



HAL
open science

Development of a target-site-based regional frequency model using historical information

Yasser Hamdi, Claire-Marie Duluc, Lise Bardet, Vincent Rebour

► **To cite this version:**

Yasser Hamdi, Claire-Marie Duluc, Lise Bardet, Vincent Rebour. Development of a target-site-based regional frequency model using historical information. *Natural Hazards*, 2019, 98 (3), pp.895-913. 10.1007/s11069-018-3237-8 . hal-02527509

HAL Id: hal-02527509

<https://hal.science/hal-02527509>

Submitted on 11 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Development of a target-site-based regional frequency model using historical information

Y. Hamdi, C-M. Duluc, L. Bardet & V. Rebour

Institute for Radiological Protection and Nuclear Safety

Correspondence to: Y. Hamdi (yasser.hamdi@irsn.fr) ORCID: 0000-0001-6329-0074

Abstract

Nuclear power plants in France are designed to withstand natural hazards. Nevertheless, some exceptional storm surges, considered as outliers, are not properly addressed by classical statistical models. Local frequency estimations can be obtained with acceptable uncertainties if the dataset containing the extreme storm surge levels is sufficiently complete and not characterized by the presence of an outlier. Otherwise, additional information such as regional and historical storm surges may be used to mitigate the lack of data and the influence of the outlier by increasing its representativeness in the sample. The objective of the present work is to develop a regional frequency model (RFM) using historical storm surges. Here we propose a RFM using historical information (HI). The empirical spatial extremogram (ESE) is used herein to form a homogenous region of interest centered on a target site. A related issue regards the reconstitution, at the target site and from its neighbors, of missed storm surges with a multiple linear regression (MLR). MLR analysis can be considered conclusive if available observations at neighboring sites are informative enough and the reconstitution results meet some criteria during the cross-validation process. A total of 35 harbors located on the French (Atlantic and English Channel) and British coasts are used as a whole region with the La Rochelle site as a target site. The Peaks-Over-Threshold frequency model is used. The results are compared to those of an existing model based on the index flood method. The use of HI in the developed RFM increases the representativeness of the outlier in the sample. Fitting results at the right tail of the distribution appear more adequate and the 100-year return level is about 30 cm higher. The 100-year return level in the initial fitting has a return period of about 30 years in the updated fitting).

Keywords: extreme surges, outlier, spatial extremogram, regional frequency model, target site, historical information.

Acknowledgements

The authors are grateful to Professor Ivan Haigh for having accepted that our research be presented in EVAN2017. We would like to extend our sincere appreciation to Yves Deville (consulting engineer) for providing us advice, technical support and guidance.

58 1 Introduction

59 Climatic hazards, such as storms and floods, have serious environmental, economic and social consequences.
60 During the last few decades, France has experienced several violent storms, such as the storm of 1987, Lothar
61 and Martin (1999), Klaus (2009) and Xynthia (2010), which induced significantly high storm surges, appearing
62 as outliers in data sets (observation much larger than the second largest). Storm surge, a rise in sea level due to
63 low atmospheric pressure and strong winds, is the main driver of coastal flood events. Extreme storms can
64 produce high sea levels, especially when they coincide with high tide. The skew surge, which is the quantity of
65 interest, is a sea level component which is a fundamental input for statistical investigations of coastal hazards. It
66 is a measure of the non-tidal contribution to peak water-level. Haigh ID et al. (2015) defined the skew surge as «
67 the difference between the maximum observed water level and the maximum predicted tidal level regardless of
68 their timing during the tidal cycle. There is one skew surge value per tidal cycle ». In the nuclear safety field, the
69 skew surge is calculated as the difference between the highest observed water-level and the highest predicted
70 one, for a same high tide. Since these maximum levels can occur at slightly different times, the calculation is
71 performed in a time-window of some hours (three hours for instance).

72 Relating extreme events to their frequency of occurrence using probability distributions, in a local context,
73 has been a common issue since the 1950s (Dalrymple 1960; Rao and Hamed 2000). The at-site frequency
74 analysis (FA), corresponding to long return periods is based on the extreme value theory (EVT) (Coles 2001).
75 The objective of a FA is to analyze past records so as to estimate the probability of future occurrence. Statistical
76 criteria, such as stationarity, randomness and homogeneity, must be satisfied. Since an exceptional event can
77 induce an exceptional observation considered as an outlier, and since the presence of this outlier often leads to
78 poor frequency estimations, the exceptional events must be properly addressed in the FA. The probabilistic and
79 statistical treatment of surge data containing outliers is limited in the literature, especially in a local FA context.

80 In the case of a sufficiently complete and long data set which is not characterized by the presence of an
81 outlier, at-site FA can be used to estimate quantiles with acceptable uncertainties. Otherwise, additional data may
82 be used to mitigate the lack of data and to improve the incorporation of the outlier in the frequency estimation.
83 The additional data may have a spatial dimension (regional) just as they may have a temporal one (making use of
84 historical data). A regional frequency model, RFM, using historical information, HI, may be used as a consistent
85 technique that makes data sets longer and increases the representativeness of the outlier in the sample. The
86 objective of a RFM is to transfer information, within a homogeneous region, from gauged sites to an ungauged
87 or poorly gauged target site. RFMs are commonly used and have been extensively studied (e.g. Dalrymple 1960;
88 Hosking and Wallis 1993, 1997; Weiss 2014). An interesting comparative study of various RFMs was presented
89 by GREHYS (GREHYS 1996a, b) by using four methods for delineating homogenous regions and seven
90 regional estimation methods.

91 One of the simplest and most popular approaches, privileged by engineers and widely used in hydrology, is
92 the index flood method (Dalrymple 1960). The stages of the standard procedure are: (i) delineation of
93 homogeneous regions; (ii) derivation of a regional flood frequency distribution; and (iii) estimation of the
94 quantiles at the site of interest. The major hypothesis of the index flood method is that the probability
95 distributions at different sites in the region are identical, except for a scale parameter. Recently, these RFMs
96 have been applied to coastal hazards (Bardet and Duluc 2012; Bernardara et al. 2011). However, these
97 applications were limited by the problem of the inter-site dependence issue and its impact on the variance of
98 local and regional quantiles and on the statistical homogeneity. Recent developments go a step further and
99 involve a procedure to consider the spatial dependence structure using copulas (Renard 2011; Weiss 2014) and
100 to form physically homogenous regions. In the model proposed by Weiss (2014), a criterion related to spatial
101 storms propagation is used in the formation of physically homogeneous groups of sites. To highlight different
102 regions, a hierarchical cluster analysis, using Ward's method, is performed on the distances between sites
103 obtained by a spatial storm propagation criterion. Likewise, a spatial extreme value copula is also considered to
104 overcome the problem of spatial dependence between surges grouped in a regional sample.

105 As mentioned above, additional information refers also to the magnitudes of the events occurring outside the
106 period of systematic measurements. HI may arise from, for instance, verbal communication from the general
107 public, written records in archives, and from high-water marks left by extreme floods. Several authors have
108 reviewed the literature on HI and the role it can play in an FA (e.g. Stedinger 1987; Salas et al. 1994; Payraastre et
109 al. 2013). More recently, Hamdi et al. (2015) conducted a local FA with HI. It was concluded that the
110 combination of both RI and HI allows us to properly address the exceptional events (including outliers) in a local
111 FA. Similarly, Bulteau et al. (2015) used the HI in a local FA (in a Bayesian framework) to estimate extreme sea
112 levels for the design of coastal defenses. Hosking and Wallis (1986a, b), Tasker and Stedinger (1987) and Jin
113 and Stedinger (1989) tried to assess the usefulness of the HI in a regional context. Their conclusions are
114 somewhat divergent. However, it seems clear that the use of HI is particularly valuable in a local context.

115 In an RFM, one is interested in assigning a weight to the target site. The existing delineation procedures do
116 not give any specific importance to the target site which may lead to a loss of some relevant local effects which
117 could not be absorbed in the standardization step using local indexes. Furthermore, the delineation of a

118 homogenous region usually leads to the problem of the so-called “border effect”. Indeed, if one is interested in a
 119 target site which is very close to the region limit, the information at the site located on the other side of the
 120 region is excluded even though both sites offer similar information and have similar asymptotic properties.
 121 Another limitation of existing RFMs is related to their applicability to coastal hazards. Indeed, most existing
 122 RFMs were developed for estimating extreme floods and in the philosophy behind these developments it was
 123 unintentionally considered that there is a physical limit between regions (watersheds).

