

Are historical stage records useful to decrease the uncertainty of flood frequency analysis ? A 200-year long case study

Mathieu LUCAS¹, Benjamin RENARD¹, Jérôme LE COZ¹, Michel LANG¹, Antoine BARD², and Gilles PIERREFEU³

¹INRAE Riverly, Villeurbanne

²ESDB, Briançon

³CNR, Lyon

December 21, 2022

Abstract

Flood frequency analysis (FFA) is a widely used method to estimate flood hazard and is affected by several sources of uncertainty. Enlarging the streamflow samples with historical continuous stage measurements can reduce sampling uncertainties, but historical flood discharges are generally more uncertain. This paper provides insights on how historical data may or may not improve design floods estimates through a chain of uncertainty estimation methods for flood frequency analysis. Uncertainties are estimated and propagated from stage and rating curves to extreme flood quantiles estimates using Monte Carlo procedures. The role of both streamflow and sampling uncertainties in design floods estimations is examined. This procedure is applied to the 205-year long continuous discharge series of the Rhône River at Beaucaire ($95\ 590\ m^2$). Total uncertainty on flood quantile is significantly reduced when the length of the series increases from 50 to 100 years. Beyond this sample size, the total uncertainty remains stable, despite 19th century flood discharges have larger uncertainties than recent flood discharges. Addition of the two largest known floods in 1840 and 1856 is increasing by 15% the flood quantiles estimates but is almost negligible regarding the total uncertainty.

1 Introduction

Flood frequency analysis (FFA) is a widely used method to estimate flood hazard. It allows linking the magnitude of a flood to its probability of occurrence (Hamed and Ramachandra Rao (2019); Jain and Singh (2019)). Design floods, estimated for various exceedance probabilities, or equivalently, return periods, are commonly used for population safety policies, land use planning, as well as industrial safety. The most standard FFA approach is to estimate a distribution using a sample of flood peaks, typically, annual maximum discharges or discharges over a given threshold. This distribution may be extrapolated to reach the desired flood quantile that commonly corresponds to a 100 or 1000-year return period.

This FFA approach is affected by various sources of uncertainty. First, the hydrological data used to estimate the FFA distribution is uncertain. Indeed, streamflow time series are generally derived from stage time series through rating curve models. This procedure leads to measurement (gaugings and stage) and model (rating curve) errors. Moreover, the estimated FFA distribution is also uncertain because of the limited size of the available streamflow data (Kjeldsen et al., 2011). Considering the decisions that rely on FFA results, a consistent treatment of uncertainty all over the data processing chain is essential but is usually not performed.

Streamflow series are affected by several sources of uncertainty as described by McMillan et al. (2012) benchmark. First, stage measurements are affected by various sources of uncertainty (Horner et al., 2018). This uncertainty is often neglected but may have a substantial impact on the final discharge series Hamilton and Moore (2012), especially for historical stage series. A large number of stage error sources is identified in

the literature (Van Der Made (1982); Petersen-Øverleir and Reitan (2005); McMillan et al. (2012), Horner et al. (2018)), such as staff gauge reading errors, levelling of the staff gauge, or stage sensor calibration. The frequency of measurement may also induce time interpolation errors. With modern automatic gauges, stage is generally measured with a small enough time step by respect to the speed of variation of the stage (e.g. between 15 min and 1 hour), to get negligible time interpolation errors. However, before the rise of automatic gauges, measurements were made by operators who read the staff gauge less frequently (e.g. once or a few times per day), thus possibly missing the flood peak. This issue is particularly important when old stage series are used. Hamilton and Moore (2012) and Kuentz et al. (2014) estimated the measurement frequency error by sub-sampling recent, sub-hourly measurements. They calculated the difference between the variable of interest (such as the daily maximum stage) derived from scarce data, and the same variable derived from high-frequency measurements. Kuentz et al. (2014) applied the (monthly averaged) calculated bias to correct old stage series. This correction aimed at taking into account the error due to the daily variability caused by snow melt during cold months. However, this type of correction has never been applied to peak stage correction during floods, especially in the case of long stage series.

Rating curve uncertainty is also a major issue when dealing with streamflow series. Transforming stage into discharge requires calibration data (gaugings) to establish the stage/discharge relationship. Gaugings uncertainty depends on the measurement method (Le Coz et al. (2014a) and Puechberty et al. (2017)). Moreover, the rating curve is also affected by uncertainties coming from the imperfection of the chosen model to represent the actual hydraulic configuration and from parameter estimation. Many methods have been proposed to quantify these uncertainties (Petersen-Øverleir et al. (2009); Juston et al. (2014); Le Coz et al. (2014b); Morlot et al. (2014); Coxon et al. (2015); McMillan and Westerberg (2015); Mansanarez et al. (2019b)). A comparison of several of these methods has been recently proposed by Kiang et al. (2018). Another important issue affecting streamflow data accuracy is rating changes. The stage/discharge relationship is frequently affected by changes coming from various origins that can be natural or anthropic, for instance: bed geometry evolution during floods or river works, aquatic vegetation growth and decay, ice cover... A regular monitoring through gaugings is essential to detect those changes (Ibbitt and Pearson, 1987), that can be transient or sudden. Several methods have been proposed to deal with rating changes: estimating rating curves on moving temporal windows (Westerberg et al. (2011) and Guerrero et al. (2012)), computing as many rating curves as there are gaugings (Morlot et al., 2014), exploring changes in the annual minimum stages (Łapuszek and Lenar-Matyas, 2015), selecting 0.5-years return period discharge as a threshold for rating changes (McMillan et al., 2010). More recently, Darienzo et al. (2021) proposed a method based on a recursive segmentation procedure, accounting for both gaugings and rating curve uncertainties. This method has a particular interest when dealing with old and thus uncertain gaugings. Following the detection of rating shifts, rating curves should be estimated for each stability period. This task may not be straightforward, as the number of gaugings available within a stability period is not always sufficient to properly estimate the stage/discharge relationship for the whole discharge range. A common way to address this problem is to artificially repeat some gaugings from other stability periods (McMillan et al. (2012); Puechberty et al. (2017)). Mansanarez et al. (2019a) proposed an alternative approach to deal with this issue. They developed a stage-period-discharge (SPD) model where some parameters of the rating curve vary across periods, while the others are supposed constant. This method has the advantage of transferring information between periods to improve the rating curve estimation for all the stable stage/discharge periods, even when few gaugings are available.

Estimating sampling uncertainty in FFA is a well-established approach. Whatever the chosen distribution and estimation method, standard statistical procedures are available (Beven and Hall, 2014). However, these standard procedures only quantify sampling uncertainty, they do not consider data uncertainty. The literature review proposed in the previous paragraphs shows that methods for quantifying individual sources of uncertainty (stage, rating curve and FFA distribution estimation) are available. However, the way these multiple uncertainties propagate through the FFA analysis chain has been less thoroughly studied. A few solutions have emerged to propagate uncertainties in stage time series through uncertain rating curves (Dymond and Christian (1982), Herschy (1998), Petersen-Øverleir and Reitan (2005)), but they assume independent stage errors and therefore neglect systematic errors. Horner et al. (2018) proposed a consistent

method for the propagation of both sources of stage uncertainty through uncertain rating curves. Therefore, it is possible to infer the contribution of each source of uncertainty in streamflow data. Petersen-Øverleir and Reitan (2009), Steinbakk et al. (2016), and Vieira et al. (2022) performed an integrated analysis in which both rating curve parameters and flood frequency distribution are estimated. All these studies highlighted the importance of considering rating curve uncertainty for design flood estimations and concluded that, under some conditions, accounting for rating curve uncertainty may notably widen the uncertainty intervals around flood quantiles. However, none of them considered stage measurement and time interpolation errors, as well as rating changes, which may constitute a major source of uncertainty for streamflow data. This is particularly the case when dealing with long streamflow series for which stage uncertainty is large and variable through time, and rating changes may have been missed. Their consideration in a flood frequency framework is therefore not avoidable. Readers should note that the "historical data" term used in this paper refers to the use of ancient, systematic and steady stage measurements, by opposition to the use of punctual/censored flood evidences, prior to steady stage measurements.

The following questions will be considered in this paper:

1. How to make the most of historical hydrometric data in flood frequency analysis while accounting for multiple and variable uncertainties at each step of the procedure ?
2. What is the contribution of each source of uncertainty to the final flood quantile uncertainty when historical data are taken into account ?
3. To what extent does enlarging streamflow samples by adding increasingly uncertain historical data improve flood quantiles estimation ? How are the relative contributions of sampling and streamflow uncertainties evolving with sample size ?

This paper illustrates the chained application of methods to quantify and propagate uncertainty from stage records (and their limited time resolution) and stage-discharge rating curves to the estimation of extreme flood quantiles (Figure 1). While most of these methods already exist, a key novelty of this work is their combination to provide an end-to-end evaluation of the uncertainty affecting FFA estimations, from stage to flood quantiles. An original method to quantify the stage uncertainty stemming from infrequent readings is also proposed.

The paper is organized as follows. First, the methodology for establishing uncertain streamflow series in a century-long context is presented (section 2.1). It goes through the detection of rating shifts (section 2.2), the estimation of rating curves (section 2.3), and the estimation and propagation of stage errors (section 2.4). Then, an approach to propagate streamflow uncertainty through the estimation of extreme flood quantiles is proposed (section 2.5). This procedure is illustrated through its application to the Beaucaire gauge on the Rhône River (section 3), for which official flood hazard assessments only used discharge series from 1920 (Rigaudière et al., 2000). The recent works of Pichard et al. (2017) and Bard and Lang (2018) provided a continuous stage series from 1816 to the present time, which makes it the ideal case study for demonstrating this procedure. The results of this application are presented in section 4, and they are discussed in section 5, where avenues for improvements are proposed.

2 Uncertainty propagation chain for flood frequency analysis

2.1 Rating shifts detection

The stage/discharge relationship is sensitive to sudden changes caused by morphogenic floods or other causes affecting the flow characteristics. Relying on residuals between the gaugings and rating curve is the most common approach to monitor the stability of this relationship over time. The method proposed by Darienzo et al. (2021) is used in this work. First, a baseline rating curve is estimated from the whole gaugings dataset, using BaRatin method (Le Coz et al., 2014b). The residuals between gaugings and the rating curve are determined, and a statistical segmentation test is applied to them. This segmentation procedure accounts for the residuals uncertainty, coming from both the gaugings uncertainty and the rating curve uncertainty. The optimal number of gaugings is determined based on the Bayesian Information criterion (BIC). Then, the

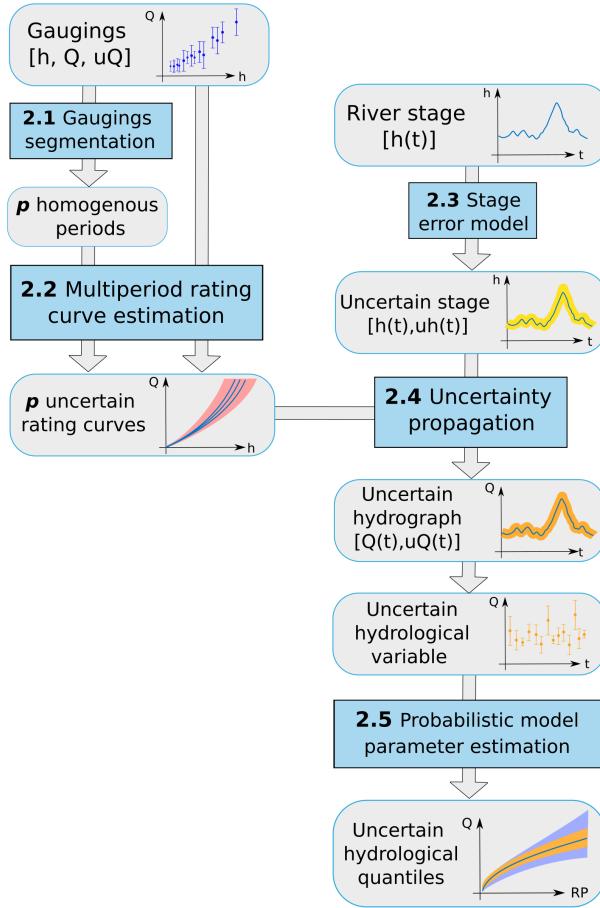


Figure 1: Block diagram of the uncertainty propagation procedure. Grey blocks represent data, blue blocks stand for analysis methods/models that correspond to the sub-sections of this article. h is the water stage, uh the stage uncertainty, Q the discharge, uQ the discharge uncertainty, t is the time and RP the return period of flood quantiles.

same steps are applied recursively to each sub-period determined earlier. The recursive procedure is stopped when the BIC indicates that a single period is optimal. The results are not only the dates of the rating shifts but the posterior probability density functions (pdf) of change point times. We assume affecting the shift time to the time of the maximum stage included in the posterior credibility interval. Prior knowledge is provided on the expected magnitude of the residuals, the number of segments at each iteration and the minimum time lag between two consecutive change points.

2.2 Multi-period rating curves estimation: stage-period-discharge model

Once the stability periods have been identified, the next step is to estimate the rating curve associated with each period. Mansanarez et al. (2019a) developed a stage-period-discharge (SPD) model "based on the physical interpretation of changes in the stage-discharge relation across a series of stability periods". The SPD model is based on the BaRatin model (Le Coz et al., 2014b).

BaRatin uses the Bayesian paradigm to estimate the parameters of the general rating curve equation, a combination of power equations: $Q = a(h - b)^c$, where Q is the discharge in m^3/s , the flow depth is the difference between the stage h (in m) and an offset b , and a and c being the coefficient and the exponent of the power function. The rating curve equation is deduced from a hydraulic analysis of the gauging station, aimed at identifying the main hydraulic controls governing the stage-discharge relation. The multiple controls can be activated successively or simultaneously. Bayesian inference allows deriving the posterior distribution of rating curve parameters by combining hydraulic information (priors for parameters of each hydraulic controls) and information from gaugings with uncertainty (likelihood).

