote this date t_{start}, and consider the difference $(t_1 - t_{start})$ between these two dates. A second possible estimate, based on the Poisson process paradox (Feller, 1971), takes advantage 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. Without any knowledge of the return period T_S of the threshold, but using the difference $(t_1 - t_{start})$, we have:

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

2.2 Models accounting for uncertainties

We first present Binomial models for historical floods known to be larger than a perception threshold, with a propagation procedure for both stage and rating curve uncertainties (*model A*), or with parameters accounting for uncertainties on perception threshold (*model B*), length of the historical period (*model C*) or both (*model D*). Table 1 summarizes which Binomial model accounts for uncertainty, and/or historical period length. A fifth *model E* considers the case when historical discharges are known within an interval.

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

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

In equation (3), the uncertainty in the AMAX discharges for the systematic period is assumed to be negligible. As this uncertainty can reach 30% at Beaucaire during the XIXth century (see following section 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), that leads to s = 500 realisations 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 equation (4), and each posterior distribution can be explored with the MCMC sampler. 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 for 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.

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 into account the imperfect knowledge of S, 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 equation (3) becomes $L(\boldsymbol{\theta}, S; \boldsymbol{q}, k)$ instead of $L(\boldsymbol{\theta}; \boldsymbol{q}, k)$, while its right-hand side remains unchanged. The *posterior* distribution of the parameters $\boldsymbol{\theta}$ and S given the data is:

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

This posterior distribution takes into account the hydrometric uncertainty of the systematic period, the sampling uncertainty and the uncertainty of the perception threshold. 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 which reflects the knowledge on this parameter, which is highly case-specific and can range from very imprecise to nearly-known.

Model C: Binomial model for historical floods, accounting for the uncertainty of 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 equation (3) becomes $L(\boldsymbol{\theta}, n; \boldsymbol{q}, k)$. The posterior distribution of the parameters $\boldsymbol{\theta}$ and n given the data is:

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

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

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

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

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

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

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). Since 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:

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

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

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

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 five centuries

We first consider the 205-year long daily discharge series of the Rhône River at Beaucaire, France, from 1816 to 2020 (catchment area: 95 590 km^2). Daily stage measurements started in 1816. The gauging station has been used until the construction of the Vallabrègues hydroelectric scheme in 1967, which led to the derivation of a part of the discharge. Consequently, a new gauging station was installed 2 km downstream from the restitution of the diverted discharges. This new station has been used ever since. The Vallabrègues Dam has no impact on the discharge at the station because it has a very limited storage capacity and it is opened during floods to cancel the backwater effect it creates for low flows. A set of 500 realisations 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% (XIXth century) to 5% (1967-2020). Secondly, a collection of historical flood testimonies from 1500 to 1815 is available from the HISTRHÔNE database (https://histrhone.cerege.fr/) (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 the *C4* class*: "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 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 highlight its impact more clearly.

3.2 Stationarity tests

As the probabilistic models described in section 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.

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 p-values of 0.15 and 0.4, respectively, indicate no significant change. The segmentation procedure proposed by Darienzo *et al.* (2021) was also applied since it allows accounting for the uncertainty around AMAX discharges. This procedure indicates that the optimum number of segments is equal to one, confirming the absence of significant change.

Historical data

The historical data used here (13 extremes floods of *C4* class) can be interpreted as peak-over-threshold (POT) values, since they correspond to all floods having exceeded the perception threshold *S* and that no year has more than one C4 flood during the 1500-1815 period. AMAX values from the continuous period larger than 9000 m^3/s can also be viewed as POT values as the 14 largest values (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 on 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. The whole sample can hence be considered as stationary.

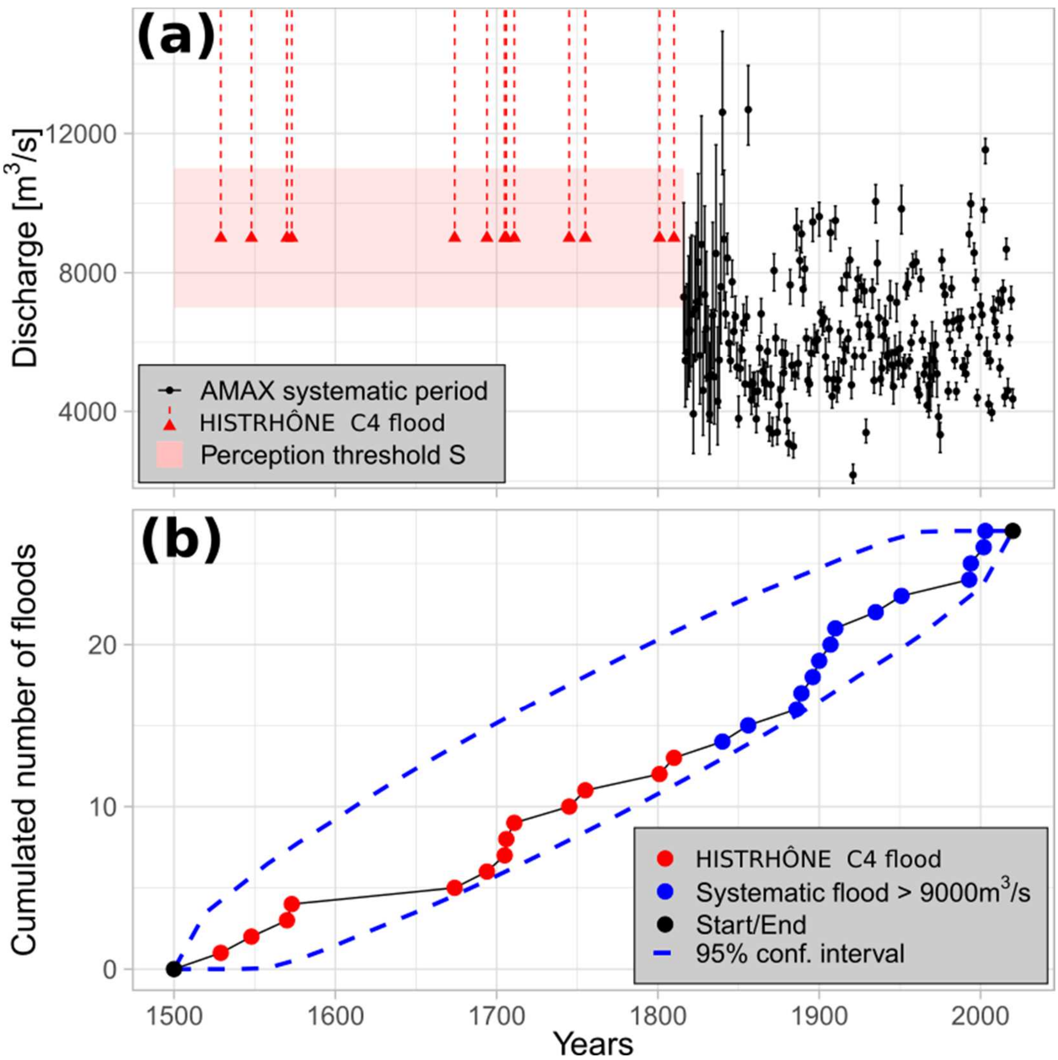

Figure 1: (a) The Rhône River at Beaucaire, AMAX flood discharges with 95% uncertainty intervals (1816-2020, systematic period, (Lucas *et al.*, 2023) and C4 class floods from 1500 to 1815 (HISTRHÔNE database); **(b)** Cumulated number of *C4* class floods and POT floods (systematic period) with 95%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

4. Flood Frequency Analysis on the 1816-2020 period

4.1 Subsampled data sets

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

- Short sample containing AMAX discharges during the 1970-2020 period, noted *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).