124 In our view, these elements motivate and justify the development of a new method that overcomes these
 125 drawbacks. In this paper we put forward a new RFM using HI for further discussion to enrich the debate on FA
 126 in a regional context. The model is presented including both the associated benefits, mainly focusing on the
 127 advantages of performing a local FA using regional and historical information, and the associated disadvantages,
 128 including the several technical challenges that still need to be overcome to ensure that the developed model
 129 becomes an appropriate frequency estimation and a design alternative for protecting structures and helping them
 130 withstand extreme coastal hazards. The developed RFM is based on the notion of asymptotic dependence and the
 131 empirical spatial extremogram (ESE) (Hamdi, et al. 2016) to define a homogenous region around a target site,
 132 the enrichment of local information at the target site (only if it was missed) with RI (from observations at the
 133 neighboring sites), and, finally, the use of any HI in the region of interest. The neighborhood between sites is
 134 measured by a degree of physical and statistical inter-site dependence. The region constituting the neighboring
 135 sites moves from one target site to another. The principle of moving regions around target sites allows us to
 136 overcome the “border effect” problem and is one of the key points of the developed model.

137 A related issue regards the transfer of information from the homogenous region to the target sites. Since
 138 surge records are often characterized by the presence of too many gaps, the regional information is used in the
 139 developed model to enrich the at-target site data set but not to dominate its global statistical behavior. A
 140 weighted regression-based approach can be used to establish a relation between storm surges at the target site
 141 and those observed in the region. A multiple linear regression (MLR) is used in the present paper as a data
 142 reconstitution model. MLR analysis can be considered conclusive only if the available observations at
 143 neighboring sites are informative enough and the reconstitution results meet some criteria during the cross-
 144 validation process.

145 Reanalysis of storm surges and extreme sea levels based on hydrodynamic modelling (generally using the
 146 ‘shallow-water equations’) is another approach that can be used to reconstitute gaps in storm surges series. For
 147 any surge model, an adequate representation of a wind field is necessary because the spatial extent of the wind
 148 and pressure fields associated with a storm is needed as input. Further, the path of the storm and its correct
 149 forward speed are necessary. In the field of nuclear safety, we may be wary of techniques that are likely to
 150 underestimate the variables of interest used in the design of protection structures for instance. In addition, most
 151 of the nearshore numerical circulation models are very computer intensive and require very small time steps (on
 152 the order of seconds) to reach steady-state solutions. This point will be a major limitation for us because
 153 thousands of simulations and meteorological forcings are needed per target site.

154 The main objective of the present work is to conduct a local frequency analysis using regional and historical
 155 information. The spatial extremogram technique to form a homogenous region of interest centered on a target
 156 site is used and the regional information is transferred to this target site with a cross-validated MLR. The paper is
 157 organized as follows. The concept of the empirical spatial extremogram is defined and presented in sections 2.
 158 The third section presents how regional information will be transferred to the target site. The POTH frequency
 159 model will be briefly introduced in section 4. Systematic and historical data and main results are presented in
 160 sections 5 and 6 with some further discussions in section 7, before the conclusion and perspectives in section 8.

161

162 **2 Definition and concept of the empirical spatial extremogram (ESE)**

163 As it was defined in Hamdi et al., (2016), the ESE compute is based on the extremal dependence between the
 164 pairwise extremal dependences (between any site and a target site). The neighborhood threshold N_t and the
 165 extremal dependence coefficient are the most determinant parameters of the ESE. These two parameters are
 166 defined for any site proposed to be neighboring to the target site and then be part of the region of interest. The
 167 extremal dependence coefficient as defined in Hamdi et al., (2016) is the probability that an observation at a site
 168 is extreme (greater than a fixed threshold u) given that the simultaneous observation at the target site belongs to
 169 an extreme set. The theoretical basis of the ESE concept is taken up herein considering a new extremal
 170 dependence coefficient (i.e. the χ coefficient).

171 Let $X = \{X_s\}_{s \in R}$ be a random field on a whole region R with the same one-dimensional marginal (i.e. X_s
 172 are storm surges at site s). The objective is to delineate a region of interest $r \subset R$ containing a homogenous
 173 group of neighboring sites. For convenience we use the following conventions:

174 - X_s belongs to an extreme set if it is greater than a fixed high threshold denoted by $u = (u_s)_{s \in r}$ where
 175 $u_s = u(s)$;

176 - The sample extreme quantile of probability p at a site s (e.g. the probability of non-exceedance $p = 99\%$)
177 can be used to define the threshold u_s ;

178 For any neighboring site s ($s \in r$), its tail dependence with the target site t is characterized by the extremal
179 coefficient χ and it is given by the following limit

$$180 \quad \chi: R \times R \rightarrow [0,1] \quad \chi_{s,t} = \lim_{x \uparrow \infty} P(X_s \geq u_s | X_t \geq u_t) \quad (1)$$

181 In the case of our study, X_s and X_t are the time series of skew storm surges at the site s and the target site
182 t , respectively. The larger the χ , the larger the extremal dependence between the sequences X_s and X_t
183 (considered asymptotically independent when $\chi = 0$ and perfectly dependent when $\chi = 1$). It is noteworthy that
184 the definition of the spatial extremogram presented herein doesn't suppose identical distributions of time series
185 X_s and X_t . On the other hand, once a physically plausible region of interest is defined, and since the RI will be
186 used to enrich the local at-target site series, it is reasonable to assess whether the proposed region is meaningful
187 and may be accepted as being homogenous (since homogeneity is a prerequisite for the FA). A visualization and
188 analysis of the matrix of correlations between the target site and its neighbors is strongly recommended. In fact,
189 the correlations associated to border sites can give us a good indication on the optimal value of neighborhood
190 threshold N_t . Indeed, the greater the number of weak correlations, the poorer the correlation matrix. It is also
191 desirable to assess whether one or more sites are consistent with the principle of similarity of at-sites frequency
192 distributions in a homogenous region and whether they should be removed from the region. The L-moment-
193 based homogeneity tests of Hosking and Wallis (1993, 1997), widely used in hydrology, are used in this paper.

194 The neighborhood threshold selection stage is a compromise between the size of the region of interest and the
195 correlations between the target site and the border sites in this region. On one hand, if a too high threshold is
196 selected, the correlation between the target site and the most distant sites in the region of interest are good
197 enough (closer to 1) while the region size decreases as there is fewer sites with extremal dependence coefficients
198 above this threshold. On the other hand, by taking a lower threshold, the region size increases as the number of
199 sites with extremal dependence coefficients above this threshold is larger and the correlations between the target
200 site and those at the edge of the region decreases significantly to reach low values close to zero (the design
201 matrix becomes poorer).

202 3 Transfer of regional information to the target site

204 One should remember here one of the basic principles of the developed model which makes it different from
205 existing RFMs: a single frequency distribution is fitted to data from a target site enriched with RI. Without
206 falling into the trap of building a too long regional series to minimize sampling variance (but often with more
207 heterogeneities, spatial dependences and problems related to the calculation of effective durations), and since
208 sites are poorly gauged, the RI is simply used to reconstruct the missing data at the target site. The enrichment of
209 the local sample at the target site is then the second step of the developed model. In order to transfer the RI to the
210 target site, the multiple relationships (at each time step) between the storm surges in the homogenous region of
211 interest are modeled. This stage of the modeling should lead to a design matrix containing all the terms of the
212 multiple regression for all the time steps at which the skew surge was missed at the target site.

213 The MLR is the technique used to transfer the information from the region to the site of interest. The weights
214 of the regression (design matrix) are calculated so that they ensure an optimal fit with respect to the least square
215 criterion for the available data set. However, these weights can no longer be optimal and there is a relatively
216 substantial shrinking in the least square criteria if the same weights are used on a second set of data. The
217 shrinkage can be estimated using a cross-validation procedure. The weights are said to be cross-validated only if
218 the design matrix calculated for the first set of data predict the second one. Furthermore, since interrelationships
219 and collinearities among at-neighboring site surges can be in some cases significant, regression becomes more
220 complex. A stepwise procedure is proposed that removes (or adds) terms from the regression model in stages. As
221 part of the hierarchical regression methods, the ANOVA technique is used. Starting from a full model (all the
222 neighboring sites are involved), at each stage, one or more terms are removed from the model (reduced model)
223 and the change in the least square criterion is calculated. A statistical test is performed as to whether the change
224 is significantly different from zero.

