Response to reviewers

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

We thank editor and reviewers for their suggestions. Please find in the following pages a detailed response for each point highlighted by reviewers.

In orange, the reviewer's comments

In **bold**, the new formulation considered in the paper

In green, the response to the reviewer

As requested by both reviewers, all the figures have been updated. Figure resolution has been improved, color palette has changed (more visible by colourblind people), and some figures have been grouped together. Please find all the new figures and captions at the end of the document.

Neil Macdonald

L28: Add reference (Macdonald and Sangster, 2017): The reference has been added to text and bibliography

L60: Change "riparian populations" → "**populations living adjacent to the river**, etc.": formulation has been changed

L78: "testing the" → "testing **of** the" & "real-life dataset" → "real-**world** dataset": formulation has been changed

L128: "In some cases, the flood inventory starts before the date t1 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)."

→ I think this is valid if we know that the bridge replaced a previous bridge lost to a flood. However, if this is not the case, the bridge may significantly alter the flood levels/channel morphology and /or over-estimate the period of time at the start of the historical record if construction occurred much earlier than the first historical account.

There is a confusion here with the expression "flood inventory", which is not the flood events collection but the surveying period (which includes flood and no-flood observations). We clarified the sentence: "In some cases, **the historical period (including flood and no-flood information)** starts before the date t1…" and the other occurrences (L65 and L337).

I agree about the channel morphology modification, but as this morphology remains stable during the whole surveying period (e.g. the whole life of the bridge), that may not be a problem. There is no over-estimation of the period if the floods are recorded since the construction of the bridge. See a very similar example in the following paper about floods in the High Rhine basin since 1268 from Wetter et al., (2011): 10.1080/02626667.2011.583613.

"The narrative information "the bridge looked like a float on the river" (Source S2) and "people standing on the bridge washed their hands in the Rhine" (Source S3) clearly tells us that the river's water level must have reached approximately the level of the bridge. According to a very accurate drawing by Emanuel Büchel showing the Greater Basel townscape of 1759 (Source S4), it can be clearly demonstrated that the bridge and the window of the Guildhouse were on approximately the same level (Fig. 8, top left: red line), and this perfectly fits the narrative information "boats needed to be boarded through the windows of the guild house" (Source S5)." As the old bridge of Basel was built in 1225, we can easily assess that if a large flood occurred before the oldest known flood in 1268, we would have some testimonies.

L243: "We use the plotting…" → "**The plotting**…" & add "**are applied**" at the end: Formulation has been changed

L244: "Appendix" → "**The** appendix": Formulation has been changed

L400: "taking into consideration" → "**to be considered**": Formulation has been changed

L406: "additional" → "**prior**": Formulation has been changed

L415: "on"→"**in**" & remove "discharge of": Corrected

"on"→"**in**": Corrected

L417: "additional" → "**prior**" & "allows reducing" → "**reduces**": Formulation has been changed

L418: "on" → "**in**" & "with" → "**for**": Formulation has been changed

L421: "proposed" → "**proposes**": Formulation has been changed

L440: remove "Note that": Formulation has been changed

L445: "In a second part…" → "**The paper also presents**": Formulation has been changed

L446: "on" → "**for**": Formulation has been changed

L447: "allows reducing"→ "**reduces**" & "was" → "**being**": Formulation has been changed

Helen Hooker

L1: Remove "more" in the title: Title has been changed

L8: "Censored nature" → Try to use plain English to make the article easy to read.

Censoring is statistics is well defined: https://en.wikipedia.org/wiki/Censoring_(statistics)

"In statistics, censoring is a condition in which the value of a measurement or observation is only partially known."

Text is changed as: "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.**"

L8: Remove "Indeed": The word has been removed

L75: I'm intrigued to know how? The following sentences have been added to clarify this point:

"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 […] 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.

L87: Rephrase or remove "Probabilistic models with a collection of historical floods" → "Probabilistic models": The end of the title has been removed to avoid miscomprehension

L96: Align equation to the left & add a full stop: Equation layout has been changed

L140: The different models could be presented more clearly in a table: We added Table 1 and the following sentence:
"Table 1 summarizes which Binomial model accounts for uncertainty, and/or historical period length."