- Mixed sample, with AMAX discharges on the 1970-2020 period and a collection of *historical* values on the 1816-1969 period, noted *Mixed A,…, Mixed E*, according to the model used. A perception threshold $S = 9000$ m^3/s leads to a collection of 10 *historical* floods. When using *model B* (or *D*), we consider a vague Normal prior distribution on the perception threshold: $N(9000; 2000)$, 2000 being the standard deviation. When using *model C* (or *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 *Q100* and *Q1000* floods (Figure 2b), parameters (ξ, *S*, *t**) (Table 2) and GEV distributions (Figure 2a). The plotting position formula proposed by Hirsch (1987) in case of a mixed sample with AMAX values from a continuous period and historical discharges larger than a perception threshold is applied. The appendix gives a procedure in case where historical flood discharges are unknown, using only exceedances of the threshold.

Table 2: *Maxpost estimation ± posterior standard deviation expressed in percentage for Q100 and Q1000 floods, and (ξ, S, 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 (m^3/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 (m^3/s)	/	/	/	9163 ± 8%	/	9332 ± 9%	/
	t^*	/	/	/	/	1833 ± 4%	1785 ± 6%	/

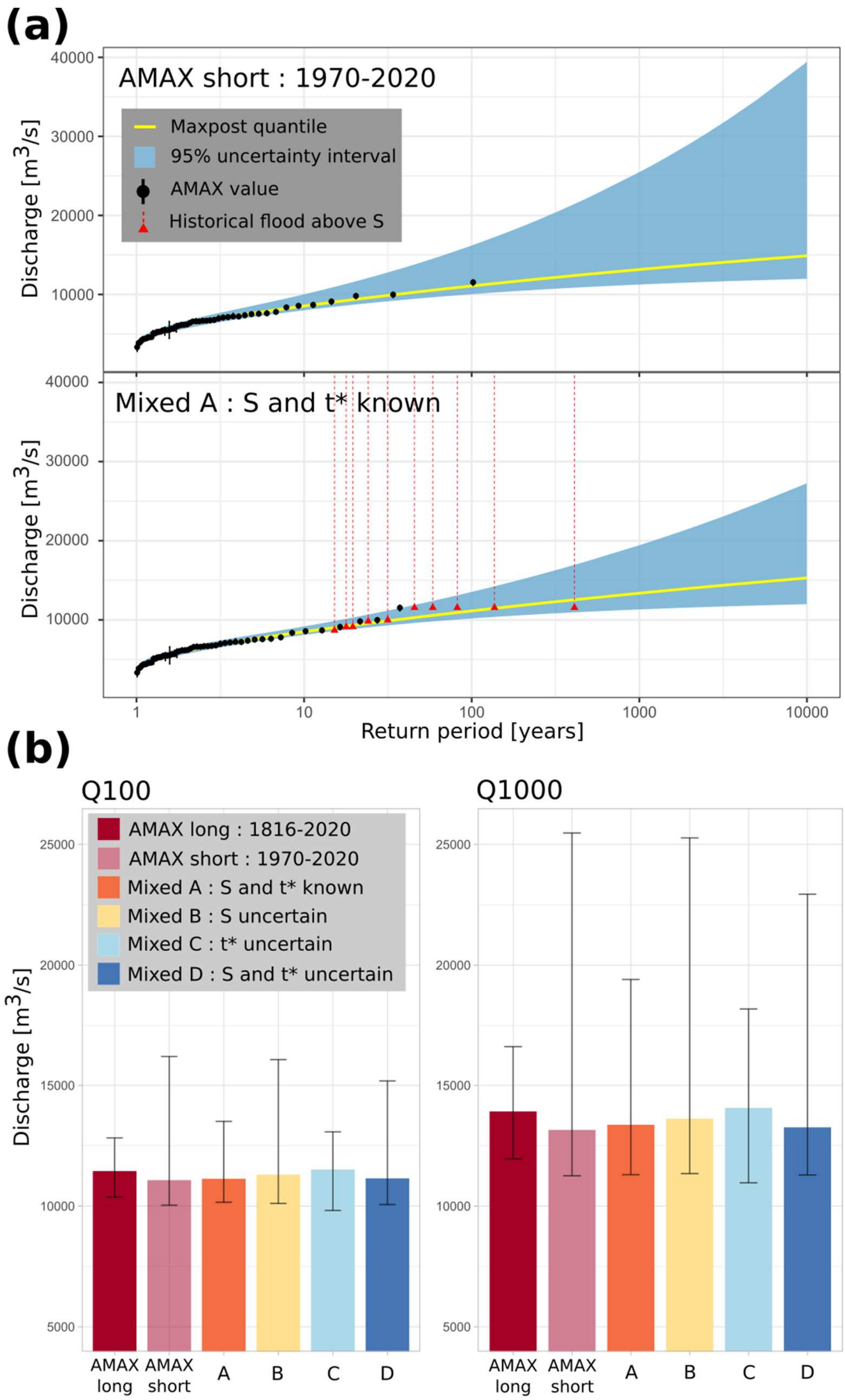

Figure 2: (a) GEV quantiles with 95% credibility intervals, example of two different models and datasets: GEV model on AMAX values (*AMAX short* 1970-2020) and binomial *Model A* on mixed sample (1816-2020); **(b)** *Q100* and *Q1000* floods with 95% credibility intervals displayed as error bars. *AMAX long* refers to the sample on the 1816-2020 period; *AMAX short* refers to the sample on the 1970-2020 period; *Mixed A-B-C-D* refers to a mixed sample ("historical" floods on the 1816-1969 period and AMAX 1970-2020) for various statistical models.