225 4 The FA model using historical information (POTH)

227 It is noteworthy that the principle of the index flood introduced by Dalrymple (1960) is not used in the developed
228 model as a natural framework to perform a regional FA. The use of a local index is restricted to normalizing

229 data for the purpose of forming the region of interest around the target site. Once RI is transferred to the target
 230 site, a long and much more complete local data set is obtained and a local FA can be conducted. The regional
 231 extreme surges, which have joined the target site series, can potentially play a useful and major role in the
 232 construction of the body of the sample of extreme surges, and then reinforce the statistical fitting at the bulk of
 233 the distribution and at the upper tail as well. The Peaks-Over-Threshold approach, POT, models the peaks
 234 exceeding a sufficiently high threshold. Of the many statistical distributions commonly used for extremes, the
 235 generalized Pareto distribution (GPD) is the most suitable theoretical distribution to fit POT series. A GPD
 236 calculates probabilities of observing extreme events which are above a sufficiently high threshold. The use of a
 237 high threshold (provided that the sample size remains reasonable) provides a sample that is more representative
 238 of extreme events.

239 Several studies show that, when the selected threshold u is sufficiently high, the asymptotic form of the
 240 distribution function of excess $F_u(y)$ converges to a GP function (Pickands 1975) which has the following
 241 cumulative distribution function:

$$242 \quad G(x) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\sigma}\right)^{-\frac{1}{\xi}} & \xi \neq 0 \\ 1 - e^{-x/\sigma} & \xi = 0 \end{cases} \quad (2)$$

243 The GPD corresponds to an exponential distribution with a medium-sized tail when $\xi = 0$, and to a long-tailed
 244 ordinary Pareto distribution for positive values of ξ and finally, when $\xi < 0$, it takes the form of a Pareto Type
 245 II distribution with a short tail. Several methods exist to estimate distribution parameters. For the distribution
 246 parameters estimation, the maximum likelihood method is considered in many investigations as an excellent
 247 option (e.g. Tremblay et al., 2008) especially when HI is used. Furthermore, HI is most often considered in the
 248 FA models as pre-gauging data. Less or no attention has been given to the HI relative to the non-recorded
 249 extreme events occurred during the systematic missing periods. The POTH model (which is a POT-based
 250 frequency estimation procedure with HI), introduced by Hamdi et al. (2015), is used in the present paper. To
 251 estimate the distribution parameters by using the maximum likelihood technique in the POTH model, let us
 252 assume a set of n POT systematic observations $X_{sys,i}$ with a set of historical HMax surges $X_{HMax,i}$. The global
 253 log-likelihood can be expressed as

$$254 \quad \ell(G|\theta) = \overbrace{\ell(X_{sys,i}|\theta)}^{\text{systematic data}} + \overbrace{\ell(X_{HMax,i}|\theta)}^{\text{HMax data}} \quad (3)$$

255 For a sufficiently high selected threshold u_s , a total duration w_s and a homogeneous Poisson process with
 256 rate λ , the log-likelihood for the systematic data can be expressed as:

$$257 \quad \ell(X_{sys,i}|\theta) = n \log(\lambda w_s) - \log(n!) - \lambda w_s + \sum_{i=1}^n \log f(X_{sys,i}, \theta) \quad (4)$$

258 A HMax data period corresponds to a time interval of known duration w_{HMax} during which n_k historical
 259 observed surges $X_{HMax,i}$ for $i = 1, 2, \dots, n_k$ are available. Periods are assumed to be potentially disjointed from
 260 the systematic period and other historical periods. The log-likelihood distribution takes the form

$$261 \quad \ell(X_{HMax,i}|\theta) = n_k \log(\lambda w_{HMax}) - \lambda w_{HMax} [1 - F(X_k, \theta)] + \sum_{i=1}^{n_k} \log f(X_{HMax,i}, \theta) \quad (5)$$

262 We refer to Hamdi et al. (2015) for the theoretical developments of these likelihood expressions.

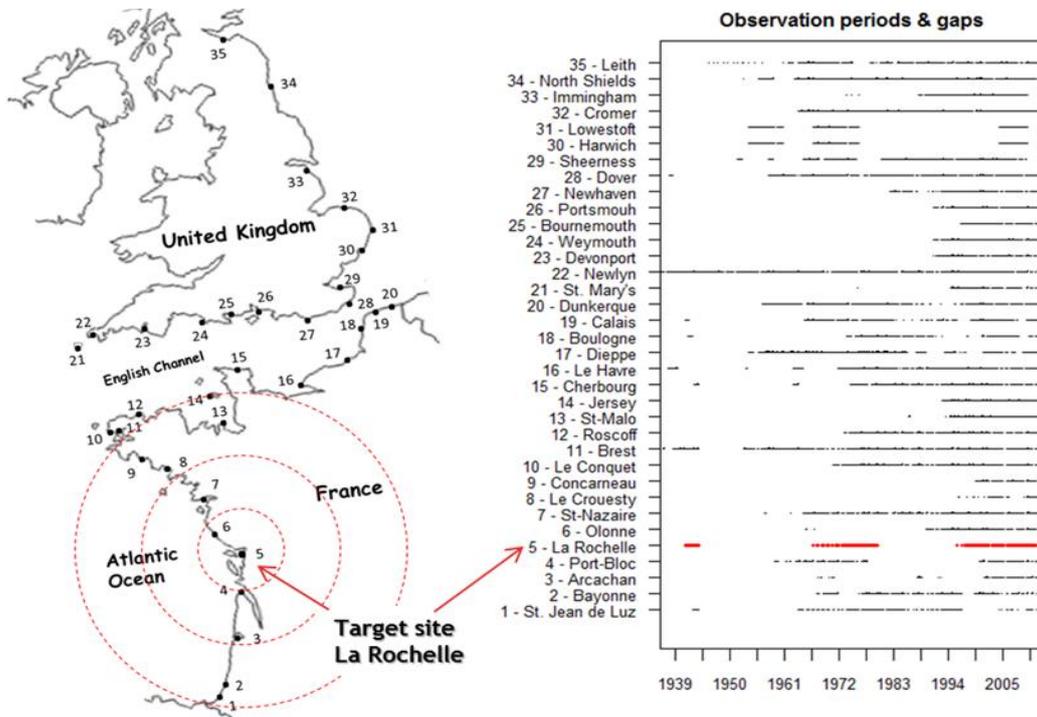
263

264 5 Data and model settings

265 Surge data sets obtained from the predicted tide levels and corrected observations, collected at a total of 35 sites
 266 located on the French (Atlantic and English Channel) and British coasts. They were chosen as the case study for
 267 this contribution (Fig. 1). The French tide gauges are managed by the French Oceanographic Service (SHOM -
 268 Service Hydrographique et Océanographique de la Marine). The British ones are managed by the British
 269 Oceanographic Data Centre (BODC). The R package OCE (Kelly et al. 2015) is used to calculate the tidal
 270 predictions. In order to remove the effect of sea level rise, the initial mean sea level (obtained by tidal analysis)
 271 is corrected for each year by using a linear regression, before calculating the predictions. This method is inspired
 272 by the method used by SHOM for its analysis on high water levels during extreme events (SHOM 2015). The
 273 regression is obtained by calculating daily means using a Demerliac Filter (Wöppelmann 1997). Monthly and
 274 annual means are calculated respecting the Permanent Service for Mean Sea Level (PSMSL) criteria (Holgate et
 275 al. 2013). The developed RFM is performed at La Rochelle (as a target site). One of the most important features
 276 of the La Rochelle site is the fact that the region in which this site is located has experienced significant storms

277 during the last two decades (Martin in 1999 and the Xynthia in 2010). Fig. 1 displays the geographic location of
 278 the whole region and the site of La Rochelle. As depicted in the right-hand panel of Fig. 1, the available surge
 279 data were generally recorded in the period from 1940 to 2010. The figure also shows that the vast majority of
 280 sites, including the target site, are poorly gauged. From 1941 to 2014, the effective length of the data series is
 281 only 30 years. There is no available data for the other years. These data sets constitute the systematic records and
 282 all the non-recorded events that occurred before or during these periods are considered in the developed model as
 283 HI. Indeed, some HI was collected as part of a previous work (Hamdi et al. 2015) in which we collected data
 284 about storm-induced coastal floods that affected the La Rochelle region over the last two centuries (Table 1).

