Nat. Hazards Earth Syst. Sci. Discuss., 2, 5647–5688, 2014
www.nat-hazards-earth-syst-sci-discuss.net/2/5647/2014/
doi:10.5194/nhessd-2-5647-2014

Natural Hazards and Earth System Sciences
Open Access
Discussions

Use of historical information in extreme surge frequency estimation: case of the marine flooding on the La Rochelle site in France

Y. Hamdi[2], L. Bardet[1], C.-M. Duluc[1], and V. Rebour[1]

[1]Institute for Radiological Protection and Nuclear Safety, BP17, 92262 Fontenay aux Roses Cedex, France
[2]National Engineering School of Gabes, University of Gabes, Rue Omar Ibn-Elkhattab 6029, Gabès, Tunis

Received: 28 May 2014 – Accepted: 23 August 2014 – Published: 3 September 2014

Correspondence to: Y. Hamdi (yasser.hamdi@irsn.fr)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Nuclear power plants located in the French Atlantic coast are designed to be protected against extreme environmental conditions. The French authorities remain cautious by adopting a strict policy of nuclear plants flood prevention. Although coastal nuclear facilities in France are designed to very low probabilities of failure (e.g. 1000 year surge), exceptional surges (outliers induced by exceptional climatic events) had shown that the extreme sea levels estimated with the current statistical approaches could be underestimated. The estimation of extreme surges then requires the use of a statistical analysis approach having a more solid theoretical motivation. This paper deals with extreme surge frequency estimation using historical information (HI) about events occurred before the systematic record period. It also contributes to addressing the problem of the presence of outliers in data sets. The frequency models presented in the present paper have been quite successful in the field of hydrometeorology and river flooding but they have not been applied to sea levels data sets to prevent marine flooding.

In this work, we suggest two methods of incorporating the HI: the Peaks-Over-Threshold method with HI (POTH) and the Block Maxima method with HI (BMH). Two kinds of historical data can be used in the POTH method: classical Historical Maxima (HMax) data, and Over a Threshold Supplementary (OTS) data. In both cases, the data are structured in historical periods and can be used only as complement to the main systematic data. On the other hand, in the BMH method, the basic hypothesis in statistical modeling of HI is that at least one threshold of perception exists for the whole period (historical and systematic) and that during a giving historical period preceding the period of tide gauging, only information about surges above this threshold have been recorded or archived. The two frequency models were applied to a case study from France, at the La Rochelle site where the storm Xynthia induced an outlier, to illustrate their potentials, to compare their performances and especially to analyze the impact of the use of HI on the extreme surge frequency estimation.

1 Introduction

France derives over 75 % of its electricity from nuclear energy. Most nuclear power plants in France are not located on the coasts. Only four are located on the Atlantic coast. Like any other installations, nuclear power plants can be subject to external influences and aggressions such as extreme environmental events (river and/or marine flooding, earthquakes, etc.). The Blayais nuclear power plant was partially flooded when the storm Martin struck the French coast in 1999. A combination of an exceptional surge (outlier), of a high tide and high waves (induced by strong winds) led to the overflow of the dikes. According to Mattéi et al. (2001), these dikes were not designed for such a concomitance of events. A guide to protection, including some fundamental changes in the assessment of flood risks, has therefore been produced by the Nuclear Safety Authority (ASN, 2013). However, some issues like the frequency estimation of extreme surges remain among the priorities of the Institute for Radiological Protection and Nuclear Safety (IRSN). During the last three decades, France has experienced several other violent storms (1987, Martin in 1999, Klaus in 2009 and Xynthia in 2010) that given rise to very high surges which can be considered as outliers.

Surges frequency estimation is an important step in the analysis of the risk associated with marine flooding. The estimation of the frequency of occurrence of extreme environmental events using probability functions has been a common issue since many decades (e.g. Chow, 1953; Dalrymple, 1960; Gringorten, 1963; Cunnane, 1978; Chow et al., 1988; Rao and Hamed, 2000). The engineer generally needs to determine the surge of a given return period T, i.e. the surge quantile X_T or design surge. Traditional methods for analyzing and estimating the frequency of extreme events have been generally based on available local observations from the systematic record alone. However, it is a common belief that these methods are not really suitable for extreme events data sets and especially those containing outliers (e.g. Hu, 1987; Stedinger, 1988; Ebrary, 1999). Indeed, the presence of outliers in data sets can lead to a bad adequacy and selection of the appropriate distribution as well as on the estimation of its parameters.

NHESSD
2, 5647–5688, 2014

An outlier can be defined as observed value which is visibly higher than the rest of the data. In order to base our statistical inference on the right tail of the selected distribution, detection and treatment of outliers are key elements to an effective frequency estimation and risk analysis (Barnett and Lewis, 1998; Chebana, 2012).

During the last four decades, several authors (Leese, 1973; US Water Resources Council Hydrology Committee (USWRC), 1982; Stedinger and Cohn, 1986; Condie, 1986; Jarrett, 1990; Salas et al., 1994; Ouarda et al., 1998; Hamdi, 2011; Payrastre et al., 2011, 2013) have recognized the value of using other sources of information in the frequency estimation of extreme events.

Indeed, to obtain both a consistent and accurate estimate of extreme events requires the use of a consistent technique which increases the representativity of outliers in data sets. The regional estimation, in which at-site observed exceptional events may become normal regional extreme observations and do not appear to be outliers any more, was considered by the scientific community as a serious track to analyze the frequency of occurrence of the surges. However, the inter-sites dependency issue in the regional framework must be revised (Bardet et al., 2012). Additional information refers also to historical events which have been experienced before the systematic period. HI may arise from verbal communication from the general public, written records in archives (books, newspapers, damage reports, unpublished written records, etc.) and from high-water marks left by extreme floods for instance. Other sources of HI such as paleoflood data (which can be obtained from the manipulation of certain types of proxy data) can be useful as well. A review of the literature on HI and the role it can play in a frequency analysis has been made by several authors (e.g. Stedinger and Baker, 1987; Salas et al., 1994; Ouarda et al., 1998). The basic reason for working on such a topic arises from the fact that, despite their significant impacts on nuclear related facilities and on economic and social activities, statistical characterization of extreme storm surges, using HI, has not been handled in the literature. The probabilistic and statistical treatment of surge data containing outliers is also limited in the literature especially in a local frequency analysis context.

A basic hypothesis in statistical modeling of HI is that at least one threshold of perception (S_0) (below which we are utterly ignorant about the magnitude of the surge) exists and that, during a giving historical period preceding the period of tide gauging, only surges above the threshold (large enough to be remembered and/or to leave a mark somewhere on the coastal region) have been recorded or archived. Typically S_0 is the surge at which significant economic damage occurs. We named this method the Block Maxima method with HI (BMH). Plotting position rules to calculate observed probabilities based on both historical and systematic information, have been proposed in the literature (Hirsch, 1987; Hirsch and Stedinger, 1987; Guo, 1990). The development proposed by Hirsch (1987) is considered herein. Another method of using HI in the frequency estimation of extreme events, which has not been frequently explored in the literature, is based on the use of two kinds of historical data: classical Historical Maxima (HMax) data, and Over a Threshold Supplementary (OTS) data. In both cases, the data are structured in historical periods and can be used only as complement to the main systematic data. This method is called Peaks-Over-Threshold method with HI (POTH) in this paper. Whatever the approach used, parametric methods based on the method of the Maximum Likelihood (ML) for estimating the distribution parameters have been developed and used (Leese, 1973; Hosking and Wallis, 1986a; Guo and Cunnane, 1991). The selection of the ML method was based on its statistical features for large samples (consistency, provides an asymptotic variance, asymptotic unbiased estimator). It was also and especially selected for its ability to easily incorporate any additional data in the estimation process. The ML method with additional non-systematic information has been used by many investigators. For instance, it was used with the GEV distribution by Phen and Fang (1989), with the Gumbel population by Leese (1973), Hosking and Wallis (1986a) and Guo and Cunnane (1991) and with other distributions (Condie and Lee, 1982; Cohn and Stedinger, 1986; Pilon and Adamowski, 1993; Francés et al., 1994, 1998; Kroll and Stedinger, 1996). HI are often imprecise, and we should consider their inaccuracy in the analysis. However, even with important uncertainty, the use of HI is a viable mean to increase the representativity of outliers in the sample (Hosking and Wallis, 1986a; Wang, 1990; Salas et al., 1994; Payrastre et al., 2011).

The objective of the present work is to examine the potential gain in estimation accuracy with the use of Historical events for the two methods described above. The La Rochelle tide gauge located in the French Atlantic coast is used in this paper as a case study. The paper is organized as follows. A brief review of the basic definition and concepts is presented in Sect. 2. The third section presents the case study with systematic and historical data. In the Sect. 4, the developed frequency analysis models will be applied to the La Rochelle site. The results of the analysis are further discussed in the same section, before the conclusion and perspectives in Sect. 5.