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

L163: "On the other hand" over-used in the paper → **"However"** other occurrences:

L165: "On the other hand" → Removed

L290: "On the other hand" → **"In contrast"**

L307: "On the other hand" → **"However"**

L390: "On the other hand" → **"Furthermore"**

L196: "After the building of the Vallabrègues Dam in 1967, the station was moved 2 km downstream"; I assume the dam did not impact the river discharge? Might be worthwhile clarifying this.

"After the building of the Vallabrègues Dam in 1967, the station was moved 2 km downstream and is still in the same place today. There is no tributary between the previous and the current station." →

This point has been developed in the article. See the following paragraph. You can also find a more detailed explanation in Lucas et al. (2023), cited in the next sentence of the article.

"**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.**"

L246: Change to a colourblind friendly palette

L249: Improve plot resolution

All the figures have been revised and the colour palette has been changed

L260: "AMAX long" → "**Amax short**": Formulation has been changed

L263: "details" → "**detail**": Formulation has been changed

L267: "A poor" → "**Poor**": Formulation has been changed

L269: Remove full stop: Removed

L269: Correct figure caption: Caption has been modified

L282: Remove "simply": Removed

L286: "a lesser" → "**less**": Formulation has been changed

L287: "a poor" → "**poor**": Formulation has been changed

L300: I'm not sure these are fairly high correlations. There is limited discussion of this Figure, so I would suggest this could be removed: We removed the figure and clarified the text about correlations. Figure numeration has also been modified.

"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."

L303: "and" non italics: Police changed

"number k of times" → "**number of times k**": Formulation has been changed

L314: Keep this figure!: Ok!

L315: It would be helpful to remind the reader and overview the aims of Section 5 here → This paragraph has been added at the beginning of the section:

"**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 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.**"

L320: add commas: Commas have been added

L321: "arbitrarly" → "**arbitrarily**": Formulation has been changed

L348-49: "poorly knowing" → "**poor knowledge of**": Formulation has been changed

L352: Correct figure caption: Caption has been changed

L371: Correct figure caption: Caption has been changed

L375: "the elicitation of more informative priors" → information from the priors?

By eliciting more informative prior, we meant determining more precise priors based on expert knowledge or other data sources. See for example: https://arxiv.org/pdf/2112.07090

Text has also been modified as follows:

"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**)."

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

L386: "poorly knowing" → "**limited knowledge of**": Formulation has been changed

L387-8: remove "a": Word has been removed

L388: "for" in italic: Police has been changed

L389: "the determination of the" → "**determining the**": Formulation has been changed

L397: Remove "an" and add "**allow us**": Formulation has been changed

L421: "explicitely" → "**explicitly**": Formulation has been changed

L440: Rephrase "The Rhone River series analysed here has the particularity of leading to a positive shape parameter, corresponding with the parameterization used in this paper to an upper-bounded GEV distribution." → "**The shape parameter estimated at Beaucaire is positive, corresponding to an upper-bounded GEV distribution (with the parametrization used in this paper).**": The sentence has been clarified

L454: Any further recommendations for FFA using historical observations in practice? How is knowledge of uncertainties useful? Some relation back to the wider flood risk picture would round off the article nicely.

The last paragraph of the conclusion has been extended in the following way:

"Although the stationarity of the data has been checked, it is likely that the long series used in this paper is impacted by ~~climatic variability and/or~~ the imperfect completeness of the historical sample, which is based on damage perception~~, which could weaken the stationarity hypothesis necessary for FFA~~. **Indeed, the damage perception has probably evolved throughout the last five centuries at Beaucaire. Directly linking the consequences of a flood to its peak discharge is risky, as the**

evolution of physical (levee failure, duration of flood...) or anthropic factors (population density, flood control policy, mediatic or political context...) could impact the stationarity and the availability of data. Therefore, it seems important to keep this in mind when using historical data, particularly during data collection. Using the whole set of available data is not always the best solution, as the exhaustiveness of data must be the first criterion. Thus, as demonstrated in this article, it is essential to carry out a complete assessment of the various sources of uncertainty in order to decide to what extent the addition of historical information is useful to improve the estimation of flood risk.

Stationarity hypothesis may also be affected 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 ~~is~~ increasingly described in the literature (see Salas *et al.*, 2018, for an overview). **It is also important to note that beyond the FFA scope, s**uch long series remain interesting for the study on the long-term variability of floods over several centuries, and they are of great value for risk awareness and memory."

L502: We added acknowledgment to the two referees.