285
 286



287
 288 Figure 1: Location of sites (left); data and gap periods at each site (right). Only Brest and Newlyn have recorded
 289 data before 1940 (from 1846 and 1915 respectively).
 290
 291

292 We are currently working with historians to build a historical database relative to high historical surges
 293 induced by exceptional events that occurred on the entire French coast (Atlantic and English Channel) before the
 294 gauging period. More than 500 historical extreme events are identified. The region of interest (around the La
 295 Rochelle site) is concerned with only some of these data. The work is ongoing and for the time being the HI
 296 presented in Hamdi et al. (2015) can be used.
 297
 298

Table 1. Historical data

Year	1866	1867	1869	1872	1924	1940	1941	1957	1999
Surge (cm)	111	80	87	96	127	139-159	137	111	217

299
 300
 301
 302 **6 Main results of the developed RFM**

303 All the simulations were carried out within the R environment (open-source software for statistical computing:
 304 <http://www.r-project.org/>). The Renext library (IRSN and Alpstat, 2013) (developed by the French Institute for
 305 Radiological Protection and Nuclear Safety - IRSN) was used for the frequency estimations. The Renext package
 306 was specifically developed for flood frequency analyses using the POT method. The main results of the
 307 developed RFM with all the diagnostics are presented in terms of tables and figures (probability plots), where the
 308 main focus was set to the storm surge quantile corresponding to the return period $T = 100$ years. Prior to the FA,

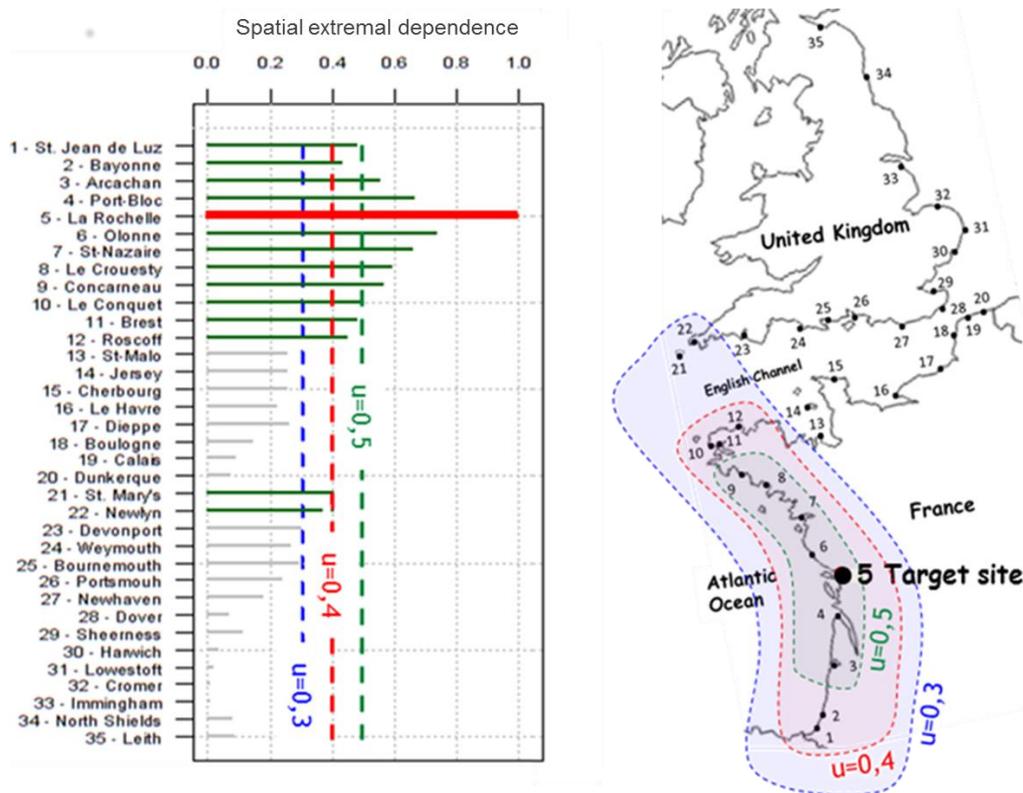
309 the results of the homogenous region delineation and those relative to the transfer of RI to the target site are first
 310 presented first.

311

312 **6.1 Results of the delineation of the homogenous region of interest**

313 The scheme for obtaining the extremal dependence between the La Rochelle site as a target site and all the other
 314 sites is applied herein. The question one can ask is from which value of χ the association between two sites
 315 becomes positive (becomes neighbors)? This brings up another question: from which value of the Spearman's
 316 rank correlation coefficient the relationship between storm surges at two sites can be described by a monotonic
 317 function. An objective answer to this question cannot in any case suggest an exact value or a statistic under or
 318 above which things are different but there is an area in the space of the parameter (i.e. the correlation coefficient)
 319 or a range of values that can be explored and depending on the sensitivity to this coefficient conclusions can be
 320 drawn. From the ESE based on the pairwise extremal dependence coefficient χ , depicted in Fig. 2, three
 321 geographically coherent regions of interest corresponding to three neighborhood thresholds ($N_i = 0.3, 0.4 \& 0.5$)
 322 are obtained. Fig. 2 presents the distributions of the ESE with these neighborhood thresholds.

323



324

325 Figure 2: The ESE centered on the target site (La Rochelle)

326

327 From a physical point of view, these extreme surges are most often induced simultaneously by the same storms.
 328 Regarding the extreme surge definition and extraction, the quantile 99% is used as a sufficiently high threshold
 329 above which a surge is considered extreme. The ESE is performed on all the available sites to define the region
 330 of interest around the target site. It is noteworthy that the propagation of a storm between sites is taken into
 331 account in the definition of simultaneous occurrences of extremes. Indeed, identifying extreme simultaneous
 332 surges is performed in a 1-day time window.

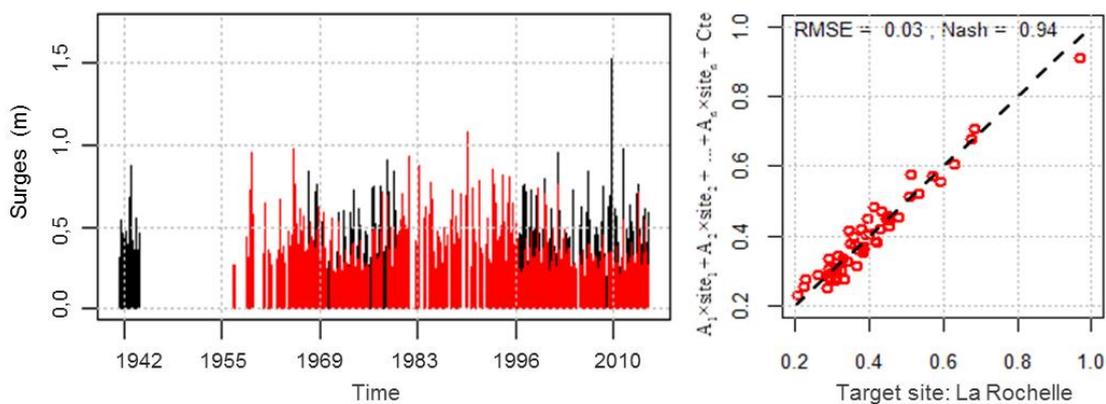
333 As stated earlier in this contribution, the prerequisite for homogenous region delineation is that the surge
 334 distributions anywhere in the region are the same except for a local index parameter. This can be verified by
 335 applying the regional heterogeneity and the at-site discordancy measures proposed by Hosking and Wallis
 336 (1986a, b). The region can be considered as acceptably homogeneous if the heterogeneity measure
 337 $H_i (i = 1, 2, 3) < 1$, possibly heterogeneous if $1 < H_i < 2$ and definitely heterogeneous if $H_i > 2$. For the
 338 discordancy statistic, the target site will be declared discordant if $D \geq 3$. The calculation and the analysis of
 339 these statistical measures confirm that the hypothesis of homogeneity of the region is accepted.