The SPD model is estimating the rating curves of each stability based on the same principle, and is considering that several identified parameters vary in time, throughout the stability periods. An important step is the identification of those varying parameters based on hydraulic analysis of the site. Generally, offsets or channel widths are suspected to change. The so-called local changes are supposed to affect the lowest control only (for instance the movement of the controlling riffle), whereas the global changes affect various controls at the same time (for instance, the scouring or filling of the channel, affecting low flows and main channel offsets at the same time). One way to follow the evolution of riverbed elevation and to help specifying priors for rating curve vertical evolution is to study the long-term evolution of the yearly lowest stages, as described by Łapuszek and Lenar-Matyas (2015). See Mansanarez et al. (2019a) for a detailed description of prior specification for time-varying rating curves.

2.3 Stage uncertainties

Many sources of error having distinct statistical properties can affect stage measurements, as described in Horner et al. (2018). Five different sources of error ($\delta_{1,\dots,5}$) affecting stage measurements are considered. Let $h(t)$ be the measured maximum stage of a day t . $\bar{h}(t)$, the unknown true maximum stage is assumed to be approximated by the following equation:

$$\bar{h}(t) = h(t) + \delta_1(t) + \delta_2(t) + \delta_3(t) + \delta_4(t) + \delta_5(t) \quad (1)$$

Staff gauge reading errors $\delta_1 \sim \mathcal{N}(0, \sigma_1)$ originate from operators reading the gauge, where σ_1 depends on the resolution of the graduations (usually 1 cm), and can be increased by waves, especially during floods (McMillan et al., 2012).

Most stage measurements nowadays are done with automatic sensors of various types such as pressure sensors, floats, radars, and they require a calibration to link the water stage to the measured proxy (respectively the pressure of the water column, the height of a float, or the air draught). Two types of errors arise from this process: sensor errors $\delta_2 \sim \mathcal{N}(0, \sigma_2)$, where σ_2 being usually estimated by the sensor manufacturer, and sensor calibration errors $\delta_3 \sim \mathcal{N}(0, \sigma_3)$ that are related to the corrections made by operators when comparing the stage measured by the sensor to the actual stage at the staff gauge reference. An operator error at this step could affect the stage measurement until the next calibration. Sensor drifts are ignored, for instance, sensor calibration error δ_3 is assumed constant between two calibrations and can be represented by drawing a new random value at each operator intervention.

Datum errors $\delta_4 \sim \mathcal{N}(0, \sigma_4)$ are related to changes in the datum reference elevation of the staff gauge zero value and possible discontinuity between successive gauges. Similarly to δ_3 , this error is constant between two gauge changes or datum reference measurements.

Measurement frequency errors δ_5 are related to the inadequacy of the frequency of measurement with respect to the speed of stage variations. Unlike other types of stage errors, this error is hence necessarily positive, which calls for using a positive distribution such as the Log-normal or Exponential distribution at each measurement time step. The parameters of this distribution can be estimated with data from the recent period, by analyzing the difference between the daily maximum stage derived from the high-frequency sensor measurement and that from an infrequent fixed-time reading. Note that the measurement frequency error for hourly (or less) measurements are considered negligible when considering large rivers with slow variations.

To sum up, δ_1 , δ_2 and δ_5 errors are drawn at each time step, while δ_3 and δ_4 errors are only drawn at specific calibration times. δ_1 to δ_4 are assumed Gaussian with known standard deviations, while δ_5 is assumed Exponential with parameter estimated by subsampling recent measurements. Monte Carlo methods are applied to combine these multiple stage uncertainty sources. The uncertainty around stage measurements from eq. 1 is represented by 500 possible realizations of the stage $h(t)$.

2.4 Propagation of stage and rating curve uncertainties to streamflow time series

Stage measurements realizations can be propagated through uncertain rating curves, following the approach described by Horner et al. (2018). To estimate the contribution of the different sources of streamflow uncertainties, four scenarii are considered:

- **Case 1: Maxpost streamflow.** Stage measurements are free of uncertainties and are assumed to be the median of the stage time series realisations. Rating curve estimation is also free of uncertainties. The unique stage time series is propagated trough the maxpost (Maximum A Posteriori, i.e. the most probable) rating curves, resulting in one set of discharge time series.
- **Case 2: Stage uncertainty.** Now, stage measurements are uncertain. This uncertainty is represented by n possible stage time series, propagated through the maxpost rating curves. Thus, n sets of discharge time series are obtained.
- **Case 3: Stage and parametric rating curve uncertainty.** Stage measurements are uncertain and rating curve estimation is affected by parametric uncertainties (represented by m sets of rating curves) but is not affected by remnant uncertainties. The n stage time series are propagated through m sets of rating curves, leading to $n \times m$ sets of discharge time series.
- **Case 4: Total streamflow uncertainty.** Stage measurements are uncertain and rating curve estimation is affected by both parametric and remnant uncertainties, represented by m sets of rating curves. The n stage time series are propagated through m sets of rating curves, leading to $n \times m$ sets of discharge time series to which remnant errors are applied. This represents the total streamflow uncertainty.

2.5 Estimation of probabilistic model parameters and flood frequency analysis

The Generalized Extreme Value (GEV) distribution is commonly used to model annual maximum discharges (AMAX) (see Hamed and Ramachandra Rao (2019) or Jain and Singh (2019)). $\boldsymbol{\theta} = (\mu, \sigma, \xi)$ is the vector of GEV parameters: μ is the location parameter, σ the scale parameter and ξ the shape parameter. The parameters can be estimated based on an independent and identically distributed (*iid*) sample of j annual maximum discharges (AMAX) $(\mathbf{q}_t)_{t=1,\dots,j}$. Bayesian-MCMC estimation is used in this work, as described in Renard et al. (2006). The computed posterior distribution quantifies sampling uncertainty and is represented by m MCMC-generated GEV parameters sets $\boldsymbol{\theta} = (\mu_{k=1,\dots,m}, \sigma_{k=1,\dots,m}, \xi_{k=1,\dots,m})$. The best set of parameters $\hat{\boldsymbol{\theta}}$ is the set that maximises the posterior distribution and is called "maxpost".

As described in the previous sections, streamflow observations are affected by uncertainties. This streamflow uncertainty is represented by n possible realisations of the AMAX series $q_t^{(i)}$, that are subsampled from the $n \times m$ streamflow sets to reduce computation time. The final uncertain flood quantiles should thus consider both sampling and streamflow uncertainties. Similarly to Steinbakk et al. (2016), the aim is to estimate the contribution of each source to the final/total uncertainty. For this purpose, three scenarii can be considered:

- **Case 1: Maxpost quantiles.** The series of AMAX floods $\hat{\mathbf{q}} = (\hat{q}_t)_{t=1,\dots,j}$ is free of streamflow uncertainties and is obtained by propagating the median stage realizations through the maxpost (Maximum A Posteriori) rating curves. The GEV distribution is estimated using this single AMAX series, and the maxpost GEV parameters $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma}, \hat{\xi})$. In this case, sampling uncertainties are not considered as well.
- **Case 2: Streamflow uncertainty.** AMAX floods are uncertain. This uncertainty is represented by n sets of possible AMAX realisations: $\mathbf{q}^{(i)} = (q_t^{(i)})_{t=1,\dots,j}; i=1,\dots,n$. Bayesian-MCMC estimation of the GEV distribution is performed n times (one for each AMAX series) leading to $n \times m$ sets of GEV parameters. However, only the maxpost set of GEV parameters is computed on each AMAX flood realisation. This results in n sets of GEV parameters $\hat{\boldsymbol{\theta}}_{i=1,\dots,n}^{(i)}$ that represent the effect of streamflow uncertainty of flood quantiles, ignoring sampling uncertainty.

- **Case 3: Total uncertainty.** AMAX floods are uncertain and sampling uncertainty is considered. Similarly to Case 2, Bayesian-MCMC estimation of the GEV distribution is performed n times (one for each AMAX series) leading to $n \times m$ sets of GEV parameters. The result, reflecting both sampling and streamflow uncertainties, is thus represented by $n \times m$ sets of GEV parameters $(\theta_k^{(i)})_{k=1,\dots,m; i=1,\dots,n}$.

3 Case study: The Rhône River at Beaucaire

3.1 Site

The Rhône River at Beaucaire ($95\ 590\ km^2$) is the most downstream gauge of the Rhône River (Figure 2). It captures all the complexity of the Rhône River hydrological regime, from the Alpine area to the French oceanic and Mediterranean influences. The annual mean discharge is around $1700\ m^3/s$ (Bard and Lang, 2018), and the maximum known discharge reached $12\ 500\ m^3/s$ (May 1856, Lang and Coeur (2014)). The station lies in a flood sensitive area, as illustrated by the recent 2003 flood, resulting in 1.1 billion euros worth of damage (Lang and Coeur, 2014). The first stage measurements started in 1816, close to the bridge linking the cities of Beaucaire and Tarascon. This station is named "Pont de Beaucaire" (Kilometric point 267.6 from Lyon). It has been used until the construction of the Vallabregues 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 original one, downstream from the restitution of the derived discharges. This station, logically named "Beaucaire Restitution" (Kilometric point 269.5), has been used ever since.



Figure 2: The French Rhône River catchment and Beaucaire gauging stations (from www.geoportail.gouv.fr and www.openstreetmap.org)

3.2 Rating curves

3.2.1 Pont de Beaucaire

At Pont de Beaucaire, the stage-discharge relationship can be approximated by two additive channel controls: a main channel and a floodway. Within the main channel (when water stage is below $k_2 \approx 2m$), the flow is

splitted in two sub-channels (figure 3b) since time immemorial (at least before 1816) as described by Armand (1907). These sub-channels are communicating upstream and downstream from the gauge location, thus, they can be considered as a unique channel. The mobile sandbars that were separating the flow were progressively fixed by dikes to ease the navigation during XIXth Century (figure 3a). The average width of this channel is the sum of both sub-channels widths (≈ 300 m). At the gauge location, the floodway width is limited to 500 m by unsubmersible levees (figure 3b), but is expanding several hundred meters downstream above k_2 . This expansion is impacting the stage at the gauge. Consequently, the floodway rating curve is additive to the main channel. The floodway width is approximately 800 m. Thus, the stage discharge relationship is estimated by the following rating curve equation:

$$Q(h) = \begin{cases} a_1(h - b_1)^{c_1}, & \text{if } k_1 < h \leq k_2 \text{ (main channel)} \\ a_1(h - b_1)^{c_1} + a_2(h - b_2)^{c_2}, & \text{if } h > k_2 \text{ (main channel + floodway)} \end{cases} \quad (2)$$

Parameters are derived from historical material retrieved in regional archives. This leads to the prior parametrisation described in table 3. Physical parameters that have a direct hydraulic meaning are here expressed in the first three lines, then some of them are used to infer the a parameter by using Monte Carlo propagation with a large number of samples of the physical parameters (B, K, S). As recommended by Le Coz et al. (2014b), log-normal priors are used for positive quantities such as slopes, channel widths and Strickler coefficients. Informative but imprecise priors are assigned to parameters such as channel widths, slopes or offsets which can be difficult to estimate precisely. For c exponents, very precise priors are used because they depend on the control type and shape (here $c = 5/3$ for rectangular channel controls based on simplified Manning-Strickler equation). As recommended, structural uncertainty parameters have uninformative priors. According to historical profiles and cross sections, we assume that main channel and floodway offsets (b_1 and b_2) may have changed independently over time, but that main channel and floodway widths have not changed. As described by Mansanarez et al. (2019a), this translates as global and local offset changes δ_g and δ_l . The successive offsets of the two controls are computed as follows:

$$\begin{cases} b_1^{(k)} = b_1^{(k-1)} - (\delta_g^{(k)} + \delta_l^{(k)}), & \text{(incremental changes in the main channel)} \\ b_2^{(k)} = b_2^{(k-1)} - \delta_g^{(k)}, & \text{(incremental changes in the floodway)} \end{cases} \quad (3)$$

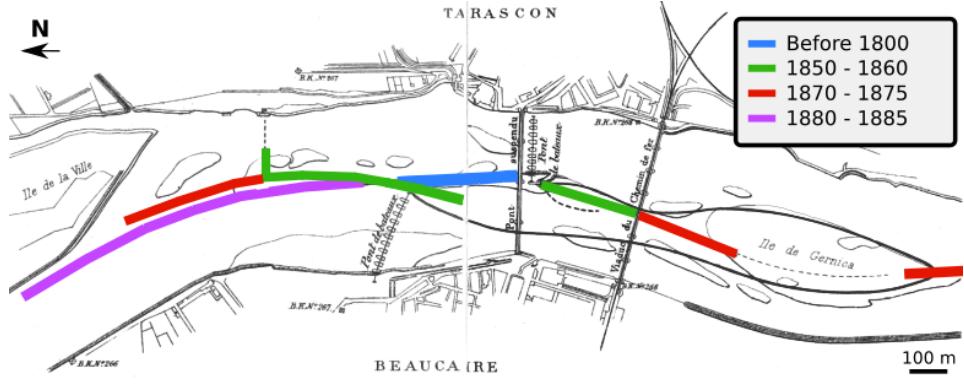
As the most recent stable stage/discharge period obtained by gaugings segmentation is assumed to be the most accurately known, it is used as the reference period. Changes are cumulative, and are therefore computed from this reference period. Prior distributions of offset changes are determined in section 3.2.3.