"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.** »
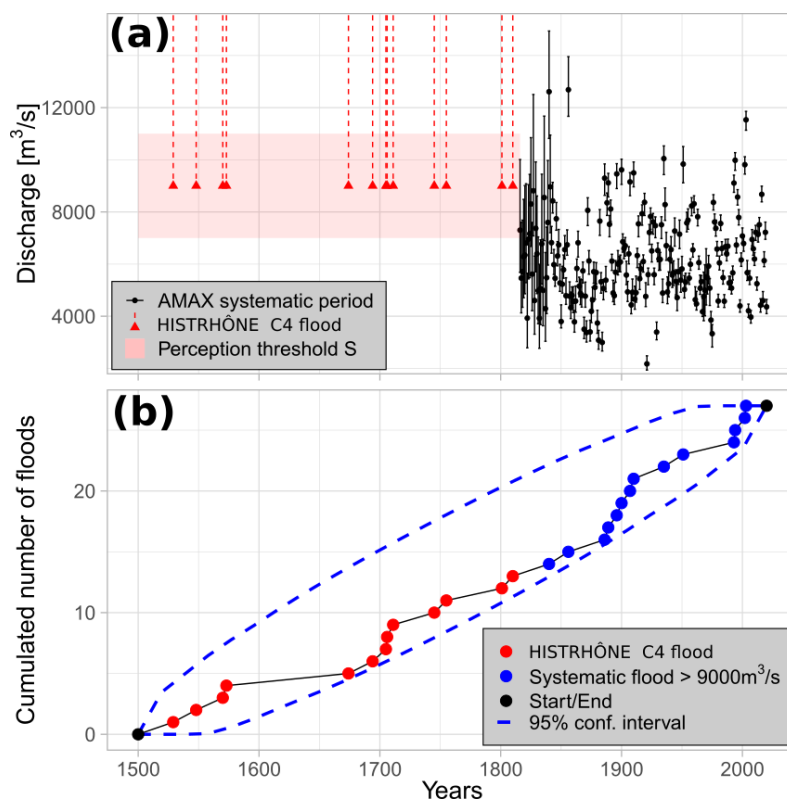
Updated figures

