Extreme sea level estimation combining systematic observed skew surges and historical record sea levels

Laurie Saint Criq [1,3], Eric Gaume [2], Yasser Hamdi [1], Taha B. M. J. Ouarda [3]

[1]Institut de Radioprotection et de Sûreté Nucléaire, France
[2]Université Gustave Eiffel, France
[3]Institut National de la Recherche Scientifique, Canada

Key Points:

- The exhaustiveness of historical sea record information is demonstrated based on French Atlantic coast data
- A comparative analysis of approaches to integrate historical information is carried out
- The efficiency of a new method for the combination of systematic skew surges and historical records is verified

Nomenclature

X_{sys}	POT sample of systematic skew surges
u	threshold value for X_{sys}, $u \geq \min(X_{sys})$ (m)
n	length of X_{sys}
w_S	systematic duration (years)
Z_{hist}	historical record sea levels
η_H	threshold value for Z_{hist}, $\eta_H \geq \min(Z_{hist})$ (m)
h_z	length of Z_{hist}
X_{hist}	corresponding skew surges of Z_{hist} and exceeding u_H
u_H	threshold value for X_{sys}, $u_H \geq \min(X_{hist})$ and $u_H \geq u$ (m)
h_x	length of X_{hist}, $h_x \leq h_z$
w_H	historical duration (years)
λ	intensity of Poisson process, $\lambda > 0$
σ	scale parameter of the GP distribution, $\sigma > 0$
ξ	shape parameter of the GP distribution, $\xi \in \mathbb{R}$
θ	parameters to estimate, $\theta = (\sigma, \xi, \lambda)$
List of abbreviations	
GP	General Pareto
POT	Peaks Over Threshold
ML	Maximum Likelihood
MCMC	Monte Carlo Markov Chain
RSD	Relative Standard Deviation
RRMSE	Relative Root Mean Square Error

Corresponding author: Laurie Saint Criq, `lauriesaintcriq@gmail.com`

Abstract

The estimation of sea levels corresponding to high return periods is crucial for coastal planning and for the design of coastal defenses. This paper deals with the use of historical observations, i.e. events that occurred before the beginning of the systematic tide gauge recordings, to improve the estimation of design sea levels. Most of the recent publications dealing with statistical analyses applied to sea levels suggest that astronomical high tide levels and skew surges should be analyzed and modelled separately. Historical samples generally consist of observed record sea levels. Some extreme historical skew surges can easily remain unnoticed if they occur at low or moderate astronomical high tides and do not generate extreme sea levels. The exhaustiveness of historical skew surge series, which is an essential criterion for an unbiased statistical inference, can therefore not be guaranteed. This study proposes a model combining, in a single Bayesian inference procedure, information of two different nature for the calibration of the statistical distribution of skew surges: measured skew surges for the systematic period and extreme sea levels for the historical period. A data-based comparison of the proposed model with previously published approaches is presented based on a large number of Monte Carlo simulations. The proposed model is applied to four locations on the French Atlantic and Channel coasts. Results indicate that the proposed model is more reliable and accurate than previously proposed methods that aim at the integration of historical records in coastal sea level or surge statistical analyses.

1 Introduction

Coastal defenses must be designed for very low probabilities of failure. Their design values, generally resulting from the statistical analyses of relatively short series of tide gauges, are particularly sensitive to inherent statistical estimation uncertainties. During the last decade, a number of coastal floods due to exceptional surges, resulted in significant damages, pointing to the importance of an appropriate design of coastal defense structures (Aelbrecht et al., 2004; Gerritsen, 2005; De Zolt et al., 2006; Kolen et al., 2013). It is now widely accepted that historical information even if partial and inaccurate, may significantly reduce statistical inference uncertainties, if properly processed (Ouarda et al., 1998; Benito et al., 2004; Reis & Stedinger, 2005; Gaal et al., 2010; Payrastre et al., 2011; Hamdi et al., 2015). This paper proposes some methodological improvements for the incorporation of historical information in coastal risk assessment studies.

The measured sea levels can be interpreted as the combination of two temporal signals: astronomical tides which can be predicted and residuals due to atmospheric and meteorological processes (see Figure 1). On average, 706 tidal cycles occur during a year. The maximum tidal sea level during a cycle can also be seen as the sum of the astronomical high tide and the skew surge - i.e. the difference between the observed maximum sea level and the predicted astronomical high tide (see Figure 1).

The common practice in extreme value statistics for coastal studies consists in fitting a theoretical statistical distribution to a sub-sample of the observed series. The sub-sample is generally a peaks over threshold (POT) sample of either maximum tidal sea levels (direct method) or skew surges or even maximum tidal residuals (indirect methods). The direct method, based directly on the analysis of maximum tidal water levels (Arns et al., 2013; Bulteau et al., 2015) does not exploit the available knowledge on the astronomical tidal component of the sea level (Tawn et al., 1989; Mazas et al., 2014). Moreover it seems to provide biased estimates of sea level quantiles corresponding to high return periods for locations with large tidal amplitudes (Haigh et al., 2010; Andreewsky et al., 2014). Indirect methods are therefore nowadays privileged. Indirect methods were first introduced based on the separate analysis of residuals and astronomical tides (Pugh & Vassie, 1978, 1980; Tawn et al., 1989; Tawn, 1992). They are nevertheless uneasy to implement, since the reconstruction of the maximum sea level statistical distribution im-

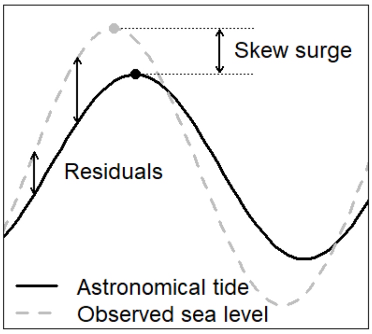

Figure 1. Definition of residuals and skew surges

plies a complex convolution between the astronomical tidal signal and the common and extreme residuals (Dixon & Tawn, 1994, 1999; Tomasin & Pirazzoli, 2008; Liu et al., 2010). Moreover, residuals and astronomical high tides may be dependent at some locations. Accounting for this dependence makes the approach even more challenging (Mazas et al., 2014). The indirect method, based on skew surges was introduced more recently in order to reduce the implementation complexity (Batstone et al., 2013; Kergadallan et al., 2014; Mazas et al., 2014; Hamdi et al., 2015). Note that the latter approach is used herein on a POT sample of skew surges X_{sys} larger than a threshold u.

Historical information, when available, is composed of a series of record sea levels Z_{hist} exceeding a threshold η_H. The corresponding historical skew surge series X_{hist} and the associated threshold u_H may be estimated for statistical inference combining systematic X_{sys} and historical X_{hist} skew surges. However, the exhaustiveness of the series of skew surges exceeding u_H during the historical period cannot be guaranteed. Indeed, some extreme historical skew surges may in fact remain unnoticed if they occur at low or moderate astronomical high tides and do not generate extreme sea levels (Outten et al., 2020). The exhaustiveness of the historical POT series is an essential criterion for an unbiased statistical inference (Gaume, 2018). Some authors have proposed to proceed with the statistical inference including historical skew surges without considering their non-exhaustiveness (Hamdi et al., 2015). We suspect that the under-sampling of historical skew surges due to their non exhaustiveness could leads to some bias. Some others have proposed to adjust (i.e. reduce) the length of the historical period to account for the non-exhaustiveness (Frau et al., 2018). Reducing the historical duration leads to a possible over-representation of extreme skew surges in a very short duration and the introduction of some bias is suspected. None of these two approaches appears to be totally satisfactory. It is therefore proposed hereafter to keep the historical information in its original form and to combine, in the same inference procedure, two different types of information: systematic skew surges X_{sys} and historical record sea levels Z_{hist}. A likelihood based inference procedure is implemented. The main idea consists in replacing the analytical form of the sea level cumulative distribution function, which is unknown, by a numerical estimate in the likelihood formulation.

This paper presents the background of the proposed approach and its performances: accuracy of the estimated skew surge quantiles and of the corresponding Bayesian credibility intervals. These performances are evaluated through Monte Carlo experiments inspired by four real-life implementation case studies. The results are compared to those of several other inference methods: without the use of any historical information, when historical skew surges are exhaustively known, the method proposed by Hamdi et al. (2015), the method proposed by Frau et al. (2018) and a modification of this last method (section 2.1). The proposed approach is then applied to the four observed data sets in order to evaluate its relevance and efficiency when implemented on real-life case studies.

The paper is structured as follows. The various tested methods and the statistical inference procedure are presented in section 2. The evaluation methodology is explained in section 3. The performances of the tested methods are compared in section 4 and some reference methods as well as the proposed method are implemented on the observed data sets in section 5. Section 6 is devoted to some discussion and conclusions.

