Use of a National Flood Mark Database to Estimate Flood Probabilities in France from 1705 to 2015

Benjamin Renard

Use of a National Flood Mark Database to Estimate Flood Probabilities in France from 1705 to 2015

Benjamin Renard

INRAE, Aix Marseille Univ, RECOVER, Aix-En-Provence, France

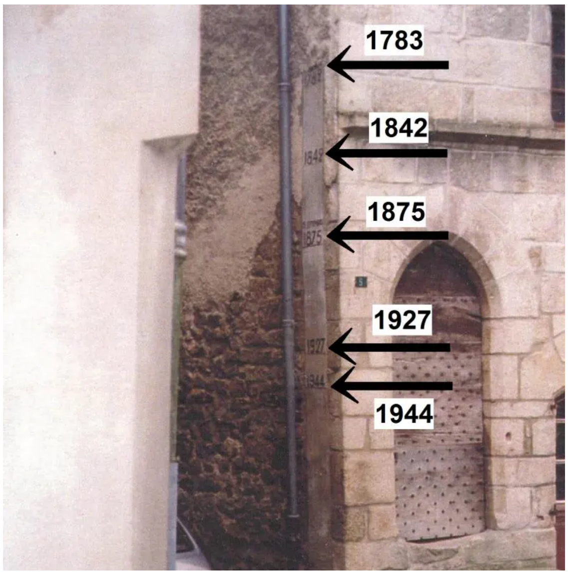
PRESENTED AT:

AGU23 WIDE. OPEN. SCIENCE.

1. INTRODUCTION

The national-scale repères de crues (https://www.reperesdecrues.developpement-durable.gouv.fr) ("flood marks") database contains thousands of geolocalized historical flood marks spanning several centuries.

The Rhône River at La Roche-De-Glun. Photo by SPC Grand Delta.



The Lot River at Entraygues-sur-Truyere. Photo by Klaus Maronna.

Flood Marks may be used to perform Historical Flood Frequency Analysis (e.g. Stedinger and Cohn, 1986 (https://doi.org/10.1029/WR022i005p00785)). This requires pairing the flood mark site with a nearby hydrometric station, and estimating flood discharges from mark elevations. This estimation is site-specific and can hardly be automated on thousands of flood mark sites.

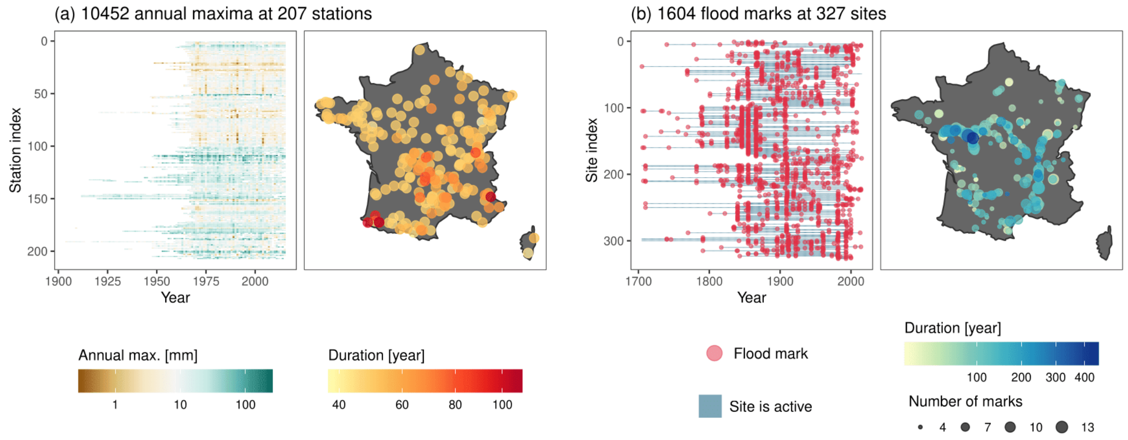
How to use this flood mark dataset along with streamflow series from the regular hydrometric network to better understand flood hazard in the distant past at the national scale?

2. DATA

Stations and sites

(*a*) 207 hydrometric **stations** forming the French reference hydrological network (≥ 40 years, undisturbed catchments with areas mostly between 20 and 2000 km^2).