3.2.2 Beaucaire Restitution

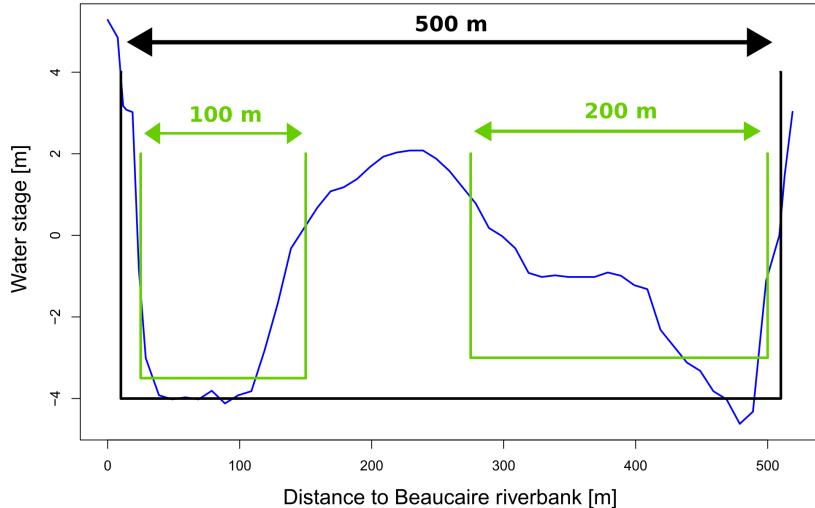
Beaucaire Restitution station has a more stable profile than Pont de Beaucaire according to 1974-2016 cross-sections (figure 3c left) as there are no mobile sandbars separating the flow. The flow is approximated by a 3-control configuration:

$$Q(h) = \begin{cases} a_1(h - b_1)^{c_1}, & \text{if } k_1 < h \leq k_2 \text{ (low flows)} \\ a_2(h - b_2)^{c_2}, & \text{if } k_2 < h \leq k_3 \text{ (main channel)} \\ a_2(h - b_2)^{c_2} + a_3(h - b_3)^{c_3} & \text{if } h > k_3 \text{ (main channel + floodway)} \end{cases} \quad (4)$$

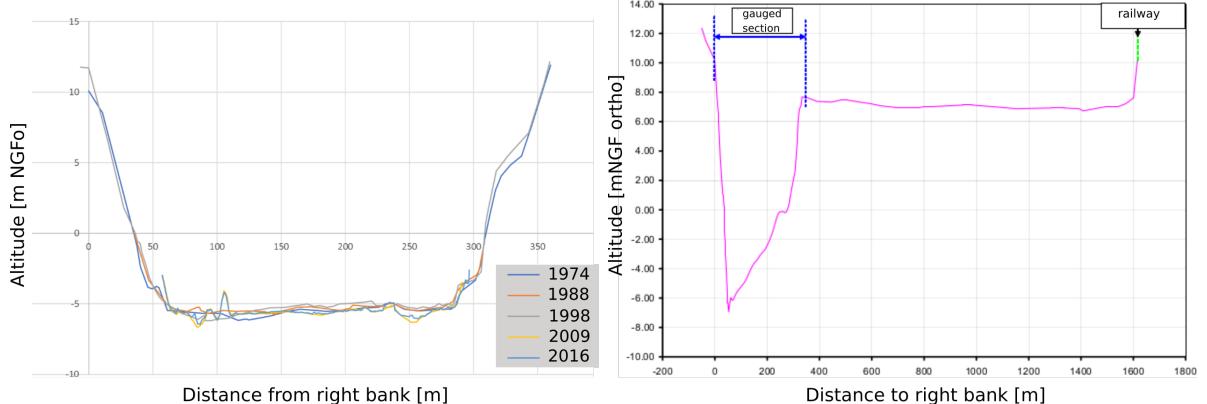
The very low flows (approximately below 0 m at the gauge) are known to be influenced by the Mediterranean Sea level variations. The first control is replaced by the same channel control with different slope when the stage is no longer influenced by sea level variations. At the gauge location, the 12 meters high banks are preventing overbank flows (figure 3c left). However, downstream from the station location, overbank flows occur at the left bank above approximately 8 m (figure 3c right). A floodway control, additive to the main channel, represents those overbank flows. Low flows and main channel offsets b_1 and b_2 are assumed changing over time, whereas floodway offset b_3 and channel widths are supposed constant. Priors of incremental bed elevation changes are computed the same as for Pont de Beaucaire in the previous paragraph and are specified in section 3.2.3 results.



(a)



(b)



(c)

Figure 3: (a) Map of dike evolution in Beaucaire between 18th and 20th centuries, adapted from Armand (1907); (b) Approximation of channel controls at Pont de Beaucaire station, based on a 1845 cross section survey; (c) Profiles from 1974 to 2016 (left) at Beaucaire Restitution station and 2.5 km downstream from the station (right) from CNR data, translated from Bard and Lang (2018) and MEDD (2005)

3.2.3 Prior estimation of bed changes

It is possible to follow the evolution of riverbed elevation through the evolution of yearly lowest changes. Here, the 5% annual quantile is considered (figure 4). At Pont de Beaucaire (1816 - 1967), the 5% quantile is oscillating with a 0.3 m standard deviation. Those variations do not seem related to the occurrences of major floods. The prior distributions of local and global offset changes defined in section 3.2 are assumed Gaussian and equal to 0.3: $\delta_g^{(k)} = \delta_l^{(k)} \sim \mathcal{N}(0, 0.3)$ (see table 3).

At Beaucaire Restitution (1970-2020), the annual 5% quantile shows a large decrease during the first 4

years (more than 1 m). This is a consequence of Vallabregues hydraulic works between 1967 and 1970 as well as important dredgings. A geomorphic adjustment after the works in the channel may have affected the riverbed level as well. After the first years, the channel bottom stabilized, however with a slight scouring trend of about 30 cm in 40 years. At Beaucaire Restitution, the standard deviation of the 5% quantiles reaches 0.5 m. Prior standard deviation of local changes is assumed larger than 0.5 m and therefore defined as 0.8 m to be more representative of the large changes that occurred during the first years: $\delta_l^{(k)} = \mathcal{N}(0, 0.8)$. Standard deviation of prior global changes is assumed equal to 0.3 m to represent the scouring trend in the main channel: $\delta_g^{(k)} = \mathcal{N}(0, 0.3)$ (see table 4).

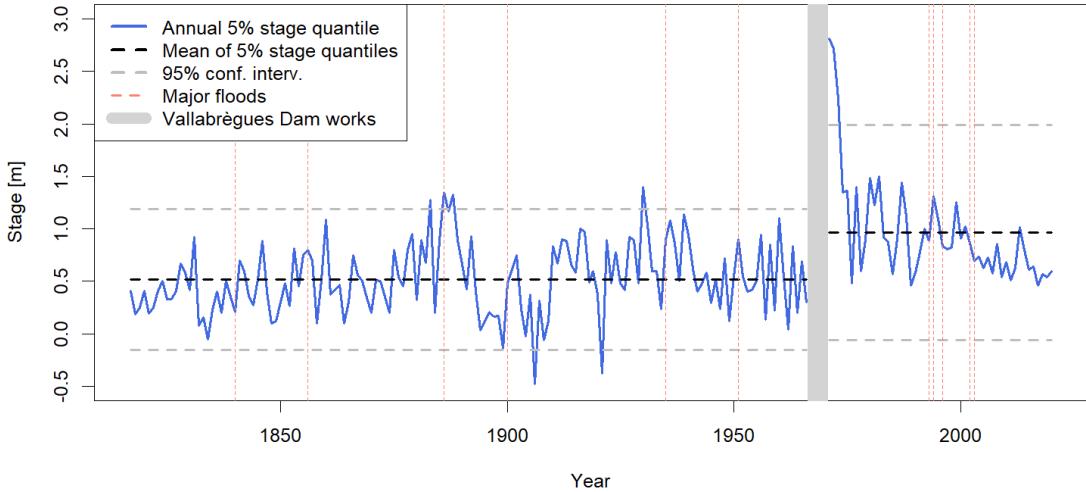


Figure 4: Time series of annual 5% stage quantile of both Pont de Beaucaire (1816-1967) and Beaucaire Restitution (1970-2020) stations used to specify priors on rating shift parameters.

3.3 Stage series

3.3.1 Pont de Beaucaire (1816 - 1967)

Thanks to the archival work of Pichard et al. (2017), a continuous stage series is available at Pont de Beaucaire from 1816 to 1967: daily stage readings from 1816 to 1840, and three stage readings per day from 1841 to 1967. The records were made visually by an operator, at noon during the first years, then at 7am, 12am and 5pm (Figure 5). When three stage readings per day are available, the maximum of the three stages is considered as the daily maximum stage, and the unique value at noon is kept as the daily maximum stage before 1840. Additionally, after 1840, when the stage was rising above 5 m, the operators made more frequent visual records (supposedly hourly measurements). When these records are available, they are of course used to establish the daily maximum stages. Annual maximum stages are considered here, as this work is focused on floods.

Stages uncertainties depend on the measurement method, as described in 2.3. Table 1 summarizes the different sources of stage uncertainties at Beaucaire, here given as standard deviations σ :

Staff gauge reading uncertainty σ_1 is taken as 5 cm. Staff gauge precision is centimetric, but as this work is focused on floods, the error is expanded because of the waves that may complicate the reading.

Sensor precision σ_2 and sensor calibration uncertainties μ_3 are not considered at Pont de Beaucaire as the stage was read by operators directly on the staff gauge.

Bard and Lang (2018) have compiled several elevation measurements of the datum reference between 1912 and 2010. Most of those measurements occurred after the decommissioning of Pont de Beaucaire station for

Bassins du Rhône et de la Saône

OBSERVATIONS HYDROMÉTRIQUES

Echelle de Beaucaire sur le Rhône

DATES	OBSERVATIONS ORDINAIRES			ÉTAT DU CIEL	VENT	RENSEIGNEMENTS DIVERS.
	7 h. matin	midz	5 h. soir			
	m.	m.	m.			
1	2.56	2.54	2.54	peu nuageux	vent fort	
2	2.50	2.44	2.38	nuageux	vent Est faible	
3	2.24	2.16	2.14	peu d.	modérément avec fort	
4	2.04	2.00	1.98	dès d.	faible	
5	1.89	1.84	1.80	nuageux	d.	
6	1.76	1.72	1.72	dès grêle	calme	
7	1.75	1.80	1.92	nuageux	modérément modérée	
8	2.62	3.02	3.36	peu d.	vent Est faible	
9	3.88	3.90	3.48	overcast	Est	T
10	4.10	4.16	4.25	dès nuageux	Est Est avec fort	T

Figure 5: Table of limnimetric surveys at Pont de Beaucaire, March 1914. Operators were supposed to provide the water level at 7am, 12am and 5pm, as well as a few meteorologic details. (Ponts&Chaussées, 1914)

Beaucaire Restitution. Datum reference uncertainty σ_4 is assumed to be equal to the standard deviation of those measurements: 6 cm. The datum reference measurement frequency during the exploitation of Pont de Beaucaire station is assumed to be 25 years (average duration between the retrieved elevation measurements). Hence, the datum reference uncertainty is drawn every 25 years.

As described in section 2.3, the distribution of measurement frequency errors δ_5 can be estimated using the frequent stage measurements at Beaucaire Restitution between 1970 and 2020 (50 AMAX values). For the "one stage reading per day" case (mimicking the 1816-1840 period), this error is corresponding to the difference between the maximum hourly stage value and the stage at noon of the same day. The "three stage readings per day" error is the difference between the maximum hourly stage of a day, and the maximum of 7am, noon and 5pm stages of the same day. At Beaucaire Restitution, the one stage/day error maximum value is 2.2 m, and the three stages/day error value is 0.9 m. An exponential distribution is fitted for both errors samples and is used to correct the annual maximum stages at Pont de Beaucaire. According to Ponts et chaussées staff gauge management instructions, after 1840 and for the stages above 5 meters, hourly measurements were made by observers. Hence, measurement frequency error δ_5 can be considered as negligible when stage is above 5 meters after 1840.

SYMADREM (2012), Pichard (2013) and Bard and Lang (2018) suggested that for the floods during which dike breaks happened downstream from Beaucaire, stage measurements should be corrected because the stage measured at the station may lead to underestimate the actual discharge of the flood. The stage corrections for the more thoroughly studied floods of 1840, 1841 and 1856 estimated by SYMADREM (2012) are adopted. The stage uncertainty of these floods replaced the previously estimated uncertainty and is assumed Gaussian, with mean the estimated discharge and standard deviation half of the applied correction, chronologically: 0.94, 0.4 and 0.4 m.

3.3.2 Beaucaire Restitution (1970 - 2020)

For Beaucaire Restitution station, most of the stage uncertainty values come from CETIAT (2005) expertise on behalf of Compagnie Nationale du Rhône. They are summarized in table 1 and detailed here:

Staff gauge reading uncertainty is considered zero (neglected) as the measurements are done by automatic sensors. However, gauge reading uncertainties affect sensor calibration uncertainties σ_3 .

Instrument precision uncertainty : $\sigma_2 = 0.01/\sqrt{3}$ m comes from the sensor manufacturer specifications.

The standard deviation of all the re-calibrations realized by the operators is equal to 5 cm according to CETIAT (2005). Calibration is also affected by staff gauge reading uncertainties, because the stage read in the staff gauge is the reference used by operators to calibrate the sensor. CETIAT (2005) estimated a 3.35 cm uncertainty for the gauge reading uncertainty. Therefore, gauge reading and calibration uncertainties are combined as follows: $\sigma_3 = \sqrt{0.0335^2 + 0.05^2} = 0.06$ m. As the average time lag between calibration is 6 months, a new value of σ_3 is drawn for each annual maximum stage.

Datum reference uncertainty σ_4 is considered negligible because of the precision of modern topographic measurements (< 1 cm).

Measurement frequency uncertainty σ_5 is considered negligible, because the sub-hourly measurement frequency is assumed adequate to capture the Rhône River stage variability.