2 Basic definitions and concepts

A frequency estimation procedure with HI must include the following steps: (i) hypothesis testing (verification of stationarity, randomness and homogeneity) and detection of outliers, (ii) selection of S_0 (for the BMH method) and the historic threshold (for the POTH-OTS), (iii) compute of empirical probabilities corresponding to systematic observations (for the BMH method, only observations below S_0 are taken into account), (iv) compute of empirical probabilities of historical data (using Weibull formula for the POTH approach and the exceeding formula for the BMH one). Because some surge events exceeding the perception threshold S_0 occur among the systematic data, these events are virtually removed from the systematic period and are treated as historical data (Bayliss and Reed, 2001) in the BMH frequency model. (v) Selecting a parameters estimation method (ML in our case) and estimating a theoretical distribution function (the GEV for the TMP model and the GP for the POTH one) to fit the observations, (vi) calculating from the theoretical distribution the T year values of interest (say 100 or 1000 years). We analyze systematic observed skew surges that are available in the period from 1941 to 2010. Data sets were extracted from the predicted tide levels and corrected observations (Bardet et al., 2012). The selected extremes for a particular year were obtained from statistically independent surges. It should be noted that the same raw data set analyzed by Bardet et al. (2012) is used in this paper.

A common problem in statistical modeling is the existence of gaps within the observation period (due to damage or failure of the measurement system, human errors, strikes, wars, etc.). Whatever the HI approach (BMH, POTH) to be used, we must take into account these missing periods in the analysis. Worth noting is that it happens quite often that failure in the measuring stations occurs because of the storm thus creating a non-independent gap. Then we should ensure that the gaps occur independently from measured variable. By means of classical extreme value distributions we are often able to analyze the statistical behavior of the maximum of a sequence of iid random variables. Such a sequence usually represents values measured on a regular time scale. In our application the extreme surge data used for the BMH frequency model are the Annual Maximum (AM) surges. As it is the maximum of a block of values, it is often denoted as block maximum (Fig. 1).

On another hand and in the context of a point process model, surge events occur at successive random times T_i when a random variable is observed. We consider that only large surge values are of interest (values exceeding a sufficiently high threshold u_s). We can assume that the times T_i corresponding to large enough surges X_i should be described by an Homogenous Poisson Process for which the number N of surges on a time interval of length w_s follows a Poisson distribution with mean λw_s. The λ parameter is the rate and it is generally given in number of events per year. Surges X_i are assumed to be independent identically distributed (iid) random variables with distribution $F_X(x)$ and density $f_X(x)$. The return period in such a case is given by $T = 1/\lambda(1 - F_X(x))$. The annual maximum surges (for the BMH model) and the peaks-over-threshold (POT) surges (for the POTH model) are used in this paper as systematic data sets. The construction of the probability plot is a key step of a frequency analysis. There are several formulas that can be used to calculate the observed probability of an event in the systematic period. Based on several studies (e.g. Alam et al., 2005; Makkonen, 2006) the Weibull plotting position rule was used ($p = i/(n+1)$).

3 Historical Information

HI refers to events which have been experienced before the systematic gauging period. Although the frequency analysis using HI often lacks precision in the data used, it has been shown to improve the frequency estimation of extreme events. As mentioned earlier in this paper, there are two methods (BMH and POTH) to integrate HI with systematic data in a frequency model. Whatever the method used and comparing with the traditional frequency models, there are two fundamental differences when using HI:

1. A different class of plotting position formulas can be used for the empirical probabilities. For instance, the exceedance formula is used when applying the BMH frequency model. The exceedance formulas are based on the fact that the frequency of the whole surges exceeding S_0 (since it subdivides the range of probabilities between groups of below and above thresholds) is a key descriptor of the data set. The development of these formulas is addressed hereafter. For the POTH model, the same plotting position formula (Weibull) is used, but the probabilities of historical information and of the systematic data are separately calculated.

2. A second fundamental difference when using HI in a frequency model is that an adapted MLE methodology should be used. To take into consideration available information on historical events, this methodology leads to use additional terms (depending on the method and the information type) in the likelihood expression.

Typically, historical surges are stated in terms of a large event (a huge submersion for instance) which took place in a particular year without knowing the characteristics (rank or magnitude) of the events that took place in other years. Also worth noting is that the completeness is not guaranteed. In fact, we only know extreme surges that have caused damages and thus participating in high levels. Extreme historical surges associated with low or medium tides are often not known. In this paper, surges are described in four types: (i) exact historical surges (magnitudes can be accurately estimated), (ii) events above a threshold (lower bound), (iii) an upper bound over a given

NHESSD
2, 5647–5688, 2014

period: this bound was never exceeded, (iv) surges whose magnitudes are known with some uncertainty and bounded by upper and lower limits (range). On the basis of these HI types, a joint likelihood function of the historical and systematic data $l(G|\theta)$ will be expressed and discussed in more detail in the rest of the paper. With the tools exploited in this study, both BMH and POTH models can use exact values as HI. However, the second and the last HI types can only be used with the BMH frequency model whereas the third type can only be used with the POTH model (particular case of OTS historical information without events above the threshold).

3.1 The Block Maxima method with HI (BMH)

In applying the BMH method one may assume that HI are available in relation to one or more fixed thresholds of perception. Typically these thresholds of perception (S_t) are surges at which important economic damage occurs. A first difficulty with this method is to quantify the magnitude of historical surge events (as it is often subject to important uncertainty) and to estimate the empirical frequencies from a sample of both systematic data and HI. A second difficulty lies in the hypothesis of the HI completeness. Indeed, historical surge data are supposed to be exhaustive and all the events exceeding S_t are reported. In other words, for the entire historical period, we assume that the thresholds have not been exceeded except for the available historical data.

3.1.1 Plotting positions of systematic and historical data for the BMH model: multi-thresholds case

For historical data, let us assume that there are m thresholds $S_1, S_2, \ldots, S_m$ such that only k_t highest observations are larger than or equal to them. Let s be the total number of systematic observed surges (annual maxima) and g the total number of surges in our data set (k of them are known to be the highest) in the total period of n years. The period of n years contains within it the s years ($s < n$) systematic period. Note that e of the k highest values are occurred during the systematic period ($g = k + s - e; e \leq k; e \leq s$).

NHESSD
2, 5647–5688, 2014

Figure 2 shows an illustration with a multi-threshold case (m thresholds), 9 exact values over thresholds ($k = 9$), 5 of them are systematic ($e = 5$), three lower bounds and one range.

A set of plotting positions can be determined in this multi-threshold case. Define a series of thresholds S_t ($t = 1,2,\ldots m$) such that $S_1 > S_2 > \ldots > S_m$ (Fig. 2). The systematic record can be considered as a special case of historical exact data with $S_{m+1} = 0$. For convenience, define $S_0 = +\infty$. To estimate the probability of exceedance $\hat{P}_i$ of each observed surge, one needs to estimate exceedance probabilities $\hat{P}_t$ of each threshold S_t. The estimates $\hat{P}_t$ and $\hat{P}_i$ must have the property $\hat{P}_t < \hat{P}_i < \hat{P}_{t-1}$ whenever $S_{t-1} < X_i < S_t$ and $\hat{P}_1 < \hat{P}_2 < \ldots < \hat{P}_n$ (n is the total number of annual surges X_i over the n years). The probability of exceedance of a threshold P_t can be defined as:

$$P_t = \Pr[X \geq S_t]. \tag{1}$$

The last equation can be re-expressed as:

$$P_t = \Pr[X \geq S_{t-1}] + \Pr[S_t \leq X < S_{t-1} | X < S_{t-1}] \cdot \Pr[X < S_{t-1}] \tag{2}$$

$$P_t = P_{t-1} + \Pr[S_t \leq X < S_{t-1} | X < S_{t-1}] \cdot (1 - P_{t-1}) \tag{3}$$

$$\begin{cases} \hat{P}_t = \hat{P}_{t-1} + C_\mathrm{p}^t \cdot (1 - \hat{P}_{t-1}) \\ P_0 = \Pr[X \geq \langle S_0 = \infty \rangle] = 0 \end{cases} \tag{4}$$

where C_p^t is the conditional probability of threshold S_t. Recursive compute is possible in this formulation. We can start by estimating the probability of exceedance of the lowest threshold and work upward. The boundary conditions become useful for this calculation. The aim now is to calculate the conditional probability C_p^t that a surge falls between the tth and the $(t-1)$th thresholds. Since m periods of N_t($t = 1,2,\cdot m$) are associated to m thresholds where $N_1 > N_2 > \ldots > N_m$, one can define a (m threshold $\times$ m periods) matrix n_{tp} of all combinations of number of surges above all the thresholds and during all the periods. And consider for each threshold S_t two variables: $\underset{\leftrightarrow t}{A}^{t-1}$ the