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
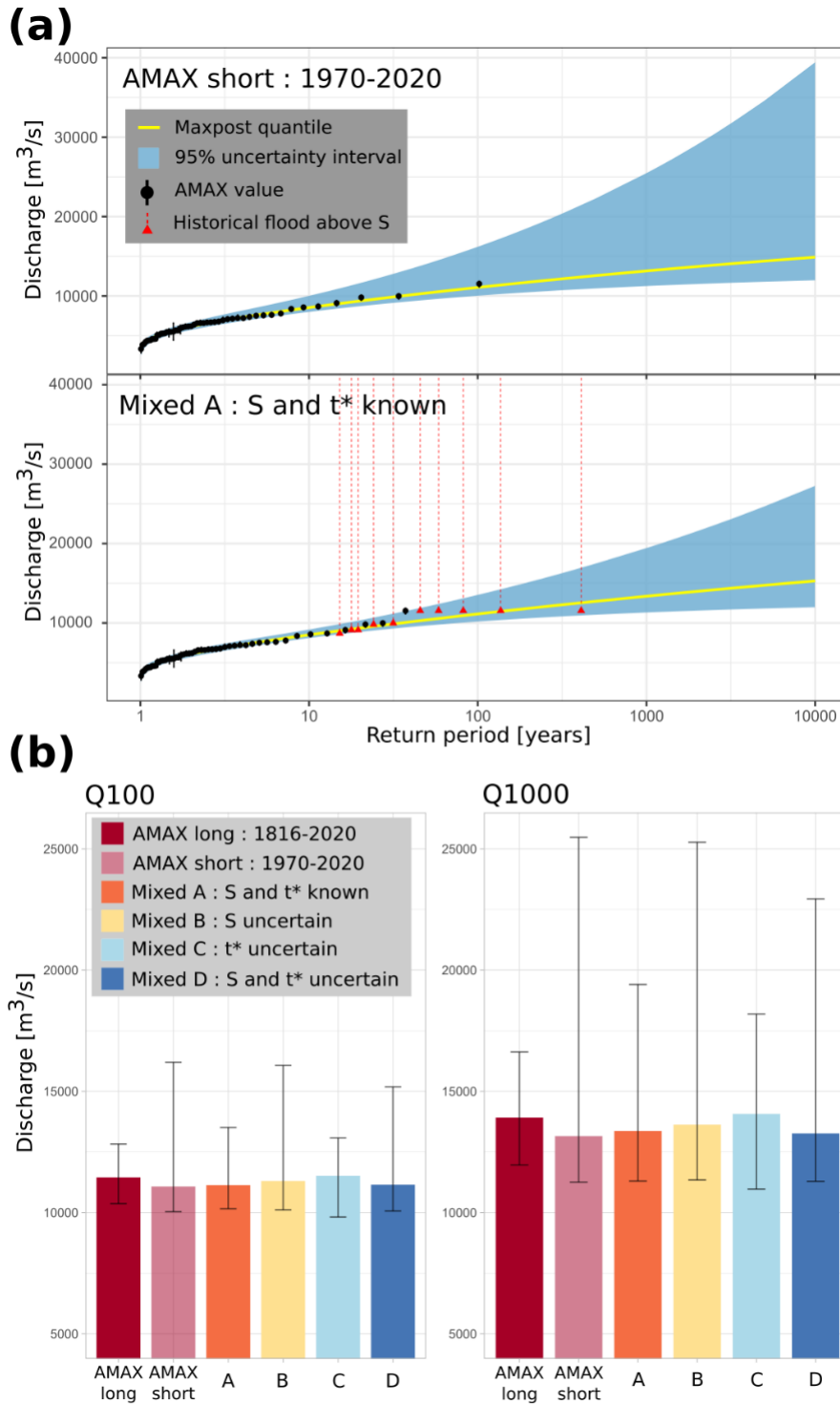
Figure 2: (a): GEV quantiles with 95% credibility intervals, example for two different models and datasets; GEV model on AMAX values (AMAX short 1970-2020) and binomial Model A on mixed sample. (b) Q100 and Q1000 floods with 95% credibility intervals displayed as error bars. AMAX long refers to the sample 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.
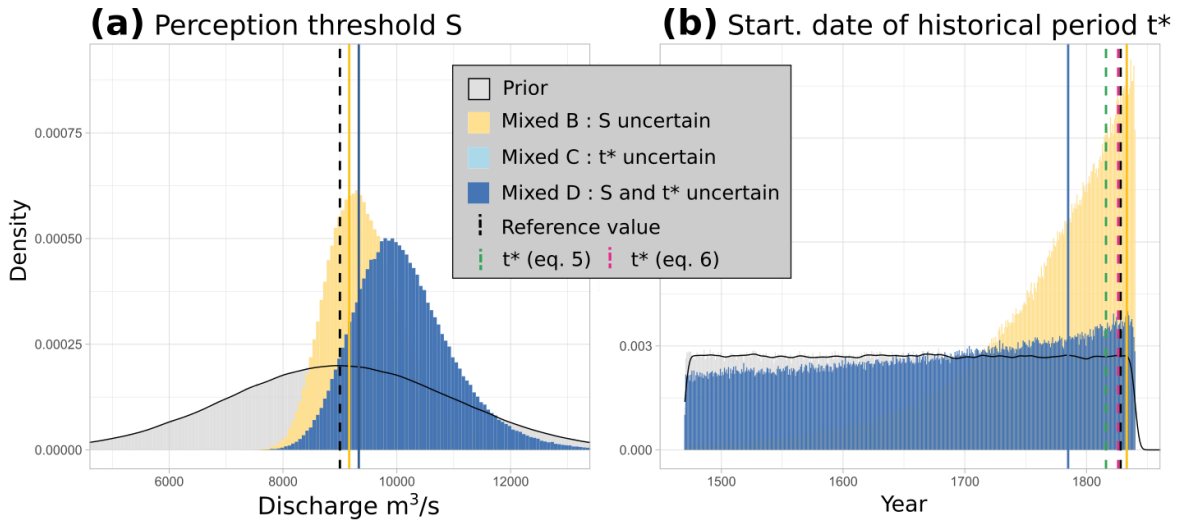