(*b*) 327 flood mark **sites** with at least 4 flood marks.



Analyzed data

At stations: annual maximum flood series (1904-2015)

At sites: flood occurrence series (1705-2015)

1. define the site's activity period as the time between the first and the last marks (included)
2. years with marks within the activity period are interpreted as flood occurrences (1)
3. years with no mark within the activity period are interpreted as flood non-occurrences (0)
4. no data outside of the activity period (NA)

Note that assumptions 1 and 3 are fairly strong (some flood events are likely to be "missed"). Sensitivity analyses will be carried out.

3. MODEL

Notation

$Q(s,t)$ denotes the annual maximum flood at station s and year t.

$O(r,t)$ denotes the occurrence of a flood mark at site r for year t.

Note that $s \neq r$ (the networks of hydrometric stations and flood mark sites are distinct)

Probabilistic model

Joint model for both variables.

Annual maxima: Generalized Extreme Value (GEV) distribution

Occurrences: Benoulli distribution (B)

Both distributions are varying in space and time (eq. 1):

$$\begin{cases} Q(s,t) \sim GEV\left(\mu(s,t), \sigma(s), \xi(s)\right) \\ O(r,t) \sim \mathcal{B}\left(\lambda(r,t)\right) \end{cases}$$

The temporal variability is induced as follows (eq. 2):

$$\begin{cases} \mu(s,t) = \mu_0(s)\left(1 + \mu_1(s)\boxed{\tau_1(t)} + \ldots + \mu_K(s)\boxed{\tau_K(t)}\right) \\ logit\left(\lambda(r,t)\right) = \lambda_0(r) + \lambda_1(r)\boxed{\tau_1(t)} + \ldots + \lambda_K(r)\boxed{\tau_K(t)} \end{cases}$$

- μ_0 and λ_0 are the "normal" parameter values.
- these "normal" values are modulated by time-varying climate indices τ_k.
- Instead of using standard climate indices (NAO, IPO etc.), we treat τ_k as unknown parameters that need to be estimated. τ_k are hence called **Hidden Climate Indices** (HCIs).
- μ_k is the effect of HCI τ_k at stations, λ_k is the effect of τ_k at sites.

Importantly, the HCIs τ_k are **the same** in both equations: they hence control the temporal variability of **both** food peaks and flood marks. This assumption of common HCIs is of prime importance since it enables the transfer of information from flood marks at sites to flood peaks at stations.

Each quantity varying in space (μ_k, σ, ξ, λ_k) is modelled with a Gaussian spatial process (constant mean, exponential variogram).

Finally, all data are assumed independent **conditionnally on** the HCIs and their effects. This does **not** mean that data are assumed independent, but rather that dependence (in space, time or between variables) is induced by the HCIs and their effects.

Estimation

The assumptions described above result in a Bayesian hierarchical model with many unknowns (μ_k, σ, ξ, λ_k, τ_k + hyperparameters of Gaussian spatial processes). Estimation is performed with a dedicated MCMC sampler. More details in Renard and Thyer (2019) (https://doi.org/10.1029/2019WR024951) and Renard et al. (2021) (https://doi.org/10.1029/2021WR030007).