NHESSD
2, 5647–5688, 2014

number of surges X, such that $S_t \leq X < S_{t-1}$ and $\overleftarrow{\underset{t-1}{B}}$ the number of surges X, where $X < S_{t-1} \cdot \underset{\leftrightarrow t}{A}^{t-1}$ and $\overleftarrow{\underset{t-1}{B}}$ can be expressed as:

$$\begin{cases} \underset{\leftrightarrow t}{A}^{t-1} = n_{tt} - n_{(t-1)t} \\ \overleftarrow{\underset{t-1}{B}} = N_t - \int_1^t n_{(p,t)} \mathrm{d}p \\ t = 2, \ldots, m \,\&\, p = 1, 2, \ldots, t \end{cases} \tag{5}$$

n_{tt} is the number of surges above threshold S_t during period N_t and for instance, n_{32} is the number of surges above threshold S_3 ($m = 3$) during the period N_2 ($p = 2$). Hirsch and Stedinger (1987) have shown that the conditional probability estimated by the method of moments (identical to the maximum likelihood estimator) as $\underset{\leftrightarrow t}{A}^{t-1} / \overleftarrow{\underset{t-1}{B}}$. The assignment of specific Weibull plotting positions to all the individual known surges which are greater than threshold S_t but below threshold S_{t-1} can be generalized to the formula:

$$\hat{P}_i = (1 - \hat{P}_t) + (\hat{P}_t - \hat{P}_{t-1}) \cdot \frac{i - a}{\underset{\leftrightarrow t}{A}^{t-1} + 1 - 2a} \tag{6}$$

where i is the rank of the ith surge among the $\underset{\leftrightarrow t}{A}^{t-1}$ surges in the range $S_t \leq X_i < S_{t-1}$ and a is a constant ($a = 0$ if the spacing between plotting positions is Weibull). By combining Eqs. (4) and (6), a different expression of exceedance probabilities can be obtained:

$$\hat{P}_i = (1 - \hat{P}_t) \left[1 + C_\mathrm{p}^t \cdot \frac{i - a}{\underset{\leftrightarrow t}{A}^{t-1} + 1 - 2a} \right]. \tag{7}$$

NHESSD
2, 5647–5688, 2014

Let the spacing between plotting position is Hazen ($a = 0.5$):

$$\hat{P}_i = (1 - \hat{P}_t)\left[1 + \frac{i - 0.5}{\overleftarrow{B}_{t-1}}\right]. \quad (8)$$

This can be re-expressed as:

$$\hat{P}_i = \hat{P}_{i-1} + \frac{1 - \hat{P}_t}{\overleftarrow{B}_{t-1}}. \quad (9)$$

3.1.2 Maximum Likelihood estimator for the BMH model

The ML estimators are obtained by maximizing the likelihood function over the parameter space $\underset{\leftrightarrow}{\theta}$ (or usually simpler, by maximizing the logarithm of this function). For a given systematic and historical surges data set G (where g is its length and n is its total period (as described in Sect. 2.3), k of the g observations are above the threshold of perception. Also remember that the n years period contains within it the systematic record period of s years ($s < n$) and e of the k values took place during the systematic period ($g = k + s - e; e \leq k; e \leq s$). Let us assume that the g surges are available with a density function $f_X(.)$. Under the assumption that the surges are iid (independent and identically distributed), the global likelihood function of the whole data sample is any function $L(G|\underset{\leftrightarrow}{\theta})$ proportional to the joint probability density function $f_X(.)$ evaluated at the observed sample and it is the product of the likelihood functions of the particular types of events and information E_i:

$$L(G|\underset{\leftrightarrow}{\theta}) = \prod_{i=1}^{n} L(E_i|\underset{\leftrightarrow}{\theta}). \quad (10)$$

NHESSD
2, 5647–5688, 2014

The events can be the $(s-e)$ systematic events (x_i), the k historical surges (y_i) above the threshold of perception S_t with exactly known magnitudes, the lb lower bounds historical events $(\overset{i}{\underset{\text{lb}}{y}})$ and the r range historical events $[y_{\text{lb}}, y_{\text{ub}}]_i$. The likelihood function is given by the joint distribution of $K, Y_1, \ldots, Y_k$, and $X_1, X_2, \ldots, X_{s-e}$. It can be expressed as:

$$L(\underset{\leftrightarrow}{\theta}, x, y, k) = \Pr[K = k; \underset{\leftrightarrow}{\theta}] \cdot \prod_{j=1}^{k} f_{X|X>S_t}(y_j, \underset{\leftrightarrow}{\theta}) \cdot \prod_{i=1}^{s-e} f_X(x_i, \underset{\leftrightarrow}{\theta}) \tag{11}$$

where y_i denotes a surge exceeding the threshold of perception S_t and observed in the total period n (as described in Sect. 2.3). Since $f_{X|X>S_t}(y) = f_X(y)/p$, the above equation simplifies to

$$L(\underset{\leftrightarrow}{\theta}, x, y, k) = \binom{n}{k} (1-p)^{n-k} \cdot \prod_{j=1}^{k} f_X(y_j, \underset{\leftrightarrow}{\theta}) \cdot \prod_{i=1}^{s-e} f_X(x_i, \underset{\leftrightarrow}{\theta}). \tag{12}$$

The joint likelihood function $L(G|\underset{\leftrightarrow}{\theta})$ of all the events (with the lower band and the range data) can then be re-expressed as:

$$L(G|\underset{\leftrightarrow}{\theta}) = \overbrace{\prod_{i=1}^{s-e} f_X(x_i, \underset{\leftrightarrow}{\theta})}^{\text{sys. below threshold}} \cdot \overbrace{\binom{n}{k} (1-p)^{n-k} \cdot \prod_{j=1}^{k} f_X(y_j, \underset{\leftrightarrow}{\theta})}^{\text{exact known historic } \bullet} \cdot \overbrace{\prod_{i=1}^{\text{lb}} \left[1 - F\left(\overset{i}{\underset{\text{lb}}{y}}, \underset{\leftrightarrow}{\theta}\right)\right]}^{\text{lower bound historic } \uparrow} \cdot \tag{13}$$

$$\overbrace{\prod_{i=1}^{r} \left[F\left(\overset{i}{\underset{\text{ub}}{y}}, \underset{\leftrightarrow}{\theta}\right) - F\left(\overset{i}{\underset{\text{lb}}{y}}, \underset{\leftrightarrow}{\theta}\right)\right]}^{\text{range historic } \updownarrow}.$$

NHESSD
2, 5647–5688, 2014

3.2 The Peaks-Over-Threshold method with HI (POTH)

The historical data (HMax and OTS data) are structured in historical periods and can be used only as complement to the main POT systematic data (Fig. 3). The OTS data is over a threshold historical data recorded on periods with known durations and known exceedances. The periods are assumed to be potentially disjoint from the systematic period and other historical periods. For each period with known duration w_{OTS}, we must have a threshold u_{h} and all observations exceeding this threshold. The historical threshold u_{h} cannot be smaller than the main threshold u_{s} (used for the POT systematic data). A HMax data period corresponds to a time interval of known duration w_{HMax} during which historical k largest values are available. Periods are also assumed to be potentially disjoint from the systematic period and other historical periods.

3.2.1 Plotting positions of systematic and historical data for the POTH model

Calculating empirical probabilities of historical data and constructing an empirical distribution function (in a POTH context) is the same as that of systematic data. A classic empirical formula can be used. Otherwise, as mentioned earlier in this paper, it was shown that Weibull plotting position formula $p = i/(\tilde{n}+1)$ is a more adequate than the other commonly used rules. In this formula, $\tilde{n}$ is a prediction of the number of events on the historical period for HMax (since this number is unknown). A natural choice is $\tilde{n} = \lambda w_{\mathrm{HMax}}$ (λ is the events rate on the systematic period). For an OTS historical period with duration w_{h} and with no surge observations, the never exceeded threshold u_{h} is shown as a horizontal segment with return period up to w_{h}.