Date	δ_1 : gauge reading	δ_2 : instr. precision	δ_3 : sensor calibration	δ_4 : datum reference	δ_5 : measurement frequency	
					Stage < 5m	Stage ≥ 5 m
Before 1840	$\mathcal{N}(0, 0.05)$	-	-	$\mathcal{N}(0, 0.06)$	exp(2.18)	
1840 - 1967	$\mathcal{N}(0, 0.05)$	-	-	$\mathcal{N}(0, 0.06)$	exp(8.86)	-
1970 - 2020	-	$\mathcal{N}(0, 0.01/\sqrt{3})$	$\mathcal{N}(0, 0.06)$	-	-	-

Table 1: Distributions used for the different sources of stage errors (in meters). \mathcal{N} and exp represent Gaussian and Exponential distributions respectively.

3.4 Gaugings

A set of 244 gaugings from 1840 to 1967 have been compiled at Pont de Beaucaire. After excluding a few gaugings which were considered dubious, 233 measurements remain. The frequency of gaugings is variable in time. No gaugings were retrieved before 1840 and there are a few 10 to 20 year gaps without gaugings, which makes the estimation of the stage/discharge relationship over time challenging. The uncertainty of the gaugings at Beaucaire depends on the gauging method according to Bard and Lang (2018) values specified in table 2.

A set of 304 gaugings is available at Beaucaire Restitution. A few of these were out of the period of stage measurements availability and were discarded. Finally, 296 gaugings were selected. As modern hydrometric developments allowed estimating the uncertainty for each individual gauging (particularly for ADCP and current meters), those values are used when available in the CNR archives. If not, values from table 2 are considered.

Gauging method	Standard uncertainty
Current meter at 0.6 h and surface	5%
Current meter point by point	3.5%
Surface current meter	7.5%
Unknown type	7.5%
ADCP	3.5%
Floats before 1936	10%
Hydrotachymeter before 1936	10%

Table 2: Gaugings uncertainty depending on the method used (hypotheses from Bard and Lang (2018)). Expressed as standard deviations of the measured discharge in %.

4 Results

4.1 Assessment of rating shifts

Darienzo et al. (2021) segmentation procedure is applied at Beaucaire (Figure 6). The expected order of magnitude of rating curve residuals is assumed Gaussian with zero mean and a $500 \text{ m}^3/\text{s}$ standard deviation for both stations. This roughly correspond to the mean of the residuals between the gaugings and the baseline rating curve. The maximum number of segments at each iteration had been fixed at six (see Darienzo et al. (2021) for details on priors and parameters specification).

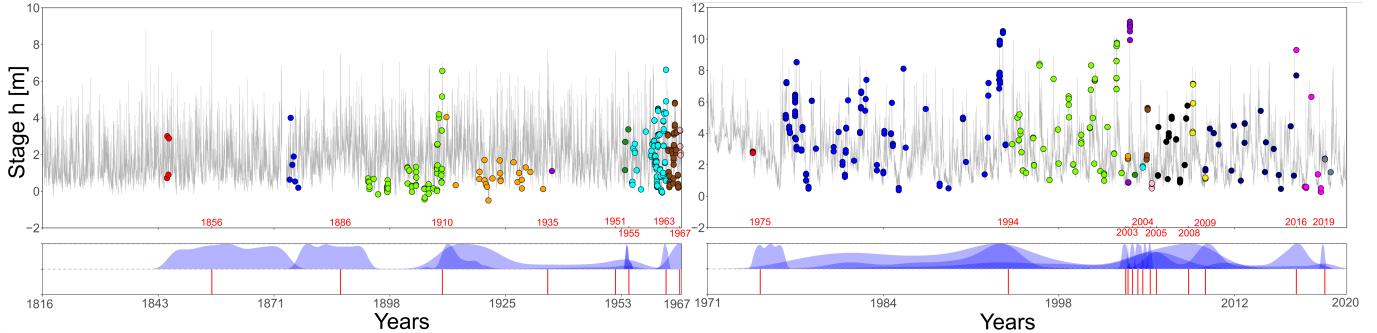


Figure 6: Gaugings segmentation of the Rhône River at Pont de Beaucaire (left) and Beaucaire Restitution (right). Dots represent gaugings with different colors for each stability period. The grey curve is the stage series. Blue ribbons represents posterior pdf of shift times and red segments are the retained shift times taken as the maximum stage included in each posterior pdf interval.

Eight shift times are detected at Pont de Beaucaire (left), located at the date of the largest flood within posterior pdfs. The gauging frequency is erratic through the history of the station and some periods include a small number of gaugings. Some gaugings may have been lost. As a consequence, the posterior pdfs of shift times span many years for the first shifts, and are similar to uniform distributions between sets of gaugings. This is reassuring about the method since there is indeed no information during long no-gauging periods to identify a shift time. It is no surprise that most of the shifts occur very close to the largest historical floods. 1935 and 1951 time shifts are respectively corresponding to the 3rd and 4th most important floods of the history of the station. There is, by construction of the segmentation model, no shift before 1845, year of the first available gaugings. However, we can consider that 1840 flood (supposedly the largest flood since 1800) is likely to have caused a shift. Therefore, we consider an additional shift time at the exact flood date. This brings the total number of stability periods to 10 (see detailed dates in table 5).

Thirteen shift times are detected at Beaucaire Restitution. As can be seen in figure 6 (right), gauging frequency is far higher than for Pont de Beaucaire station (except during the 5 first years), resulting in a better determination of the rating shift times. Due to the lack of gaugings at the beginning of the series, only one rating shift is detected but many shifts potentially took place in those first four years as morphological readjustments and dredging occurred (see previous section). This first shift is assumed to be assigned to the first important flood of the station in 1976, after which the channel is stabilized. The following shift occurred during 1994 flood, one of the biggest of the station. The most notable flood at Beaucaire Restitution occurred in 2003 ($11\,500 \text{ m}^3/\text{s}$, with a return period of about 100 years according to MEDD (2005)). Unsurprisingly, the stage/discharge relationship is considerably disrupted by this event, as reflected by the six rating shifts detected from 2003 to 2005. Two out of these six shifts were discarded because the shift amplitude is considered minor based on further analysis of the corresponding rating curve change. The largest flood within posterior pdf of those shifts is almost always corresponding to the 2003 flood. This is also the case for 2005, 2008 and 2009 shifts, for which the posterior pdf spans many years and include 2003 flood. Therefore, the shift dates are assumed to be located to the maxpost shift times, as several shifts can not be located at the same date. Lastly, the last shift of 2019 is also discarded because the shift amplitude is considered minor based on further analysis. Finally, ten rating shifts were retained. This brings the number of stability periods to eleven for Beaucaire Restitution, as seen in table 5.

4.2 Multiperiod rating curves estimation

Uncertain rating curves are estimated using Mansanarez et al. (2019a) SPD model, for each stability period detected previously. For Pont de Beaucaire, this leads to ten rating curves that show a good adequacy with gaugings (figure 7a). The evolution of the low flows control offset (b_1) gives indications on the evolution of bed elevation (figure 7c). Important changes occurred during the third and fourth stability periods with a successive increase and decrease of the offset. Those changes may be related to the channel works that occurred during the end of the XIXth Century. Afterwards, the offset is stabilized, along with a slight increase trend which may be a consequence of the filling of the channel noticed in figure 4. The widest uncertainty interval belongs to the first period (1816-1840: dark red) for which no gaugings are available (figure 7a). The expected range of rating curve uncertainties for floods discharges (above 6 m) varies from around 20% for the first period, to less than 10% after 1840. Static parameters are precisely estimated and are presented in figure 7e. The c posterior distributions are as wide as priors, due to the fact that c priors are already very precise because they come from simplified Manning-Strickler formula for which the exponent is exactly 5/3.

Eleven uncertain rating curves were computed at Beaucaire Restitution (figure 7b). The rating curve uncertainty intervals are smaller than Pont de Beaucaire overall because of a larger number of gaugings and a smaller gaugings uncertainty: around 5% of uncertainty is expected for floods above 6 m. However, low stages uncertainty is greater than at Pont de Beaucaire, because of the unavailability of low flows gaugings. The sustained flows of the Rhône River limits the exploration of the low flows hydraulic control. The first period rating curve (dark red) is shifted with respect to the other curves due to the channel readjustments and dredging operations after Vallabregues works (1967-1970). The first control offsets (b_1) are globally decreasing over time, showing a slight scouring trend of the channel (figure 7d). Static parameters (figure 7f) appear precisely estimated, except for the 3rd control offset b_3 for which the posterior distribution is as wide as the prior.

4.3 Stage errors

The error sources described in table 1 are applied to both Beaucaire stations and propagated via Monte Carlo procedures. As seen in figure 8, measured stage is outside of stage uncertainty interval before 1840 at Pont de Beaucaire. This is due to the exponential distribution used to model measurement frequency error which is predominant here and has, by definition, only positive results. The upper uncertainty bound is sometimes 1.5 meters higher than the measured stage. Therefore, considering this source of uncertainty may have important consequences on the final results. Apart from that, the uncertainty of AMAX stages decreases over time as the measurement frequency and precision improve. The width of the 95% uncertainty interval is 1.7 m before 1840, 0.3 m between 1840 and 1967, and 0.24 m at Beaucaire Restitution (1970-2020). The 5 m threshold above which hourly measurements were done after 1840 explains an important reduction of the uncertainty. After this date, the uncertainty is controlled by the exceedance of this 5 m threshold, the AMAX below 5 m being penalized by measurement frequency error δ_5 that is considered negligible for hourly measurements made above 5 m. Beaucaire Restitution uncertainty is smaller than Pont de Beaucaire uncertainty. The difference between uncertainty bounds and originally measured stages is presented in the bottom part of figure 8. The median of stage realisations shows some fluctuations that are mainly due to datum reference errors that are drawn every 25 years at Pont de Beaucaire.

4.4 Total streamflow uncertainties

Stage uncertainties were propagated through uncertain rating curves. The contributions of each source of uncertainty in streamflow time series uncertainty are represented in three categories: parametric and remnant uncertainties (from rating curves), and stage uncertainties. The result is presented in figure 9 for the two Beaucaire stations from 1816 to 2020. Streamflow uncertainty, although fluctuating, appears to decrease over time. From $\pm 30\%$ before 1840, and $\pm 10\%$ before 1967, to $\pm 5\%$ at Beaucaire Restitution (1967-2020). Stage uncertainty appears dominant at Pont de Beaucaire, as well as rating curve parametric uncertainty, originating from the difficulty to estimate rating curve parameters with only a few gaugings. Thus, parametric uncertainty is reduced for properly gauged periods. During the Vallabregues hydraulic

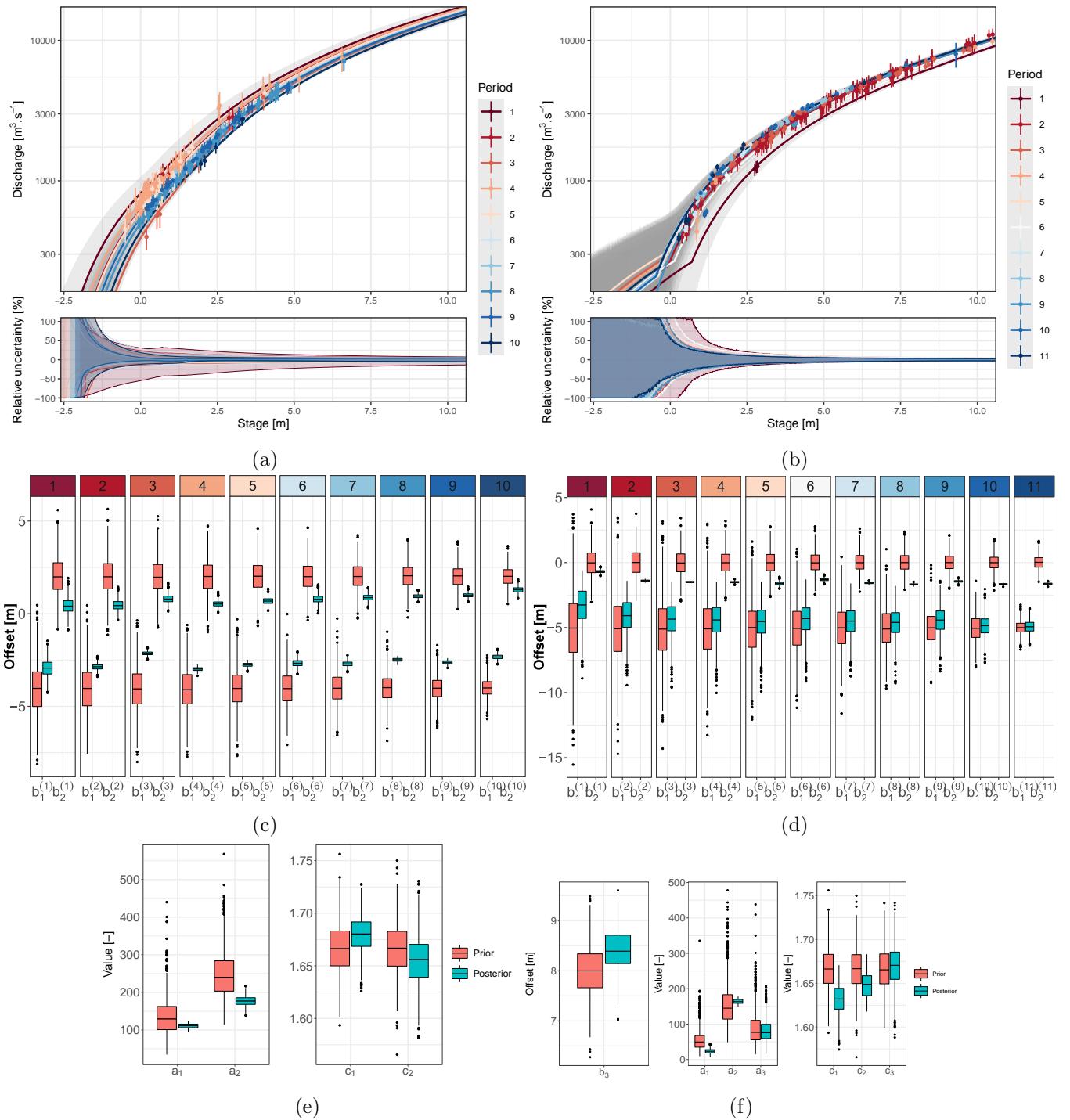


Figure 7: Pont de Beaucaire (a) and Beaucaire Restitution (b) rating curves and relative 95% uncertainty by respect to maxpost, offsets priors and posteriors (c and d) and static parameters priors and posteriors (e and f). Rating curves are in logarithmic scale, solid lines represent maxpost values, grey transparent envelops represent 95% uncertainty intervals and dots with error bars represent gaugings with 95% uncertainty. Stable stage/discharge periods are numbered from the oldest to the latest.

system construction (1967 - 1970), the waters of the Durance River, one of the major tributaries, were diverted from the Rhône River course. AMAX discharges of these missing years were reconstructed by CNR with upstream gauging stations. The uncertainty around these reconstructed discharges is assumed Gaussian distribution with mean being the reconstructed values and standard deviation being 10% of those values (represented in grey).