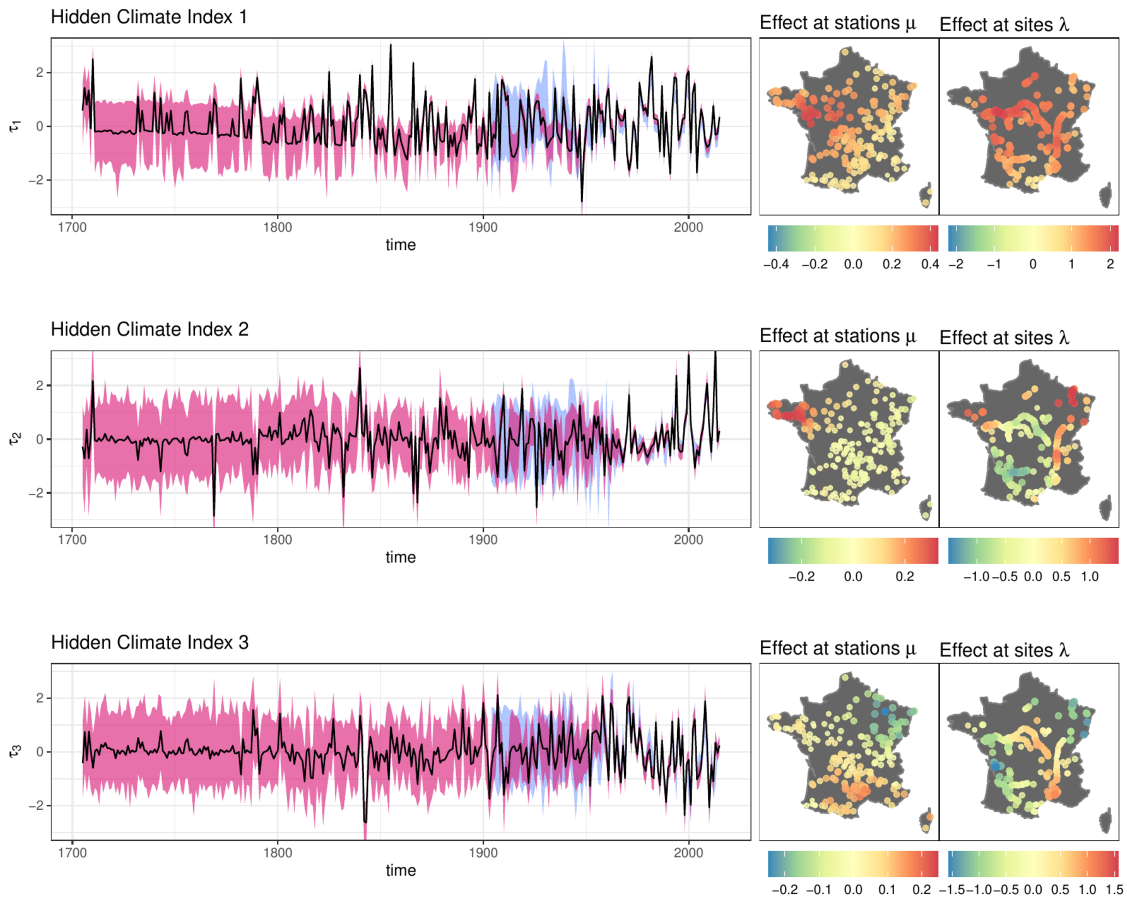
4. RESULTS

Estimated HCIs and their effects

The first 3 HCI time series (out of 6) and their spatial effects are shown below and can be interpreted as follows:

- when the HCI is high, stations / sites in red tend to have larger flood peaks / more frequent flood marks; stations / sites in blue tend to have smaller flood peaks / less frequent flood marks; stations / sites in yellow "don't care".
- opposite when the HCI is low.
- HCI ≈ 0 means "normal" conditions.

Note that HCI time series start in 1705, but uncertainty grows as one goes back in time (due to decreasing data availability).

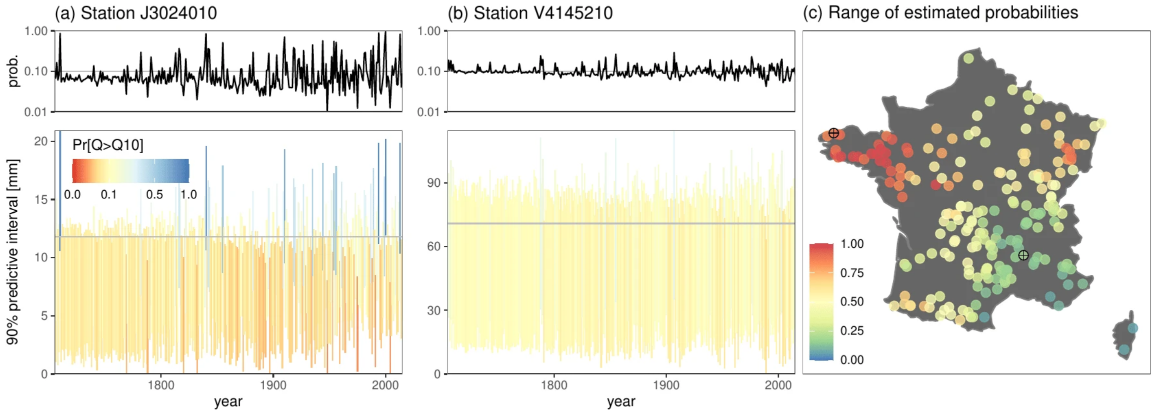


Time-varying GEV since 1705

Having estimated HCIs, equations 1-2 allow estimating the time-varying GEV at stations since 1705, i.e. before the stations even existed!