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 Figure 2). A Binomial model exploiting historical flood events notably reduces uncertainty when the perception threshold S is known (Figure 2b, *Mixed A* and *Mixed C*), although not achieving the precision obtained with 205 years of systematic records (Figure 2b, *AMAX long*). Accounting for the uncertainty on threshold *S* (Figure 2b, *Mixed B* and *Mixed D*) increases the uncertainty of 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 the estimates are close to zero and slightly positive (Table 2), which corresponds 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 two centuries than with a short series (*AMAX short*) of 5 decades. The use of historical data through a Binomial model is not very efficient in reducing uncertainty on the shape parameter ξ (Table 2). Overall, the *maxpost* estimate of *Q100* and *Q1000* quantiles are very close for all models (Figure 2b). In the next sections, the interest 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 the perception threshold uncertain

The use of *model B* reflects a lack of knowledge of the perception threshold, which becomes a parameter of the model. Figure 2b shows that the quantile uncertainty estimated with *model B* is much greater than with *model A*, and is close to the one obtained with systematic data only (*AMAX short*). Poor knowledge of the perception threshold therefore has major consequences for the quantile estimates, since it greatly reduces the value of using historical occurrences. The true value of the perception threshold is $S = 9000\ m^3/s$. The prior and posterior distributions of the threshold S are shown in Figure 3a. It can be seen that the posterior estimate for *model B* (9163 m^3/s) is close to the true value (9000 m^3/s), and that the model has effectively improved the knowledge of the threshold compared with the prior distribution $N(9000; 2000)$. The posterior uncertainty of 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 as a priority objective for historical FFA.

4.4 Impact of considering the historical period length uncertain

Model C is used to represent the lack of knowledge on the length of the historical period. In Figure 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 Figure 3b. The *maxpost* date is 1833, whereas the series actually begins in 1816. This underestimation by 17 years 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 section 3.2. The posterior distribution of the starting date *t** for *model C* (Figure 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* (Figure 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.

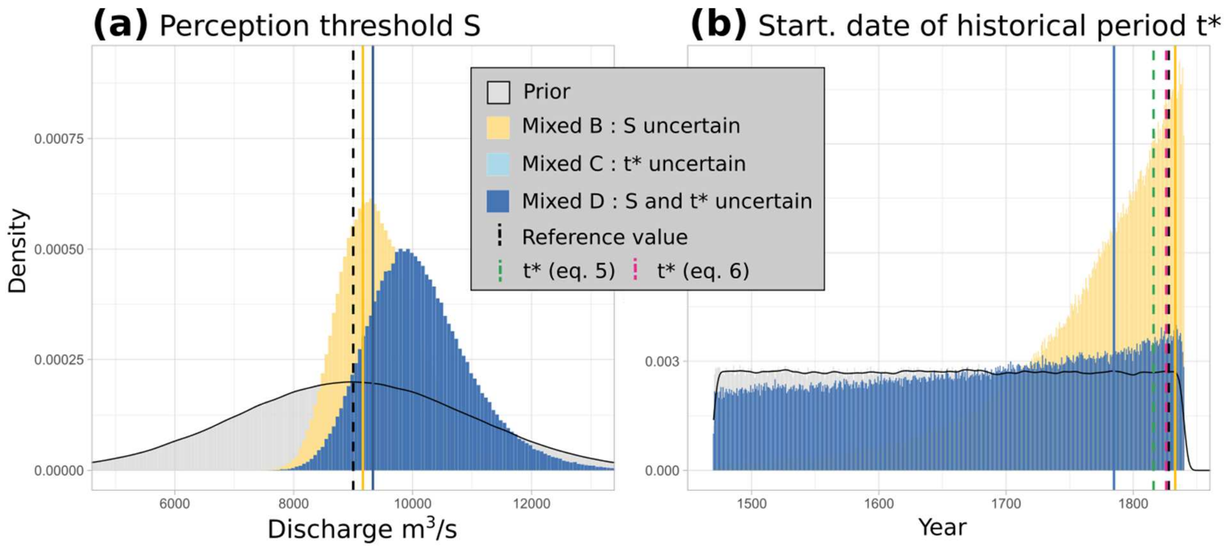

Figure 3: Prior and posterior distributions of: **(a)** the perception threshold *S*; **(b)** the starting date *t** of the historical period (1816-2020 period). The solid vertical lines represent the maxpost estimate of the parameter for each of the models, and the black dashed lines represent the reference values (*S* = 9000 m3/s and *t** = 1816). The green and pink dashed vertical lines (b) represent the estimates of *t** from equations (5) and (6).

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

Model D assumes that both *S* and *n* are uncertain in the probabilistic model. The *maxpost* quantiles estimated in Figure 2b are close to the reference values. In contrast, the width of the credibility interval is large and lies between that of *models B* and *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 understanding the origin of this large uncertainty. The posterior distribution of the perception threshold *S*, although with a *maxpost* value (9332 m^3/s) close to the true value (9000 m^3/s), is very imprecise with a large standard deviation (883 m^3/s). The perception threshold *S* appears to be slightly less precisely estimated than with *model B* (Table 2), with respectively posterior standard deviation of 9% and 8%. 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 (Figure 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 this might be due to some correlations between parameters. In particular, the Pearson correlation coefficient ρ is respectively equal to 0.44 and 0.42 between the length of the historical period n and the perception threshold S, as well as between the perception threshold S and the shape parameter ξ.

4.6 Value of estimating the peak discharge of historical flood

Binomial *models A, B, C* and *D* only use information on the number of times k 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 Figure 4. There is a reduction in uncertainty of around 25% for *Q1000* with *model E* compared to Binomial *model A* (posterior standard deviations of 2255 and 3019 m^3/s, respectively). However, the uncertainty of *model E* remains around 65% greater than that of the *GEV 1816-2020* model for *Q1000*. 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 reduce 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 have better information on historical floods.