2 Models and statistical inference procedure

2.1 The tested methods

The proposed method is compared to several methods including methods of reference (with only systematic data sets or in the case of perfect knowledge of historical skew surge series) and previously published methods integrating historical information (Hamdi et al., 2015; Frau et al., 2018). In total, six different methods are implemented and tested herein for the estimation of the 100-year skew surge quantile:

- **Method 1:** The inference is only based on the series of systematic skew surges X_{sys} exceeding a threshold value u (see Section 2.2.1). This method with no historical information included is considered herein as the reference one.
- **Method 2:** All historical skew surges exceeding the threshold u are known for the systematic and historical period. This is the ideal situation.
- **Method 3:** The series of systematic skew surges X_{sys} exceeding u and historical record sea levels Z_{hist} exceeding a threshold value η_H are combined in a single likelihood formulation (see Section 2.2.3). This is the proposed method.
- **Method 4:** The series of historical skew surges exceeding u_H, corresponding to the record sea levels exceeding η_H is supposed to be exhaustive. This method proposed by Hamdi et al. (2015) (see Section 2.2.2) will be called "naive", as the exhaustiveness of the historical skew surge series can never be guaranteed.
- **Method 5:** The FAB method proposed by Frau et al. (2018) adjusts the duration of the historical observation period, assuming that the mean annual frequency of a skew surge exceeding the threshold value u is the same during the historical and the systematic periods (see Section 2.2.2).
- **Method 6:** A modification of the FAB method accounting for the fact that the real skew surge sampling threshold u_H for the historical period may be much larger than u and that the mean annual frequency of exceedance should therefore be adjusted (see Section 2.2.2).

The likelihood formulations for all of these methods are provided in the next section.

2.2 Likelihood formulations

Let us denote $X_{sys} = \{x_{sys,1}, x_{sys,2}, ..., x_{sys,n}\}$ the POT series of n skew surges exceeding a threshold value u during the systematic observation period w_S (years). $Z_{hist} = \{z_{hist,1}, z_{hist,2}, ..., z_{hist,h_z}\}$ are the h_z record historical sea levels. It is assumed - ideally cross-checked with available archives - that the sample of record sea levels exceeding a threshold η_H is exhaustive over the considered historical period. η_H is often chosen equal to the minimum historical value: $\min(Z_{hist})$. Finally, $X_{hist} = \{x_{hist,1}, x_{hist,2}, ..., x_{hist,h_x}\}$ is the series of h_x historical skew surges, corresponding to the historical record levels and in the same time, exceeding the threshold u. Note that $h_x \leq h_z$. Let us also note θ the parameters of the skew surge statistical distribution to be estimated using the available observed data set.

Depending on whether the historical record sea levels or the historical skew surges are considered, the combined likelihood of the systematic and historical data sets may have two distinct formulations:

$$L(X_{sys}, X_{hist}|\theta) = L(X_{sys}|\theta)\ .\ L(X_{hist}|\theta) \tag{1}$$

$$L(X_{sys}, Z_{hist}|\theta) = L(X_{sys}|\theta)\ .\ L(Z_{hist}|\theta) \tag{2}$$

The likelihood terms $L(X_{sys}|\theta)$, $L(X_{hist}|\theta)$ and $L(Z_{hist}|\theta)$ are described in the next sections.

2.2.1 Likelihood of the systematic skew surge sample: $L(X_{sys}|\theta)$

The General Pareto (GP) distribution is usually selected as the statistical distribution of skew surges exceeding u. Indeed, according to the extreme value theory, it has been proven that a POT independent random sample converges to a GP distribution (Coles, 2001). The GP cumulative distribution function F_θ is given by:

$$\forall\ x > u,\ F_\theta(x) = \begin{cases} 1 - \left[1 + \xi\left(\frac{x-u}{\sigma}\right)\right]^{-\frac{1}{\xi}} & \text{if } \xi \neq 0, \\ 1 - \exp\left(-\frac{x-u}{\sigma}\right) & \text{if } \xi = 0. \end{cases} \tag{3}$$

with $\sigma > 0$ the scale parameter and $\xi \in \mathbb{R}$ the shape parameter.

The number of skew surges exceeding the threshold u per year is generally assumed to follow a Poisson process (Coles, 2001) with parameter λ (average number of skew surges exceeding the threshold u per year). The probability of observing n skew surges exceeding u during a systematic observation period of duration w_S years is then equal to:

$$\mathbb{P}_\theta(N = n) = \frac{(\lambda w_s)^n}{n!}\ \exp\left(-\lambda w_S\right) \tag{4}$$

If the observed systematic skew surges $x_{sys,j=\{1,\ldots,n\}}$ are considered independent and identically distributed, the likelihood of the systematic sample is given by equation (5) where f_θ is the GP probability density function.

$$L(X_{sys}|\theta) = \mathbb{P}_\theta(N = n)\ .\prod_{j=1}^{n} f_\theta(x_{sys,j}) \tag{5}$$

The parameters to be estimated through the inference procedure are the scale and shape parameters of the GP distribution and the intensity of the Poisson process: $\theta = (\sigma, \xi, \lambda)$.

2.2.2 Likelihood of the historical skew surge sample: $L(X_{hist}|\theta)$

Considering the h_x historical skew surges exceeding a threshold value $u_H \geq u$ over a historical period of w_h years as independant and identically distributed, the likelihood of the historical skew surge sample is:

Table 1. Likelihoods of the historical skew surge samples for methods 4, 5 and 6. $\hat{\theta}$ and $\hat{\lambda}$ represent the parameter set and the Poisson process intensity estimated with the maximum likelihood based on the systematic skew surges only and $R_\lambda(u_H) = h_x \dfrac{\lambda\,[1-F_\theta(u_H)]}{\hat{\lambda}\,[1-F_{\hat{\theta}}(u_H)]}$.

Method	u_H	w'_H	$L(X_{hist}\|\theta)$
4	$\min(X_{hist})$	w_H	$\dfrac{[\lambda w_H]^{h_x}}{h_x!}\ \exp\left(-\lambda w_H\,[1-F_\theta(u_H)]\right)\prod_{j=1}^{h_x} f_\theta(x_{hist,j})$
5	u	$\dfrac{h_x}{\hat{\lambda}}$	$\dfrac{\left(h_x\lambda/\hat{\lambda}\right)^{h_x}}{h_x!}\ \exp\left(-h_x\dfrac{\lambda}{\hat{\lambda}}\right)\prod_{j=1}^{h_x} f_\theta(x_{hist,j})$
6	$\min(X_{hist})$	$\dfrac{h_x}{\hat{\lambda}\left[1-F_{\hat{\theta}}(u_H)\right]}$	$\dfrac{R_\lambda(u_H)^{h_x}}{h_x!}\ \exp\left(-R_\lambda(u_H)\right)\prod_{j=1}^{h_x}\dfrac{f_\theta(x_{hist,j})}{1-F_\theta(u_H)}$

$$L(X_{hist}|\theta) = \mathbb{P}_\theta(H_X = h_x)\ .\prod_{j=1}^{h_x}\frac{f_\theta(x_{hist,j})}{1-F_\theta(u_H)} \tag{6}$$

where $\mathbb{P}_\theta(H_x = h_x)$ is given by the following equation:

$$\mathbb{P}_\theta(H_x = h_x) = \frac{[\lambda w_H(1-F_\theta(u_H))]^{h_x}}{h_x!}\ \exp\left(-\lambda w_H\,[1-F_\theta(u_H)]\right) \tag{7}$$

Methods 4, 5 and 6 differ by the estimation of the threshold value u_H and the considered effective duration of the historical period w'_H. The various proposed estimates and the final formulation of the likelihood $L(X_{hist}|\theta)$ are provided in Table 1.

In the naive method (method 4), the threshold u_H is the minimum value of the historical skew surge sample $\min(X_{hist})$. But, due to the sampling approach based on record sea levels, there is a risk that this sample represents a partial and not the exhaustive record of all skew surges that have exceeded the threshold u_H during the historical period. A statistical inference based on the hypothesis of exhaustiveness and conducted on a partial sample will provide biased quantile values. To avoid this problem, the FAB method (method 5), proposes to introduce a corrected duration for the historical period w'_H. This duration is chosen to be perfectly consistent with the average number $\hat{\lambda}$ of skew surges exceeding the threshold per year and with the number of recorded historical skew surges h_x : $w'_H = h_x/\hat{\lambda}$. In the initial version of the FAB method (Frau et al., 2018), the historical sampling threshold was considered equal to the systematic threshold u. Since the minimum value of historical sampled skew surges appears often much larger than u, this *a priori* choice may be a source of significant biases as will be illustrated hereafter. A modified version of the FAB method is therefore tested here (method 6), where the historical threshold is adapted to the available sample and the corrected duration w'_H is adjusted accordingly (see Table 1).