340

341 6.2 Results of the transfer of regional information to the target site

342 The transfer of the RI from the homogenous region to the target site (La Rochelle) is performed using the MLR
343 model and the ANOVA technique. Once the regression matrix is extracted for each configuration of effective
344 sites and prior to performing the transfer, correlation coefficients between the target site and its neighbors are
345 analyzed to evaluate the influence of surges at each neighboring site on the desired surges at the target site.
346 These coefficients provide a measure of the linear relation between two sites' surges and also indicate the
347 existence of collinearities between the explanatory surges (at neighboring sites). This additional step gives us a
348 preliminary idea about which sites are consistent. A site will be kept or removed, in the transfer procedure, by
349 the ANOVA technique.

350 The statistical significance of the regional regressions is also analyzed by calculating the RMSE and the Nash
351 coefficient for each configuration of effective sites. Fig. 3 (left) shows the cross-validation result of the full MLR
352 model when all the neighboring sites (obtained with a neighborhood threshold $N_i = 0.3$) are involved (full
353 model). The scatterplot demonstrates a tendency for reconstructed surges to go with significantly low errors and
354 this is confirmed by the RMSE (3%) and Nash coefficient (94%) criteria. The results show that the MLR may
355 well reconstruct surges at La Rochelle when the full model is used. Before 1941, observed surges are only
356 available at the Brest site. The configuration in which the observations are available only at the Brest site led to a
357 low Nash coefficient and a high RMSE. Furthermore, the ANOVA analysis has also led to reducing the MLR
358 model by removing the Brest site when it is the only one providing RI. Fig. 3 (right) shows the results of the RI
359 transfer from the region of interest to the target site (La Rochelle). These results suggest that 5,259 surges were
360 used to enrich the some 3,148 initial surges at La Rochelle. Note that only relevant configurations of effective
361 sites (according to the RMSE and the Nash coefficient) are used.
362
363



364
365 Figure 3: Cross-validation of the full MLR model (left) and local series at-target site (La Rochelle) with 3,148
366 local observations (black bars) enriched with 5,259 regional extreme surges (red bars) (right).
367

368
369 This means that we have much less variability in the sample and the width of the confidence interval, calculated
370 with the delta method, may also decrease significantly.
371

372 6.3 Results of the developed RFM

373 The FA using the initial data set at the La Rochelle site (before transferring the RI), comparison of the results
374 (with and with no HI included) and the application of the maximum likelihood method to estimate the GPD
375 parameters was performed in a previous work (Hamdi et al. 2015). Initial systematic surges and 9 HMax data
376 ($n_k = 9$) are presented in Fig. 4. Since historical surges tend to be the most extreme and since those occurred in
377 1867, 1869 and 1872 are not high enough, only 6 HMax data are used by the POTH model. The application of
378 the developed RFM on the French coast (Atlantic and English Channel) with La Rochelle as a target site leads to
379 the results presented and analyzed herein. This section is divided into two parts: (i) the results of the FA when
380 systematic data (local and RI) alone are used; and (ii) the main results of the developed RFM with systematic
381 (local and regional) data and HI.

382

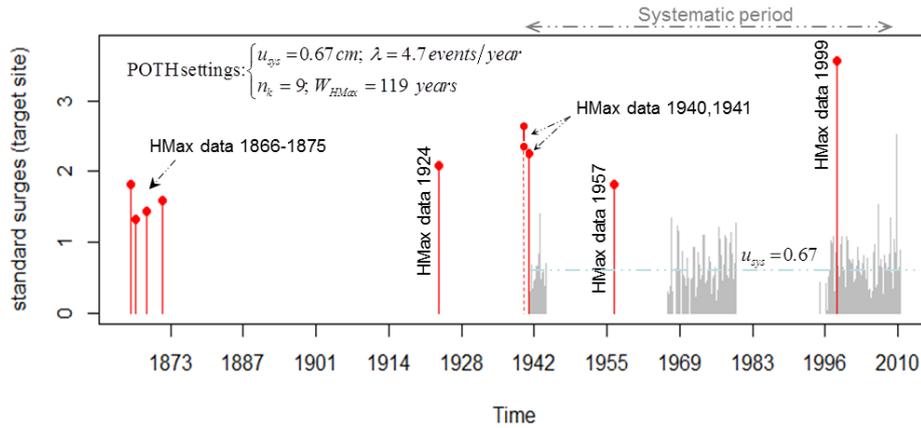


Figure 4: Systematic and historical data with the POTH model settings.

All the results are presented in terms of Return level plots, estimates of the quantiles of interest S_T ($T=100$ years) and relative confidence interval width $\Delta CI/S_T$. The results of the FA, performed with no HI included, obtained for La Rochelle (target site) are summarized in Fig. 5 and Table 2. The presented results correspond to the cases: with no RI included in the sample, and with both local and regional data. The distribution of extreme surges converges to the exponential one when systematic data alone are used. The inclusion of the RI leads to a positive shape parameter and consequently higher quantiles are obtained for return periods higher than 100 years. Indeed, the results given in Table 2 indicate that the effect on the 100-year quantile remains nevertheless limited.

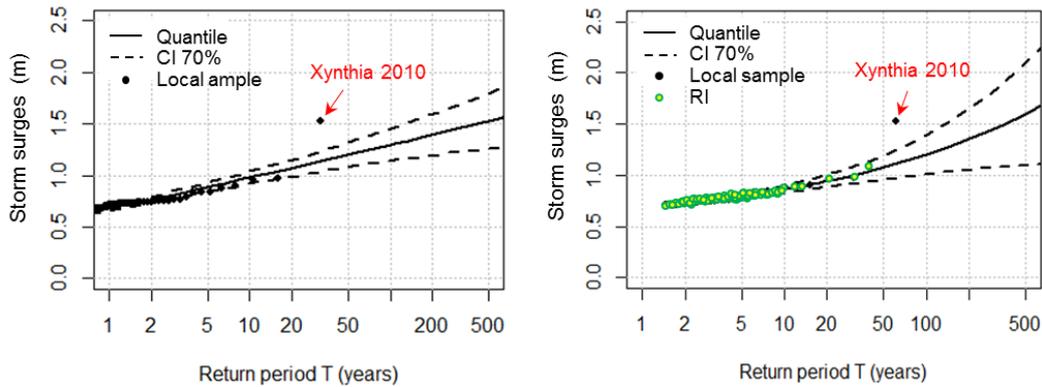
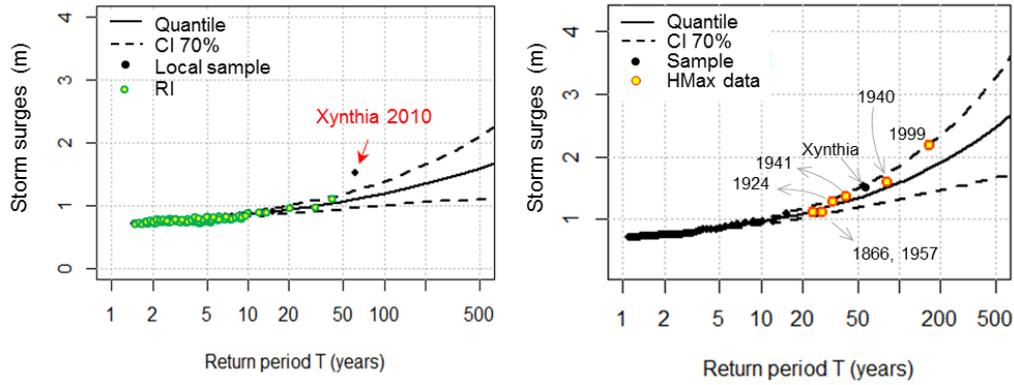


Figure 5: The GPD fitted to the extreme surges: with no HI included (left), local data enriched with RI (right).