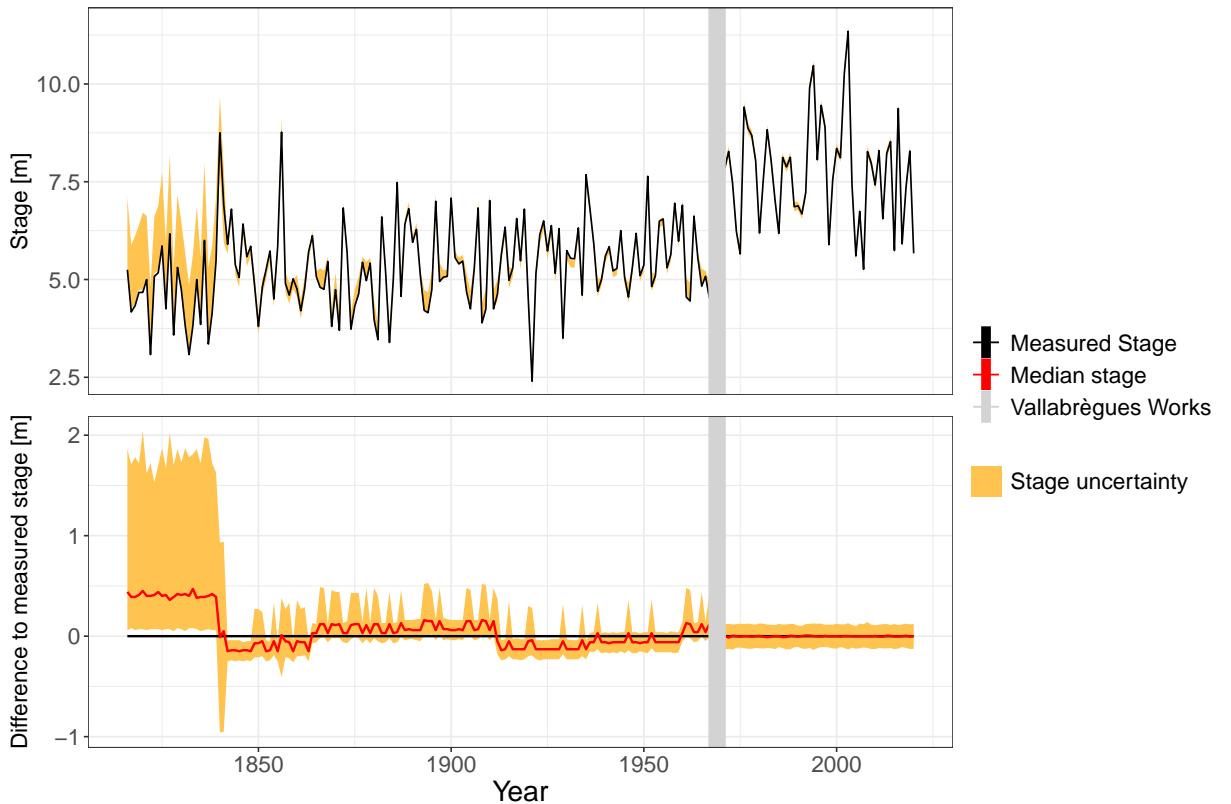


Figure 8: Top : AMAX stages and uncertainty at Beaucaire (1816-2020). Bottom : Difference between stage uncertainty bounds and originally measured stage.

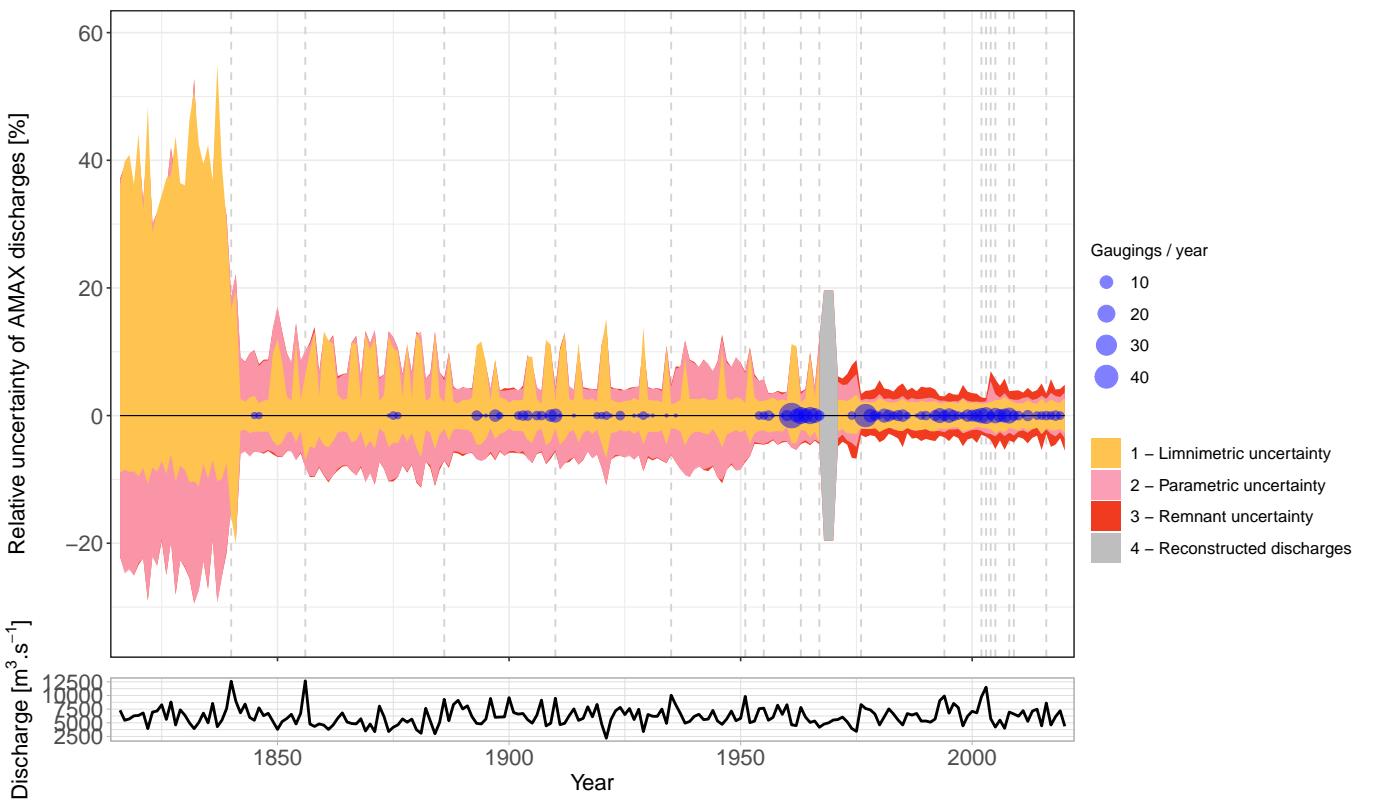


Figure 9: Relative uncertainty of AMAX discharges by respect to the maxpost discharges (black solid line) for the three sources of streamflow uncertainties, at Beaucaire (1816 - 2020). Vertical dotted lines represent rating shifts and the area of the blue dots is proportional to the number of gaugings by year.

4.5 Flood frequency analysis

4.5.1 Streamflow series homogeneity

Streamflow series homogeneity is an essential prerequisite as FFA is based on the hypothesis of iid (independent and identically distributed) random variables. In order to check this hypothesis, the Mann-Kendall non-parametric test (Mann (1945), Kendall (1948)) is applied to AMAX series at Beaucaire. As streamflow uncertainty is represented by $n = 500$ AMAX discharge realisations, the homogeneity test is applied on each of the n AMAX series. Among the n Mann-Kendall tests, 83% concluded in the non-rejection of the null hypothesis (there is no trend in the series) with a 0.05 significance level. We assume that this is enough to consider this series as homogeneous and to proceed to FFA.

4.5.2 Flood frequency analysis

The GEV distribution estimation procedure described in section 2.5 is applied to the 205-year long AMAX discharges series at Beaucaire accounting for uncertainties. Vague priors are given to GEV parameters : flat priors for location and scale parameters, and Gaussian with zero mean and 0.2 standard deviation for the shape parameter. This shape parameter prior is coherent with Martins and Stedinger (2000) suggestions. Flood quantiles results (figure 10a) show that streamflow uncertainty dominates for the lowest return periods, but sampling uncertainty is taking over when the return period tends toward 1000 years (see figure 10c bottom right for a better understanding of the respective part of each source of uncertainty in this 205 years case). The AMAX observed discharges display a large variability of streamflow uncertainties. The three largest floods of this 205 years sample (1840, 1856 and 2003, by chronological order and from the most precise to the most uncertain) illustrate this point. Thus, not considering 1840 and 1856 years could have a large effect on the estimation of the maxpost quantiles value, as well as the uncertainty bound values. This is explored next by varying the sample size.

4.5.3 Sample size influence on quantiles uncertainty

With an exceptionally long sample at Beaucaire, the influence of sample size on flood quantiles estimation in a real case can be quantified, hence assessing the interest of using old hydrometric data when available. Four sample sizes are tested, taken as the last 50 years, 100 years, 150 years, and the largest available sample of 205 years. GEV distributions are estimated and the contribution of both sources of uncertainties is computed for each case, following section 2.5 procedure. Total uncertainty is clearly reduced between the 50 years sample and the other samples for the three return periods: 10, 100 and 1000 years (figure 10b). Surprisingly, for the 1000-year flood estimation, the total uncertainty is not reduced between the 100 and 205 years samples. This illustrates that the reduction of sampling uncertainty induced by increasing the sample size is compensated by the increased streamflow uncertainty when going back in time. Figure 10c is a good illustration of this phenomenon, showing the augmentation of the relative part of streamflow uncertainty when increasing the sample size.

The maxpost values of the 205 years sample is higher than the 100 and 150 years samples (figure 10b), probably because of the inclusion of the two largest floods of the history in the 205 years sample (1840 and 1856 floods). Thus, without using those old hydrometric data although very uncertain (1816-1870), the 1000-year flood could have been 15% lower in this specific case. The sample size impact on flood quantiles estimation is further explored in Figure 10e. The 1000-year flood (maxpost) and the part of both sources of uncertainties are estimated for several sample sizes, from 20 to 205 years, with a two-year step. A large reduction of sampling uncertainty (and of the total uncertainty as a consequence) appears between 20 and 100 years, along with the reduction of the maxpost value. Then, the maxpost and uncertainty bounds are constant between 100 and 160 years of sample size, until the increase due to the inclusion of 1856 and 1840 major floods. The total uncertainty interval width is not much changed by those flood inclusions but the respective contribution of streamflow uncertainty appears more important.