2.2.3 Likelihood of the historical sea level sample: $L(Z_{hist}|\theta)$

The likelihood formulation of the historical sea levels comprises (a) the probability associated to the $N - h_z$ ($N = 706 \times w_H$) maximum tidal levels that did not exceed the historical threshold η_H and (b) the probability associated to the h_z extreme historical maximum tidal levels that exceeded η_H during the historical period of duration of w_H years (equation (8)).

$$L(Z_{hist}|\theta) = \underbrace{\tilde{G}_\theta(\eta_H)^{N-h_z}}_{\text{(a)}} \cdot \underbrace{\Big[1-\tilde{G}_\theta(\eta_H)\Big]^{h_z} \cdot \prod_{j=1}^{h_z} \frac{\tilde{g}_\theta(z_{hist,j})}{1-\tilde{G}_\theta(\eta_H)}}_{\text{(b)}} \tag{8}$$

$\tilde{g}_\theta$, $\tilde{G}_\theta$ are respectively the probability density and cumulative distribution functions of maximum tidal levels which result from the combination of (1) the statistical distribution of the maximum astronomical tidal levels, (2) the statistical distribution of skew surges lower than the threshold u, and (3) the calibrated statistical distribution (f_θ, F_θ) of the skew surges exceeding u. The proposed numerical approximations of the functions $\tilde{g}_\theta$ and $\tilde{G}_\theta$ are presented in Appendix A.

3 Test and evaluation methodology

The comparison of the different methods is based on samples generated through Monte Carlo simulations inspired by four real world case studies in order to verify the accuracy of the maximum likelihood (ML) estimates and the posterior credibility intervals.

3.1 Monte Carlo experiments

1000 synthetic series are randomly generated with characteristics corresponding to each of the four observed data sets: duration of the systematic and historical observation periods w_S and w_H, systematic and historical sampling thresholds u and η_H, parameters of the GP distribution and Poisson intensity for the skew surges exceeding u and empirical statistical distributions of the astronomical high tides and of the ordinary skew surges (lower than u) as well as the astronomical high tide/skew surge relation (see Section 3.2 and Appendix C).

Each synthetic sample is generated as follows:

- For the systematic period, n systematic skew surges X_{sys} are drawn from the Poisson process (intensity λw_S) and GP distribution.
- For the historical period, n_2 skew surges X_{hist} larger than u are drawn from the Poisson process (intensity λw_H) and GP distribution (series used for the implementation of method 2) and complemented with ($w_H \times 706 - n_2$) ordinary skew surges (lower than u), drawn from the empirical ordinary skew surge distribution. $w_H \times 706$ astronomical high tides are drawn form the empirical high tide distribution. Astronomical high tides and skew surges, assumed to be independent (see Appendix C), are summed to generate $w_H \times 706$ maximum tidal levels. The subset of h_z sea levels Z_{hist} exceeding η_H is then extracted (series used for the implementation of method 3), as well as the corresponding subset of h_x skew surges larger than u for the implementation of methods 4 to 6.

3.2 Case study

Four tide gauges located on the French Atlantic and Channel coasts are used as examples for the configuration of the Monte Carlo experiment: Brest, Dunkerque, La Rochelle and Saint Nazaire. These tide gauges are selected because of the availability of historical information, but also because they cover a variety of situations: i) statistical distributions of the skew surges and tidal levels, ii) tide/surge ratio (Table 2), iii) tidal am-

Table 2. Characteristics of the systematic data set and selected values for the Monte Carlo simulations ($\hat{\sigma}$, $\hat{\xi}$, $\hat{\lambda}$).

Site	Period	w_S (years)	u (m)	n	Tide surge ratio*	$\hat{\sigma}$	$\hat{\xi}$	$\hat{\lambda}$
Brest	1953-2017	63.57	0.50	81	22.50	0.09	0.19	1.29
Dunkerque	1959-2016	47.75	0.74	58	15.58	0.14	0.34	1.23
La Rochelle	1941-2016	32.58	0.62	34	17.11	0.08	0.36	1.08
Saint Nazaire	1957-2014	47.56	0.66	53	15.45	0.11	0.12	1.14

* Ratio of the 98% astronomical high tide to the 98% skew surge quantile (Dixon & Tawn, 1999).

Table 3. Characteristics of the historical data sets.

Site	Period	w_H (years)	$\frac{h_x}{\hat{\lambda}}$ (years)	$\frac{h_x}{\hat{\lambda}\left[1-F_{\hat{\theta}}(u_H)\right]}$ (years)	η_H (m)	h_z	u_H (m)	h_x
Brest	1846-1952	120	2.33	13.72	8.02	10	0.69	3
Dunkerque	1720-1953	250	6.50	108.42	7.60	8	1.40	8
La Rochelle	1866-1940	80	3.70	13.31	7.15	4	1.00	4
Saint Nazaire	1863-1956	100	4.40	17.62	7.09	5	0.82	5

plitude, iv) historical perception threshold level and number of documented historical events.

The hourly tide gauge data were retrieved from Shom, the French hydrographical and oceanographical service (data.shom.fr), harmonic analyis is applied on these data with the R package *TideHarmonics* (Stephenson, 2015), as well as a correction of sea level rise. Then, hourly astronomical tide levels were processed to extract the series of corresponding astronomical high tides and systematic skew surge series.

The threshold u for the POT sampling is selected according to the GP parameter stability criterion (Coles, 2001).

Historical sea levels were extracted from Hamdi et al. (2018); Giloy et al. (2018, 2019) for Dunkerque (Table B2) and from Breilh et al. (2014) for La Rochelle (Table B3). At La Rochelle, the sampling threshold η_H had to be raised to ensure the exhaustiveness of the historical record levels and two reported record levels were ignored (see Table B3). In fact, the systematic observations started in 1846 and 1863 respectively at Brest and Saint Nazaire. The complete observed samples were split into systematic and historical samples for the sake of illustration. To test the proposed method, censored samples of historical record sea levels were extracted at these two stations setting a threshold value of 8m at Brest and 7m at Saint Nazaire (Tables B1 and B4).

Table 3 presents the characteristics of the historical samples as well as the considered duration for the implementation of the various methods. As suggested by Schendel and Thongwichian (2017), the historical duration w_H is larger than the time laps between the first record and the start of the systematic period. The duration considered for the FAB method $h_x/\hat{\lambda}$ appears to be extremely reduced. For Dunkerque, the reported historical skew surges are extremely high if compared to the systematic data: 8 values exceeding $u_H = 1.40$m, when the largest measured value during the systematic period is 1.30m. Some inconsistencies between the historical and systematic data sets at Dunkerque may be suspected and will be discussed further on in section 4. The observed historical series are the result of a random drawing. The simulated historical series, based on

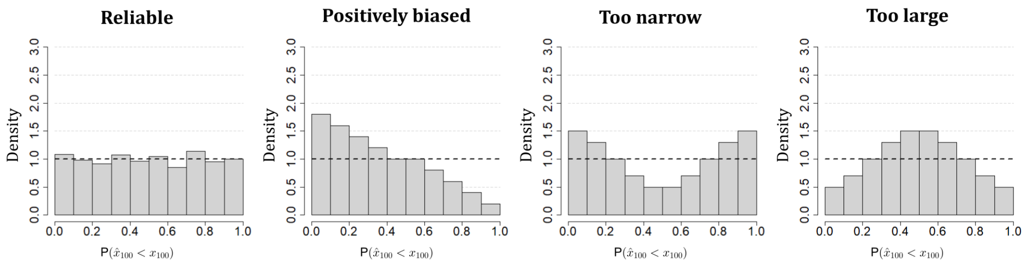

Figure 2. Possible distributions of $\mathbb{P}(\hat{x}_{100} < x_{100})$ and conclusions on the reliability of the posterior densities and corresponding credibility intervals.

the parameters calibrated on the observed series, may have slightly different characteristics on average, especially different numbers of record events (see Table 4).

3.3 Evaluation methods

The RStan package was used to conduct Bayesian MCMC (Monte Carlo Markov Chain) inferences based on the formulated likelihood with non-informative priors. The results of the inference procedure consist in the posterior densities for the calibrated parameters $\theta = (\sigma, \xi, \lambda)$ and of the corresponding skew surge quantiles, including the maximum likelihood estimates. The evaluation of the various tested methods (see Section 2.1) was conducted in two steps. The accuracy of the maximum likelihood estimator was first verified based on the 100-year quantile estimate (comparison between the quantile values $\hat{x}_{100}^{ML}$ and the real quantile value x_{100} for the 1000 generated series). The evaluation will be based on boxplots of the ratio $\hat{x}_{100}^{ML}/x_{100}$ and classical average performance estimation criteria: relative bias, relative standard deviation (RSD) and relative root mean square error (RRMSE).

In a second step, the average widths of the computed posterior credibility intervals for the 100-year quantile are compared and their reliability is evaluated based on the rank histogram diagnosis method (Bellier, 2018; Nguyen et al., 2014). For each of the 1000 inferences, the exceedance probability $\mathbb{P}(\hat{x}_{100} < x_{100})$ of the real quantile value x_{100} is computed according to the estimated posterior density for the quantile. If the estimated posterior densities are reliable, $\mathbb{P}(\hat{x}_{100} < x_{100})$ should be uniformly distributed over $[0, 1]$ (Halbert et al., 2016; Gaume, 2018). Figure 2 illustrates how the rank histogram can be interpreted.