The matter at hand is that, as shown in Table 2, the relative widths of the confidence intervals are slightly larger when the regional data are involved. In addition to the fact that, unlike the GPD distribution, the exponential one naturally leads to narrower confidence intervals, there is more variability (measured by the standard deviation) since the sample size ($N_s = 53$ exceedances) is the same in both cases and likewise the effective duration has increased (from 30 years to 60 years when the RI is included). Moreover, on the basis of the visual inspection, we observe that the fit of the large empirical frequencies at the right tail of the distribution (excluding the outlier) is much better than that obtained without RI. This illustrates the consistency of the regional extreme surges in the analysis. Even though the outlier is slightly better represented in the sample, and is closer to the central body of the distribution, it has remained outside the 70% confidence interval.

Fig. 6 and Table 2 summarize the results of the developed RFM. The systematic and historical surges are indicated by black and red filled circles respectively. The results correspond to the cases: when systematic (local and regional) data alone are used, and with both systematic data and HI. In the latter case, the estimated quantile has appreciably changed, owing to the presence of many additional extreme surges (regional data) in the central part of the distribution (i.e. increased probability mass) and many exceptional events (HI) in the upper tail of the distribution especially in the vicinity of the outlier. As a matter of fact, the inclusion of both regional data and HI allow a much better representativeness of the outlier in the sample. The fit with empirical frequencies of both systematic and historical data is satisfactory. Furthermore, given the shape suggested by these observed frequencies, it seems likely that no curve would fit these points more convincingly.



417
418
419
420

Figure 6: The GPD fitted to local extreme surges enriched with RI: with no HI included (left) and with HI included (right).

421
422

Table 2: 100-year quantile and relative width (in %) of its 70% confidence interval (using the local FA, RFM with no HI included and RFM with HI).

	<i>Model settings</i>				$S_{T=100} (m)$	$\Delta CI/S_T$
	$w_s (years)$	$u (m)$	N_s	$w_{HMax} (years)$		
Sys. data	30	0.6	53	---	1.32	25%
Sys. + regional	60	0.7	53	---	1.29	32%
Sys. + regional + HI	60	0.7	53	85	1.59	32%

423
424
425
426
427
428
429
430
431
432
433

From Fig. 6, we observe that the fitting results at the right tail of the distribution appear more adequate and from Table 2, we observe that combining regional data and HI leads to an increase of the 100-year quantile of about 30 cm since the shape parameter is positive and greater than that obtained when regional data alone is used. In other words, the 100-year return level in the initial fitting has a return period of only 30 years in the new fitting with regional and historical information included. Lastly, despite the facts that there is more variability in the sample (the sample size is the same but the effective duration has doubled), and that the 100-year quantile has significantly increased, the relative width of the confidence interval has remained almost constant.

7 Further discussion and performance study

434
435
436
437
438
439
440

The results of the developed RFM are compared with those obtained by Bardet and Duluc (2012) by applying the index flood principle and the regional FA approach proposed by Hosking and Wallis (1993, 1997). The models are applied on the French coast (Atlantic and English Channel). The basics of the approach proposed by Hosking and Wallis (1993, 1997) are presented in the introductory section. The comparison is mainly based on the 100-year quantile and its relative confidence interval width (with and without the use of HI).

7.1 Comparison with the model proposed by Hosking and Wallis

441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456

Table 3 summarizes the results obtained by the developed model and those obtained by the reference model (Bardet and Duluc, 2012). The FA is performed at the La Rochelle site. The return level plot obtained by Bardet and Duluc (2012) for the same site is presented in Figure 7 (storm surges are normalized with the empirical annual value). The comparison is also based on the 100-year quantile and relative widths of their 70% confidence intervals and carried out with and without the use of HI. From Table 3, we observe that the quantile of interest resulting from the proposed RFM performed with the regional data and with no HI included are close to those obtained by the reference model. Conversely and unsurprisingly, the relative width of the 70% confidence interval is significantly greater when the proposed model is applied. Indeed, this is directly related to the so called “bias-variance tradeoff” in selecting the POT threshold. Obviously, when the threshold is chosen very low as was the case in the reference model (41 cm for the La Rochelle site), the sample size may increase significantly. This means that we have much less variability in the sample and the width of the confidence interval, calculated with the delta method, may also decrease significantly. Since the threshold chosen to implement the proposed model has allowed the extraction of extreme storm surges, minimizing at the same time errors related to the bias and the variance, and since the use of this threshold led to a sample size 2 times smaller than that used by the reference model, uncertainty (expressed herein in terms of confidence interval) has increased accordingly. Comparison of the associated uncertainties becomes somewhat meaningless.

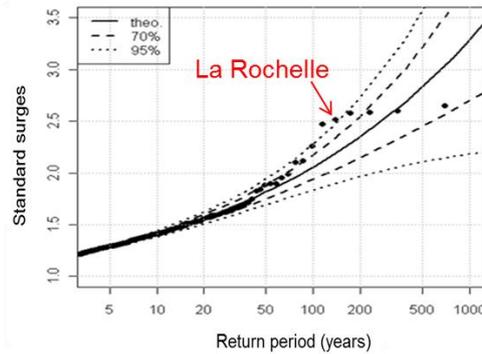
457 Both models use the delta method to calculate confidence intervals. The delta method is a classic technique in
 458 statistics for computing confidence intervals based on maximum likelihood theory. The variance of return level
 459 estimates is calculated using this method and an asymptotic approximation to the normal distribution. The delta
 460 method is used in the present work because, unlike the other methods (such as the profile likelihood and the
 461 bootstrap), it makes it possible to integrate the heterogeneous information (systematic, historical and regional)
 462 more easily. Notwithstanding the fact that it provides symmetric intervals, which is not quite credible, it can be
 463 quite misleading so, in practice, tricks based on transformations, bias-corrections, etc., are often used to improve
 464 their accuracy. The bootstrap confidence intervals automatically incorporate such tricks without requiring
 465 thinking them through for each new application, at the price of an important increase in computational effort.

466 On the other hand, the inclusion of HI in addition to the regional data in the inference has allowed increasing
 467 the quantile of interest by about 35 cm. Moreover, on the basis of the visual inspection of the return level plots,
 468 we observe that the fit of the large empirical frequencies at the right tail of the distribution, including the outlier,
 469 is much better than that obtained by the reference model. The outlier (Xynthia 2010) is also much better
 470 represented in the sample and closer to the central body of the data. The reader is referred to figure 7 and to
 471 section 6 of the present paper for the return level plots.
 472
 473

Table 3: The 100-year quantile and relative widths of its 70% confidence interval (proposed & reference model).

Bardet and Duluc [9]		Proposed model			
		Without HI		With HI	
$S_{T=100}(m)$	$\Delta CI/S_T$	$S_{T=100}(m)$	$\Delta CI/S_T$	$S_{T=100}(m)$	$\Delta CI/S_T$
1.25	9%	1.29	32%	1.59	32%

474
 475



476
 477
 478
 479
 480

Figure 7: The GPD fitted to normalized regional extreme surges (Bardet and Duluc, 2012).

7.2 Advantages and disadvantages of the developed model

481 Unlike the reference model, the one developed here allows us to keep all the raw data at the target site in the
 482 final local data set to be declustered and then used in the POT frequency model. Indeed, the RI is used in the
 483 proposed model to complete the data available at the target site (often poorly gauged). This means that the FA is
 484 performed directly with the at-target site sample enriched with the RI. A first main advantage lies then in the
 485 possibility to conduct the FA directly on the improved local sample at the target site (longer than the initial one;
 486 body and upper tail of data with much more extreme surges, outliers better represented) while avoiding the
 487 problems associated with the use of a local index as a framework of the FA in a regional context. It is commonly
 488 known that the index flood (presented in the reference model as the index surge), which is the core of the
 489 reference model, leads to strong constraints concerning at-site distribution parameters (Ribate et al. 2007).
 490 Indeed, in addition to the issue of wrongly using the same distribution shape parameter throughout a possibly
 491 heterogeneous region and a possibly mixed statistical population (in relation to the meteorological and climatic
 492 mechanisms of extreme storm generation), the index surge model assigns the same weight to all the observations
 493 in the regional sample. This is debatable since the most relevant information is certainly and obviously the target
 494 site information as it is the only information which is “really” distributed as the target site. Thus, in the index
 495 surge approach the available at-target site information is not efficiently used. Furthermore, the use of the local
 496 sample allows us to avoid the problems often encountered in the existing RFMs in dealing with the spatial
 497 dependence and the estimation of the effective duration to be used in the inference.