NHESSD
2, 5647–5688, 2014

3.2.2 Maximum likelihood estimator for the POT systematic data

Let us assume a set of n POT systematic observations X_i and a selected threshold u_s and consider w_s the total duration. The likelihood $L(X_{sys,i}|\theta)$ is:

$$L(X_{sys,i}|\theta) = \Pr(N = n) \cdot \prod_{i=1}^{n} f(X_{sys,i}, \theta). \tag{14}$$

For a Homogeneous Poisson Process with rate λ, the above equation can be re-expressed as:

$$L(X_{sys,i}|\theta) = \frac{(\lambda w_s)^n}{n!} \exp(-\lambda w_s) \cdot \prod_{i=1}^{n} f(X_{sys,i}, \theta). \tag{15}$$

The log-likelihood $\ell(X_{sys,i}|\theta)$ is then:

$$\ell(X_{sys,i}|\theta) = n \log(\lambda w_s) - \log(n!) - \lambda w_s + \sum_{i=1}^{n} \log f(X_{sys,i}, \theta). \tag{16}$$

3.2.3 Maximum likelihood estimator for the OTS historical data

For HI, consider an OTS period with threshold u_h, duration w_{OTS} and with given k observed surges $X_{OTS,i}$ for $i = 1, 2, \ldots k$. The events with $X_{OTS} > u_h$ from an Homogenous Poisson Process with a rate $\lambda \cdot (1 - F(x, \theta))$ and their number K follow a Poisson distribution. Conditional on $\{K = k\}$, the k observed surges are independent with density $f(x)/(1 - F(u_h, \theta))$ for $x > u_h$. Hence

$$L(X_{OTS,i}|\theta) = \Pr(K = k) \cdot \prod_{i=1}^{n} \frac{f(X_{OTS,i}|\theta)}{[1 - F(u_h, \theta)]} \tag{17}$$

NHESSD
2, 5647–5688, 2014

$$L(X_{\mathrm{OTS},i}|\theta) = \frac{(\lambda w_{\mathrm{OTS}}[1-F(u_\mathrm{h},\theta)])^k}{k!} \cdot \exp(-\lambda w_{\mathrm{OTS}}[1-F(u_\mathrm{h},\theta)]) \cdot \prod_{i=1}^{k} \frac{f(X_{\mathrm{OTS},i}|\theta)}{[1-F(u_\mathrm{h},\theta)]} \quad (18)$$

and by taking the log, the log-likelihood $\ell(X_{h,i}|\theta)$ takes the form

$$\ell(X_{\mathrm{OTS,i}}|\theta) = k\log(\lambda w_{\mathrm{OTS}}) - \log(k!) - \lambda w_{\mathrm{OTS}}[1-F(u_\mathrm{h},\theta)] + \sum_{i=1}^{k} \log f(X_{\mathrm{OTS,i}},\theta). \quad (19)$$

3.2.4 Maximum likelihood estimator for the HMax historical data

Consider an HMax period with duration w_{HMax} and with given k observed surges $X_{\mathrm{HMax},i}$ for $i = 1,2,\ldots k_\mathrm{h}$. For $\{N = n\}$ and conditional on $\{N = n\}$, the probability of observing the surge $X_{\mathrm{HMax},i}$ is:

$$\Pr[X_1, X_2, \ldots, X_k | N = n] = \frac{n!}{(n-k)!} \cdot F(X_k)^{n-k} \cdot \prod_{i=1}^{k} f(X_{\mathrm{HMax},i},\theta). \quad (20)$$

Among the n observations, k values must be equal to the observed ones while the $n-k$ remaining values must be less or equal to $X_{\mathrm{HMax,\,k}}$. The likelihood can be obtained by using the total probability formula:

$$L(X_{\mathrm{HMax},i}|\theta) = \sum_{n=k}^{\infty} \Pr[N=n] \cdot \frac{n!}{(n-k)!} \cdot F(X_k)^{n-k} \cdot \prod_{i=1}^{k} f(X_{\mathrm{HMax},i},\theta). \quad (21)$$

After using the Homogenous Poisson Process with rate λ

$$L(X_{\mathrm{HMax},i}|\theta) = \prod_{i=1}^{k} f(X_{\mathrm{HMax},i},\theta) \cdot (\lambda w_{\mathrm{HMax}})^k . \exp(-\lambda w_{\mathrm{HMax}}) \cdot \sum_{n=k}^{\infty} \frac{(\lambda w_{\mathrm{HMax}})^{n-k}}{(n-k)!} \cdot F(X_k)^{n-k} \quad (22)$$

$$L(X_{\mathrm{HMax},i}|\theta) = (\lambda w_{\mathrm{HMax}})^k.\exp(-\lambda w_{\mathrm{HMax}}[1-F(X_k)]) \cdot \prod_{i=1}^{k} f(X_{\mathrm{HMax},i},\theta) \quad (23)$$

and by taking the log, the log-likelihood $\ell(X_{h,i}|\theta)$ takes the form

$$\ell(X_{\mathrm{HMax},i}|\theta) = k\log(\lambda w_{\mathrm{HMax}}) - \lambda w_{\mathrm{HMax}}[1-F(X_k,\theta)] + \sum_{i=1}^{k} \log f(X_{\mathrm{HMax},i},\theta). \quad (24)$$

3.2.5 Global likelihood estimator

The global log-likelihood can be expressed as

$$\ell(G|\theta) = \overbrace{\ell(X_{\mathrm{sys},i}|\theta)}^{\text{systematic data}} + \overbrace{\ell(X_{\mathrm{OTS,i}}|\theta)}^{\text{OTS data}} + \overbrace{\ell(X_{\mathrm{HMax},i}|\theta)}^{\text{HMax data}}. \quad (25)$$

The three terms of this equation are developed in the previous three sections.

4 Data

The frequency analysis is performed at the La Rochelle site, which is located on the French Atlantic Coast. Independent surges time series were obtained from the predicted tide levels and corrected observations (Bardet et al., 2012). Surge database was provided by the French Oceanographic Service SHOM (Service Hydrographique et Océanographique de la Marine). One of the most important features of the station La Rochelle is the fact that the region, in which this station is located, has experienced important storms during the last two decades (the storm Martin in 1999 and Xynthia in 2010). Figure 4 displays the geographic location of the La Rochelle site.

NHESSD
2, 5647–5688, 2014

4.1 Systematic data and statistical tests

In the case of the BMH model, the annual maximum surges available in the period from 1941 to 2010 constitute the systematic record and all events occurred before 1941 are considered as HI. The choice of systematic data for the POTH approach is different. Indeed, to use all available data separated by periods of missing data, we concatenated all the surges available since 1941 in a one systematic record. It was shown in previous researches (e.g. Bardet, 2011; Hamdi et al., 2013) that a threshold u_s equal to 41 cm is an adequate choice that gives a rate λ equal to 4 events/year. BMH and POTH systematic data from 1941 to 2010 are represented by bar plots in Figs. 5 and 6 respectively.

Stationarity, homogeneity and randomness of time series are required conditions in a frequency analysis (Rao and Hamed, 2001). The Kendall test for stationarity (Mann, 1945), the Wald–Wolfowitz Test for randomness (Wald and Wolfowitz, 1943) and the Wilcoxon test for homogeneity (Wilcoxon, 1945) are the three non-parametric tests used as a prerequisite for frequency analysis. The La Rochelle station passed these tests at the significance level of 5 %. The Grubbs–Beck Test (GBT) for the detection of outliers is also used in the present work. The results of this test show clearly that the 2010 event is undoubtedly an outlier. The reader can be referred to Grubbs and Beck (1972) for more details on this test.

4.2 Settings of the BMH and POTH frequency models

Some historical known events experienced by the La Rochelle site are available. The oldest information was collected as part of the Ph.D. thesis of Gouriou (2012) at the University of La Rochelle. A 13 years surges record (between 1863 and 1875) in the La Rochelle site was treated in this thesis. Only the four highest values are considered in the present work (Table 1). To verify the validity of the obtained historical data, we basically searched through Garnier and Surville (2010). We could identify that there is at least one extreme event happened in 1957 (during a gap in the systematic period exploited in the present work) and which has not been quantified in terms of surge. This event is comparable in terms of consequences to the Xynthia event. This is worrying for us since exhaustiveness in HI has been assumed in the developed BMH model and in the POTH model with OTS data. A possible solution is to choose a threshold of perception S_t (in the case of the BMH method) or a historical threshold u_{h} (in the case of the POTH method) as high as possible. This would allow us to get closer to the completeness of the HI. Care must be taken when using the available HI (Table 1), it is not necessary to use all the available historical data and depending on the selected threshold, one or more historical data can be excluded from the analysis. This is indeed the case here, only the historical surge of 111 cm induced by the event that took place in 1866 was used. Some elements of justification will be presented in the next section. The other data are considered quite small and taking a threshold of perception below these values further exacerbate the problem of the assumed completeness.

We introduced in a previous section some parameters characterizing data sets (containing both a HI and a systematic record) that we apply here on the La Rochelle records.

4.2.1 Settings of the BMH frequency model

As mentioned in Sect. 4.1, events occurred before 1941 are considered as HI. The threshold of perception is not easy to estimate. In addition to the fact that it was not exceeded during all the historical period except when information is available, the value of the threshold of perception must meets the following criteria: (i) should be as high as possible, (ii) should be exceeded by the selected historical values introduced in the estimation, (iii) may result in serious economic damage and disruption. In the case of marine surges, the last criterion is not always taken into account for the simple reason that high surges can occur at the same time with low tide and may not cause damage.