3.4 Characteristics of the Monte Carlo simulations

Table 4 summarizes the characteristics of the 1000 simulated samples for each case study. It seems that the parameters of the Monte Carlo simulations, adjusted on the observed series, lead to generated series with contrasted characteristics like the number of sampled record sea levels or the sampling rate of the historical skew surges exceeding the threshold u. The selected threshold η_H at Brest leads to a large number of sampled historical sea levels. But due to a large tide/surge ratio, the corresponding samples of skew surges exceeding u represent only a small proportion of the total number of generated skew surge exceeding u for the historical period - on average less than 10%. Dunkerque and La Rochelle are considered intermediate cases where smaller average amounts of historical sea levels are sampled, but the skew surge sampling rate is higher due to a more favorable tide/surge ratio - i.e. due to a higher contribution of the skew surges to the record levels. Finally, Saint Nazaire appears to be an extreme case, where, due to a relatively high threshold value η_H, a limited number of record sea levels and skew surges

Table 4. Characteristics of the generated historical series of sea levels and skew surges.

	Brest	Dunkerque	La Rochelle	Saint Nazaire
Generated historical sea levels				
Sampling threshold η_H (m)	8.02	7.60	7.15	7.09
Minimum generated value (m)	8.02	7.70	7.24	7.17
Average number of record values	22	7	3	1
Duration of the historical period (years)	120	250	80	100
Generated historical skew surges				
Sampling threshold u (m)	0.50	0.74	0.62	0.66
Minimum sampled value u_H (m)	0.55	1.63	0.90	0.93
Average number of skew surges $> u$	156	308	86	116
Average number of skew surges $> u_H$	26	18	24	28
Average number of sampled values $> u_H$	2	6	2	1
Average skew surge sampling rate (%)	7.63	33.33	8.33	3.57

are sampled. A high proportion of the generated historical samples at Saint Nazaire does not contain record sea levels exceeding η_H (33%) or skew surges exceeding u (45%).

3.5 Maximum likelihood estimates

The evaluation of the various tested inference procedures confirms some anticipated results, but also provides some satisfactions and surprises. The hypothesis of exhaustiveness for the sample of skew surges exceeding u_H during the historical period, on which the naive method (method 4) is based, is clearly not reached for the four test cases. The average skew surge sampling rates appear largely lower than 100% in Table 4. As a consequence, method 4 underestimates the 100-year skew surge quantile x_{100} (see Figures 3 and 4). Table S1 in the supplementary information provides the numeric values corresponding to Figure 4 for a more detailed analysis. The magnitude of the bias affecting the estimation of the parameter λ (i.e. average number of skew surges exceeding u per year) seems clearly dependent on the skew surge sampling rate for the historical period (see Figure 5 as well as S1, S2, S3 in the supplementary information). The estimation of the two parameters of the GP distribution is also biased since these parameters control the probability of exceedance of the threshold value u_H appearing in the likelihood formulation for the historical period in method 4 (see Table 1). The increase of the amount of information used for the inference in method 4 leads nevertheless to a significant decrease of the standard deviation of the x_{100} estimator, if compared to the method based on the systematic data only (method 1). Surprisingly, the balance between bias and reduced standard deviation appears positive for the naive method : for the four test cases, the RRMSE of the x_{100} estimator is significantly lower for the naive method than for the method based on the systematic data only (see Figure 4). This remains true, even for the Saint Nazaire case study, where a high proportion of historical generated series does not contain any recorded skew surges exceeding u. This issue will be addressed later.

The results also confirm the suspected biases introduced by the FAB method (method 5) and reveal other important anomalies. In fact, since an equivalent duration of the historical period is estimated, the information about the non-exceedances of the threshold u during the historical period, which is an important part of the historic information as shown by Payrastre et al. (2011), is not evaluated. The historical information is therefore only partly used and limited to the set of a few skew surges reported to have exceeded u, that complement the rich series of systematic skew surges. The possible added value of the historic data is hence extremely limited in the FAB method. Moreover, the sam-

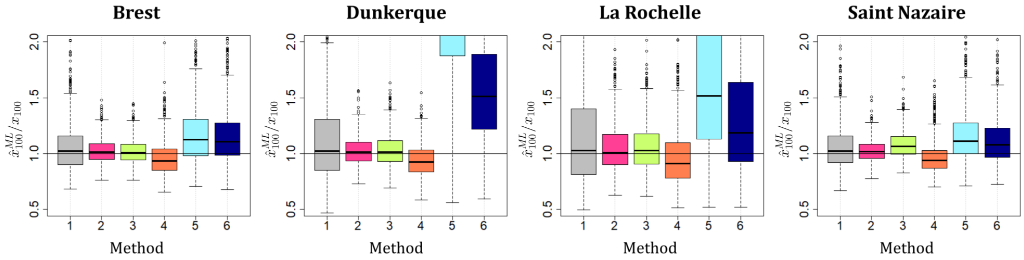

Figure 3. Dispersion of the 100-year quantile estimated with the maximum likelihood (divided by the real value), obtained from simulations.

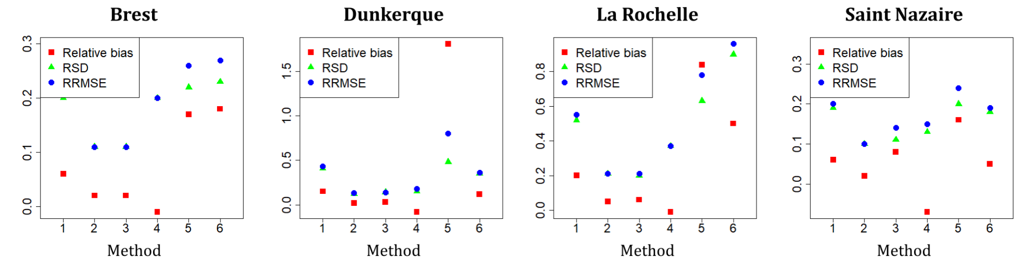

Figure 4. Relative bias, RSD and RRMSE on the 100-year quantile estimated with the maximum likelihood at the 4 study sites with the different tested methods.

pling process for the historic and systematic surges are different: the sampling threshold is higher for the historic surges, especially for locations with low tide/surge ratios and highly skewed GP distributions (i.e. large ξ values). Merging the historic skew surges with the systematic sample without further adjustments introduces significant biases in the estimates of the parameters (σ, ξ) of the GP distribution (see Figure 5). As a conclusion, the FAB method cannot really contribute to reduce significantly the inference uncertainties and introduces some biases. Its implementation leads to an increase of the x_{100} estimation RRMSE if compared to the analyses of the sole systematic data (method 1). The proposed adjusted FAB method reduces partly the estimation biases but the effect on the estimation RSD remains limited if compared to method 1 (Figure 4). The principles of the FAB method appear as inefficient and statistically inconsistent. Its implementation leads to a deterioration in the inference results, if compared to the analyses of the systematic data only.

In contrast, the proposed method (method 3) appears to perform almost as well as the ideal method (method 2). In details, the gain, if compared to method 1, seems to be mainly related to a more accurate estimation of the GP shape parameter ξ (Figures 5, S1, S2 and S3 in the supplementary information). These excellent performances may be surprising at first sight since many more historical events are evaluated in method 2 (about 80 to 300 additional historical skew surges) than in method 3 (1 to 22 record sea levels) (see Table 4). Moreover, the historical samples used in methods 2 and 3 are partly or totally dissociated - i.e. corresponding to different events (see Figure C1). The record sea levels included in the inference of method 3 do not necessarily involve the most extreme skew surges of the historical period. To understand this surprising result, it must be firstly considered that the high frequency of skew surges observed during the historical period does not provide significant additional information to the one contained in the systematic data set. The historical information is mainly encapsulated in the largest observed values, that will help constraining the skew surge distribution tail. Payrastre et al. (2011) have shown that when including historical information in a statistical in-

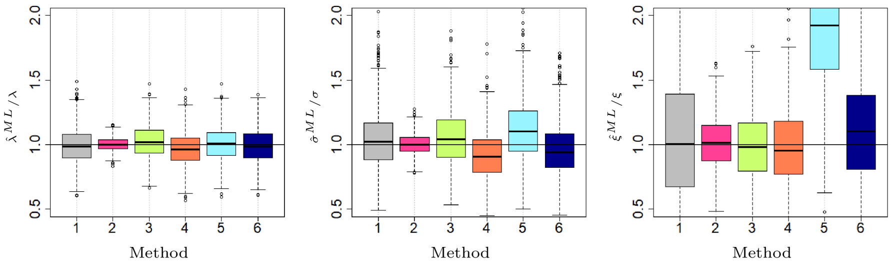

Figure 5. Dispersion of the parameters estimated with the maximum likelihood (divided by the real values), obtained from simulations at Dunkerque with different tested methods.

ference procedure, the length of the documented historical period is a predominant factor: "accurate estimates of the values having exceeded the perception threshold are not necessarily needed when historical data is used in combination with systematic measurements ; provided that the theoretical return period of the perception threshold is sufficiently high, censored (only the values exceeding the threshold are known) or binomial censored (only the number of values having exceeded the threshold is known) historical data lead to similar inference results". This explains also why the results obtained with the proposed method for the Saint Nazaire case study, where a special case of binomial censored historic data set is frequently generated (no exceedance of the threshold η_H or in other words $h_z = 0$), are also satisfactory. It is worth noting that the maximum likelihood estimates of the GP parameters and quantiles appear slightly positively biased for all methods except method 4. This bias appears to be more pronounced when inference is conducted on a binomial censored sample (method 3 at Saint Nazaire). This appears to be a general feature for the ML estimates of the parameters of a GP distribution. Indeed, (Hosking & Wallis, 1987) indicated that the ML method leads to biased GP parameters estimates when the sample size is not large.