Figure 3: Prior and posterior distributions of: (a) the perception threshold S; (b) the starting date t∗ of the historical period (1816-2020 period). The solid vertical lines represent the maxpost estimate of the parameter for each of the models, and the black dashed lines represent the reference values (S = 9000 m3/s and t* = 1816). The green and pink dashed vertical lines (b) represent the estimates of t∗ by equations (5) and (6).
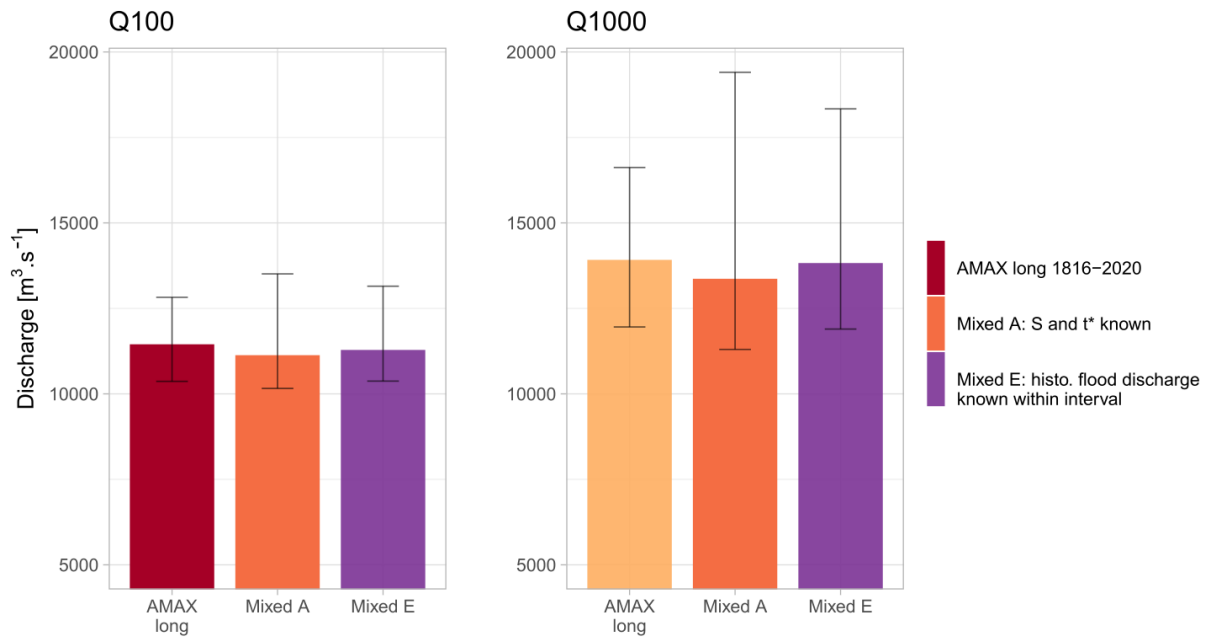