The 2010 event (Xynthia) is exceptional (comparing to the other observed systematic surge peaks) and affected not only the French coasts but also the Western Europe. The surge value (153 cm) caused by the Xynthia event has largely exceeded two times the average of the other surge records (61 cm). This event is considered as historic in the BMH frequency model. Thereby changing its empirical probability but not the value of the overall likelihood. Consequently, the theoretical probabilities remain unchanged. All other systematic surges seem ordinary extreme ones with relatively low return periods. The choice of the threshold of perception becomes then a goal easy to achieve, we simply need to choose a threshold that allows us to consider as historic only the Xynthia event among the systematic ones. It should range between the Xynthia event surge (153 cm) and the maximum value of all the other systematic observations (90 cm). A value of $S_t = 110$ cm with a total of two exact values ($k = 2$) is then utilized in the present paper. Only one of these values is systematic ($e = 1$) and there are no lower bounds (lb = 0) and ranges ($r = 0$) that can be used for this case study. Figure 5 illustrates well these model settings.

4.2.2 Settings of the POTH frequency model

The parameters characterizing a POTH model data set are u_h and w_h. The events occurred before 1941 are considered as HI. The HI is used in this work as OTS data. As shown in Fig. 6, HI complements the systematic record with these OTS data recorded on one block (1866–1940) with known duration ($w_\mathrm{OTS} = 75$ years) and one exceedance ($k = 1$). As was the case for the BMH frequency model and to ensure exhaustiveness of HI, a threshold as high as possible should be chosen. A value of $u_\mathrm{h} = 110$ cm is then utilized. Unlike the BMH frequency model, systematic events that exceeded the historical threshold are not considered as historical in the POTH model. Consequently their empirical probabilities will be different (compared to their equivalents in the BMH model) but theoretically the likelihood remains insensitive to the fact that these observations are considered as systematic or historic. Worth noting is that theoretically, with the proposed BMH and POTH models settings, the portion of the likelihood related to the addition of the HI must converge to the same value and the impact of the use of the HI must be the same for the two concepts.

NHESSD
2, 5647–5688, 2014

The POTH frequency model is applied with the Renext tool developed by the French Institute for Radiological Protection and Nuclear Safety (IRSN). The Renext software proposes fits and diagnostics for the POT method. The software is programmed in the R environment (open-source software for statistical computing).

5 Results and discussion

In this section we report the results of the BMH and POTH frequency models for the extreme storm surges frequency estimation and analysis applied to the La Rochelle station. All the developments and simulations were carried out within the R environment. The packages "evd" (Stephenson, 2012) and "Renext" (Deville, 2013), which have specifically been developed for extreme events frequency analysis, was used. As shown in Figs. 5 and 6, the surge frequency analysis was performed considering only one scenario for each method: 30 systematic surges and 1 exact historical observation for the BMH frequency model and 106 systematic POT surges and 1 OTS data for the POTH model. Whatever the frequency model settings, the historical period considered in these two cases starts when the first historical data was observed (75 years long). For each model, the evaluation of the impact of additional HI (by comparing quantiles with no historical data and those obtained using the whole 1866–2010 data set, including both systematic and HI) constitutes the main result of this paper and it is highlighted later in the present section. These results are presented in form of probability and diagnostic plots (Figs. 7 and 8) and in form of tables summarizing the numerical estimates of the return levels of interest and uncertainty parameters.

By fitting the GEV distribution to the data yields the following parameters, the absolute and relative widths (ΔCI and $\Delta \mathrm{CI}/S_t$, respectively) of the confidence intervals of the representative surge quantiles which are presented in Tables 2 and 3. Since high return periods are needed for the safety of nuclear installations, the main focus was set to the 100, 500 and 1000 year surge quantiles. The uncertainty corresponding to the return level is closely related to that of the model parameters. The Delta method was used to calculate the variance of the return level estimates. The distribution functions (GEV for the BMH model and GP for the POTH model) was fitted to data. Figures 7 and 8 consists of 4 subplots. In each figure are displayed the fitting and the Q-Q plots (top left panel and top right panel, respectively) excluding the HI. The graphics with HI are presented in the bottom. The solid line in the middle is the fitted distribution function, while the dashed lines are the lower and the upper bounds of the 70 % confidence intervals of surge quantiles. In the plots at the bottom (with HI), the systematic data are represented by empty circles, while the HI is depicted by a solid red circle. In the next sub-sections, we present the results of each frequency model separately. However, we can afford to analyze first in the present sub-section the results when no HI is included. It can be seen that the fit with no historical data included, presented in Figs. 6 and 7, is poor at the right tail of the distribution and more particularly at the outlier. It is observed that the POT model gives lower surge estimates for the La Rochelle station than the AM model. Quantiles corresponding to 100, 500 and 1000 year return periods given by the POT model are approximately 6 cm lower than those obtained by the AM method. The correspondent confidence intervals are very large when the annual maxima data are used and it is much narrower in the case the POT frequency model is used. This was expected because the robustness of the results greatly depends on the amount of available information and since the POT method generally uses more data than the AM one. Also note that this difference between the AM and the POT models should decrease and the estimated values become fairly close when additional data (given by historical information in our case) is used. This also has been proven in a theoretical way by the scientific community (e.g. Langbein, 1947; Chow et al., 1988; Hamdi et al., 2013).

5.1 Results of the BMH frequency model

In a classical frequency analysis (with systematic data only), some goodness-of-fit criteria and tests (e.g. χ^2, Kolmogorov–Smirnov and Anderson–Darling tests) can be applied. As soon as any HI is used, testing the goodness of the fit becomes a more complicated business. The way additional HI affects the quantile and uncertainty estimates can even though be analyzed using a visual inspection of diagnostic plots (Fig. 7) and the numerical results presented in Table 2. As shown in Fig. 7, the use of HI appreciably changes not only the empirical probabilities and the estimated quantiles but also the uncertainty associated to these estimates. The observed probabilities of surges that exceeded the threshold of perception were calculated using the exceedance formula. It can be seen that this formula allows better positioning of the outliers in the systematic data and gives to these events more reasonable return periods. It can then be concluded that the fit with empirical frequencies of large surges (at the right tail of the distribution and especially at the outlier) is more satisfactory when using HI.

It is discernible that the inclusion of the HI reduces the relative width of the confidence bounds. This finding is underpinned by the values of the absolute and relative widths of the confidence intervals of surge quantiles presented in Table 2. This is the case when more data is included in the parameters estimation (if the extra data is consistent with previous one). The relative widths of the confidence bounds for the BMH frequency model with no HI involved are 2.4 times larger than those obtained with both systematic and historical surges. This holds for all the return periods of interest (100, 500 and 1000 years). This fact is also graphically underpinned by the confidence intervals plots presented in Fig. 7. The BMH model with both systematic and historical data gives return levels systematically lower than those given with systematic data only.

5.2 Results of the POTH frequency model

The visual inspection of diagnostic plots is complicated when the POTH frequency model is used. Remember here that the empirical probabilities are calculated with the Weibull plotting position rule (not with the exceedance formula), but in a separate way for the historical information and the systematic data. Their positions are therefore not the same as those given by the BMH model. We can quickly see from Fig. 8 that the visual fit at the outlier is almost the same as that obtained with no HI included. The fit of the large empirical frequencies, at the right tail of the distribution excluding the outlier,

NHESSD
2, 5647–5688, 2014

is slightly better than that obtained without HI. The improvement is not significant and as shown in Fig. 7 several observations have remained outside the confidence interval.

The results given in Table 3 indicate that the values of relative confidence intervals widths with no HI involved are 1.65 times narrower than those obtained with both systematic and historical surges. In other words, the user of this method is more confident in its parameters and quantiles estimates when using HI. As with the BMH model, return levels given by the POTH frequency model with HI are also lower than those obtained with no historical data included.

6 Further discussions on the impact of the historical information

One of the most important questions still remains is about the robustness: what is the impact of the use of the HI on the BMH and POTH models results (for high return levels)? And what are the shortcomings and the advantageous of the two frequency models?

From Tables 2 and 3, the convergence of the AM quantiles towards the POT ones is clearly observed when the HI is used. 100, 500 and 1000 years quantiles obtained by the POTH method are close to those given by the BMH one (with a difference of only 1.8, 1.6 and 1.4 % respectively). This convergence can be seen as tangible evidence of the robustness of the two models results. Unlike the POTH model, the BMH one gave a relatively better visual fit not only at the outlier but also at the other large empirical frequencies. The confidence intervals obtained by the POTH method are much more advantageous (much tighter) and the relative width is only 17 % for the 100 years return surge (against 61 % for the BMH model).