The implemented Bayesian inference procedure generates not only best-estimates for the quantile values, but also credibility intervals and posterior distributions. The next section compares this computed intervals for methods 1 to 4.

3.6 Posterior credibility intervals

The computed credibility intervals confirm the trends observed on the ML estimators. The added value of the historical information is confirmed by the reduced averaged widths of the posterior credibility intervals (Table 5). Without surprise, the widths of the posterior credibility intervals for the proposed method (method 3) are larger than those for the "ideal" method (method 2), but hence of similar magnitudes, confirming that the loss of historical information for proposed method if compared to the ideal case is limited, even for the Brest case study with a high tide/surge ratio. Some posterior intervals based on the naive method (method 4) may have lower widths than the intervals based on the proposed method -especially at Dunkerque, but the estimation bias related to method 4 should be considered (see next paragraph).

Figure 6 shows the rank histograms of the 100-year skew surge quantiles for methods 1 to 4 and all of the case studies. The histograms confirm the conclusions drawn from the ML estimates. The naive method (method 4) has a clear tendency to underestimate

Table 5. Average width of the posterior credibility interval for the 100-year quantile with the Bayesian MCMC procedure for methods 1, 2, 3 and 4.

	Average width of posterior credibility interval for x_{100}			
Site	Method 1	Method 2	Method 3	Method 4
Brest	1.15	0.48	0.55	0.95
Dunkerque	6.05	1.10	1.31	1.05
La Rochelle	10.48	1.47	1.60	2.56
Saint Nazaire	1.37	0.46	0.67	0.52

the quantile value x_{100} for all case studies. A slight over-estimation tendency is detectable for methods 1 and 2, but the computed posterior distributions and the corresponding credibility intervals for x_{100} appear overall reliable. As far as the proposed method 3 is concerned, the over-estimation tendency is clearly marked for the Saint Nazaire case study. This suggests that the method should ideally be implemented on historical samples including some documented historical sea levels. The rank histograms also reveal that the estimated posterior credibility intervals based on method 3 are too large (the uncertainty affecting the estimated value is overrated) at stations with large tide/surge ratios: i.e. stations where the historical record sea level sample does not coincide with the historical record skew surges. This is visible on the histogram obtained for the Brest case study and to a lower extend for the La Rochelle case study. The outcome of the Bayesian-MCMC inference provides a pessimistic assessment of the accuracy of the estimated quantile values.

As a partial conclusion, the conducted tests indicate that the proposed method combining skew surges for the systematic period and sea levels for the historic period is reliable and provides inference results that are almost as accurate as those obtained through in the ideal situation with an inference based on systematic and historical skew surges (method 2). This is a satisfactory result, but it is important to keep in mind that these conclusions are valid provided that the underlying statistical model is valid: i.e. skew surges and astronomical high tides are independent and the distribution of the skew surges is a GP distribution. It is therefore interesting as a conclusion to evaluate how the proposed approach behaves when implemented on real-world data sets. The next section presents and analyses the implementation of the method on the data sets available at the considered tide gauges.

4 Application of the proposed method to the observations

At Brest and Saint Nazaire, complete observed data sets of sea levels and estimated tides are available. It is hence possible to compare the results of method 3 with those of methods 1 and 2 at these two stations. At Dunkerque and La Rochelle, the historical data sets are composed of the observed record sea levels then, only methods 1 and 3 will be implemented. The hypothesis of independence between astronomical high tides and skew surges was tested and seems to be reasonably valid for all four stations (see Appendix C).

The implementation results of the methods at Brest and Saint Nazaire appear fully consistent with the conclusions previously drawn (Figure 7). The adjusted credibility intervals with the proposed method are very similar to those obtained with method 2, even if they are slightly larger. This is particularly striking for Brest where the historical sea levels do not represent the events with the largest skew surges. This confirms the con-

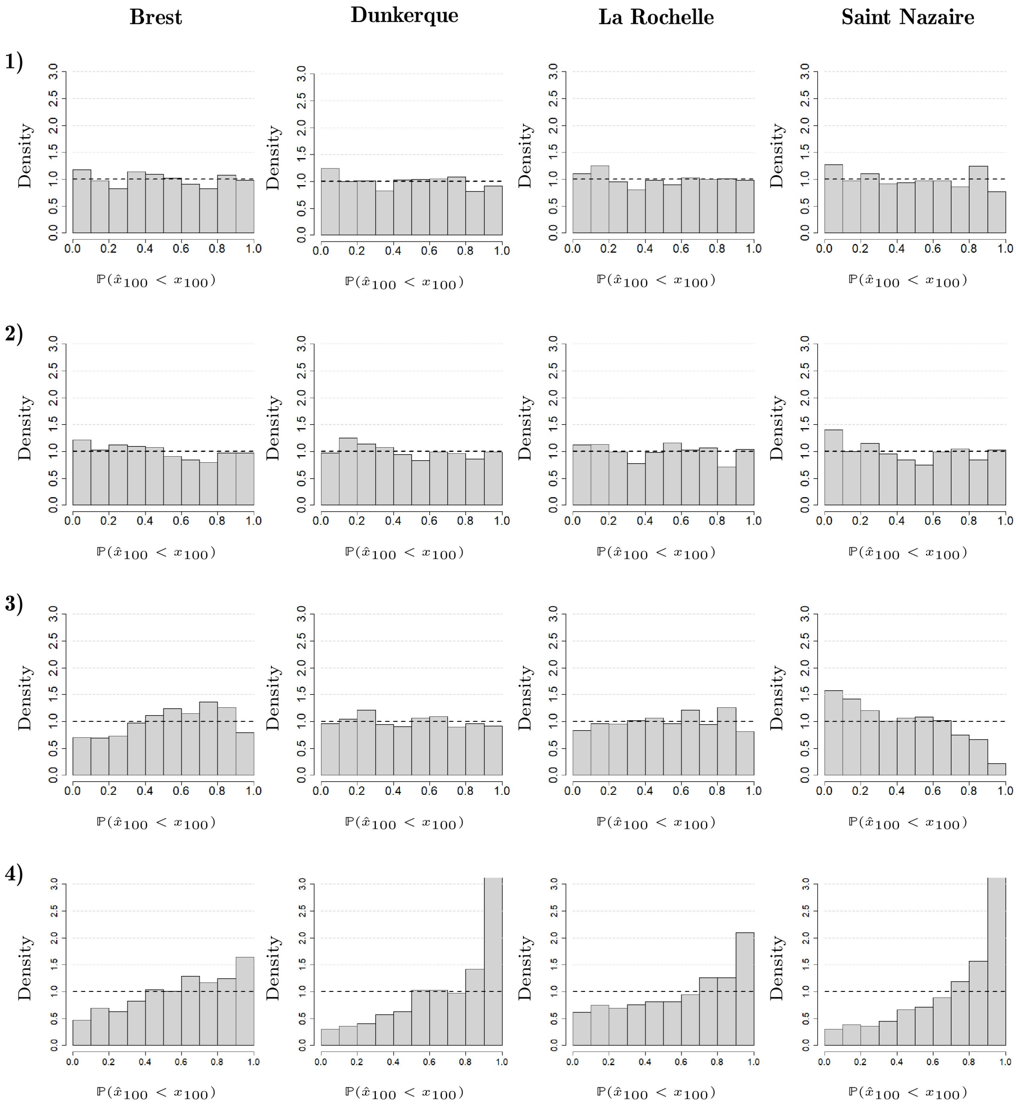

Figure 6. Uniformity test for the credibility intervals computed with the Bayesian MCMC procedure for methods 1, 2, 3 and 4.