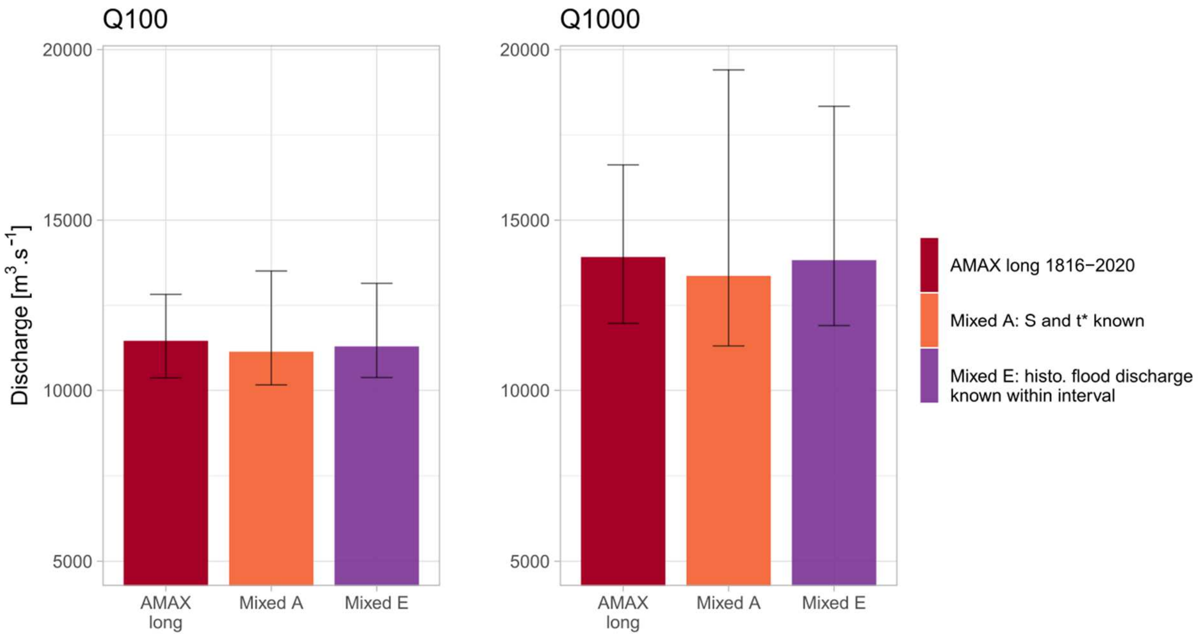

Figure 4: *Q100* and *Q1000* floods with 95% credibility intervals displayed as error bars. *AMAX long* refers to an annual maximum sample on the 1816-2020 period; Mixed refers to a mixed sample (“historical” floods on the 1816-1969 period and AMAX 1970-2020). *Model A* uses only the number of times the perception threshold has been exceeded, while *Model E* considers the peak discharge (and its uncertainty) of each historical flood that exceeded threshold *S*. Perception threshold *S* and start date of historical period *t** are considered perfectly known (*models A* and *E*).

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, E) with known parameters (*S* and *n*). The systematic record has been artificially subsampled into a mixed data set, containing 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, 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.

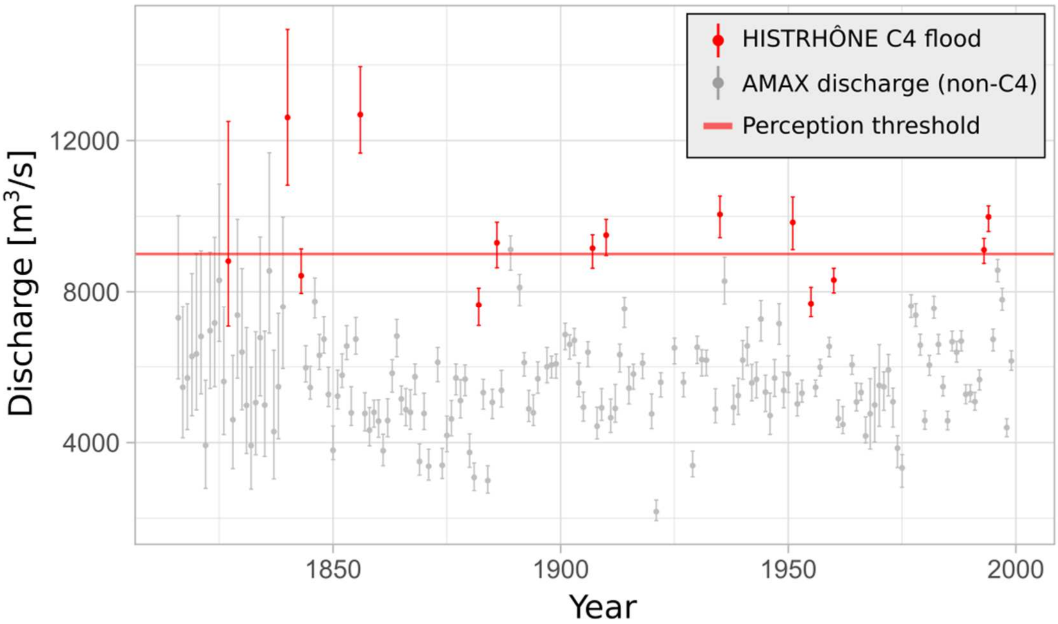

Figure 5: AMAX flood discharges (1816-2000) from Lucas *et al.* (2023) (in grey) cross-referenced with C4 floods from HISTRHÔNE database (in red). The horizontal line corresponds to the estimated perception threshold *S* = 9000 m^3/s.

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 on a mixed sample over the period 1500-2020, with AMAX values for the systematic period 1816-2020 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 of *t**.

A second analysis will refine 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* (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). Five amongst fourteen *C4* floods are below the threshold $S = 9000$ m^3/s. Even accounting for discharge uncertainty, three *C4* floods are still fully below the threshold *S*. As the flood ranking of the HISTRHÔNE database is based on observed damages, it is therefore not possible to have a direct match between *C4* floods and an exact discharge threshold. We refine the prior distribution $N(9000; 500)$, with a standard deviation of 500 m^3/s (instead of 2000 m^3/s with *model D*). No C4 flood is fully below the 95% prior interval [8000; 10 000].

Considering the thirteen *C4* floods of the HISTRHÔNE database (1500-1815) and the fourteen floods higher than a threshold $S = 9000$ m^3/s 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 the 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 the knowledge of the starting date of the surveying period (t_{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 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 *Q100* and *Q1000* floods (Figure 6) and parameters (ξ, *S*, t^*) (Table 3).

Table 3: Maxpost estimation ± posterior standard deviation expressed in percentage for *Q100* and *Q1000* floods, and (ξ, *S*, 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 (m^3/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 (m^3/s)	/	/	9628 ± 5%	/	9613 ± 6%	9386 ± 4%
	t^*	/	/	/	1527 ± 3%	1529 ± 4%	1526 ± 1%

The results with a mixed sample on the 1500-2020 period show that the uncertainty on *Q100* and *Q1000* floods (Figure 6a) is lower than with AMAX values on the 1816-2020 period for models assuming a known perception threshold (*models A* and *C*). For these two *models A* and *C*, the *maxpost* quantiles are also slightly lower (by around 5%) than with AMAX values on the

1816-2020 period (Figure 6a). In the same way as with subsamples on section 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^* (Figure 6b).