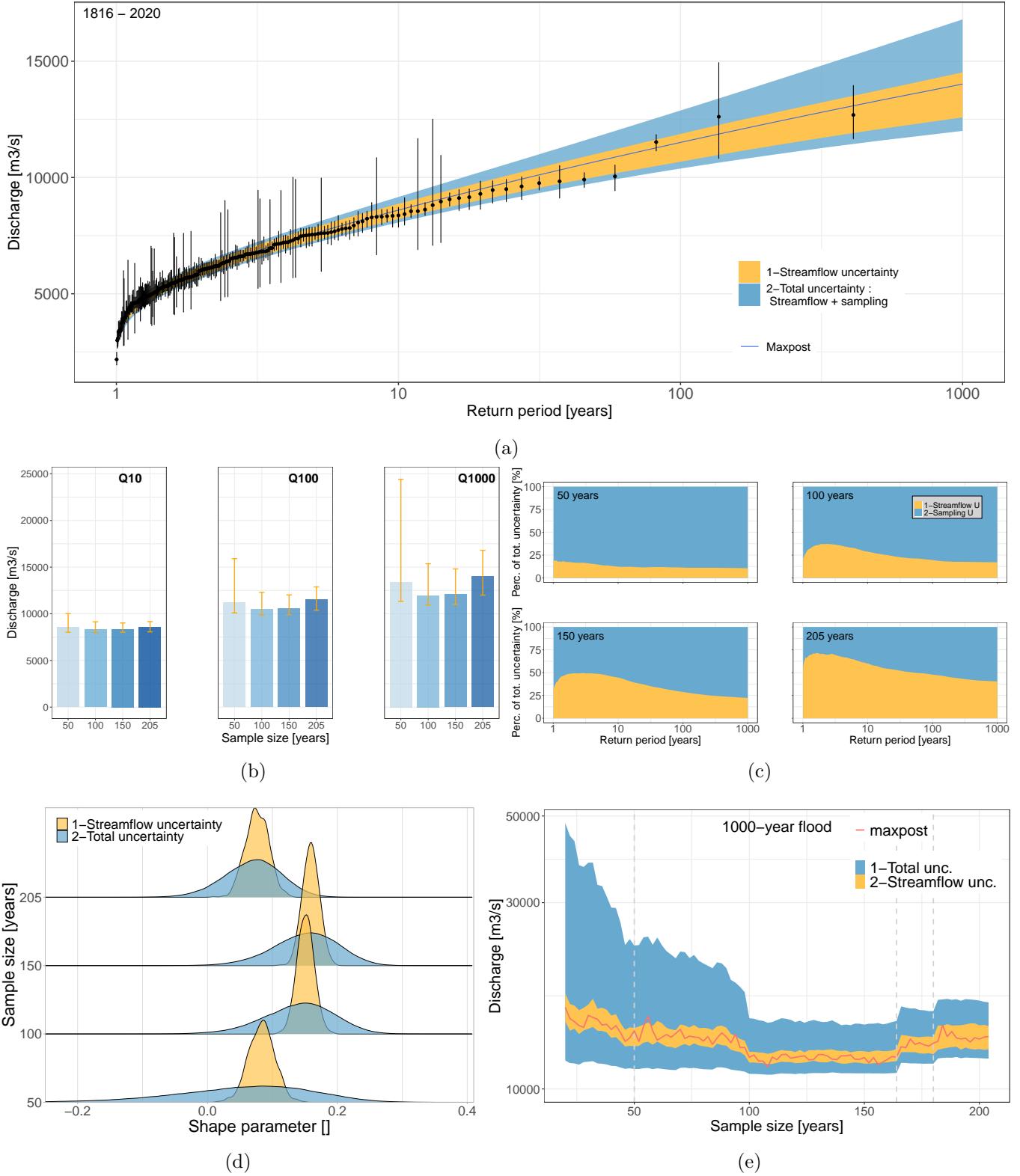


Figure 10: FFA results of the Rhône River at Beaucaire. (a) Fitted GEV distribution on the full sample (1816-2020). Error bars represent the AMAX observed discharges with 95% streamflow uncertainty. (b) Quantiles estimation of three return periods for four sample sizes. 95% total uncertainty envelop is represented by the yellow error bars. (c) Contribution of streamflow and sampling uncertainties to the total uncertainty for four sample sizes. (d) GEV shape parameter distributions considering streamflow or total uncertainty for four sample sizes. (e) 1000-year flood estimations for various sample sizes. Grey dotted lines are representing important changes in the sample data (the change from Pont de Beaucaire to Beaucaire Restitution gauge en 1967, the inclusion of 1856 flood and the inclusion of 1840 flood to the sample).

5 Discussion and limitations of the procedure

5.1 Consequences of including several uncertainty sources in FFA with historical stage records

Even if sampling uncertainty is a major concern while conducting FFA, historical continuous stage records are often forgotten. The laborious process to gather data and the concerns about its reliability may be the potential reasons for this oversight. The proposed approach achieves a complete decomposition of the uncertainty through the FFA chain. Uncertainties are propagated from stage to extreme quantiles. A similar approach has been proposed by Steinbakk et al. (2016), who combined a Bayesian multi-segment rating curve model and FFA. However, their approach is not adapted to century-long series because stage measurements, gaugings and remnant rating curve uncertainties have been neglected. This could lead to the underestimation of extreme quantiles uncertainty. The proposed approach merges recent works on streamflow uncertainty estimation (Horner et al. (2018); Mansanarez et al. (2019a)) and rating changes detection (Darienzo et al., 2021)), into an end-to-end uncertainty propagation, extended up to FFA. A particular attention is given to stage uncertainty, with the consideration of 5 major sources of errors including measurement frequency error that is often omitted. Thus, the contribution of the different sources of uncertainty to flood quantiles uncertainty is calculated. With the availability of a 205 years long streamflow series at Beaucaire, FFA uncertainty for different sample sizes is explored. Petersen-Øverleir and Reitan (2009) highlighted that considering rating curve uncertainty in FFA is enlarging extreme quantiles uncertainty estimates. We show similar results when considering not only rating curve uncertainty, but a larger number of uncertainty sources. We concluded that the contribution of both streamflow and sampling uncertainties varies with sample size (10c). Similarly to Steinbakk et al. (2016) who have used sample sizes up to 100 years, we found that sampling uncertainty is dominant for 50 and 100 years samples, for which the streamflow uncertainty is relatively small. However, the sampling uncertainty reduction when increasing sample size above 100 years is offset by the increase of streamflow uncertainty when considering older and thus more uncertain floods. Although the respective contribution of both sources of uncertainty on the extreme quantiles vary, the width of the total uncertainty interval does not change much from 100 to 205 years samples 10b. Nevertheless, in this case study, as the two largest floods since 1816 occurred during the first 50 years of records, the maxpost value of extreme quantiles is increased by 15% when considering 205 years rather than 150 or 100 years (10e). Not considering the full 205 years sample could lead to a different estimation of the flood hazard, but this 15% difference is slight within the total uncertainty margin of flood quantiles estimates. Thus, the interest of using samples over 100 years in flood hazard evaluation is questionable in this particular case.

5.2 Limitations and potential improvements of the method

The streamflow uncertainty analysis procedure proposed in this work is affected by several limitations. For simplicity purposes, gaugings stage uncertainty is assumed negligible, which is inconsistent with the consideration of stage measurement errors of section 2.3. Horner et al. (2018) proposed a model that accounts for this source of uncertainty in a streamflow uncertainty propagation procedure. The gauging segmentation model proposed by Darienzo et al. (2021) is an efficient tool to detect change points in the stage/discharge relationship, but is limited when dealing with old gauges for which gaugings are often scarce. Selecting a date within the posterior pdf is dubious when very few information are available about rating changes and channel morphology. Additional tests could be used to investigate about stage/discharge relationship, such as the analysis of flood recessions and morphogenic events occurrences proposed by Darienzo (2021). Another limitation of the method comes from the elicitation of hydraulic priors in historical context, for which information about hydraulic configuration is scarce. This lack of knowledge is impacting when considering gauges where works inside the main channel were frequent. The Rhône River is part of those, even if it holds the interest of many historians who retrieved many pieces of hydraulic and morphological data. As described by Petersen-Øverleir and Reitan (2009), this definitely has an impact on flood quantile estimates as the extrapolation of rating curves in this context is more uncertain. Another drawback comes from the fact that, for simplification purposes, we considered the days of annual maximum discharge as the days of annual maximum stage. For a given year, if $h_j > h_k$, it is not guaranteed that $Q_j > Q_k$ if there is an important rating change during this given year. The BaRatin (Le Coz et al., 2014b) based propagation

model choice may also have an important impact on uncertainty estimation, as the seven rating curve uncertainty estimation methods compared by Kiang et al. (2018) concluded in a very large differences in uncertainty intervals.

GEV distribution have been chosen because of its flexibility to describe several tail behaviours. However, many other distributions, as legitimate as GEV in our case-study, might give different results, as explored by Kochanek et al. (2014). Shape parameter is of great importance as it determines the tail behaviour of the GEV distribution. Several regional or local-regional methods have been proposed to reduce the uncertainty around shape parameter estimation. Renard et al. (2013) have proposed a framework to compare those different possibilities. They concluded in particular that the best results were given by local-regional approach, as proposed by Ribatet et al. (2007). This could be a source of improvement, but the definition of regional approach could be difficult for a catchment as large as the Rhône River at Beaucaire. Likewise, the specific conclusions of the Rhône River at Beaucaire may not be generalized in other climatic regions for which the tail behaviour of flood distribution is influenced by different processes (Merz et al., 2022).

5.3 Are historical stage records useful or dangerous for flood frequency analysis?

The interest of including historical stage records to reduce the flood quantiles sampling uncertainty can be balanced by the streamflow uncertainty of those records. For the specific case of Beaucaire, the use of historical stage records have shown a real interest up to a 100 years sample size, but their use are questionable above. Evaluating the procedure on several different gauges with historical stage records to have a better estimation of the added value of long streamflow series in FFA could be an interesting point in order to generalize the results. The reduction of flood quantiles uncertainties when increasing sample size is not specific to the Rhône River at Beaucaire, however, the respective contribution of streamflow and sampling uncertainties could be different for gauges with better or worse historical hydrometric data. Likewise, although the homogeneity of AMAX discharges has been checked, 82% of the 500 AMAX series representing streamflow uncertainty conducted to reject the hypothesis of trend in the series. The underlying assumption of stationarity required for FFA may have been corrupted by anthropogenic climate change, as described by Milly et al. (2008). Trends have been identified in several climatic regions of Europe (Hall et al. (2014), Blöschl et al. (2017) and France (Renard et al., 2008), but no general trend was found. The Rhône River catchment at Beaucaire is at the crossroad of several climatic regions for which trends are opposed. Also, Madsen et al. (2014) underlined that no particular guideline for climate change adjustment factors on design floods are given in France.

Finally, promising improvements could come from the further valorization of historical data by investigating flood evidences before systematic stage measurements. Various procedures have been developed in the literature, through the use of perception threshold and censored data as summarized by Kjeldsen et al. (2014) or Brázdil et al. (2006). Applications of this kind of approach have emerged in Europe, including France (Naulet et al. (2005); Lang et al. (2010); Neppel et al. (2010); Payrastre et al. (2011)) and could be interesting for the Lower-Rhône Valley for which many climatic and flood evidences have been gathered. Pichard and Roucaute (2014) and Pichard et al. (2017) have identified more than 1500 hydro-climatic events in the Lower-Rhône Valley since the XIVth Century, synthesized in the HISTRHONE database (histrhone.cerege.fr). The flood events are classified by magnitude of damages, and further investigations are required to identify the perception threshold corresponding to those different magnitudes.

6 Conclusion

Flood hazard estimation is affected by several sources of uncertainties, including sampling uncertainty that is dominant for usual sample sizes (less than 100 years). It is sometimes possible to gather historical continuous stage measurements in order to enlarge flood samples above usual sizes. This process sounds intuitive for reducing the uncertainties of flood quantiles estimates. Nevertheless, historical streamflow series are generally affected by much greater uncertainties than modern series. This paper investigates the following questions: to what extent including historical (and thus uncertain) hydrometric data is useful

to improve FFA estimates, and what is the contribution of streamflow and sampling uncertainties to the total FFA uncertainty ? Those questions are explored through a general framework for FFA accounting for uncertainties in the specific case of long discharge series. This uncertainty propagation chain is applied to a 205 years long continuous stage series of the Rhône River at Beaucaire.

The estimated streamflow uncertainty varies from $\pm 30\%$ (XIXth Century) to $\pm 5\%$ (1967-2020). This uncertainty is propagated towards flood quantiles estimates. For the full flood sample (205 years), streamflow uncertainty is dominant below the 100-year flood and sampling uncertainty is dominant above. Sample size influence on flood quantiles estimates is explored through the use of samples with varying sizes. The total uncertainty of flood quantiles is substantially decreasing from 20 years to 100 years samples. This decrease is directly induced by sample uncertainty reduction. Between 100 years and 205 years sample sizes, the total uncertainty is nearly constant because the sampling uncertainty reduction is offset by the increase of streamflow uncertainty (older floods discharges are more uncertain). In contrast, central flood quantiles estimates are increasing of 15% when increasing sample size between 160 and 205 years, because of the inclusion of the two largest floods that occurred during the first 40 years of measurements. However, this 15% increase is almost negligible regarding the total uncertainty. Finally, in the specific case of the Rhône River at Beaucaire, the added value of lengthening the flood sample size with historical stage measurements is confirmed at least up to a 100 years long sample size. Beyond 100 years, the benefit is questionable because of the major increase of streamflow uncertainties and the non-reduction of sampling uncertainty. However, those results promote the use of historical stage records to improve design flood estimates and underline the particular importance of estimating and propagating all sources of uncertainty through the estimation processp.

Several steps of the procedure could be improved, such as as further investigations of rating changes through the analysis of floods recessions or morphogenic events, as well as the consideration of gaugings stage uncertainty. Further, the probabilistic model for FFA could be improved by considering a regional or local-regional approach. Finally, the added value of historical data in FFA could be more deeply explored by the combined use of flood evidences and historical continuous stage measurements. The treatment of mixed samples (continuous stage measurements and punctual flood evidences) in FFA has already been studied but the uncertainty around perception thresholds and time coverage is less well-known.

7 Acknowledgements

H2O + ENS Lyon + CNR ? OSR ?