The biggest disadvantage of the BMH frequency model comes from the completeness assumption. It is assumed that during the historical period the perception threshold was never exceeded except in years when the information is available. This hypothesis can quickly become annoying if we have evidence that one or more significant events occurred during the historical period and we cannot get more details on these

NHESSD
2, 5647–5688, 2014

events. This is precisely the case of the La Rochelle site that we used as a case study in the paper. It is shown in the literature that an extreme event took place 15 February 1957 in the same region (Garnier and Surville, 2010). It is also said that this event is comparable (in terms of damage) to Xynthia event (which combined large tide with the surge outlier in this study) and unfortunately the sources give no information on the value of the surge. The issue on which we are working now is the way we can consider the historical data that cannot be accurately estimated. Such situation requires a different treatment of the data wherein one may choose to use only the fact that during the historical period, the threshold of perception was exceeded at least k times. This is called "binomial censored data" (Stedinger and Cohn, 1986).

In order to compare the two approaches, the two models settings were chosen to be as close as possible. This is why in the POTH model we have considered one block ending where the systematic period began (the same period as that used in the BMH model). The POTH historical information can be handled differently; one could use the four available historical data (Table 1) over the corresponding period of 13 years without making assumptions about the non-gauging (1873–1940).

Another issue that has arisen in the present paper is about the historical data that were not used because they were below the threshold of perception (Table 1). Assuming that these data are of valuable information, we strongly believe that these data must be exploited. Theoretically, nothing prevents us to consider them as systematic information even if it occurred 50 years before the beginning of the systematic period. In other words, systematic data that have exceeded the threshold of perception are considered as historical information and historical data that have not exceeded this threshold would be considered as systematic information, subject to an adapted length of the whole systematic period. This work is ongoing.

NHESSD
2, 5647–5688, 2014

7 Conclusion

In this paper, two methods how to use historical surges into the local frequency analysis have been presented and applied on the La Rochelle site located on the French Atlantic Coast. The first method is based on the presence of a perception threshold using annual maxima data (BMH model) and the second method is rather based on the use of historical data periods and POT surges (POTH model). Two adapted likelihood functions and different class of plotting positions formulas are then built to properly handle the information on historical surges. Several types of HI can be considered in the frequency analysis when using these methods. Unlike the BMH frequency model, systematic events that exceeded the historic threshold are not considered as historical data in the POTH model. Consequently their empirical probabilities are different but the likelihood remains the same with similar hypotheses for the historical information. Although the two approaches are physically different, the use of historical information has then close impacts when using the two methods with the available data and the applied settings.

Overall adding information on historical surges to the local frequency models has reduced the variance of the distribution function parameters estimates. It improved the fitting to the observations especially when the BMH frequency model was used. However, this does not imply that using historical information will provide an improved model fit as more data may invalidate the exhaustiveness assumption. Despite of the fact that relatively subjective choices of the perception thresholds (or historical thresholds) cannot be avoided, the two methods (BMH and POTH) give robust results for the La Rochelle case study when considering a threshold as high as possible. Indeed, the comparison of the theoretical distributions fitted to AM and POT extreme events have indicated that, at high return periods (greater than 100 years), the two approaches give comparable predictions of surge magnitudes when historical information is used. This can be interpreted differently; the amount of information is more critical and important than the quality of the approach used.

NHESSD
2, 5647–5688, 2014

Care is always needed when modeling data with the aim of extrapolating whatever the model used (BMH or POTH) and regardless of whether or not HI is used. As indicated by the results of the case study and reasoning in terms of high return levels and associated uncertainty, we have two models that handle differently the impact of the use of HI on high quantiles estimation. In such an exercise in which extreme values are used with the aim of extrapolating, it is very important to account for the underlying error with which we make our inferences. Despite the improvement of the uncertainty degree resulting from the use of HI in the BMH model, this uncertainty remains high. The distribution parameters and quantiles estimates are qualified with much more appropriate degree of uncertainty when using the POTH frequency model, because more systematic data are available with this model. On the other hand, how the theoretical distribution is representative of the observations is also an important issue. Overall observations are relatively well fitted except at the outlier when the POTH frequency model is used. Unlike this model, the BMH one provides a much better visual fit at large frequencies and tends to similar return levels when HI is used. Indeed, the BMH frequency model has the advantage to allow the repositioning in history of the systematic observations that exceeded the threshold of perception (the outlier for instance). However, the BMH frequency model needs to assume the exhaustiveness of HI whereas the POTH method is more faithful to the observations and needs fewer hypotheses with regards to the empirical distributions.

During our research that led to the writing of this paper we noted three perspectives. The first one is related to the kinds of historical information, which are different in the two models (lower and/or upper bounds, threshold of perception ...). The BMH and POTH models could be adapted to consider similar HI and the plotting positions for the POTH model could be calculated with the exceedance formula. In addition, another kind of HI considering "at least *k* events above the threshold of perception" would be very interesting for the two models, then partially answering to the exhaustiveness issue. The second perspective is relative to the spatial extension of extreme surges data including HI. It can be envisaged using the two methods presented herein may improve the frequency estimation and uncertainty in a regional framework (by increasing the amount of information), especially when annual maxima are used. Finally, the last perspective is no less important and may be complementary to the first and the second ones. The question that has arisen is about the historical data that were not used because they were below the threshold of perception. Assuming that these data are of valuable information, we think that these data have to be exploited.

Acknowledgements. The authors are grateful to two referees and the Editor for their helpful comments, corrections, and suggestions. The authors thank Luc Hamm and Frank Mazas from ARTELIA for providing the historical data for the La Rochelle station.

References

Alam, M. J. B. and Matin, A.: Study of plotting position formulae for Surma basin in Bangladesh, J. Civ. Eng. Manage., 33, 9–17, 2005.

ASN – Nuclear Safety Authority: Protection des installations nucléaires de base contre les inondations externes, guide No. 13, France, 2013.

Baker, V. R.: Paleoflood hydrology and extraordinary flood events, J. Hydrol., 96, 79–99, 1987.

Bardet, L. and Duluc, C.-M.: Apport et limites d'une analyse statistique régionale pour l'estimation de surcotes extrêmes en France, Congrès de la SHF, 1–2 février 2012, Paris, France, 2012.

Bardet, L., Duluc, C.-M., Rebour, V., and L'Her, J.: Regional frequency analysis of extreme storm surges along the French coast, Nat. Hazards Earth Syst. Sci., 11, 1627–1639, doi:10.5194/nhess-11-1627-2011, 2011.

Barnett, V. and Lewis, T.: Outliers in Statistical Data, 3rd Edn., John Wiley, Chichester, UK, 1998.

Bayliss, A. C. and Reed, D. W.: The use of historical data in flood frequency estimation, Rep. to MAFF, CEH Wallingford, available at: http://nora.nerc.ac.uk/8060/1/BaylissRepN008060CR.pdf (last access: 14 May 2014), 2001.

Chebana, F., Dabo-Niang, S., and Ouarda, T. B. M. J.: Exploratory functional flood frequency analysis and outlier detection, Water Resour. Res., 48, W04514, doi:10.1029/2011WR011040, 2012.

Chow, V. T., Maidment, D. R., and Mays, L. R.: Applied Hydrology, McGraw-Hill, New York, 1988.

Condie, R.: Flood samples from a three-parameter lognormal population with historic information: the asymptotic standard error of estimate of the T year flood, J. Hydrol., 85, 139–150, 1986.

Cunnane, C.: Unbiased plotting positions – a review, J. Hydrol., 37, 205–222, 1978.

Cunnane, C.: Review of statistical models for flood frequency estimation, in: Hydrologic Frequency Modeling, edited by: Singh, V. P., D. Reidel, Dordrecht, the Netherlands, 49–95, 1987.

Dalrymple, T.: Flood Frequency Analyses, Manual of Hydrology: Part 3, “Flood Flow Techniques” USGS Water Supply Paper 1543-A, US GPO., 1960.

Deville, Y.: Renext: Renewal Method for Extreme Values Extrapolation, availabe at: http://cran.r-project.org/web/packages/Renext/ (last access: 22 May 2014), 2013.

Ebrary, I.: Improving American River Flood Frequency Analyses, National Academy Press, Washington, D.C., 1999.

Garnier, E. and Surville, F.: La tempête Xynthia face à l’histoire. Submersions et tsunamis sur les littoraux français du Moyen Âge à nos jours, Ed. Le croît Vif. Saintes, Saintes, France, 2010.

Gouriou, T.: Evolution des composantes du niveau marin à partir d’observations de marégraphie effectuées depuis la fin du 18e siècle en Charente-Maritime, Ph.D. Thesis, University of La Rochelle, La Rochelle, France, 493 pp., 2012.

Grubbs, F. E. and Beck, G.: Extension of sample sizes and percentage points for significance tests of outlying observations, Technometrics, 14, 847–854, 1972.

Guo, S. L.: Unbiased plotting position formulae for historical floods, J. Hydrol., 121, 45–61, 1990.

Guo, S. L. and Cunnane, C.: Evaluation of the usefulness of historical and paleological floods in quantile estimation, J. Hydrol., 129, 245–262, 1991.