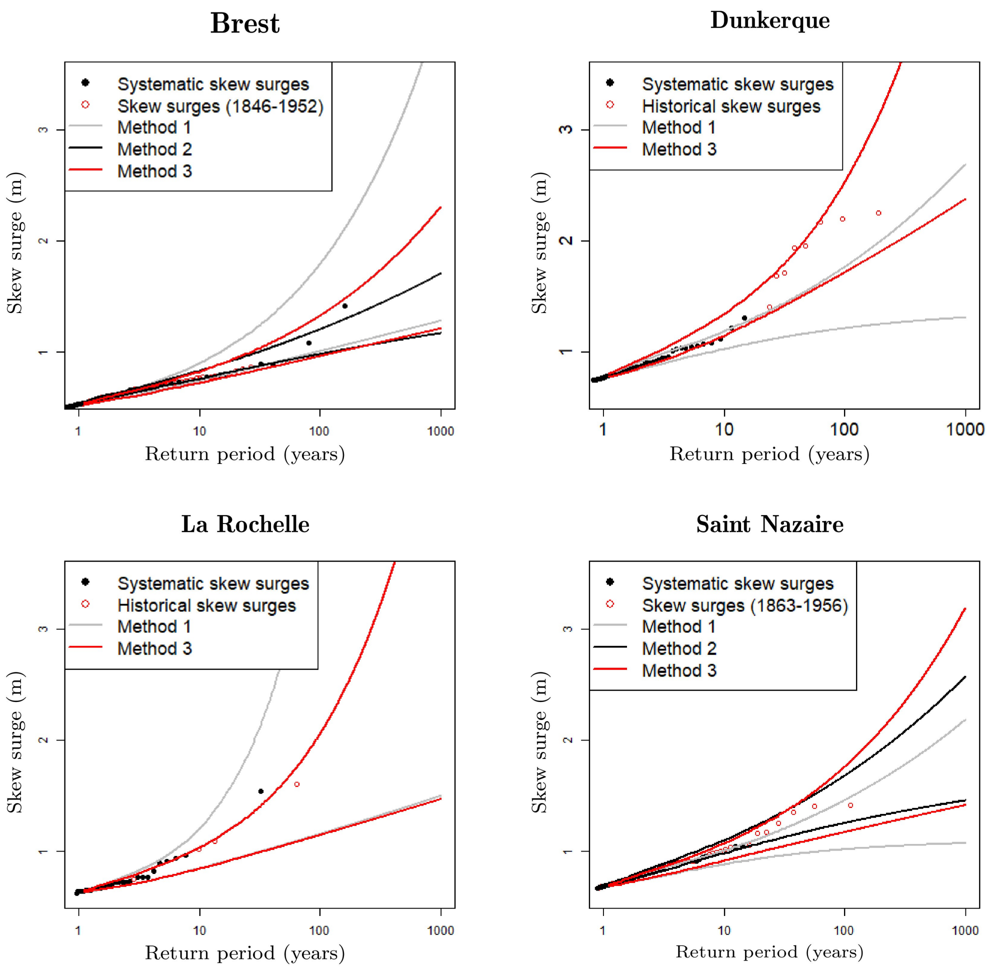

Figure 7. 90% posterior skew surge credibility intervals based on the systematic data (grey) and on the historic data with the proposed method (red) and in the ideal case (black). The empirical return periods of the historical records at Dunkerque and La Rochelle were corrected (reduced) according to the skew surge estimated sampling rates (see Table 4).

sistency between the observations and the calibrated statistical model: GP distribution for the skew surges and independence between skew surges and astronomical high tides.

The inclusion of the historical information appears to have contrasted impacts between the case studies. For Brest and La Rochelle, the posterior credibility intervals accounting for the historical information are significantly reduced and totally coherent with the intervals based on the sole systematic data sets (Figure 7). This is the expected result which reveals an overall good consistency between (a) the systematic observations, (b) the historical data sets and (c) the calibrated statistical model. In the case of Saint Nazaire, the historical data do not help to reduce the estimation credibility intervals, but lead to a modification of the calibrated statistical skew surge distribution. Note that this modification remains consistent with the systematic sample - i.e. the observations are contained in the revised posterior credibility intervals. This result may be explained by the peculiarities of the short systematic sample available at Saint Nazaire, which does not contain large skew surges - i.e. skew surges greater than 1m (Figure 7). Since the estimated uncertainties (i.e. widths of the posterior credibility intervals) are also related to the estimated variability of the skew surge distribution and especially to the magnitude of the parameter $\hat{\xi}$, the inclusion of the historical information at Saint Nazaire, leading to an increased $\hat{\xi}$ estimated values, does not result in a reduction of the inference estimation uncertainties. The case of Dunkerque is completely different: even if the length of the historical period is considered, the historical record levels and corresponding skew surges appear strongly inconsistent with the systematic data set. This inconsistency, revealed by the inference trials presented herein, remains to be explained.

As a conclusion, a final inference test was conducted to confirm the robustness of the proposed approach, even in cases where limited information about historical record sea levels is available and to verify if the conclusions drawn by Payrastre et al. (2011) based on historical river record discharges are also valid for historical record sea levels. For the considered case studies, the historical threshold η_H was selected such as there is no remaining documented record level exceeding the threshold (i.e. $h_Z = 0$, case 3* in Table S2 in the supplementary information). The resulting credibility intervals appear to be only moderately affected by this simplification of the historical information if compared to case 3. Even the knowledge that a given sea level has not been exceeded over a considered historical period (i.e. a given coastal defence structure has never been over-topped for instance) is a valuable information, that can efficiently processed with the new inference procedure presented herein. This opens new perspectives in coastal risk assessments.

5 Conclusions

A new statistical inference procedure is proposed and evaluated to properly integrate historical sea levels in coastal risk assessment studies. This procedure enables the combined analysis of data sets of different nature: skew surges for the recent period and sea levels for the historical period. It overcomes a major limitation in the previously proposed methods to include historical information in sea level frequency analyses. The key idea of this new method consists in replacing, in the likelihood formulation, the analytic expression of the probability density or cumulative distribution functions related to the historical sea level observations, by their numerical approximations (see Appendix A). The related R source codes as well as the data files corresponding to the test cases are available at: `doi.org/10.5281/zenodo.6260203`. Based on the results presented herein, some major conclusions can be drawn.

1. The suggested numerical scheme for the estimation of the historical sea level likelihood as well as its incorporation in the statistical inference procedure are effective and reliable. This is particularly well illustrated by the comparison with the results of the "ideal" method (method 2).

2. Unlike the previously published approaches which appear to be biased, the proposed method allows for accurate and reliable estimates of the maximum likelihood quantiles, as well as of their posterior distributions in a Bayesian MCMC inference framework.
3. The proposed method is almost as accurate as the ideal method - i.e. method based on a perfect knowledge of the historical skew surges - even in places exhibiting high tide/surge ratios. This is valid if the hypotheses on which the calibrated statistical model is based, especially the independence between high tides and skew surge, are reasonably consistent with the observations. It seems to be the case at Brest.
4. This last conclusion may appear surprising, since the data set used in the "ideal" method contains apparently much more information on skew surges, but it is consistent with the conclusions of previous studies dealing with statistical inferences based on historical records (Payrastre et al., 2011). It seems that the length of the documented historical period is more decisive than the number or the accuracy of the documented record events.

The proposed approach could be further improved in several ways. First, even if moderate, some estimation biases remain present: over-estimated credibility intervals in cases with large tide/surge ratios and over-estimations in the case of binomial censored historical samples with no exceedance. It would be satisfying if the origin of these biases were understood and if they could be corrected. Moreover, the possible dependence between high tides and skew surges, as well as some seasonal features may be considered in the inference procedure, to increase its pertinence and application range. In fact, the largest skew surges often occur during winter storms while high tides are observed around the equinoxes (Tomasin & Pirazzoli, 2008).

The method could also be implemented on a larger number of case studies and results should be compared to those of previous assessments. The possible implementation of the method on samples with no documented record sea level exceeding the threshold seems to lead to satisfactory results (see the concluding paragraph of Section 4). This opens new perspectives, especially at sites where little or no historical records are available. Indeed, any coastal structure with known altitude that has not been submerged during a considered historical period, may provide valuable information for the statistical inference.

Finally, the method was developed for the analysis of coastal sea levels, but the same principles could certainly be adapted for the statistical analysis of other geophysical variables.

Appendix A Estimation of $\tilde{g}_\theta$ and $\tilde{G}_\theta$

The maximum sea level Z is the sum of a skew surge X and an astronomical high tide Y. Both components are supposed to be independents (see Section Appendix C). Hence,

$$\mathbb{P}(Z < z) = \int_{\min(Y)}^{\max(Y)} q(y)\mathbb{P}(X < z - y)\ dy \tag{A1}$$

where $q(y)$ is the probability density function of Y, $\min(Y)$ and $\max(Y)$ represent respectively the lowest and the highest astronomical high tide. The skew surge X may either be smaller or larger than the systematic threshold u. Therefore,

$$\tilde{G}_\theta(z) = \mathbb{P}(Z < z) = \mathbb{P}(X \leq u)\ \mathbb{P}_{X \leq u}(Z < z) + [1 - \mathbb{P}(X \leq u)]\, \mathbb{P}_{X>u}(Z < z) \tag{A2}$$

Considering that $\mathbb{P}_{X>u}(X < x) = F_\theta(x)$ and $\mathbb{P}(X > u) = \hat{\lambda}/706$ and combining equations (A1) and (A2) leads to:

$$\begin{aligned}\tilde{G}_\theta(z) = &\left(1 - \frac{\hat{\lambda}}{706}\right) \int_{\min(Y)}^{\max(Y)} q(y)\ \mathbb{P}_{X \leq u}(X < z - y)\ dy \\ &+ \frac{\hat{\lambda}}{706} \int_{\min(Y)}^{\max(Y)} q(y)\ F_\theta(z - y)\ dy\end{aligned} \tag{A3}$$

The two terms $q(y)$ and $\mathbb{P}_{X \leq u}(X < z-y)$ can be estimated based on the observed systematic data set, prior to the implementation of the statistical inference procedure. The distribution of astronomical high tides is defined by the analysis of the predicted high tide values over a saros cycle (18,6 years). To enable the numeric computation of equation (A3), the range of possible values for Y is split into n_T intervals $Y_{k=\{1,\ldots,n_T\}}$ of 0.01m width. The vector of length n_T including the probability values $\mathbb{P}(Y \in Y_k)$ is computed and the integrals in equation (A3) are approximated by finite sums, leading to:

$$\begin{aligned}\tilde{G}_\theta(z) \approx &\left(1 - \frac{\hat{\lambda}}{706}\right) \sum_{k=1}^{n_T} \mathbb{P}(Y \in Y_k)\ \mathbb{P}_{X \leq u}(X < z - \mathrm{Med}(Y_k)) \\ &+ \frac{\hat{\lambda}}{706} \sum_{k=1}^{n_T} \mathbb{P}(Y \in Y_k)\ F_\theta(z - \mathrm{Med}(Y_k))\end{aligned} \tag{A4}$$

where $\mathrm{Med}(Y_k)$ represents the median high tide value for interval k.