498 The second main advantage of the developed model lies in the use of HI and how to properly address the
 499 outlier issue in the FA. Unlike the reference model and many other existing RFMs, the developed model allows
 500 using the HI without making “crude” assumptions (especially those relative to the effective durations).
 501 Furthermore, the existing RFMs are not really suitable for data sets containing an outlier. Indeed, as mentioned
 502 in the introductory section and as depicted in Fig. 5 (left), the presence of an outlier in a sample leads to poor
 503 estimation of the distribution parameters and must be addressed properly in the FA. Without surprise, the use of
 504 HI increases the representativeness of exceptional surges in the sample. It is commonly known that the GPD
 505 parameter estimates are strongly influenced by the upper observations and by the body of the observations as
 506 well (e.g. Coles, 2001; Serinaldi and Kilsby, 2014). With respect to this issue, one should be aware that the historical and
 507 the RI work well together to meet this requirement. Indeed, the main feature of the regional observations added
 508 to the local sample at the target site is that they have reinforced much more the lower tail and the body of the
 509 data than the right tail. Alternatively, the latter is reinforced by the historical events because they tend to be the
 510 most extreme.

511 The matter at hand is that the representativity of the outlier has increased substantially, in part because we are
 512 probably somewhat lucky since historical surges are available at the target site. An in-depth analysis to illustrate
 513 what can happen when no HI is available at the target site, and/or the sites, at which HI is available, are removed
 514 because they are not consistent. However, when one is interested in a particular site, the HI can be harvested
 515 around this site. Two more issues that should be kept in mind for a fair comparison of the models are related the
 516 regional regression (the MLR model) used to transfer the RI from the region to the target site. The first issue
 517 regards the bias that could be introduced in the frequency estimates. Indeed, surges used in the full regression did
 518 not exceed 1 m. As a consequence, high surges (including the exceptional ones) are less represented in the
 519 regression matrix. A second related issue when the region of interest is small concerns the quality of the MLR.
 520 Indeed, a larger regional MLR model (with a larger number of predictors) is likely to lead to a more global and a
 521 better regression (this corresponds to a large number of potential neighboring sites in our case). The challenge is
 522 then to find a compromise between increasing the degree of similarity between neighbors (by increasing the
 523 neighborhood threshold N_t , this leads to a smaller number of neighbors) and increasing the number of neighbors
 524 (by decreasing N_t) for a better regression.

525 On another hand, further description of how the skew storm surge time-series were determined and the likely
 526 sources of uncertainty need to be assessed further. Indeed, an in-depth work on the evaluation of uncertainties
 527 associated with the calculation of skew storm surges from observed and predicted water-levels should be
 528 conducted to appreciate for example how this uncertainty propagates in the RFM and how it is likely to be
 529 reflected in the final results (the 100-year return levels, for instance).

530 Finally, a restrictive issue regards the selection of the thresholds (the POT threshold u_{sys} , the neighborhood
 531 threshold N_t and the threshold u_p defined by the sample extreme quantile of probability p). As a matter of fact, it
 532 is difficult to choose a threshold level and this introduces subjectivity in what should be taken as a reasonable
 533 threshold (e.g. Coles, 2001). The reader is referred herein to Bernardara et al. (2014) for a literature review about
 534 this topic.

535 - The threshold u_{sys} to extract the POT observations: the use of a too low threshold introduces a bias in the
 536 estimation by using data which may not be extreme and this violates the principle of the EVT. On another
 537 hand, the use of a too high threshold reduces the size of the sample. This constraint is even more difficult to
 538 avoid when the number of extremes in the data is small (as is the case herein with the initial local series at La
 539 Rochelle) where one is led to mistakenly lower the threshold to collect a sufficient number of exceedances.
 540 As mentioned in a previous section, a large number of regional extremes have joined the at-target site surges.
 541 These additional extremes allowed us to select a higher threshold than that used for the initial sample while
 542 maintaining at least the same sample size. This has the advantage of strengthening the statistical fitting at the
 543 bulk and at the upper tail of the distribution. Unfortunately, even when increasing the number of extremes in
 544 the sample, the problem of simultaneously minimizing the bias and the variance (bias-variance tradeoff)
 545 remains a source of uncertainty when performing a POT frequency model (Coles, 2001).

546 - The thresholds N_t and u_p to compute the ESE and form the region of interest: the asymptotic properties of the
 547 marginal $X_{s \in R}$ (i.e. the fact that X_s belongs to an extreme set is a convention of the ESE) can be violated
 548 when a too low threshold u_p (i.e. $p < 70\%$, for instance) is used to compute the extremal dependence
 549 coefficient of a given site. The use of a too high threshold u_p (i.e. $p > 95\%$, for instance) can significantly
 550 reduce the number of the pairs of simultaneous extreme events to be used in the compute of the χ
 551 coefficient. On the other hand, and even if the up threshold is chosen correctly, a neighborhood threshold too
 552 high will limit the number of neighbors and decrease the size of the region, while a threshold too low will
 553 erroneously increase the region. It is for this reason that we have calculated the extremogram for three values
 554 of N_t . Moreover, even if the u_p is adequately selected, a too high neighborhood threshold N_t (i.e. $N_t > 0.7$,
 555 for instance) will limit the number of neighboring sites and decrease the size of the region of interest while a

556 too low threshold will erroneously increase this region. Although, it is for this reason that we have calculated
557 the extremogram for three values of N_t .

558 **8 Conclusion and future work**

559 A new approach, aiming at incorporating both regional and historical information in a local FA, has been
560 proposed and applied to extreme surges at a poorly gauged target site (La Rochelle), which is located on the
561 French coast (Atlantic and English Channel). Another approach (based on the spatial extremal dependence and
562 the spatial empirical extremogram) to form a homogenous group of sites is also developed. A regional MLR
563 model is used to transfer the RI from the homogenous region to the target site. The RMSE, the Nash coefficient
564 and the ANOVA technique allowed removing the non-consistent terms (i.e. sites) from the regression. Lastly, the
565 POTH frequency model (with adapted likelihood functions to properly handle the HI) is used to perform the FA
566 of extreme surges.

567 Return level plots and estimates of the 100-year quantile with the associated confidence interval are presented
568 and analyzed. Overall, extreme surges (including the outlier) are well fitted when both the regional data and the
569 HI are simultaneously included in the inference. The 100-year quantile is 30 cm higher when the FA is carried
570 out with the developed model.

571 The approach used to form the region of interest (based on the notion of the asymptotic dependence and the
572 ESE) leads to a homogenous region centered on the target site and avoids the problem of the “border effect”.
573 Likewise, the proposed RFM is based on the local data at the site of interest. Indeed, it strengthens the local
574 distribution and assigns a significant weight to the data at the target site (often poorly gauged) which allow us to
575 consider some relevant local effects in the inference. The approach presented in this contribution is thus a
576 promising approach for the estimation of high quantiles at sites of interest (Nuclear Power Plants for instance)
577 with short flood records characterized by the presence of outliers and too many gaps. One question that
578 nevertheless remains open is related to the robustness of the frequency estimations beyond the 100-year return
579 levels and the associated uncertainties.

580 An in-depth study could help to thoroughly analyze the results of the extrapolation to higher return levels (the
581 1000-year quantile for instance) and the associated uncertainties. This could generate several important
582 perspectives, namely:

- 583 1. The assessment of the reliability of storm surges reconstituted with the regional MLR model. Indeed, the
584 estimation of the uncertainties related to these reconstitutions and comparing them with those related to a
585 reconstitution with other approaches (e.g. hydrodynamic modelling) would allow a better appreciation of
586 the results.
- 587 2. In the conducted FA, using the proposed RFM, it was assumed that the largest historical surges were
588 measured or simulated without errors;
- 589 3. In the case study used in the present contribution, HI is available at the target site. It could be helpful to
590 explore what would happen if HI is not available at the target site or available at less consistent sites;
- 591 4. The POT threshold selection remains subjective;
- 592 5. How the theoretical distribution (GPD) is representative of both the systematic and the historical surges is
593 also an important issue.