Hamdi, Y.: Frequency analysis of droughts using historical information – new approach for probability plotting position: deceedance probability, Int. J. Global Warm., 3, 203–218, 2011.

Hamdi, Y., Bardet, L., Duluc, C.-M., and Rebour, V.: Extreme storm surges: a comparative study of frequency analysis approaches, Nat. Hazards Earth Syst. Sci., 14, 2053–2067, doi:10.5194/nhess-14-2053-2014, 2014.

NHESSD
2, 5647–5688, 2014

Use of historical information in extreme surge frequency estimation

Y. Hamdi et al.

Hirsch, R. M.: Probability plotting position formulas for flood records with historical information, J. Hydrol., 96, 185–199, 1987.
Hirsch, R. M. and Stedinger, J. R.: Plotting positions for historical floods and their precision, Water Resour. Res., 23, 715–727, 1987.
Hosking, J. R. M. and Wallis, J. R.: The value of historical data in flood frequency analysis, Water Resour. Res., 22, 1606–1612, 1986.
Hu, S.: Problems with outlier test methods in flood frequency analysis, J. Hydrol., 96, 375–383, 1987.
Jarrett, R. D.: Paleohydrologic techniques used to define the spatial occurrence of floods, Geomorphology, 3, 181–195, 1990.
Langbein, W. B.: Annual floods and the partial duration flood series, T. Am. Geophys. Un., 30, 879–881, 1947.
Leese, M. N.: Use of censored data in the estimation of Gumbel distribution parameters for annual maximum flood series, Water Resour. Res., 9, 1534–1542, 1973.
Makkonen, L.: Plotting positions in extreme value analysis, J. Appl. Meteorol. Clim., 45, 334–340, 2006.
Mann, H. B.: Nonparametric tests against trend, Econometrica, 3, 245–259, 1945.
Mattéi, J. M., Vial, E., Rebour, V., Liemersdorf, H., and Türschmann, M.: Generic Results and Conclusions of Re-evaluating the Flooding in French and German Nuclear Power Plants Eurosafe Forum 2001, available at: http://www.eurosafe-forum.org/files/semb1_7.pdf (last access: 21 March 2011), 2001.
Ouarda, T. B. M. J., Rasmussen, P. F., Bobée, B., and Bernier, J.: Utilisation de l'information historique en analyse hydrologique fréquentielle, Rev. Sci. Eau, 11, 41–49, 1998.
Payrastre, O., Gaume, E., and Andrieu, H.: Usefulness of historical information for flood frequency analyses: developments based on a case study, Water Resour. Res., 47, W08511, doi:10.1029/2010WR009812, 2011.
Payrastre, O., Gaume, E., and Andrieu, H.: Historical information and flood frequency analyses: which optimal features for historical floods inventories?, La Houille Blanche, 3, 5–11, 2013.
Rao, A. R. and Hamed, K. H.: Flood Frequency Analysis, CRC Press, Boca Raton, FL, USA, 2000.
Rao, A. R. and Hamed, K. H.: Flood Frequency Analysis, CRC Press, New York, USA, 2001.

NHESSD
2, 5647–5688, 2014

Use of historical information in extreme surge frequency estimation

Y. Hamdi et al.

Salas, J. D., Wold, E. E., and Jarrett, R. D.: Determination of flood characteristics using systematic, historical and paleoflood data, in: Coping with Floods, edited by: Rossi, G., Harmoncioglu, N., and Yevjevich, V., Kluwer, Dordrecht, 111–134, 1994.

Stedinger, J. R. and Baker, V. R.: Surface water hydrology: historical and paleoflood information, Rev. Geophys., 25, 119–124, 1987.

Stedinger, J. R. and Cohn, T. A.: Flood frequency analysis with historical and paleoflood information, Water Resour. Res., 22, 785–793, 1986.

Stephenson, A. G.: evd: extreme value distributions, R News, 2, 31–32, available at: http://CRAN.R-project.org/doc/Rnews/ (last access: 29 April 2014), 2002.

US Water Resources Council Hydrology Committee – USWRC: Guidelines for Determining Flood Flow Frequency, Bull. 17B, US Gov. Print. Off., Washington D.C., USA, 1982.

Wald, A. and Wolfowitz, J.: An exact test for randomness in serial correlation, Ann. Math. Stat., 14, 378–388, 1943.

Wang, Q. J.: Unbiased estimation of probability weighted moments and partial probability weighted moments from systematic and historical flood information and their application to estimating the GEV distribution, J. Hydrol., 120, 115–124, 1990.

Weibull, W.: A statistical theory of strength of materials, Ing. Vetensk. Akad. Handl., 151, 1–45, 1939.

Wilcoxon, F.: Individual comparisons by ranking methods, Biometrica, 1, 80–83, 1945.

Table 1. Historical known values.

Year	1866	1867	1869	1872
Surge (cm)	111	80	87	96

NHESSD
2, 5647–5688, 2014

Table 2. The T year quantiles (in cm) estimated on the basis of the systematic and historical periods and absolute and relative widths of their 70 % confidence interval (the BMH frequency model).

T (years)	No historical data included			Systematic and Historical data		
	S_T (cm)	$\Delta CI_{70\%}$ (cm)	$\Delta CI/S_T$ (%)	S_T (cm)	$\Delta CI_{70\%}$ (cm)	$\Delta CI/S_T$ (%)
100	128.16	183.95	144	119.11	72.59	61
500	157.15	348.25	222	141.64	131.55	93
1000	170.14	437.39	257	151.27	162.14	107

NHESSD
2, 5647–5688, 2014

Table 3. The T year quantiles (in cm) estimated on the basis of the systematic and historical periods and absolute and relative widths of their 70 % confidence intervals (the POTH frequency model).

T (years)	No historical data included			Systematic and Historical data		
	S_T (cm)	$\Delta CI_{70\%}$ (cm)	$\Delta CI/S_T$ (%)	S_T (cm)	$\Delta CI_{70\%}$ (cm)	$\Delta CI/S_T$ (%)
100	123.63	34.23	28	116.96	19.89	17
500	150.95	60.54	40	139.32	34.07	24
1000	163.37	74.96	46	149.10	41.77	28

NHESSD
2, 5647–5688, 2014

NHESSD
2, 5647–5688, 2014

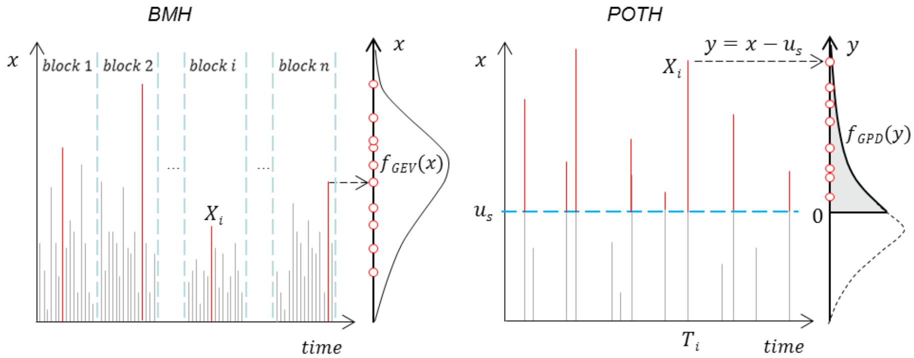

Figure 1. A sketch of a systematic record. To the left: the block maxima data set (for the BMH method). To the right: the POT data set (for the POTH method). In POT only surges $X_i > u_s$ are modeled through exceedances $Y_i = X_i - u_s$. The distribution $F(X)$ is known only for the upper part $x > u_s$.

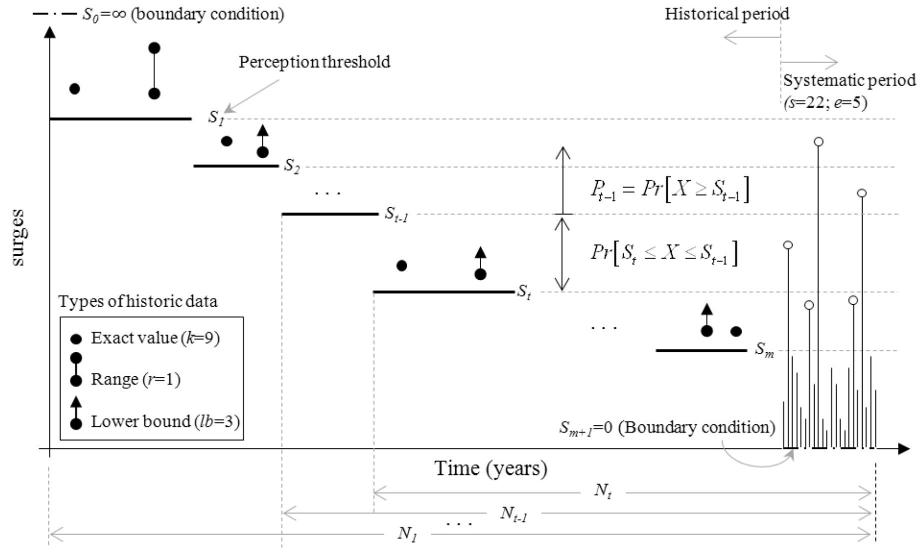

Figure 2. A sketch of systematic records and historical information: a multiple-threshold case.