Example at two stations (*a*-*b*): 90% probability intervals from the GEV, and probability of exceeding a 10-year threshold.

Sharpness strongly varies (*c*): good in Northwestern France (station *a*), poor in the Mediterranean (station *b*).

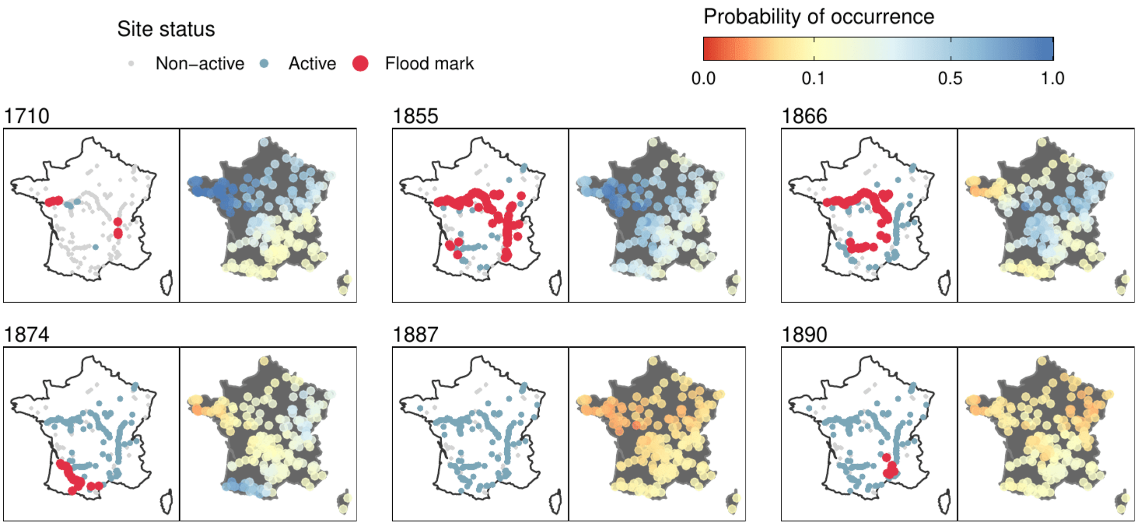


Reconstructed flood probability maps

The probability of exceeding a 10-year threshold on any given year can be mapped as shown below:

- the map on the left shows active sites and presence / absence of flood marks.
- the map on the right shows how this transfers to the probability of exceeding the threshold at hydrometric stations.

For the full movie, see panel 5.



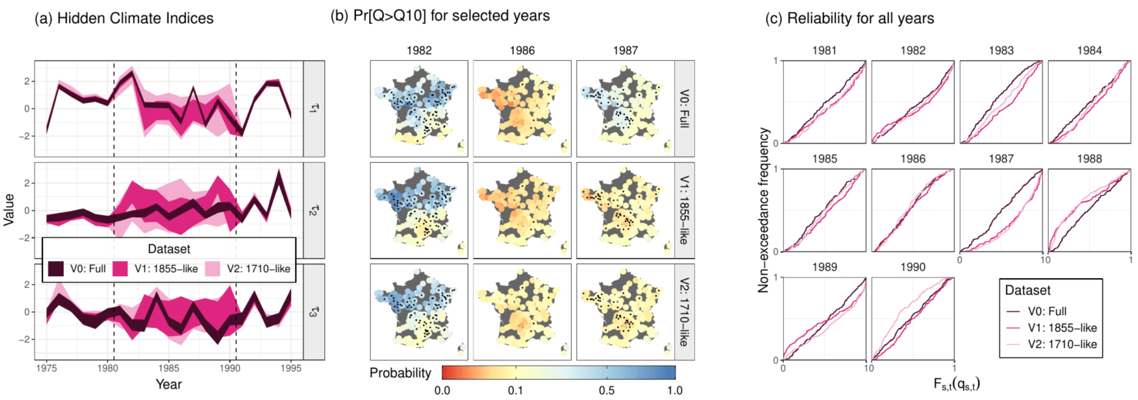
Can we trust these reconstructions?