Figure 4: Q100 and Q1000 floods with 95% credibility intervals displayed as error bars. AMAX long refers to an annual maximum sample on the 1816-2020 period; Mixed refers to a mixed sample ("historical" floods on the 1816-1969 period and AMAX 1970-2020). Model A uses only the number of times the perception 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).
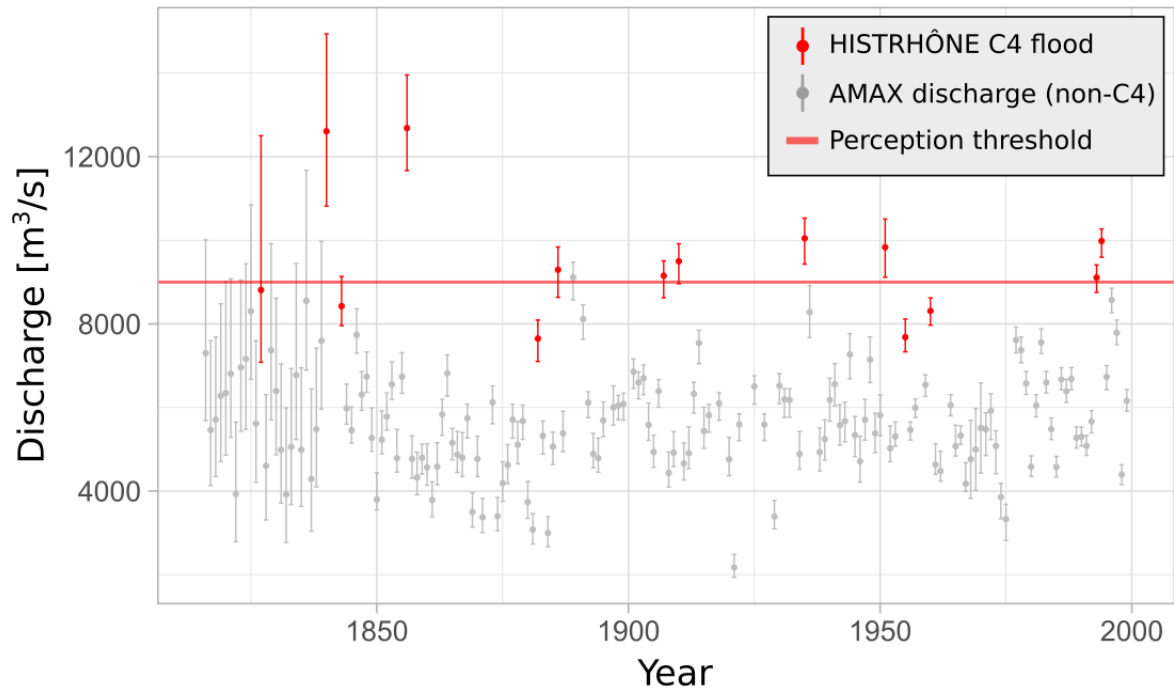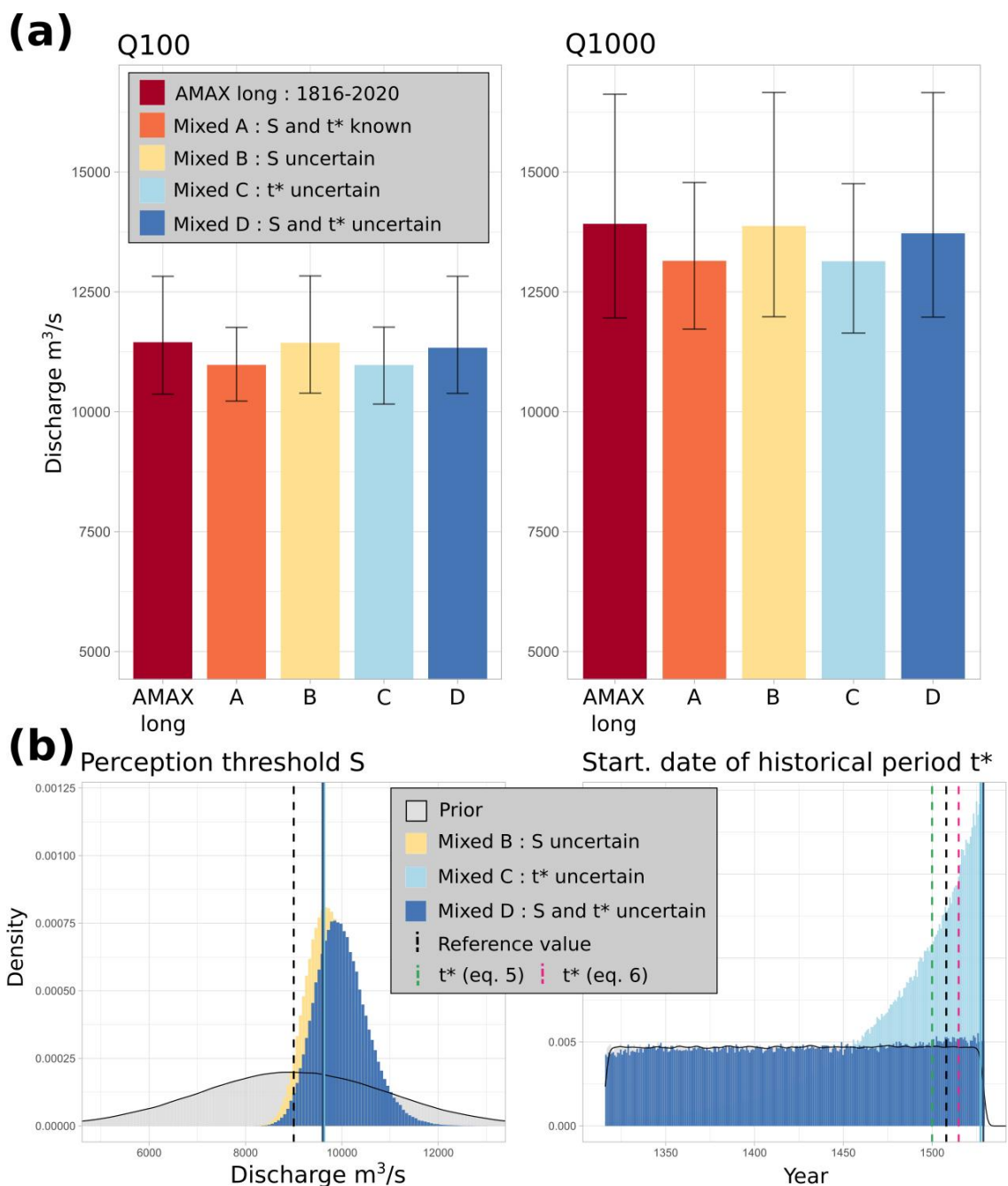
Figure 5: AMAX flood discharges (1816-2000) from Lucas et al. (2023) (grey) cross-referenced with C4 floods from HISTRHÔNE database (red). The horizontal line corresponds to the estimated perception threshold S = 9000 m³/s.