The term $\mathbb{P}_{X \leq u}(X < z - \mathrm{Med}(Y_k))$ is estimated based on the empirical distribution of the measured sample of ordinary skew surges (i.e. skew surges lower than the threshold u). It is simply equal to the ratio of the number of observed ordinary skew surges lower than $(z - \mathrm{Med}(Y_k))$ to the total number of observed skew surges lower than u. Finally, an approximate value of the sea level z probability density function $\tilde{g}_\theta(z)$ is deduced from the cumulative density function $\tilde{G}_\theta(z)$:

$$\tilde{g}_\theta(z) \approx \left[\frac{\tilde{G}_\theta(z+h) - \tilde{G}_\theta(z)}{h}\right] \tag{A5}$$

For the computations, h is set equal to $0.01z$.

Appendix B Available historical information

Table B1. Historical information at Brest. In parenthesis, skew surges not exceeding u.

Date	1856	1877	1882	1888	1899	1913	1928	1936	1939	1940
Sea levels (m)	8.03	8.05	8.03	8.14	8.04	8.02	8.10	8.10	8.07	8.05
Skew surges (m)	(0.44)	0.91	(0.33)	0.72	(0.37)	0.69	(0.48)	(0.38)	(0.48)	(0.32)

Table B2. Historical information at Dunkerque.

Date	1720	1763	1767	1807	1808	1846	1846	1953
Sea levels (m)	7.68	7.60	7.76	7.60	8.10	7.96	7.86	7.90
Skew surges (m)	1.68	1.94	1.71	1.40	2.20	1.95	2.25	2.17

Table B3. Historical information at La Rochelle. In parenthesis, sea levels not exceeding η_H.

Date	1866	1872	1890	1895	1924	1940
Sea levels (m)	(5.70)	(6.34)	7.30	7.15	7.15	7.40
Skew surges (m)	1.15	1.00	1.02	0.75	1.09	1.60

Table B4. Historical information at Saint Nazaire.

Date	1864	1877	1894	1937	1940
Sea levels (m)	7.16	7.24	7.09	7.16	7.12
Skew surges (m)	0.90	1.25	1.35	0.82	1.41

Appendix C Settings of the Monte Carlo runs

The independence between skew surges and astronomical high tides has to be verified to consider the sea levels as the sum of both components randomly sampled independently. To evaluate the interactions between astronomical high tides and skew surges, Williams et al. (2016) proposed to i) visually analyse the scatter plot of observed astronomical high tides versus the corresponding skew surges (Figure C1), and ii) conduct a Kendall test (Table C1, the test is conducted on the largest skew surge values that are of particular interest here). Both indicate that there is no obvious correlation between astronomical high tides and skew surges. Especially, the skew surges exceeding u, correspond to diverse levels of high tides.

It is worth noting that the sample of historical events valuated in method 3 is a sub-set of the sample of events used in method 2 at Saint Nazaire. It furthermore includes 3 of the 4 largest observed skew surge events. At Brest, a station with a large tide/surge ratio, the samples used for the implementation of the two methods are almost totally different : they have only two events in common including only one of the largest observed skew surges.

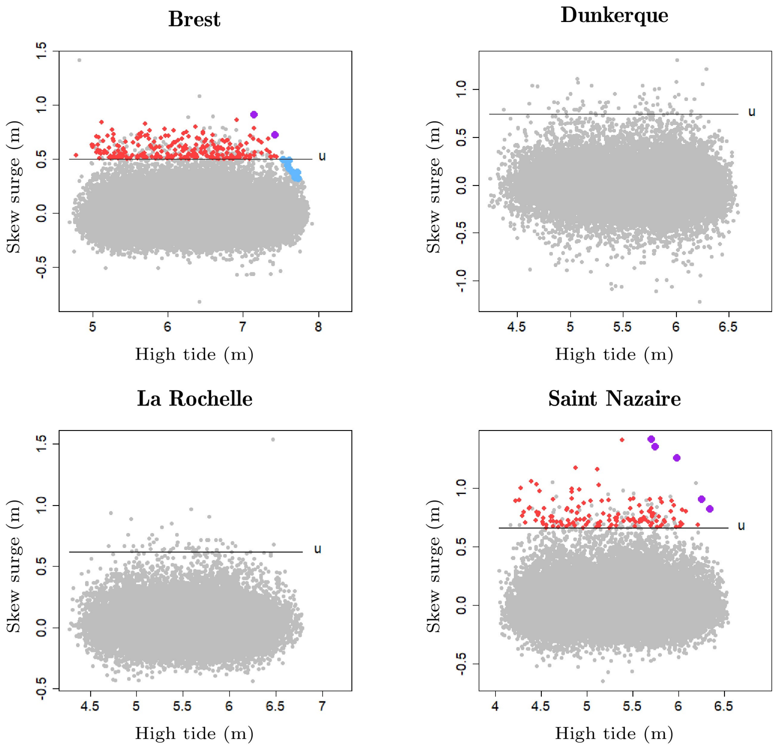

Figure C1. Scatter plot of the high tide / skew surge samples. For Brest and Saint Nazaire, the red points represent the historical sample used in method 2 (ideal case), the blue points represent the historical sample used in method 3 (proposed method) and the purple points represent the observations common to both historical samples.

Table C1. Kendall's τ and p-value (5%) for the top 1% skew surges.

Site	τ	p-value
Brest	-0.023	0.257
Dunkerque	-0.009	0.806
La Rochelle	-0.021	0.628
Saint Nazaire	-0.45	0.65

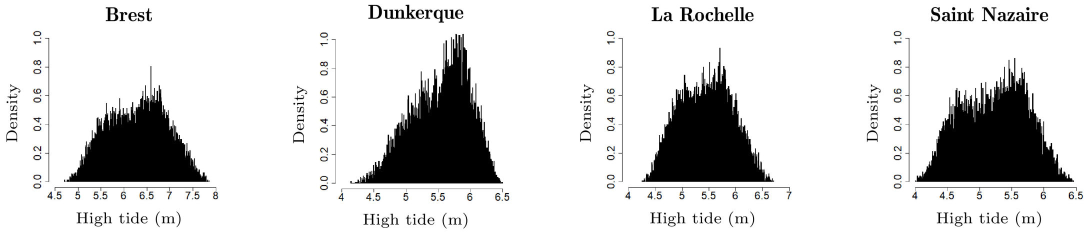

Figure C2. Empirical distributions of astronomical high tides.

The empirical distributions of astronomical high tides for the four case studies are shown in Figure C2. The number n_T intervals used to describe these distributions in the numerical implementation (see Appendix A) depends on the range of high tide values at each station: 4.70m to 7.86m at Brest (317 intervals), 4.14m to 6.49m at Dunkerque (237 intervals), 4.26m to 6.71m at La Rochelle (247 intervals), 4.00m to 6.46m at Saint Nazaire (247 intervals).

Acknowledgments

The authors would like to thank the Editor, Madan Kumar Jha, the Associate Editor and three anonymous reviewers for all their efforts and for their comments which helped improve the quality of the manuscript. Data available at: `doi.org/10.5281/zenodo.6260203`.

References

Aelbrecht, D., Benoit, M., & Allilaire, J. (2004). Renforcement de la protection contre l'inondation du front de Gironde sur le site du Blayais : apports conjoints des modélisations physique et numérique. *La Houille Blanche*(3), 37–44. doi: 10.1051/lhb:200403004

Andreewsky, M., Kergadallan, X., Bernardara, P., Benoit, M., Gaufres, P., & Trmal, C. (2014). Comparaison de différentes méthodes d'estimation des niveaux extrêmes en site à fort et à faible marnage. *La Houille Blanche*, 26–36. doi: 10.1051/lhb/2014035

Arns, A., Wahl, T., Haigh, I. D., Jensen, J., & Pattiaratchi, C. (2013). Estimating extreme water level probabilities: A comparison of the direct methods and recommendations for best practise. *Coastal Engineering*, *81*, 51–66. doi: 10.1016/j.coastaleng.2013.07.003

Batstone, C., Lawless, M., Tawn, J., Horsburgh, K., Blackman, D., McMillan, A., . . . Hunt, T. (2013). A UK best-practice approach for extreme sea-level analysis along complex topographic coastlines. *Ocean Engineering*, *71*, 28–39. doi: 10.1016/j.oceaneng.2013.02.003

Bellier, J. (2018). *Prévisions hydrologiques probabilistes dans un cadre multivarié: quels outils pour assurer fiabilité et cohérence spatio-temporelle?* (PhD Thesis).