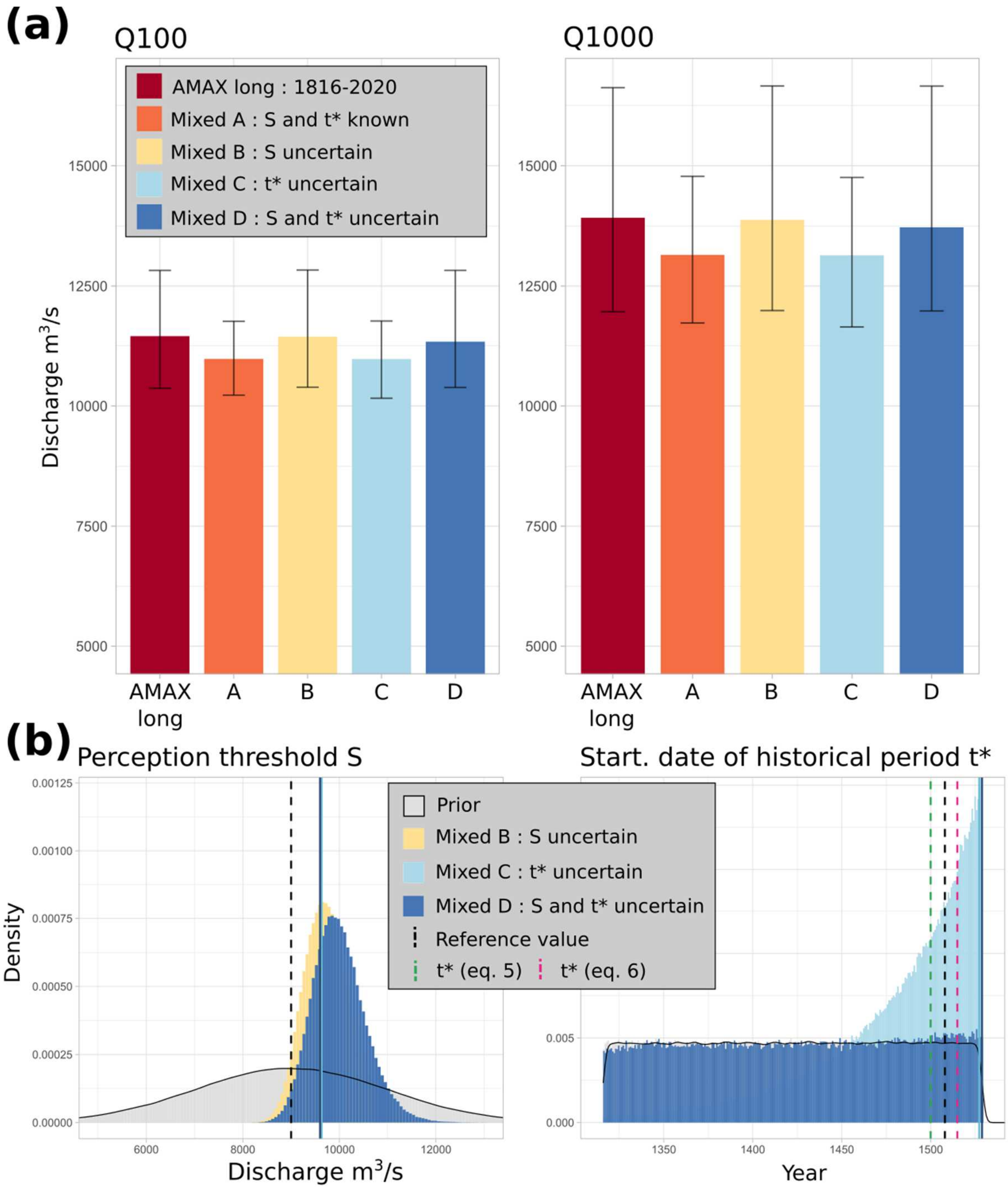

Figure 6: (a) *Q100* and *Q1000* floods with 95% credibility intervals displayed as error bars. *AMAX long* refers to the annual maximum sample on the 1816-2020 period; *Mixed A-B-C-D* refer to a mixed sample ("historical" floods on the 1500-1815 period and AMAX for 1816-2020) for various statistical models. **(b)** Posterior distribution of: **(left)** the perception threshold S; **(right)** the starting date t^* of the historical period (1500-2020 period). The solid vertical lines represent the parameter maxpost estimates for each model and the black dashed lines represent the reference values (S = 9000 m^3/s and t^* = 1500). The green and pink dashed vertical lines **(right)** represent the estimates of t^* by equations (5) and (6).

The posterior standard deviations for the perception threshold (*models B* and *D*) are relatively small (around 500 m^3/s for both models) and the distributions are lying mostly above the prior value of 9000 m^3/s (*maxpost* values around 9600 m^3/s, 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) of the HISTRHÔNE database, despite the fact that the stationarity hypothesis of the Poisson test over the 1500-2020 period was not rejected (Figure 1b). Once again, we can compare the rate of occurrence of floods above the threshold S = 9000 m^3/s 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 Figure 7a. It can be seen that the uncertainty of the quantiles is smaller by about 15% compared to the reference for *Q100* and *Q1000*. *Maxpost* estimates are also reduced by approximately 3% for both return periods. The use of historical floods therefore appears relevant to reduce the uncertainty of 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 *Q1000* by about 25% (comparison of model D with vague priors on S and t^*, and model $D*$ with refined priors).

The posterior distributions of S and t^* are shown in Figure 7b. Once again, the posterior distribution of the perception threshold is shifted towards values higher than the assumed value of 9000 m^3/s, with a *maxpost* threshold at 9386 m^3/s. 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. The *maxpost* estimate of t^* is here 1526, i.e. a length of the historical period 26 years shorter than expected. Therefore, a doubt remains as to the completeness of the historical sample or the inter-sample stationarity as described in the previous section.