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 m³/s and t* = 1500). The green and pink dashed vertical lines (right) represent the estimates of $t*$ by equations (5) and (6).
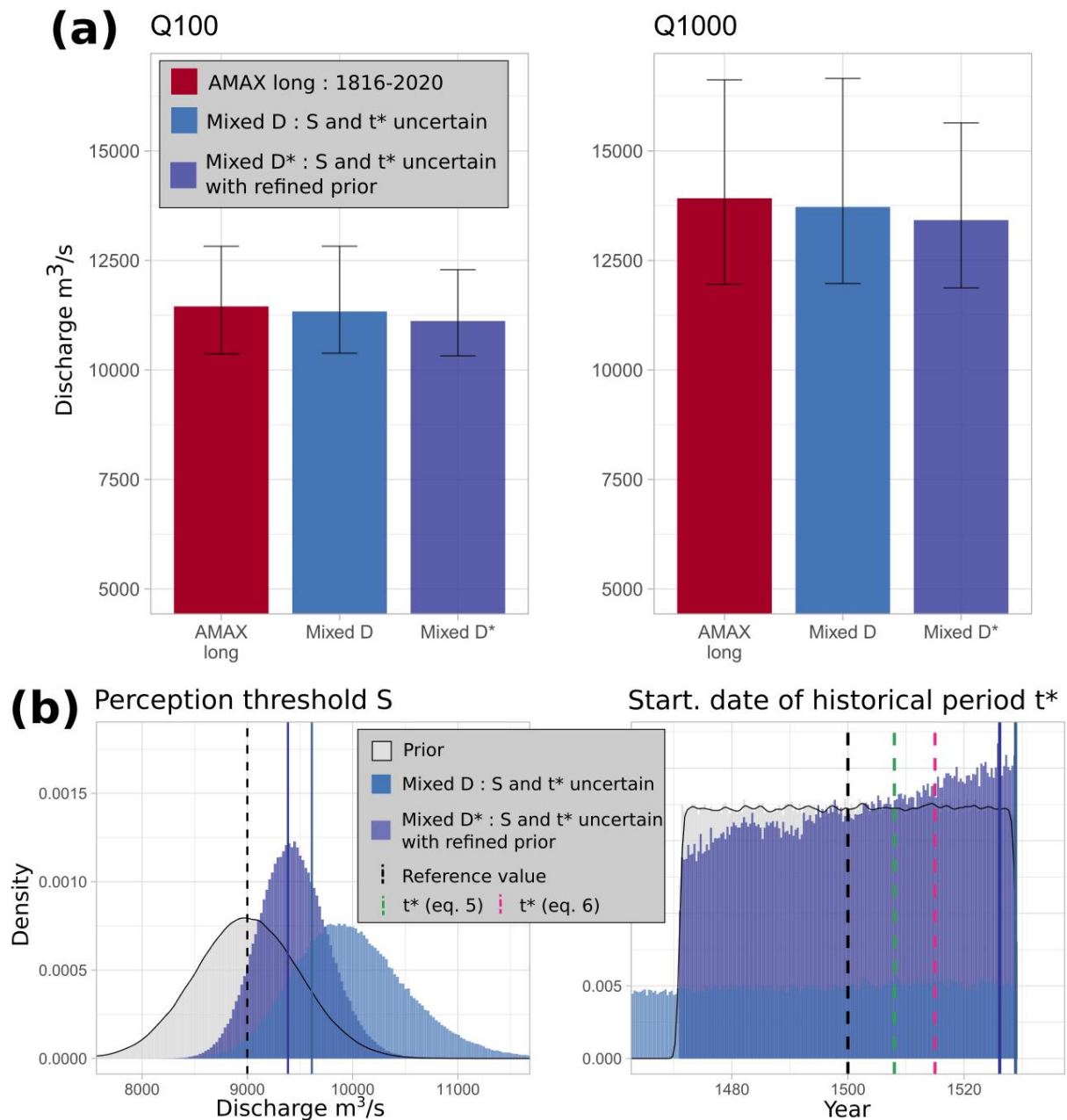
Figure 7: (a) Q100 and Q1000 floods with 95% credibility intervals displayed as error bars. AMAX long refers to an annual maximum sample on the 1816-2020 period; Mixed D* refers to a mixed sample ("historical" floods on the 1500-1815 period