Benito, G., Lang, M., Barriendos, M., Llasat, M. C., Francés, F., Ouarda, T., . . . Bobée, B. (2004). Use of systematic, palaeoflood and historical data for the improvement of flood risk estimation. review of scientific methods. *Natural Hazards*, *31*(3), 623–643. doi: 10.1023/B:NHAZ.0000024895.48463.eb

Breilh, J.-F., Bertin, X., Chaumillon, e., Giloy, N., & Sauzeau, T. (2014). How frequent is storm-induced flooding in the central part of the bay of biscay? *Global and Planetary Change*, *122*, 161–175. doi: 10.1016/j.gloplacha.2014.08.013

Bulteau, T., Idier, D., Lambert, J., & Garcin, M. (2015). How historical information can improve estimation and prediction of extreme coastal water levels: application to the Xynthia event at La Rochelle (France). *Natural Hazards and Earth System Sciences*, *15*, 1135–1147. doi: 10.5194/nhess-15-1135-2015

Coles, S. (2001). *An introduction to statistical modeling of extreme values.*

De Zolt, S., Lionello, P., Nuhu, A., & Tomasin, A. (2006). The disastrous storm of 4 November 1966 on Italy. *Natural Hazards and Earth System Sciences*, *6*(5), 861–879. doi: 10.5194/nhess-6-861-2006

Dixon, M. J., & Tawn, J. A. (1994). Extreme sea-levels at the UK A-class sites: site-by-site analyses.

Dixon, M. J., & Tawn, J. A. (1999). The effect of non stationarity on extreme sea level estimation. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *48*, 135–151. doi: 10.1111/1467-9876.00145

Frau, R., Andreewsky, M., & Bernardara, P. (2018). The use of historical information for regional frequency analysis of extreme skew surge. *Natural Hazards and Earth System Sciences*, *18*(3), 949–962. doi: https://doi.org/10.5194/nhess-18-949-2018

Gaal, L., Szolgay, J., Kohnov, S., Hlavcova, K., & Viglione, A. (2010). Inclusion of historical information in flood frequency analysis using a Bayesian MCMC technique: a case study for the power dam Orlk, Czech Republic. *Contributions to Geophysics and Geodesy*, *40*, 121–147. doi: 10.2478/v10126-010-0005-5

Gaume, E. (2018). Flood frequency analysis: The bayesian choice. *Wiley Interdisciplinary Reviews: Water*, *5*(4). doi: 10.1002/wat2.1290

Gerritsen, H. (2005). What happened in 1953? The Big Flood in the Netherlands in retrospect. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *363*(1831), 1271–1291. doi: 10.1098/rsta.2005.1568

Giloy, N., Duluc, C.-M., Frau, R., Ferret, Y., Bulteau, T., Mazas, F., & Sauzeau, T. (2018). La base de données TEMPETES : un support pour une expertise collégiale et interdisciplinaire des informations historiques de tempêtes et de submersions.. doi: 10.5150/jngcgc.2018.093

Giloy, N., Hamdi, Y., Bardet, L., Garnier, E., & Duluc, C.-M. (2019). Quantifying historic skew surges: an example for the Dunkirk Area, France. *Natural Hazards*, *98*(3), 869–893. doi: 10.1007/s11069-018-3527-1

Haigh, I. D., Nicholls, R., & Wells, N. (2010). A comparison of the main methods for estimating probabilities of extreme still water levels. *Coastal Engineering*, *57*, 838–849. doi: 10.1016/j.coastaleng.2010.04.002

Halbert, K., Nguyen, C. C., Payrastre, O., & Gaume, E. (2016). Reducing uncertainty in flood frequency analyses: A comparison of local and regional approaches involving information on extreme historical floods. *Journal of Hydrology*, *541*, 90–98. doi: 10.1016/j.jhydrol.2016.01.017

Hamdi, Y., Bardet, L., Duluc, C.-M., & Rebour, V. (2015). Use of historical information in extreme-surge frequency estimation: the case of marine flooding on the la rochelle site in france. *Natural Hazards and Earth System Sciences*, *15*, 1515–1531. doi: 10.5194/nhess-15-1515-2015

Hamdi, Y., Garnier, E., Giloy, N., Duluc, C.-M., & Rebour, V. (2018). Analysis of the risk associated with coastal flooding hazards: a new historical extreme storm surges dataset for Dunkirk, France. *Natural Hazards and Earth System Sciences*, *18*, 3383–3402. doi: 10.5194/nhess-18-3383-2018

Hosking, J. R., & Wallis, J. R. (1987). Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics*, *29*(3), 339–349. doi: 10.1080/00401706.1987.10488243

Kergadallan, X., Bernardara, P., Benoit, M., & Daubord, C. (2014). Improving the estimation of extreme sea levels by a characterization of the dependence of skew surges on high tidal levels. *Coastal Engineering Proceedings*, *1*, 48. doi: 10.9753/icce.v34.management.48

Kolen, B., Slomp, R., & Jonkman, S. N. (2013). The impacts of storm Xynthia February 2728, 2010 in France: lessons for flood risk management. *Journal of Flood Risk Management*, *6*(3), 261–278. doi: https://doi.org/10.1111/jfr3.12011

Liu, J. C., Lence, B. J., & Isaacson, M. (2010). Direct joint probability method for estimating extreme sea levels. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, *136*, 66–76. doi: 10.1061/(ASCE)0733-950X(2010)136:1(66)

Mazas, F., Kergadallan, X., Garat, P., & Hamm, L. (2014). Applying POT methods to the Revised Joint Probability Method for determining extreme sea levels. *Coastal Engineering*, *91*, 140–150. doi: 10.1016/j.coastaleng.2014.05.006

Nguyen, C., Gaume, E., & Payrastre, O. (2014). Regional flood frequency analyses involving extraordinary flood events at ungauged sites: further developments and validations. *Journal of Hydrology*, *508*, 385–396. doi: 10.1016/j.jhydrol.2013.09.058

Ouarda, T., Rasmussen, P., Bobée, B., & Bernier, J. (1998). Utilisation de l'information historique en analyse hydrologique fréquentielle. *Revue des sciences de l'eau / Journal of Water Science*, *11*, 41–49. doi: 10.7202/705328ar

Outten, S., Wolf, T., Mangini, F., Chen, L., & Nilsen, J. E. (2020). *Re-assessing extreme sea level events through interplay of tides and storm surges* (Tech. Rep.). doi: 10.5194/egusphere-egu2020-8000

Payrastre, O., Gaume, E., & Andrieu, H. (2011). Usefulness of historical information for flood frequency analyses: Developments based on a case study. *Water Resources Research*, *47*. doi: 10.1029/2010WR009812

Pugh, D. T., & Vassie, J. M. (1978). Extreme sea levels from tide and surge probability. *Coastal Engineering Proceedings*, *1*, 52. doi: 10.9753/icce.v16.52

Pugh, D. T., & Vassie, J. M. (1980). Applications of the joint probability method for extreme sea level computations. *Proceedings of the Institution of Civil Engineers*, *69*, 959–975. doi: 10.1680/iicep.1980.2179

Reis, D., & Stedinger, J. (2005). Bayesian MCMC Flood Frequency Analysis With Historical Information. *Journal of Hydrology*, *313*, 97–116. doi: 10.1016/j.jhydrol.2005.02.028

Schendel, T., & Thongwichian, R. (2017). Considering historical flood events in flood frequency analysis: Is it worth the effort? *Advances in Water Resources*, *105*, 144–153. doi: 10.1016/j.advwatres.2017.05.002

Stephenson, A. (2015). *Tideharmonics: harmonic analysis of tides.*

Tawn, J. A. (1992). Estimating probabilities of extreme sea levels. *Applied Statistics*, *41*, 77. doi: 10.2307/2347619

Tawn, J. A., Vassie, J. M., & Gumbel, E. J. (1989). Extreme sea levels; the joint probabilities method revisited and revised. *Proceedings of the Institution of Civil Engineers*, *87*, 429–442. doi: 10.1680/iicep.1989.2975

Tomasin, A., & Pirazzoli, P. A. (2008). Extreme Sea Levels in the English Channel: Calibration of the Joint Probability Method. *Journal of Coastal Research*, *4*, 1–13. doi: 10.2112/07-0826.1

Williams, J., Horsburgh, K. J., Williams, J. A., & Proctor, R. N. F. (2016). Tide and skew surge independence: New insights for flood risk. *Geophysical Research Letters*, *43*, 6410–6417. doi: 10.1002/2016GL069522