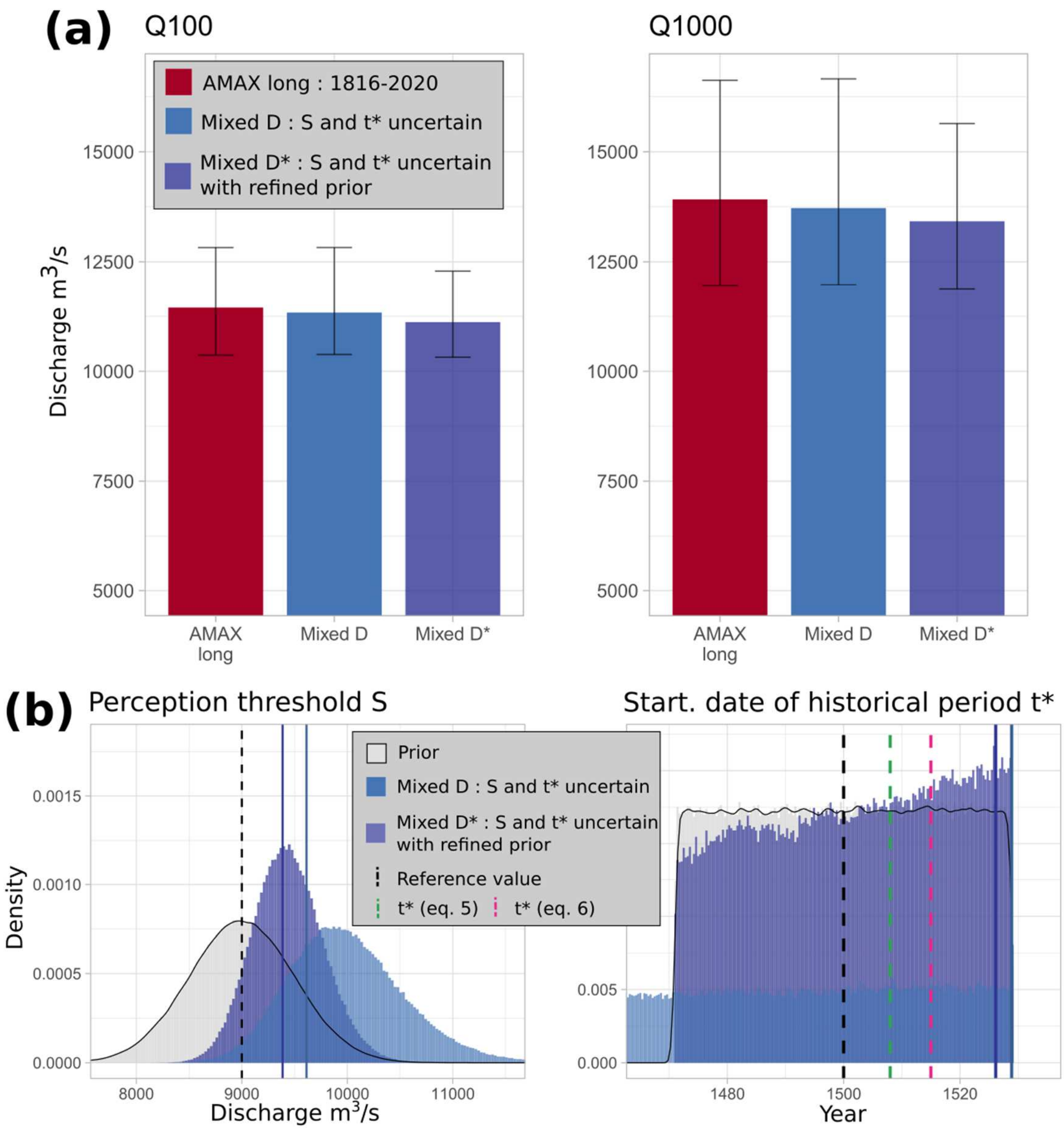

Figure 7: (a) *Q100* and *Q1000* floods with 95% credibility intervals displayed as error bars. *AMAX long* refers to an annual maximum sample on the 1816-2020 period; *Mixed D** refers to a mixed sample ("historical" floods on the 1500-1815 period and AMAX for 1816-2020), with refined priors on S and t^*. **(b)** Posterior distributions of **(left)** the perception threshold S; **(right)** the starting date t^* of the historical period for the two mixed *models D* and *D**. The solid vertical lines represent the maxpost estimate of the parameter for each of the models and the black dashed lines represent the reference values (threshold S = 9000 m^3/s; starting date t^* = 1500).

6. Discussion

6.1 Main findings on the 1816-2020 period

By using the probabilistic models described in section 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 threshold 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 m^3/s), the uncertainty resulting from determining the threshold has a strong impact on quantiles 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 of 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 on the 1816-2020 period suggest that the quantiles 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.

6.2 Main findings on the 1500-2020 period

Application of the Binomial *model D* to a mixed sample with discharge estimate of AMAX values on 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 on the 1816-2020 period is really long (205 years), it is possible to reduce the uncertainty of the flood quantiles (Figure 7a) by adding information of 13 exceedances of a threshold $S = 9000$ m^3/s during three prior centuries (period 1500-1815) and prior knowledge about *S* and t^*;

- The refinement of the prior distributions on the threshold *S* and the starting date t^*, with *model D**, gives a more precise assessment of flood quantiles than with *model D* (Figure 7a). Posterior standard deviation (expressed in %) of *Q1000* quantile decreases from 11% to 8% (Table 3). In both cases, considering the perception threshold *S* as being uncertain has much more impact on the uncertainty of the results than considering a lack of knowledge about the length of the historical period;

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

- Flood distribution and 95% credibility interval of *model D** are represented in Figure 8. AMAX values are reported with their uncertainty (from 5% to 30%) and historical floods as exceedances. Information on floods during three 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 Figure 8, instead of a 400-year return period in Figure 2a for the 1816-2020 period).

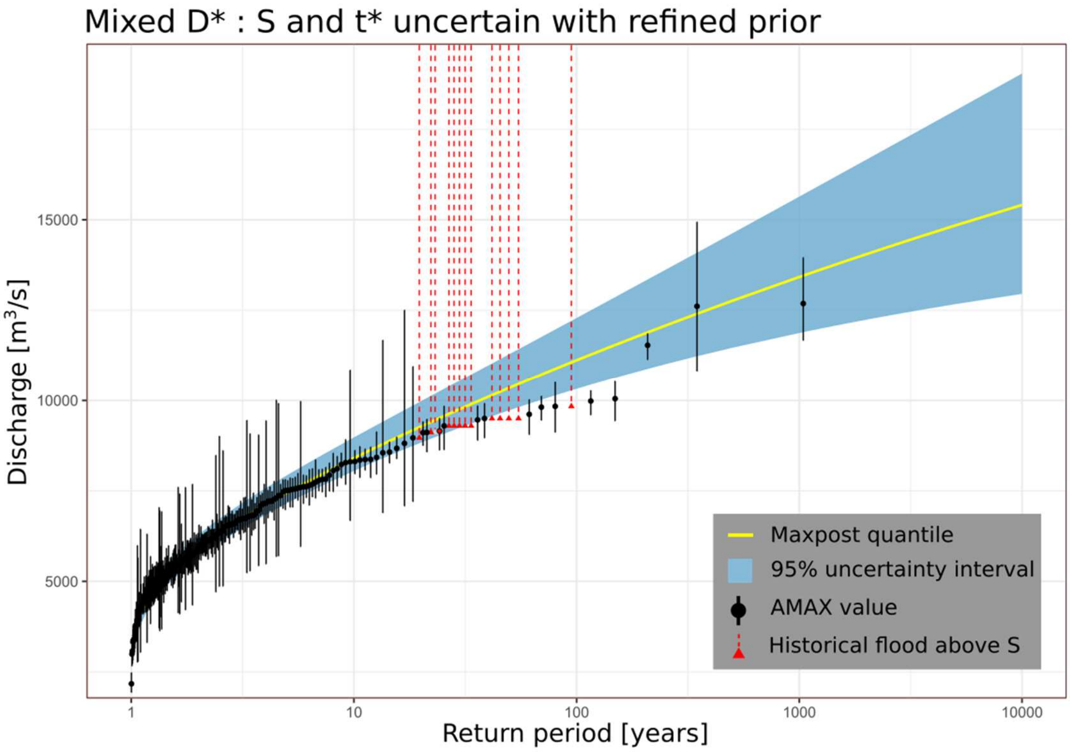

Figure 8: Flood distribution and 95% confidence interval of *model D** (Mixed sample: systematic period 1816-2020 + 13 historical exceedances on 1500-1815, refined prior on *S* and *t**). Experimental distribution in black (AMAX values) or red (exceedances of the perception threshold).

7. Conclusion

This paper proposes Binomial models for the inclusion of historical data into FFA, which 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. It has been artificially subsampled in order to mimic a historical context, considering AMAX values on 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 of flood quantiles. This also holds for the length of the historical period but to a much lesser extent. In the case of this subsample, the use of historical data makes it possible to reduce the uncertainty of the quantiles compared to the sole use of the short systematic sample (1970-2020), considering uncertainties on 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 three centuries reduces the uncertainty of *Q100* and *Q1000* 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.

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 taking into account of the impact of climate change on flood risk estimates. However, it is still possible to integrate temporal changes in climate processes or watershed characteristics within 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 out of the FFA scope, such long series remain interesting for the study on the long-term variability of floods over several centuries, and their value for risk awareness and memory.