and AMAX for 1816-2020), with refined priors on S and t*. (b) 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 m3/s; starting date t* = 1500).
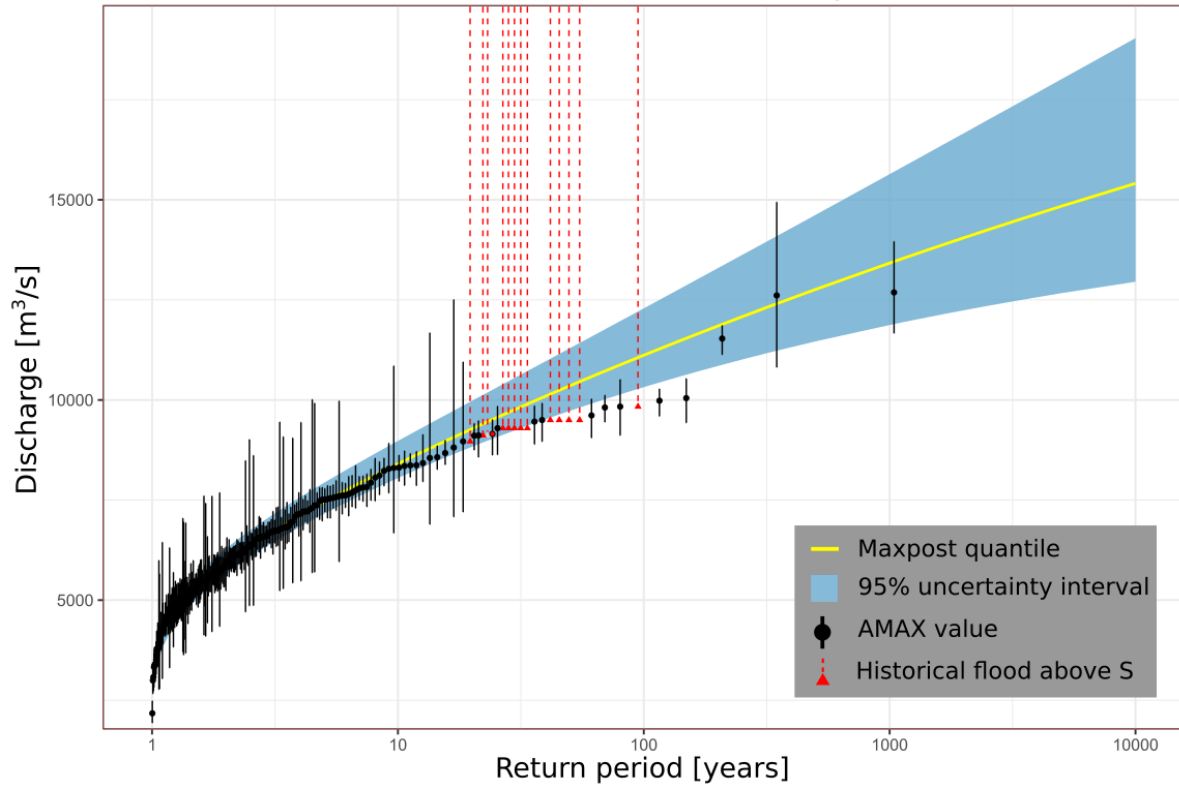
Figure 8: Flood distribution and 95% confidence interval of model D* (mixed sample: systematic period 1816-2020 + 13 historical exceedances on 1500-1815, refined prior on S and t*). Experimental distribution in black (AMAX values) or red (exceedances of the perception threshold)