A cross-validation exercise has been performed as follows:

- During the decade 1981-1990, all flood peak data at hydrometric stations are removed.
- During the same decade, only 138 active flood mark sites (1855-like) or only 10 sites (1710-like) are used.
- The model and resulting probabilities are re-estimated on these reduced datasets.

The key results are :

- HCI time series estimated with these reduced datasets are consistent with the original ones, albeit much more uncertain (*a*).
- Flood-rich year 1982 is well-identified even with few active sites, but 1987 is somehow "missed" (*b*).
- Overall reliability diagrams remain acceptable (*c*).



Sensitivity analyses have also been performed to assess the influence of strong assumptions regarding the interpretation of flood marks (points 1 and 3 in panel 2):

- Site-specific parameters λ_k are highly sensitive to these assumptions; station-specific parameters μ_k are completely insensitive.
- The HCIs τ_k are weakly sensitive to the definition of the activity period or to randomly missed flood events; they are, however, sensitive to systematically-missed events (e.g. a period of time where **all** floods would be missed for some historical reasons).
- Since reconstructed probabilities only depend on μ_k and τ_k (eq. 2), they are also weakly sensitive.

5. BE KIND, REWIND!

Reconstructed flood probabilities, 1705-2015

[VIDEO] https://player.vimeo.com/video/815008124?app_id=122963&byline=0&badge=0&portrait=0&title=0

6. CONCLUSION

A national flood mark database could be used to reconstruct flood probabilities from 1705 to 2015, providing information on the intensity that historical floods might have reached at hydrometric stations, well before these stations even existed.

Several improvements are possible:

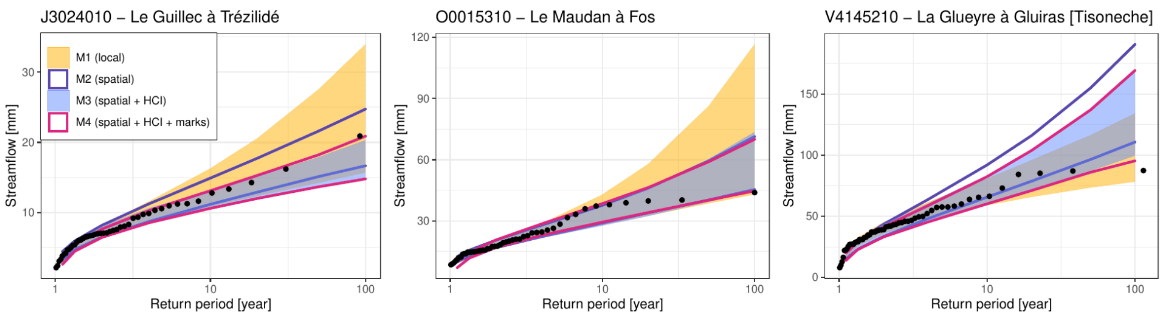
- Sharpness is weak in the Mediterranean area: is it due to the small density of flood marks in the area or to the specifities of highly-localized Mediterranean floods?
- The repères de crues (https://www.reperesdecrues.developpement-durable.gouv.fr) database is a collaborative citizen science project and is hence still growing.
- Can marks' elevation be used when available?
- How to use long climatic reconstructions such as the Twentieth Century Reanalysis (https://psl.noaa.gov/data/20thC_Rean/) (1836-2015) as an additional source of information to reconstruct flood hazard over several centuries?

And what about Flood Frequency Analysis (FFA)?

Flood marks and other paleo-historical information have traditionnaly been used to improve FFA (in particular, reduce estimation uncertainty). The marginal FFA distribution can be derived at each site using the model proposed in this poster, by "integrating out" HCIs.

Do flood marks improve FFA with the proposed model? In short, ***no***:

- models with or without flood marks (M3 and M4) lead to almost identical FFA distributions.
- they are similar to the FFA distribution obtained with a standard regional approach (M2).
- a purely local FFA (M1) leads to a much more uncertain distribution.