8. Acknowledgments

The PhD fellowship of Mathieu Lucas is funded by INRAE, the Compagnie Nationale du Rhône (CNR), and EUR H2O'Lyon (ANR-17-EURE-0018) from the University of Lyon, France. This study was conducted within the Rhône Sediment Observatory (OSR), a multi-partner research program funded through the 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 HISTRHONE database from the CEREGE (Georges Pichard). Finally, we thank Neil Macdonald and Helen Hooker for their constructive comments that helped us improve the paper.

9. Appendix: Plotting position for unknown historical floods

The exceedance probability of the i^{th} value $q(i)$ of a sample ($q(1) \geq \ldots \geq q(N)$) sorted by decreasing value is:

$$p'_i = Prob[Q > q(i)] = \frac{i-a}{N+1-2a} \qquad \text{(A1)}$$

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

Hirsch (1987) proposed to split a mixed sample, formed by $(q_1, \ldots, q_{NY_C})$ AMAX values during NY_C years (continuous period) and NE_H historical values larger than a perception threshold S during NY_H years (historical period), into two sub-samples:

- NE exceedances of the threshold S on the whole period (divided into NE_H and NE_C exceedances on the historical and continuous periods), during NY years, with $NY = NY_H + NY_C$ years:

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

- floods lower than S on the continuous period:

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

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 NE_C floods larger than S during the continuous period:

- Step 1: randomly sample without replacement the rank of the NE_C floods of the continuous period within the whole period: *sample*(*x* = 1:*NE*, *size* = NE_C, *replace* = *FALSE*) (R code);
- Step 2: as we know the values of the 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 NE_H floods larger than S during the historical period.

As we assigned ranks to all exceedances of the threshold S on the whole period, we are able to compute their plotting position with equation (A2). Let denote $q_1 \geq \cdots \geq q_{NE_C} \geq S$ the known discharges of the continuous period larger than the threshold S, with their corresponding ranks $r_1 < \cdots < r_{NE_C}$. We now assign an interval to the unknown historical discharges (see Figure 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 ½ ($q_i + q_{i+1}$);
- If $r_{NE_C} < NE$, we have (NE - r_{NE_C}) historical flood discharges within the interval [S; q_{NE_C}]. They will be plotted with vertical dashed lines larger ½ ($S + q_{NE_C}$).

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

10. References

Apel, H., Thieken, A. H., Merz, B., Blöschl, G. (2004). Flood risk assessment and associated uncertainty. In : *Natural Hazards and Earth System Sciences* 4.2. Publisher: Copernicus GmbH, p. 295-308. issn : 1561-8633. doi : 10.5194/nhess- 4-295-2004.

Benito, G., Lang, M., Barriendos, M., Llasat, M. C., Francés, F., Ouarda, T., Thorndycraft, V., Enzel, Y., Bardossy, A., Cœur, D., Bobée, B. (2004). Use of Systematic, Palaeoflood and Historical Data for the Improvement of Flood Risk Estimation. Review of Scientific Methods. In : *Natural Hazards* 31.3, p. 623-643. issn : 1573-0840. doi: 10.1023/B:NHAZ.0000024895.48463.eb.

Benson, M. A. (1950). Use of historical data in flood-frequency analysis. In : *Transactions, American Geophysical Union* 31.3, p. 419. issn : 0002-8606. doi : 10.1029/TR031i003p00419.

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 et O. Wetter (2020). Current European flood-rich period exceptional compared with past 500 years". In : *Nature* 583.7817, p. 560-566. issn : 0028-0836, 1476-4687. doi : 10.1038/s41586-020-2478-3

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

Darienzo, M., Renard, B., Le Coz J., Lang M. (2021). Detection of Stage-Discharge Rating Shifts Using Gaugings : A Recursive Segmentation Procedure Accounting for Observational and Model Uncertainties. In : *Water Resources Research* 57.4. issn : 0043-1397, 1944-7973. doi : 10.1029/2020WR028607.

Dezileau, B., Terrier, L., Berger, J., Blanchemanche, P., Latapie, A., Freydier, R., Bremond, L., Paquier, A., Lang, M., Delgado, J. (2014). A multidating approach applied to historical slackwater flood deposits of the Gardon River, SE France. In : *Geomorphology* 214, p. 56-68. issn : 0169555X. doi : 10.1016/j.geomorph.2014. 03.017.

Engeland, K., Aano, A., Steffensen, I., Støren, E., Paasche, Ø. (2020). *New flood frequency estimates for the largest river in Norway based on the combination of short and long time series*. preprint. Catchment hydrology/Instruments et observation techniques. doi : 10.5194/hess-2020-269.

European Union (2007). *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, 12p.

Falconer, J.R., Frank, E., Polaschek, D.L.L., Joshi, C. (2022). Methods for Eliciting Informative Prior Distributions: A Critical Review. In : *Decision Analysis*, 2022, 19(3), p. 189–204,10.1287/DECA.2022.0451

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*. Vol. II (2nd ed.), New York: Wiley.

Gerard, R., Karpuk, E. W. (1979). Probability Analysis of Historical Flood Data. In : *Journal of the Hydraulics Division* 105.9. Publisher : American Society of Civil Engineers, p. 1153-1165. doi : 10.1061/JYCEAJ.0005273.

Giuntoli, I., Renard, B., Lang, M. (2019). Floods in France. In: *Changes in Flood Risk in Europe*, 1st edn. CRC Press, p 13, isbn : 978-0-203-09809-7.

Hall, J., B. Arheimer, M. Borga, R. Brázdil, P. Claps, A. Kiss, T. R. Kjeldsen, J. Kriaučiunienė, Z. W. Kundzewicz, M. Lang, M. C. Llasat, N. Macdonald, N. McIntyre, L. Mediero, B. Merz, R. Merz, P. Molnar, A. Montanari, C. Neuhold, J. Parajka, R. A. P. Perdigão, L. Plavcová, M. Rogger, J. L. Salinas, E. Sauquet, C. Schär, J. Szolgay, A. Viglione et G. Blöschl (2014). Understanding flood regime changes in Europe: a state-of-the-art assessment". In : *Hydrology and Earth System Sciences* 18.7, p. 2735-2772. issn : 1607-7938. doi : 10.5194/hess-18-2735-2014.

Hirsch, R. M. (1987). Probability plotting position formulas for flood records with historical information". In : *Journal of Hydrology* 96.1, p. 185-199. issn : 00221694. doi : 10.1016/0022-1694(87)90152-1.

Kendall, M. (1948). *Rank Correlation Methods*. 1. London : Charles Griffin & Company

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., Wilson D. (2014). Documentary evidence of past floods in Europe and their utility in flood frequency estimation. In : *Journal of Hydrology* 517, p. 963-973. issn : 00221694. doi : 10.1016/j.jhydrol.2014.06.038.

Kuczera, G. (1999). Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference. In : *Water Resources Research* 35.5, p. 1551-1557. issn : 1944- 7973. doi : 10.1029/1999WR900012.

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

Lang, M., Ouarda, T., Bobée T. (1999). Towards operational guidelines for over-threshold modeling. In : *Journal of Hydrology* 225.3, p. 103-117. issn : 00221694.doi : 10.1016/S0022-1694(99)00167-5.