Fournir un lien GITHUB vers code et données ? <https://github.com/MatLcs/PropagMaxAn>

References

- Armand (1907). “II. Le Rhône à Tarascon (Planche I)”. In: *Revue des Études Anciennes* 9.1, pp. 19–21. ISSN: 0035-2004. DOI: [10.3406/rea.1907.1469](https://doi.org/10.3406/rea.1907.1469).
- Bard, A. and M. Lang (2018). *Actualisation de l'hydrologie des crues du Rhône. Analyse de la station de Beaucaire*. Irstea pour la DREAL Auvergne-Rhône-Alpes, p. 56.
- Beven, K. and J. Hall (2014). *Applied Uncertainty Analysis for Flood Risk Management*. Imperial College Press. ISBN: 978-1-84816-270-9. DOI: [10.1142/p588](https://doi.org/10.1142/p588).
- Blöschl, G., J. Hall, J. Parajka, R. A. P. Perdigão, B. Merz, B. Arheimer, G. T. Aronica, A. Bilibashi, O. Bonacci, M. Borga, I. Čanjevac, A. Castellarin, G. B. Chirico, P. Claps, K. Fiala, N. Frolova, L. Gorbachova, A. Gül, J. Hannaford, S. Harrigan, M. Kireeva, A. Kiss, T. R. Kjeldsen, S. Kohnová, J. J. Koskela, O. Ledvinka, N. Macdonald, M. Mavrova-Guirguinova, L. Mediero, R. Merz, P. Molnar, A. Montanari, C. Murphy, M. Osuch, V. Ovcharuk, I. Radevski, M. Rogger, J. L. Salinas, E. Sauquet, M. Šraj, J. Szolgay, A. Viglione, E. Volpi, D. Wilson, K. Zaimi, and N. Živković (2017). “Changing climate shifts timing of European floods”. In: *Science* 357.6351. Publisher: American Association for the Advancement of Science, pp. 588–590. DOI: [10.1126/science.aan2506](https://doi.org/10.1126/science.aan2506).
- Brázdil, R., Z. W. Kundzewicz, and G. Benito (2006). “Historical hydrology for studying flood risk in Europe”. In: *Hydrological Sciences Journal* 51.5, pp. 739–764. ISSN: 0262-6667, 2150-3435. DOI: [10.1623/hysj.51.5.739](https://doi.org/10.1623/hysj.51.5.739).
- CETIAT (2005). *Conférence de consensus sur le débit du Rhône à Beaucaire pour la crue de Décembre 2003. Annexe V : Estimation des incertitudes des débits calculés à partir des relations hauteur/débit*. pour le compte de la CNR. Estimation des incertitudes des débits calculés à partir des relations hauteur/débit.
- Coxon, G., J. Freer, I. K. Westerberg, T. Wagener, R. Woods, and P. J. Smith (2015). “A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations”. In: *Water Resources Research* 51.7, pp. 5531–5546. ISSN: 0043-1397, 1944-7973. DOI: [10.1002/2014WR016532](https://doi.org/10.1002/2014WR016532).
- Darienzo, M., B. Renard, J. Le Coz, and M. Lang (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](https://doi.org/10.1029/2020WR028607).
- Darienzo, M. (2021). “Detection and estimation of stage-discharge rating shifts for retrospective and real-time streamflow quantification”. These de doctorat. Université Grenoble Alpes.
- Dymond, J. R. and R. Christian (1982). “Accuracy of discharge determined from a rating curve”. In: *Hydrological Sciences Journal* 27.4, pp. 493–504. ISSN: 0262-6667, 2150-3435. DOI: [10.1080/02626668209491128](https://doi.org/10.1080/02626668209491128).
- Guerrero, J.-L., I. K. Westerberg, S. Halldin, C.-Y. Xu, and L.-C. Lundin (2012). “Temporal variability in stage-discharge relationships”. In: *Journal of Hydrology* 446-447, pp. 90–102. ISSN: 00221694. DOI: [10.1016/j.jhydrol.2012.04.031](https://doi.org/10.1016/j.jhydrol.2012.04.031).
- Hall, J., B. Arheimer, M. Borga, R. Brázdil, P. Claps, A. Kiss, T. R. Kjeldsen, J. Kriaučičūnienė, 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, and G. Blöschl (2014). “Understanding flood regime changes in Europe: a state-of-the-art assessment”. In: *Hydrology and Earth System Sciences* 18.7, pp. 2735–2772. ISSN: 1607-7938. DOI: [10.5194/hess-18-2735-2014](https://doi.org/10.5194/hess-18-2735-2014).
- Hamed, K. H. and A. Ramachandra Rao (2019). *Flood Frequency Analysis*. 1st ed. CRC Press. ISBN: 978-0-429-12881-3. DOI: [10.1201/9780429128813](https://doi.org/10.1201/9780429128813).
- Hamilton, A. and R. Moore (2012). “Quantifying Uncertainty in Streamflow Records”. In: *Canadian Water Resources Journal / Revue canadienne des ressources hydriques* 37.1, pp. 3–21. ISSN: 0701-1784, 1918-1817. DOI: [10.4296/cwrj3701865](https://doi.org/10.4296/cwrj3701865).
- Herschy, R. (1998). *Hydrometry Principles and Practices*. 2nd ed. John Wiley. 384 pp.
- Horner, I., B. Renard, J. Le Coz, F. Branger, H. K. McMillan, and G. Pierrefeu (2018). “Impact of Stage Measurement Errors on Streamflow Uncertainty”. In: *Water Resources Research* 54.3, pp. 1952–1976. ISSN: 0043-1397, 1944-7973. DOI: [10.1002/2017WR022039](https://doi.org/10.1002/2017WR022039).
- Ibbitt, R. P. and C. P. Pearson (1987). “Gauging frequency and detection of rating changes”. In: *Hydrological Sciences Journal* 32.1, pp. 85–103. ISSN: 0262-6667, 2150-3435. DOI: [10.1080/02626668709491164](https://doi.org/10.1080/02626668709491164).

- Jain, S. K. and V. P. Singh (2019). "Design Flood Estimation". In: *Engineering Hydrology: An Introduction to Processes, Analysis, and Modeling*. First edition. New York: McGraw-Hill Education. ISBN: 978-1-259-64197-8.
- Juston, J., P.-E. Jansson, and D. Gustafsson (2014). "Rating curve uncertainty and change detection in discharge time series: case study with 44-year historic data from the Nyangores River, Kenya". In: *Hydrological Processes* 28.4, pp. 2509–2523. ISSN: 08856087. DOI: [10.1002/hyp.9786](https://doi.org/10.1002/hyp.9786).
- Kendall, M. (1948). *Rank Correlation Methods*. 1. London: Charles Griffin & Company.
- Kiang, J. E., C. Gazoorian, H. McMillan, G. Coxon, J. L. Coz, I. K. Westerberg, A. Belleville, D. Sevrez, A. E. Sikorska, A. Petersen-Øverleir, T. Reitan, J. Freer, B. Renard, V. Mansanarez, and R. Mason (2018). "A Comparison of Methods for Streamflow Uncertainty Estimation". In: *Water Resources Research* 54.10, pp. 7149–7176. DOI: <https://doi.org/10.1029/2018WR022708>.
- Kjeldsen, T., N. Macdonald, M. Lang, L. Mediero, T. Albuquerque, E. Bogdanowicz, R. Brázdil, A. Castellarin, V. David, A. Fleig, G. Gül, J. Kriauciuniene, S. Kohnová, B. Merz, O. Nicholson, L. Roald, J. Salinas, D. Sarauskiene, M. Šraj, W. Strupczewski, J. Szolgay, A. Toumazis, W. Vanneuville, N. Veijalainen, and D. Wilson (2014). "Documentary evidence of past floods in Europe and their utility in flood frequency estimation". In: *Journal of Hydrology* 517, pp. 963–973. ISSN: 00221694. DOI: [10.1016/j.jhydrol.2014.06.038](https://doi.org/10.1016/j.jhydrol.2014.06.038).
- Kjeldsen, T. R., R. Lamb, and S. D. Blazkova (2011). "Uncertainty in Flood Frequency Analysis". In: *Applied Uncertainty Analysis for Flood Risk Management*. Imperial college press, pp. 153–197. ISBN: 978-1-84816-270-9. DOI: [10.1142/9781848162716_0008](https://doi.org/10.1142/9781848162716_0008).
- Kochanek, K., B. Renard, P. Arnaud, Y. Aubert, M. Lang, T. Cipriani, and E. Sauquet (2014). "A data-based comparison of flood frequency analysis methods used in France". In: *Natural Hazards and Earth System Sciences* 14.2, pp. 295–308. ISSN: 1684-9981. DOI: [10.5194/nhess-14-295-2014](https://doi.org/10.5194/nhess-14-295-2014).
- Kuentz, A., T. Mathevret, D. Cœur, C. Perret, J. Gailhard, L. Guérin, Y. Gash, and V. Andréassian (2014). "Hydrométrie et hydrologie historiques du bassin de la Durance". In: *La Houille Blanche* 100.4, pp. 57–63. ISSN: 0018-6368, 1958-5551. DOI: [10.1051/lhb/2014039](https://doi.org/10.1051/lhb/2014039).
- Lang, M. and D. Coeur (2014). *Les inondations remarquables en France. Inventaire 2011 pour la directive Inondation*. 1st ed. Hors Collection. Editions Quae. 640 pp. ISBN: 978-2-7592-2260-5.
- Lang, M., K. Pobanz, B. Renard, E. Renouf, and E. Sauquet (2010). "Extrapolation of rating curves by hydraulic modelling, with application to flood frequency analysis". In: *Hydrological Sciences Journal* 55.6, pp. 883–898. ISSN: 0262-6667, 2150-3435. DOI: [10.1080/02626667.2010.504186](https://doi.org/10.1080/02626667.2010.504186).
- Lapuszek, M. and A. Lenar-Matyas (2015). "Methods of analysis the riverbed evolution. a case study of two tributaries of the upper Vistula river". In: *Infrastruktura i Ekologia Terenów Wiejskich / Infrastructure and ecology of rural areas* (IV/3/2015), pp. 1313–1327. ISSN: 1732-5587. DOI: [10.14597/infraeco.2015.4.3.095](https://doi.org/10.14597/infraeco.2015.4.3.095).
- Le Coz, J., P.-M. Bechon, B. Camenen, and G. Dramais (2014a). "Quantification des incertitudes sur les jaugeages par exploration du champ des vitesses". In: *La Houille Blanche* 100.5, pp. 31–39. ISSN: 0018-6368, 1958-5551. DOI: [10.1051/lhb/2014047](https://doi.org/10.1051/lhb/2014047).
- Le Coz, J., B. Renard, L. Bonnifait, F. Branger, and R. Le Boursicaud (2014b). "Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves: A Bayesian approach". In: *Journal of Hydrology* 509, pp. 573–587. ISSN: 00221694. DOI: [10.1016/j.jhydrol.2013.11.016](https://doi.org/10.1016/j.jhydrol.2013.11.016).
- Madsen, H., D. Lawrence, M. Lang, M. Martinkova, and T. R. Kjeldsen (2014). "Review of trend analysis and climate change projections of extreme precipitation and floods in Europe". In: *Journal of Hydrology* 519, pp. 3634–3650. ISSN: 0022-1694. DOI: [10.1016/j.jhydrol.2014.11.003](https://doi.org/10.1016/j.jhydrol.2014.11.003).
- Mann, H. B. (1945). "Nonparametric Tests Against Trend". In: *Econometrica* 13.3. Publisher: [Wiley, Econometric Society], pp. 245–259. ISSN: 0012-9682. DOI: [10.2307/1907187](https://doi.org/10.2307/1907187).
- Mansanarez, V., B. Renard, J. L. Coz, M. Lang, and M. Darienzo (2019a). "Shift Happens! Adjusting Stage-Discharge Rating Curves to Morphological Changes at Known Times". In: *Water Resources Research* 55.4, pp. 2876–2899. ISSN: 0043-1397, 1944-7973. DOI: [10.1029/2018WR023389](https://doi.org/10.1029/2018WR023389).
- Mansanarez, V., I. K. Westerberg, N. Lam, and S. W. Lyon (2019b). "Rapid Stage-Discharge Rating Curve Assessment Using Hydraulic Modeling in an Uncertainty Framework". In: *Water Resources Research* 55.11, pp. 9765–9787. ISSN: 0043-1397, 1944-7973. DOI: [10.1029/2018WR024176](https://doi.org/10.1029/2018WR024176).

- Martins, E. S. and J. R. Stedinger (2000). "Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data". In: *Water Resources Research* 36.3, pp. 737–744. ISSN: 00431397. DOI: [10.1029/1999WR900330](https://doi.org/10.1029/1999WR900330).
- McMillan, H. K. and I. K. Westerberg (2015). "Rating curve estimation under epistemic uncertainty". In: *Hydrological Processes* 29.7, pp. 1873–1882. ISSN: 08856087. DOI: [10.1002/hyp.10419](https://doi.org/10.1002/hyp.10419).
- McMillan, H., J. Freer, F. Pappenberger, T. Krueger, and M. Clark (2010). "Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions". In: 24. DOI: <https://doi.org/10.1002/hyp.7587>.
- McMillan, H., T. Krueger, and J. Freer (2012). "Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality". In: *Hydrological Processes* 26.26, pp. 4078–4111. ISSN: 08856087. DOI: [10.1002/hyp.9384](https://doi.org/10.1002/hyp.9384).
- MEDD (2005). *Débit du Rhône à Beaucaire pour la crue de Décembre 2003*. Conférence de consensus. Minitère français de l'écologie et du développement durable.
- Merz, B., S. Basso, S. Fischer, D. Lun, G. Blöschl, R. Merz, B. Guse, A. Viglione, S. Vorogushyn, E. Macdonald, L. Wietzke, and A. Schumann (2022). "Understanding Heavy Tails of Flood Peak Distributions". In: *Water Resources Research* 58.6. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021WR030506>, e2021WR030506. ISSN: 1944-7973. DOI: [10.1029/2021WR030506](https://doi.org/10.1029/2021WR030506).
- Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer (2008). "Stationarity Is Dead: Whither Water Management?" In: *Science* 319.5863, pp. 573–574. ISSN: 0036-8075, 1095-9203. DOI: [10.1126/science.1151915](https://doi.org/10.1126/science.1151915).
- Morlot, T., C. Perret, A.-C. Favre, and J. Jalbert (2014). "Dynamic rating curve assessment for hydrometric stations and computation of the associated uncertainties: Quality and station management indicators". In: *Journal of Hydrology* 517, pp. 173–186. ISSN: 00221694. DOI: [10.1016/j.jhydrol.2014.05.007](https://doi.org/10.1016/j.jhydrol.2014.05.007).
- Naulet, R., M. Lang, T. B. Ouarda, D. Coeur, B. Bobée, A. Recking, and D. Moussay (2005). "Flood frequency analysis on the Ardèche river using French documentary sources from the last two centuries". In: *Journal of Hydrology* 313.1, pp. 58–78. ISSN: 00221694. DOI: [10.1016/j.jhydrol.2005.02.011](https://doi.org/10.1016/j.jhydrol.2005.02.011).
- Neppel, L., B. Renard, M. Lang, P.-A. Ayral, D. Coeur, E. Gaume, N. Jacob, O. Payrastre, K. Pobanz, and F. Vinet (2010). "Flood frequency analysis using historical data: accounting for random and systematic errors". In: *Hydrological Sciences Journal* 55.2, pp. 192–208. ISSN: 0262-6667, 2150-3435. DOI: [10.1080/02626660903546092](https://doi.org/10.1080/02626660903546092).
- Payrastre, O., E. Gaume, and H. Andrieu (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](https://doi.org/10.1029/2010WR009812).
- Petersen-Øverleir, A. and T. Reitan (2005). "Uncertainty in flood discharges from urban and small rural catchments due to inaccurate head measurement". In: *Hydrology Research* 36.3, pp. 245–257. ISSN: 0029-1277, 2224-7955. DOI: [10.2166/nh.2005.0018](https://doi.org/10.2166/nh.2005.0018).
- Petersen-Øverleir, A. and T. Reitan (2009). "Accounting for rating curve imprecision in flood frequency analysis using likelihood-based methods". In: *Journal of Hydrology* 366.1, pp. 89–100. ISSN: 00221694. DOI: [10.1016/j.jhydrol.2008.12.014](https://doi.org/10.1016/j.jhydrol.2008.12.014).
- Petersen-Øverleir, A., A. Soot, and T. Reitan (2009). "Bayesian Rating Curve Inference as a Streamflow Data Quality Assessment Tool". In: *Water Resources Management* 23.9, pp. 1835–1842. ISSN: 0920-4741, 1573-1650. DOI: [10.1007/s11269-008-9354-5](https://doi.org/10.1007/s11269-008-9354-5).
- Pichard, G., G. Arnaud-Fassetta, V. Moron, and E. Roucaute (2017). "Hydro-climatology of the Lower Rhône Valley: historical flood reconstruction (AD 1300–2000) based on documentary and instrumental sources". In: *Hydrological Sciences Journal* 62.11, pp. 1772–1795. ISSN: 0262-6667, 2150-3435. DOI: [10.1080/02626667.2017.1349314](https://doi.org/10.1080/02626667.2017.1349314).
- Pichard, G. (2013). *Hauteurs et altitudes aux échelles du bas Rhône*, p. 48.
- Pichard, G. and E. Roucaute (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>.
- Ponts&Chaussées (1914). *Observations hydrométriques du Rhône à Beaucaire, 1914*.
- Puechberty, R., C. Perret, S. P. Pitsch, P. Battaglia, A. Belleville, P. Bompard, G. Chauvel, J. Cousseau, G. Dramais, G. Glaziou, A. Hauet, S. Helouin, M. Lang, F. Larrarte, J. L. Coz, P. Marchand, P. Moquet, O.