This inability of flood marks to have an impact on FFA is not a general result, and only holds ***with the specific model used here***. It is a consequence of keeping the site network and the hydrometric station network separated, with no attempt at pairing sites and stations. This could be improved in future work.

Full paper

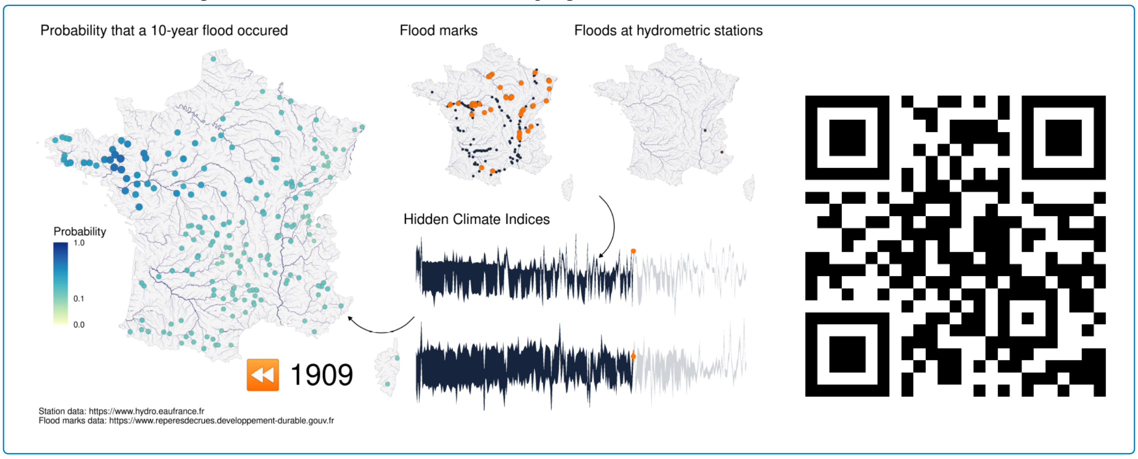
Benjamin Renard (2023). Use of a national flood mark database to estimate flood hazard in the distant past. *Hydrological Sciences Journal*, DOI: 10.1080/02626667.2023.2212165 (https://doi.org/10.1080/02626667.2023.2212165)

TRANSCRIPT

ABSTRACT

The national-scale 'repères de crues' database ('flood marks', www.reperesdecrues.developpement-durable.gouv.fr (https://www.reperesdecrues.developpement-durable.gouv.fr/)) contains thousands of geolocalized historical flood marks spanning several centuries. This constitutes a valuable source of information complementing data from the network of hydrometric stations in order to characterize ancient flood events and, possibly, to improve flood hazard assessment. However, this information is not obvious to use, in particular because the network of flood mark sites is distinct from the network of hydrometric stations. The aim of this work is to address this issue by means of a probabilistic model jointly describing the occurrence of flood marks at sites and annual maximum flood peaks at stations. A Bernoulli distribution is used for the former and a Generalized Extreme Value distribution for the latter. Some parameters of these distributions are allowed to vary in time following a set of unobserved latent time series called Hidden Climate Indices (HCI). Importantly, the same HCIs drive the temporal evolution of both flood marks and peaks: this allows transferring information between the two variables despite them being measured on distinct networks.

The model is applied to about 300 flood mark sites (1707-2015) and 200 hydrometric stations (1904-2015) in France. Results illustrate how the information carried by flood marks at sites can be used to estimate the time-varying probability of exceeding some high discharge threshold at stations during the whole period 1707-2015, which largely predates the existence of stations. In general, these estimations are sharp in the Oceanic part of the country (probabilities may approach 0 or 1), much less in the Mediterranean area (probabilities remain near-constant). Probability maps can also be derived and provide a quantitative information on the extent and spatial structure of ancient flood events. The model also allows deriving the marginal distribution used in Flood Frequency Analysis, but results indicate that the model does not make an efficient use of flood marks to improve the estimation of this distribution. Overall, this work illustrates the interest of deriving flexible probabilistic models to make the best use of several existing sources of information and data while adapting to their varied characteristics.



(https://agu.confex.com/data/abstract/agu/fm23/6/1/Paper_1308216_abstract_1145206_0.png)