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

Macdonald, N., Kjeldsen, T. R., Prosdocimi, I., Sangster, H. (2014). Reassessing flood frequency for the Sussex Ouse, Lewes : the inclusion of historical flood information since AD 1650. In : *Natural Hazards and Earth System Sciences* 14.10, p. 2817-2828. issn : 1684-9981. doi : 10.5194/nhess-14-2817-2014.

Macdonald, N., Sangster, H. (2017). High-magnitude flooding across Britain since AD 1750. In : *Hydrology and Earth System Sciences* 21: p. 1631–1650. issn : 1027-5606. doi : https://doi.org/10.5194/hess-21-1631-2017

Machado, M. J., Botero, B. A., López, J., Francés, F., Díez-Herrero, A., Benito, G. (2015). Flood frequency analysis of historical flood data under stationary and non-stationary modelling. In : *Hydrology and Earth System Sciences* 19.6, p. 2561- 2576. issn : 1607-7938. doi : 10.5194/hess-19-2561-2015.

Mann, H. B. (1945). Nonparametric Tests Against Trend. In : *Econometrica* 13.3.Publisher : [Wiley, Econometric Society], p. 245-259. issn : 0012-9682. doi : 10.2307/1907187.

Martins, E. S., Stedinger, J. R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. In : *Water Resources Research,* 36.3, p. 737-744. issn : 00431397. doi : 10.1029/1999WR900330.

Merz, R., Blöschl, G. (2008). Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information. *Water Resources Research*, vol. 44, W08432, doi:10.1029/2007WR006744

METS (2023). *Repères de crues, plateforme collaborative de référence pour le recensement des repères de crues en France.* url : https://www.reperesdecrues.developpementdurable.gouv.fr

Neppel, L., Renard, B., Lang, M., Ayral, P.A., Coeur, D., Gaume, E., Jacob, N., Payrastre, O., Pobanz, K., Vinet, F. (2010). Flood frequency analysis using historical data : accounting for random and systematic errors. In : *Hydrological Sciences Journal* 55.2, p. 192-208. issn: 0262-6667, 2150-3435. doi : 10.1080/02626660903546092

Parkes, B., Demeritt, D. (2016). Defining the hundred year flood : A Bayesian approach for using historic data to reduce uncertainty in flood frequency estimates. In : *Journal of Hydrology* 540, p. 1189-1208. issn : 00221694. doi : 10 . 1016 / j . jhydrol.2016.07.025.

Payrastre, O., Gaume, E., Andrieu, H. (2011). Usefulness of historical information for flood frequency analyses : Developments based on a case study. In : *Water Resources Research* 47.8. issn : 00431397. doi : 10.1029/2010WR009812.

Pettitt, A.N. (1979). A non-parametric approach to the change-point problem. *Appl. Stat.* 28 (2), 126–135

Pichard, G. (1995). Les crues sur le bas Rhône de 1500 à nos jours. Pour une histoire hydro-climatique. In : *Méditerranée* 82.3, p. 105-116. issn : 0025-8296. doi : 10.3406/medit.1995.2908.

Pichard, G., Arnaud-Fassetta, G., Moron, V., Roucaute E. (2017). Hydroclimatology of the Lower Rhône Valley : historical flood reconstruction (AD 1300–2000) based on documentary and instrumental sources. In : *Hydrological Sciences Journal* 140 62.11, p. 1772-1795. issn : 0262-6667, 2150-3435. doi : 10.1080/02626667.2017. 1349314.

Pichard, G., Roucaute, E. (2014). Sept siècles d'histoire hydroclimatique du Rhône d'Orange à la mer (1300-2000). Climat, crues, inondations. In : *Presses Universitaires de Provence* (Hors-série de la revue Méditerranée), p. 194. doi : https://doi.org/ 10.4000/geocarrefour.9491.

Piotte, O., Boura, C., Cazaubon, A., Chaléon, C., Chambon, D., Guillevic, G., Pasquet, F., Perherin, C., Raimbault ; E. (2016). Collection, storage and management of high-water marks data : praxis and recommendations. In : *E3S Web of Conferences* 7. Publisher : EDP Sciences, p. 16003. issn : 2267-1242. doi : 10.1051/ e3sconf/20160716003.

Prosdocimi, I. (2018). German tanks and historical records: the estimation of the time coverage of ungauged extreme events. In : *Stochastic Environmental Research and Risk Assessment* 32.3, p. 607-622. issn : 1436-3240, 1436-3259. doi : 10.1007/s00477- 017-1418-8.

Reis, D. S., Stedinger, J. R. (2005). Bayesian MCMC flood frequency analysis with historical information. In : *Journal of Hydrology* 313.1, p. 97-116. issn : 00221694. doi : 10.1016/j.jhydrol.2005.02.028.

Renard, B. (2023). Use of a national flood mark database to estimate flood hazard in the distant past. *Hydrol. Sci. J.* (ISSN: 0262-6667) null. http://dx.doi.org/10.1080/02626667.2023.2212165.

Renard, B., Garreta, V., Lang, M. (2006). An application of Bayesian analysis and Markov chain Monte Carlo methods to the estimation of a regional trend in annual maxima. In : *Water Resources Research* 42.12. issn : 00431397. doi : 10 . 1029 / 2005WR004591.

Salas, J. D., Obeysekera, J., Vogel R. M. (2018). Techniques for assessing water infrastructure for nonstationary extreme events : a review. In : *Hydrological Sciences Journal* 63.3, p. 325-352. issn : 0262-6667. doi : 10.1080/02626667.2018.1426858.

Shang X., Wang D., Singh V.P., Wang Y., Wu J., Liu J., Zou Y., He R. (2021). Effect of Uncertainty in Historical Data on Flood Frequency Analysis Using Bayesian Method. In: J. *Hydrol. Eng.*, 2021, 26(4): 04021011. doi 10.1061/(ASCE)HE.1943-5584.0002075

Sharma S., Ghimire G.R., Talchabhadel R., Panthi J., Lee B.S., Sun F., Baniya R., Adhikari T.R. (2022) Bayesian characterization of uncertainties surrounding fluvial flood hazard estimates. In *Hydrological Sciences Journal*, 67:2, 277-286, doi: 10.1080/02626667.2021.1999959

St. George, S., Hefner, A. M., Avila, J. (2020). Paleofloods stage a comeback. In : *Nature Geoscience* 13.12, p. 766-768. issn : 1752-0894, 1752-0908. doi : 10.1038/ s41561-020-00664-2.

Stedinger, J. R., Cohn, T. A. (1986). Flood Frequency Analysis With Historical and Paleoflood Information. In : *Water Resources Research* 22.5, p. 785-793. issn : 1944-7973. doi : 10.1029/WR022i005p00785

Viglione, A., Merz, R., Salinas, J. L., Blöschl, G. (2013). Flood frequency hydrology : 3. A Bayesian analysis. In : *Water Resources Research* 49.2, p. 675-692. issn : 1944-7973. doi : 10.1029/2011WR010782

Data and codes availability: https://github.com/MatLcs/HistoFloods