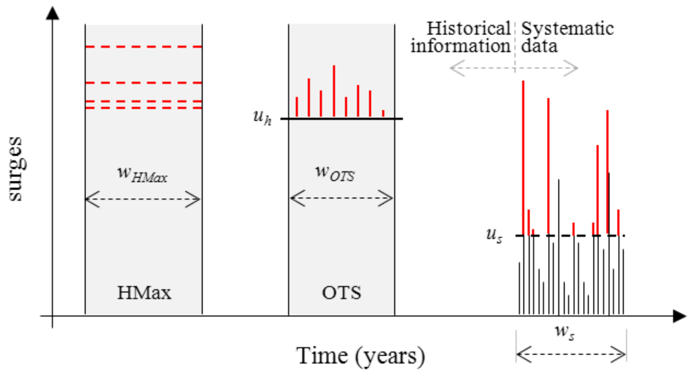

Figure 3. POTH: a sketch of systematic records and historical information (HMax and OTS data).

NHESSD
2, 5647–5688, 2014

Use of historical information in extreme surge frequency estimation

Y. Hamdi et al.

Figure 4. Geographic location of the La Rochelle site.

NHESSD
2, 5647–5688, 2014

Use of historical information in extreme surge frequency estimation

Y. Hamdi et al.

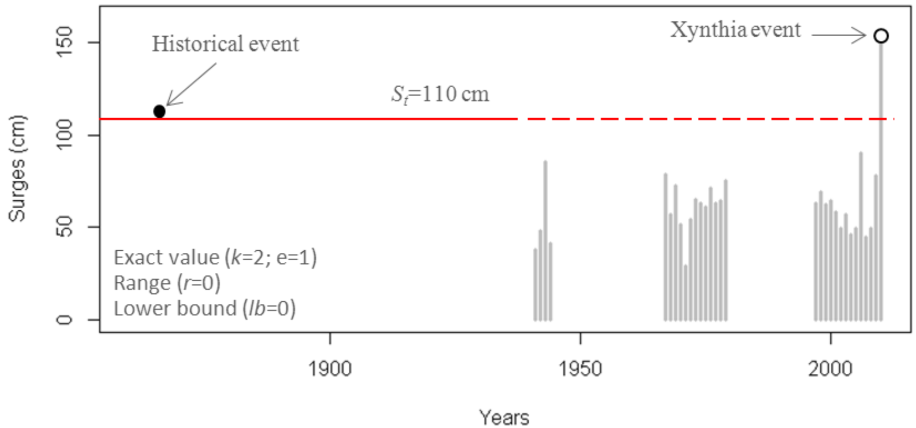

Figure 5. La Rochelle site – systematic and historical data and settings of the BMH frequency model.

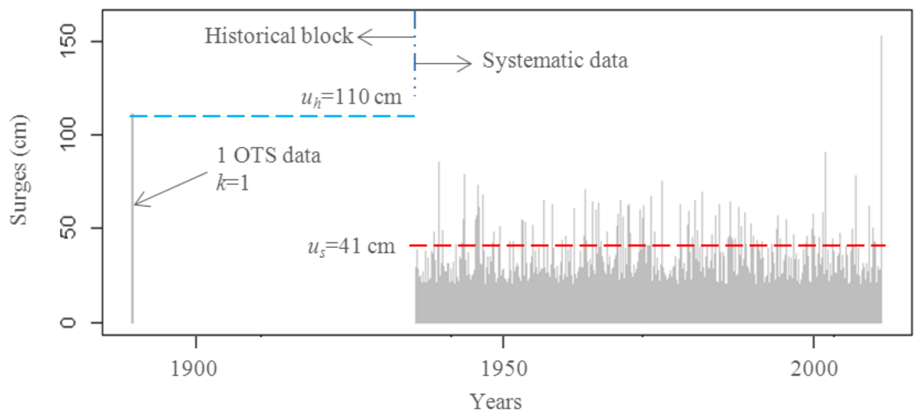

Figure 6. La Rochelle site – systematic and historical data and settings of the POTH frequency model.

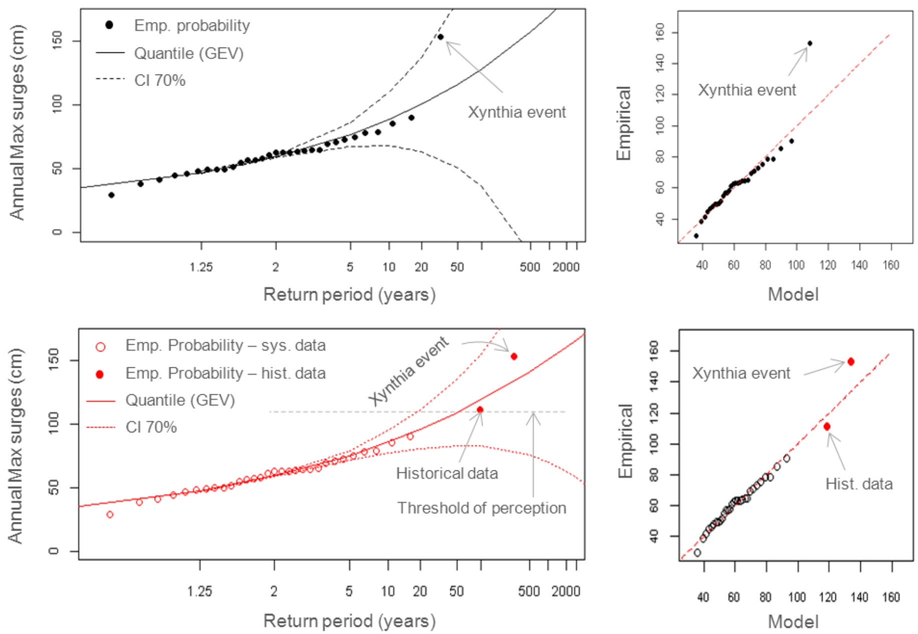

Figure 7. The GEV distribution fitted to the AM of peak surges at the La Rochelle station: with no historical information included (top left panel) and the corresponding Q–Q plot (top right panel); with both systematic and historical data (bottom left panel) and the corresponding Q–Q plot (bottom right panel).

NHESSD
2, 5647–5688, 2014

Use of historical information in extreme surge frequency estimation

Y. Hamdi et al.

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◀| |▶
◀ ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

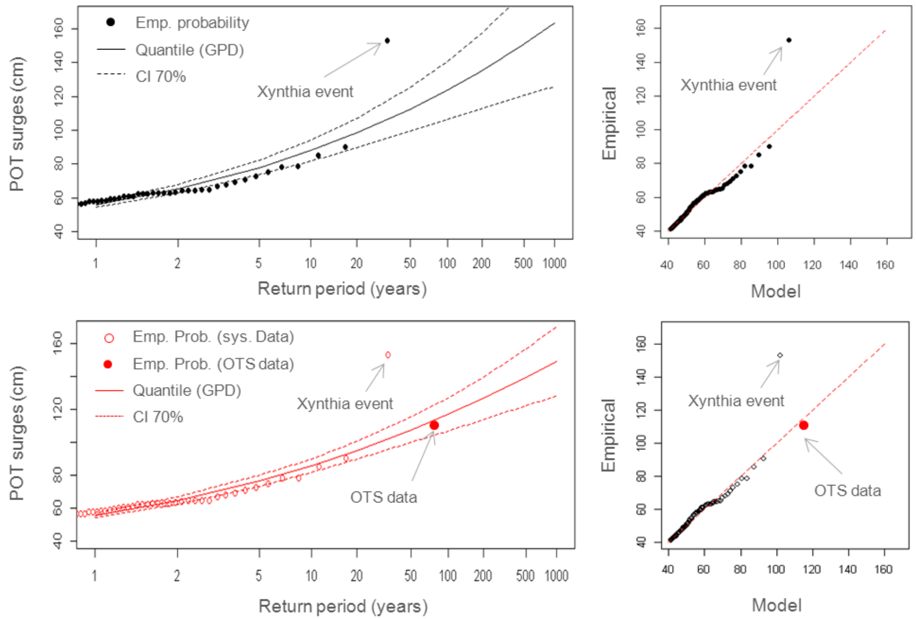

Figure 8. The GP distribution fitted to the POT surges at the La Rochelle station: with no historical information included (top left panel) and the corresponding Q–Q plot (top right panel); with both systematic and historical data (bottom left panel) and the corresponding Q–Q plot (bottom right panel).

NHESSD
2, 5647–5688, 2014