- Payrastre, P. Pierrefeu, and G. Rauzy (2017). *Charte qualité de l'hydrométrie. Guide de bonnes pratiques*. Ministère de l'environnement, de l'énergie et de la mer, France. 86 pp.
- Renard, B., V. Garreta, and M. Lang (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](https://doi.org/10.1029/2005WR004591).
- Renard, B., K. Kochanek, M. Lang, F. Garavaglia, E. Paquet, L. Neppel, K. Najib, J. Carreau, P. Arnaud, Y. Aubert, F. Borchi, J.-M. Soubeyroux, S. Jourdain, J.-M. Veysseire, E. Sauquet, T. Cipriani, and A. Auffray (2013). “Data-based comparison of frequency analysis methods: A general framework”. In: *Water Resources Research* 49.2, pp. 825–843. ISSN: 00431397. DOI: [10.1002/wrcr.20087](https://doi.org/10.1002/wrcr.20087).
- Renard, B., M. Lang, P. Bois, A. Dupeyrat, O. Mestre, H. Niel, E. Sauquet, C. Prudhomme, S. Parey, E. Paquet, L. Neppel, and J. Gailhard (2008). “Regional methods for trend detection: Assessing field significance and regional consistency”. In: *Water Resources Research* 44.8. ISSN: 00431397. DOI: [10.1029/2007WR006268](https://doi.org/10.1029/2007WR006268).
- Ribatet, M., E. Sauquet, J.-M. Grésillon, and T. B. M. J. Ouarda (2007). “A regional Bayesian POT model for flood frequency analysis”. In: *Stochastic Environmental Research and Risk Assessment* 21.4, pp. 327–339. ISSN: 1436-3240, 1436-3259. DOI: [10.1007/s00477-006-0068-z](https://doi.org/10.1007/s00477-006-0068-z).
- Rigaudière, P., A. Dekergariou, C. Laroche, N. Caze, J. Brun, J. Laborde, and W. Berolo (2000). *Etude globale des crues du Rhône*. Tech. rep. SAFEGE Cetiis and Université de Nice.
- Steinbakk, G. H., T. L. Thorarinsdottir, T. Reitan, L. Schlichting, S. Hølleland, and K. Engeland (2016). “Propagation of rating curve uncertainty in design flood estimation”. In: *Water Resources Research* 52.9, pp. 6897–6915. ISSN: 00431397. DOI: [10.1002/2015WR018516](https://doi.org/10.1002/2015WR018516).
- SYMADREM (2012). *Programme de sécurisation des ouvrages de protection contre les crues du Rhône du barrage de Vallabrégnes à la mer*, p. 490.
- Van Der Made, J. (1982). “Determination of the accuracy of water level observations”. In: *IAHS Publications. Proceedings of the Exeter Symposium* 134, pp. 172–184.
- Vieira, L. M. d. S., J. C. L. Sampaio, V. A. F. Costa, and J. C. Eleutério (2022). “Assessing the effects of rating curve uncertainty in flood frequency analysis”. In: *RBRH* 27, e11. ISSN: 2318-0331, 1414-381X. DOI: [10.1590/2318-0331.272220220012](https://doi.org/10.1590/2318-0331.272220220012).
- Westerberg, I., J.-L. Guerrero, J. Seibert, K. J. Beven, and S. Halldin (2011). “Stage-discharge uncertainty derived with a non-stationary rating curve in the Choluteca River, Honduras”. In: *Hydrological Processes* 25.4, pp. 603–613. ISSN: 08856087. DOI: [10.1002/hyp.7848](https://doi.org/10.1002/hyp.7848).

8 Annexes

8.1 Rating curve hydraulic priors

Physical param.	Meaning	Prior	Inferred param.	Prior
Control 1: main channel				
$b_1[m]$	Offset	$\mathcal{N}(-4, 0.5)$	$b_1[m]$	$\mathcal{N}(-4, 0.5)$
$B_1[m]$	Channel width	$\mathcal{LN}(ln(300), 0.16)$	$a_1[m^{3/2}/s]$	$\mathcal{LN}(ln(128.6), 1.8 \cdot 10^{-2})$
$K_1[m^{1/3}/s]$	Strickler coeff.	$\mathcal{LN}(ln(35), 0.14)$		
$S_1[m/m]$	Bed slope	$\mathcal{LN}(ln(1.5 \cdot 10^{-4}), 0.55)$		
$c_1[-]$	Exponent	$\mathcal{N}(5/3, 0.025)$	$c_1[-]$	$\mathcal{N}(5/3, 0.025)$
Control 2: floodway				
$b_2[m]$	Offset	$\mathcal{N}(1.5, 0.5)$	$b_2[m]$	$\mathcal{N}(1.5, 0.5)$
$B_2[m]$	Channel width	$\mathcal{LN}(ln(500), 0.1)$	$a_2[m^{3/2}/s]$	$\mathcal{LN}(ln(241.9), 1.10^{-2})$
$K_2[m^{1/3}/s]$	Strickler coeff.	$\mathcal{LN}(ln(30), 0.16)$		
$S_2[m/m]$	Bed slope	$\mathcal{LN}(ln(2.6 \cdot 10^{-4}), 0.34)$		
$c_2[-]$	Exponent	$\mathcal{N}(5/3, 0.025)$	$c_2[-]$	$\mathcal{N}(5/3, 0.025)$
Structural uncertainty parameters				
$\gamma_1[m^3/s]$	Intercept	$\mathcal{U}(0, 1000)$	$\gamma_1[m^3/s]$	$\mathcal{U}(0, 1000)$
$\gamma_2[-]$	Slope	$\mathcal{U}(0, 100)$	$\gamma_2[-]$	$\mathcal{U}(0, 100)$
Multiperiod RC parameters				
$\delta l[m]$	Local change	$\mathcal{N}(0, 0.3)$	$\delta l[m]$	$\mathcal{N}(0, 0.3)$
$\delta g[m]$	Global change	$\mathcal{N}(0, 0.3)$	$\delta g[m]$	$\mathcal{N}(0, 0.3)$

Table 3: Priors elicitation for Pont de Beaucaire. \mathcal{U} stands for continuous uniform distribution, \mathcal{N} for Normal distribution and \mathcal{LN} for Log Normal distribution

8.2 Gaugings segmentation detailed results

Physical param.	Meaning	Prior	Inferred param.	Prior
Control 1: Low flows channel				
$b_1[m]$	Offset	$\mathcal{N}(-5, 0.5)$	$b_1[m]$	$\mathcal{N}(-5, 0.5)$
$B_1[m]$	Channel width	$\mathcal{LN}(\ln(200), 0.42)$	$a_1[m^{3/2}/s]$	$\mathcal{LN}(\ln(49.50), 3.2 \cdot 10^{-2})$
$K_1[m^{1/3}/s]$	Strickler coeff.	$\mathcal{LN}(\ln(35), 0.14)$		
$S_1[m/m]$	Bed slope	$\mathcal{LN}(\ln(5.10^{-5}), 0.20)$		
$c_1[-]$	Exponent	$\mathcal{N}(5/3, 0.025)$	$c_1[-]$	$\mathcal{N}(5/3, 0.025)$
Control 2: Main channel				
$b_1[m]$	Offset	$\mathcal{N}(0, 0.5)$	$b_1[m]$	$\mathcal{N}(0, 0.5)$
$B_1[m]$	Channel width	$\mathcal{LN}(\ln(300), 0.32)$	$a_2[m^{3/2}/s]$	$\mathcal{LN}(\ln(148.49), 2.4 \cdot 10^{-2})$
$K_1[m^{1/3}/s]$	Strickler coeff.	$\mathcal{LN}(\ln(35), 0.14)$		
$S_1[m/m]$	Bed slope	$\mathcal{LN}(\ln(2.10^{-4}), 0.25)$		
$c_1[-]$	Exponent	$\mathcal{N}(5/3, 0.025)$	$c_2[-]$	$\mathcal{N}(5/3, 0.025)$
Control 3: Floodway				
$b_3[m]$	Offset	$\mathcal{N}(8, 0.5)$	$b_3[m]$	$\mathcal{N}(8, 0.5)$
$B_3[m]$	Channel width	$\mathcal{LN}(\ln(200), 0.47)$	$a_3[m^{3/2}/s]$	$\mathcal{LN}(\ln(241.9), 1.10^{-2})$
$K_3[m^{1/3}/s]$	Strickler coeff.	$\mathcal{LN}(\ln(25), 0.20)$		
$S_3[m/m]$	Bed slope	$\mathcal{LN}(\ln(2.4 \cdot 10^{-4}), 0.21)$		
$c_3[-]$	Exponent	$\mathcal{N}(5/3, 0.025)$	$c_3[-]$	$\mathcal{N}(5/3, 0.025)$
Structural uncertainty parameters				
$\gamma_1[m^3/s]$	Intercept	$\mathcal{U}(0, 1000)$	$\gamma_1[m^3/s]$	$\mathcal{U}(0, 1000)$
$\gamma_2[-]$	Slope	$\mathcal{U}(0, 100)$	$\gamma_2[-]$	$\mathcal{U}(0, 100)$
Multiperiod RC parameters				
$\delta l[m]$	Local change	$\mathcal{N}(0, 0.8)$	$\delta l[m]$	$\mathcal{N}(0, 0.8)$
$\delta g[m]$	Global change	$\mathcal{N}(0, 0.3)$	$\delta g[m]$	$\mathcal{N}(0, 0.3)$

Table 4: Priors elicitation for Beaucaire reconstitution. \mathcal{U} stands for continuous uniform distribution, \mathcal{N} for Normal distribution and \mathcal{LN} for Log Normal distribution

Maxpost shift time	Largest flood within <i>post. pdf</i>	Final choice	Period number	Number of gaugings
No gaugings	No gaugings	1840-11-02	1	0
1860-02-20	1856-06-01	1856-06-01	2	4
1887-05-11	1886-10-29	1886-10-29	3	6
1910-11-21	1910-12-09	1910-12-09	4	58
1921-06-22	1935-11-14	1935-11-14	6	22
1954-08-08	1951-11-23	1951-11-23	6	1
1954-03-30	1955-01-23	1955-01-23	7	3
1963-03-23	1963-11-08	1963-11-08	8	91
1967-01-31	1967-01-31	1967-01-31	9	43
1967-12-31	End of stage series	End of stage series	10	5
1975-02-06	1976-11-11	1976-11-11	1	3
1994-06-10	1994-01-08	1994-01-08	2	122
2003-08-05	2002-11-27	2002-11-27	3	65
2003-10-09	2003-12-04	2003-12-04	4	17
2004-02-16	2003-12-04	No shift	X	X
2004-07-15	2003-12-04	2004-07-15	5	1
2004-12-02	2004-12-02	2004-12-02	6	2
2005-07-03	2004-12-02	No shift	X	X
2005-12-24	2003-12-04	2005-12-24	7	14
2008-06-28	2003-12-04	2008-06-28	8	28
2009-10-20	2003-12-04	2009-10-20	9	7
2016-11-21	2016-11-22	2016-11-22	10	26
2019-02-11	2018-11-24	No shift	X	X
2020-01-01	End of stage series	End of stage series	11	11

Table 5: Beaucaire rating shifts dates