594 Further research on other target sites could provide us with solutions and answers to these problems and
595 questions.

596

597 **References**

- 598 • Bardet L, Duluc C-M (2012) Apport et limites d'une analyse statistique régionale pour l'estimation de
599 surcotes extrêmes en France. Proc. of the SHF Conf., 1-2 février 2012, Paris, France.
- 600 • Bernardara P, Andreewsky M, Benoit, M (2011) Application of regional frequency analysis to the
601 estimation of extreme storm surges, *J. Geophys. Res.*, 116, C02008, doi:10.1029/2010JC006229.
- 602 • Bulteau T, Idier D, Lambert J, Garcin M (2015) How historical information can improve estimation and
603 prediction of extreme coastal water levels: application to the Xynthia event at La Rochelle (France).
604 *Nat. Hazards Earth Syst. Sci.*, 15:1135-1147. doi:10.5194/nhess-15-1135-2015
- 605 • Chavez DV, Davison AC (2012) Modelling time series extremes. *Statistical J.*, 10(1):109-133.
- 606 • Coles S (2001). *An Introduction to Statistical Modeling of Extreme Values*, Springer, Berlin.
- 607 • Cho YB, Davis RA, Ghosh S (2016) Asymptotic Properties of the Empirical Spatial Extremogram.
608 *Scand J Statist.* 43(3):757-773. doi: 10.1111/sjos.12202
- 609 • Dalrymple T (1960) Flood-Frequency Analyses. *Manual of Hydrology: Flood-Flow Techniques*,
610 Geological Survey Water-Supply, Washington.
- 611 • Groupe de Recherche en Hydrologie Statistique (1996a) Presentation and review of some methods for
612 regional flood frequency analysis, *J. Hydrol.*, 186:63-84. doi:10.1016/S0022-1694(96)03042-9

- 613 • Groupe de Recherche en Hydrologie Statistique (1996b) Inter-comparison of regional flood frequency
614 procedures for Canadian rivers, *J. Hydrol.*, 186:85-103. doi: 10.1016/S0022-1694(96)03043-0
- 615 • Haigh, I.D., Wadey, M.P., Gallop, S.L., Loehr, H., Nicholls, R.J., Horsburgh, K., Brown, J.M.,
616 Bradshaw, E. (2015). A user-friendly database of coastal flooding in the United Kingdom from 1915–
617 2014. *Scientific Data*, 2, 1-13. [150021]. DOI: 10.1038/sdata.2015.21
- 618 • Hamdi Y, Duluc C-M, Bardet L, Rebour V (2016) Use of the spatial extremogram to form a
619 homogeneous region centered on a target site for the regional frequency analysis of extreme storm
620 surges. *J. of Safety and Security Eng.*, 6(4):777–781. doi : 10.2495/SAFE-V6-N4-777-781
- 621 • Hamdi Y, Bardet L, Duluc C-M, Rebour V (2015) Use of historical information in extreme-surge
622 frequency estimation: the case of marine flooding on the La Rochelle site in France. *Nat. Hazards Earth*
623 *Syst. Sci.*, 15:1515-1531. doi: 10.5194/nhess-15-1515-2015
- 624 • Holgate S.J., et al. (2013) New data systems and products at the permanent service for mean sea level.
625 *Coastal Research*, 29:493-504. doi : 10.2112/JCOASTRES-D-12-00175.1
- 626 • Hosking JRM, Wallis JR (1986a) Paleoflood hydrology and flood frequency analysis, *Water Resour.*
627 *Res.*, 22(4):543-550. doi: 10.1029/WR022i004p00543
- 628 • Hosking JRM, Wallis JR (1986b) The value of historical data in flood frequency analysis, *Water*
629 *Resour. Res.*, 22(11):1606-1612. doi: 10.1029/WR022i011p01606
- 630 • Hosking JRM, Wallis JR (1993) Some statistics useful in regional frequency analysis, *Water Resour.*
631 *Res.*, 29:271-282. doi: 10.1029/92WR01980
- 632 • Hosking JRM, Wallis JR (1997) *Regional Frequency Analysis: An Approach Based on L-Moments*, 240
633 pp., Cambridge Univ. Press, Cambridge, U.K.
- 634 • IRSN, Alpst, 2013. Renext: Renewal method for extreme values extrapolation. Online. [http://cran.r-](http://cran.r-project.org/web/packages/Renext/)
635 [project.org/web/packages/Renext/](http://cran.r-project.org/web/packages/Renext/)
- 636 • Jin M. & Stedinger JR, (1989) Flood frequency analysis with regional and historical information. *Water*
637 *Resources Research*, 25(5):925-936. doi: 10.1029/WR025i005p00925
- 638 • Kelly D et al. (2015) OCE: Analysis of Oceanographic Data. Online. <http://dankelley.github.com/oce/>.
- 639 • Ledford AW, Tawn JA (2003) Diagnostics for dependence within time series extremes. *J. R. Statist.*
640 *Soc. Ser. B*, 65:521–543.
- 641 • Payrastré O, Gaume E, Andrieu H (2013) Information historique et étude statistique des crues extrêmes:
642 quelles caractéristiques souhaitables pour les inventaires de crues historiques ? *La Houille Blanche*, 3:5-
643 11. doi: 10.1051/lhb/2013019
- 644 • Pickands J (1975) Statistical inference using extreme order statistics, *ANN STAT*, 3:119-131.
- 645 • Rao, AR, Hamed KH, (2000). *Flood Frequency Analysis*. CRC Press, Boca Raton, Florida, USA.
- 646 • Renard B (2011) A Bayesian hierarchical approach to regional frequency analysis, *Water Resources*
647 *Research*, 47(11), W11513. doi: 10.1029/2010WR010089
- 648 • Ribatet M, Sauquet E, Gresillon J-M, Ouarda TBMJ, (2007) A regional Bayesian POT model for flood
649 frequency analysis, *Stochastic Environ. Res. Risk Assess.* 21(4):327-339. doi: 10.1007/s00477-006-
650 0068-z
- 651 • Davis R, Mikosch T (2009) The extremogram: a correlogram for extreme events. *Bernoulli*, 15(4):977-
652 1009. doi: 10.3150/09-BEJ213
- 653 • Salas JD, Wold EE, Jarrett RD, (1994). Determination of flood characteristics using systematic,
654 historical and paleoflood data, in: Rossi, G., Harmoncioglu, N., Yevjevich, V. *Coping with floods*
655 *Kluwer, Dordrecht*, pp. 111-134.
- 656 • Serinaldi, F., and C. G. Kilsby (2014), Rainfall extremes: Toward reconciliation after the battle of
657 distributions, *Water Resour. Res.*, 50, 336-352, doi:10.1002/2013WR014211.
- 658 • SHOM (2015), Rapport technique du Projet NIVEXT: Niveaux marins extrêmes. Camille DAUBORD.
659 Contributeurs associés: Gaël André, Virginie Goirand, Marc Kerneis. 444 pp. France.
- 660 • Stedinger JR, Baker VR (1987) Surface water hydrology: Historical and paleoflood information, *Rev.*
661 *Geophys.*, 25(2):119-124. doi: 10.1029/RG025i002p00119
- 662 • Tasker G, Stedinger JR (1987) Regional regression of flood characteristics employing historical
663 information, *J. Hydrol.*, 96:255-264. Doi: 10.1016/0022-1694(87)90157-0
- 664 • TRAMBLAY Y., ST-HILAIRE A. and OUARDA, T.B.M.J. (2008). Frequency analysis of maximum
665 annual suspended sediment concentrations in North America, *Hydrological Sciences Journal*, 53:1, 236-
666 252, doi:10.1623/hysj.53.1.236.
- 667 • Weiss, J., 2014. Analyse régionale des aléas maritimes extrêmes. Thèse de doctorat, Sciences de
668 l'ingénieur, Université Paris-Est, 256 p. France.
- 669 • Wöppelmann G, (1997) Rattachement géodésique des marégraphes dans un système de référence
670 mondial par techniques de géodésie spatiale. Thèse de Doctorat. Observatoire de Paris / Ecole Nationale
671 Supérieure des Sciences Géophysiques, 263 p. France.